Decision making and efficient surrogate-assisted optimisation techniques under uncertainty

Application to CO_2 sequestration

Rachares Petvipusit

A thesis presented for the degree of
Doctor of Philosophy

IMPERIAL COLLEGE LONDON

Dept. of Earth Science and Engineering
Qatar Carbonates and Carbon Storage Research Centre
United Kingdom
October 3, 2015

DECLARATION

I declare that this thesis, **Decision making and efficient surrogate-assisted optimisation techniques under uncertainty: Application to CO_2 sequestration**, is entirely my own work under the supervision of Prof. Martin Blunt, Prof. Peter King, and Dr. Ahmed Elsheikh. The work was performed in the Department of Earth Science and Engineering at Imperial College London. All published and unpublished material used in this thesis has been given full acknowledgment. This work has not been previously submitted, in whole or in part, to any other academic institution for a degree, diploma, or any other qualification.

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution-Non Commercial-No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or distribution, researchers must make clear to others the licence terms of this work.

Rachares Petvipusit
Department of Earth Science and Engineering
Imperial College London
Email: Kurt.Petvipusit@gmail.com
October 3, 2015

ABSTRACT

One promising proposal to mitigate the effect of climate change as a result of high atmospheric CO_2 concentrations is carbon capture and storage. Deep saline aquifers provide the most storage capacity for CO_2 storage. However, their geological characteristics are often poorly defined, which makes it difficult to design CO_2 sequestration operations. This thesis considers how CO_2 sequestration can be performed effectively and efficiently under geological uncertainty by allocating injection rates among injectors to maximise economic performance criteria and simultaneously minimise the risk of CO_2 leakage. To this end, we consider the following four challenges in this thesis.

- Optimisation of CO_2 injection strategies is a time-consuming operation since it requires multiple evaluations of expensive black-box functions. This computational cost increases rapidly when accounting for the impact of geological uncertainty. In this thesis, we propose the use of adaptive sparse grid interpolation (ASGI) to speed up the CO_2 optimisation process. Based on numerical results, the ASGI is an efficient surrogate technique and will be used throughout the thesis.
- Injection designs for CO_2 sequestration is difficult when the geology of the storage formation is not well-defined. In the thesis, we propose a utility function to find the injection strategy that is insensitive to the impact of geological uncertainty. Numerical results show potential benefits of using the utility function for the optimisation of CO_2 under geological uncertainty.
- Optimisation of CO_2 sequestration for several conflicting criteria is a challenging task. In the thesis, we use a non-dominated sorting genetic algorithm (NSGA-II) coupled with the ASGI surrogate to maximise both economic performance and sweep efficiency of CO_2 flooding while simultaneously minimising the risks of CO_2 leakage. Solutions obtained from the NSGA method help decision makers to manage multiple conflicting criteria for CO_2 sequestration.
- The ASGI surrogate is an efficient surrogate-assisted optimisation technique for a small to moderate number of input variables, but for problems with 10s of control parameters, it suffers from the curse of dimensionality — the computational time increases exponentially with the number of input variables. In the thesis, we develop a high-dimensional model representation (HDMR) technique to efficiently map the input-output relationship of a function. Several numerical examples show that the HDMR approach is suitable for CO_2 storage optimisation with many input variables.

LIST OF PUBLICATIONS

Peer Reviewed Journal Articles

1. Petvipusit KR, Elsheikh AH, Laforce TC, King PR, Blunt MJ (2014) Robust optimisation of CO_2 sequestration strategies under geological uncertainty using adaptive sparse grid surrogates. Computational Geosciences pp 1–16, DOI 10.1007/s10596-014-9425-z
2. Petvipusit KR, Elsheikh AH, King PR, Blunt MJ Efficient Optimisation Technique Using Spectral High-Dimensional Model Representation: Application to CO_2 Sequestration in Deep Saline Aquifers (In review)
3. Petvipusit KR, Elsheikh AH, King PR, Blunt MJ Adaptive Spectral High-Dimensional Model Representation: Application to Optimisation of CO_2 Sequestration in Deep Saline Aquifers (In review)

Conference Talks

1. Petvipusit KR, Elsheikh AH, Laforce TC, King PR, Blunt MJ (2012) Optimal CO_2 injection strategies under geological uncertainty, SPE Mathematical Methods in Fluid Dynamics and Simulation of Giant Oil and Gas Reservoirs Conference, 3-5 Sep 2012, Istanbul, Turkey.
2. Petvipusit KR, Elsheikh AH, Laforce TC, King PR, Blunt MJ (2013) A robust multi-criterion optimisation of CO_2 sequestration under model uncertainty. Sustainable Earth Sciences Conference, 30 Sep-4 Oct 2013, Pau, France
3. Petvipusit KR, Elsheikh AH, Laforce TC, King PR, Blunt MJ (2013) Multiple Criteria Optimisation of CO_2 Sequestration Strategy Under Uncertainty, SIAM Conference on Mathematical and Computational Issues in the Geosciences, 17-20 June 2013, Padua, Italy
4. Petvipusit KR, Trani M, Bergey P (2014) Does the reduction in structural uncertainty with EnKF improve predictability ?, 9th International EnKF Workshop, 23-25 June 2014, Bergen, Norway
5. Petvipusit KR, Elsheikh AH, King PR, Blunt MJ (2014) Robust Optimisation Using Spectral High Dimensional Model Representation - An Application to CO_2 Sequestration Strategy, ECMOR XIV - 14th European Conference on the Mathematics of Oil Recovery, 8 - 11 September 2014, Catania, Italy
6. Petvipusit KR, Elsheikh AH, King PR, Blunt MJ (2015) An Efficient Optimisation Technique Using Adaptive Spectral High-Dimensional Model Representation: Application to CO_2 Sequestration Strategies, SPE Reservoir Simulation Symposium, 23-25 February 2015, Texas, USA
7. Petvipusit KR, Elsheikh AH, King PR, Blunt MJ (2015) Adaptive Spectral High-Dimensional Model Representation: Application to Optimisation of CO_2 Sequestration Strategies, SIAM Conference on Mathematical and Computational Issues in the Geosciences, 29 June-2 July 2015, California, USA

ACKNOWLEDGEMENT

First of all, I would like to thank my supervisors: Prof. Martin Blunt, Prof. Peter King, and Dr. Ahmed Elsheikh for the continuous support of my study, for his patience, inspiration, and motivation. Often, many discussions from them stimulated me to research new idea and explore exciting topics. The thesis work has been an inspiring and thoroughly enjoyable experience which would not have been possible without their valuable input.

I would also like to acknowledge Dr. Tara Laforce for her support and advice during the early stage of my study before she moved to Australia. I am grateful to Prof. Geoffrey Maitland, Dr. Iain Macdonald, and Qatar Carbonates and Carbon Storage Research Centre (QCCSRC) provided jointly by Qatar Petroleum, Shell, and the Qatar Science & Technology Park for studentship, supporting, and the permission to publish this research.

Similar profound gratitude goes to Prof. Ann Muggeridge, Prof. Matthew Jackson, Prof. Olivier Gosselin, Dr. Jonathan Carter, Dr. Chris Farmer, and Dr. Sam Krevor for their support and feedbacks. I also thank them for their guidance and also serving as my examiners for different stages of PhD examination. Special thank goes to Prof. Ann Muggeridge and Dr. Chris Farmer for their comments and suggestions for the final version of the thesis.

I thank Surinder Dio, Zacharias Zachariadis, and Theo Papadopoulos for assistance with IT support. I thank my fellows: Dr. Allan Leal, Dr. Irina Kostyleva, Nithipoom Durongwattana, and all other friends for their stimulating discussions, encouragement, and lots of fun for day and night.

Last but not the least, I would like to thank my family: my parents and sister for supporting me spiritually throughout writing this thesis. At the end I would like to thank Yuqing Chang for all her love, support, and work-life balance. Without her I probably could have finished my PhD a year earlier ☺.

CONTENTS

List of Tables **ix**

List of Figures **x**

1 Introduction and Objectives **1**
1.1 Thesis objectives and contributions . . . 2
1.2 Organisation of the thesis . . . 3

2 Introduction to Carbon Storage in Deep Saline Aquifers **5**
2.1 Climate change and carbon capture and storage . . . 5
2.2 General aspects of deep saline aquifers for CO_2 storage . . . 7
2.3 Trapping mechanisms . . . 7
2.4 Challenges . . . 10

3 Uncertainty Quantification and Robust optimisation **11**
3.1 Uncertainty Quantification . . . 11
3.1.1 Sources of uncertainty . . . 12
3.2 Robust optimisation . . . 13
3.2.1 Background . . . 13
3.3 Literature review . . . 14
3.3.1 Sensitivity studies and uncertainty assessment related to CO_2 injection . . . 14
3.3.2 A survey of CO_2 sequestration management using optimisation approaches 17

4 Surrogate-Assisted Optimisation **21**
4.1 Introduction . . . 21
4.1.1 A surrogate model . . . 21
4.1.2 Polynomial interpolation . . . 22
4.1.3 Lagrange polynomial interpolation . . . 22
4.1.4 Barycentric Lagrange interpolation . . . 24
4.2 Basis functions in one-dimensional interpolation . . . 25
4.2.1 Definitions . . . 26
4.2.2 Nodal basis function . . . 26
4.2.3 Hierarchical basis functions . . . 26
4.3 Tensor product grid and multivariate interpolation . . . 28
4.4 Sparse grid . . . 30
4.4.1 Hierarchical basis functions . . . 31
4.4.2 Smolyak algorithm . . . 32
4.4.3 Hierarchical surplus . . . 33
4.4.4 Illustration of sparse grid Interpolation . . . 33
4.4.5 Chebyshev-based grid . . . 34

4.4.6 Dimensional adaptivity . . . 35
4.4.7 Implementation and software . . . 36
4.5 Applications . . . 37
4.5.1 Numerical evaluations of mathematical functions . . . 37
4.5.2 Deep saline storage aquifers . . . 41
4.6 Summary . . . 61

5 Multi-objective Optimisation 63
5.1 Introduction . . . 63
5.2 Multi-objective optimisation (MOO) . . . 63
5.2.1 Concept of dominance and a Pareto-optimal set . . . 64
5.2.2 Concept of weak and strong dominance . . . 65
5.2.3 Concept of efficiency and efficient frontier . . . 66
5.2.4 Globally and locally Pareto-optimal front . . . 66
5.2.5 Dependency of criteria . . . 67
5.3 A survey of multi-objective optimisation techniques . . . 68
5.3.1 Prior articulation of preference methods . . . 68
5.3.2 Posterior articulation of preference methods . . . 70
5.3.3 No articulation of preference method . . . 72
5.3.4 Discussion . . . 72
5.4 Multi-criterion optimisation under uncertainty . . . 73
5.4.1 MOO under input uncertainty . . . 74
5.4.2 MOO under model uncertainty . . . 74
5.5 Surrogate-assisted multi-criterion optimisation . . . 77
5.6 Applications . . . 78
5.6.1 Robust optimisation based on the MOO approach . . . 78
5.6.2 Robust multi-criterion optimisation . . . 88
5.7 Summary . . . 100

6 High-Dimensional Model Representation 101
6.1 Introduction . . . 101
6.2 High-dimensional model representation techniques . . . 101
6.3 ANOVA high-dimensional model representation . . . 102
6.4 Cut high-dimensional model representation . . . 103
6.5 Truncated high-dimensional model representation . . . 104
6.6 Hierarchical adaptive sparse grid using high-dimensional model representation . . . 105
6.7 Surrogate building techniques for optimisation . . . 106
6.7.1 A walking Cut-HDMR technique . . . 107
6.8 Applications : hierarchical adaptive sparse grid using high-dimensional model representation . . . 110
6.8.1 Convergence studies using simple mathematical test functions . . . 110
6.8.2 The curse of dimensionality . . . 110
6.8.3 Mathematical function structures . . . 112
6.8.4 Subdomain high-dimensional model representation . . . 116
6.8.5 The PUNQ-S3 reservoir model . . . 122
6.9 Summary . . . 130

7 Adaptive High-Dimensional Model Representation 132
7.1 Introduction . . . 132

7.2 Brief background: adaptive schemes for high-dimensional model representation . . 132
7.3 Sobol indices for sensitivity analysis 133
7.4 Algorithm to compute Sobol indices 135
7.5 Adaptive decomposition ASGI-HDMR based on the Sobol indices. 136
7.6 An adaptive walking Cut-HDMR technique 136
7.7 Applications : adaptive high-dimensional model representation 138
7.7.1 Numerical studies using mathematical functions 138
7.7.2 The PUNQ-S3 reservoir model 143
7.8 Summary 152

8 Summary and recommendations **153**
8.1 Summary 153
8.2 Research directions for future work 156
8.2.1 Decision making and robust optimisation 156
8.2.2 Surrogate building techniques: adaptive sparse grid 158
8.2.3 WCut-AHDMR algorithm 160
8.2.4 Global sensitivity analysis 160

A Error analysis of Lagrange polynomial interpolation **163**

Bibliography **166**

LIST OF TABLES

4.1 Description of the rectangular model and fluid properties 42
4.2 Total mole fraction and diffusion coefficient for initial conditions of the rectangular model. 42
4.3 Description of the PUNQ-S3 model and fluid properties 43
4.4 Geostatistical parameters used to generate Gaussian random fields for the rectangular and the PUNQ-S3 models 45
4.5 Well configuration parameters for both the rectangular and the PUNQ-S3 model 46

5.1 Dispersion estimator using bootstrapping method 82

6.1 A detailed procedure of the Walking Cut-HDMR method 109
6.2 Geostatistical parameters used to generate Gaussian random fields for the PUNQ-S3 model 123
6.3 Well configuration parameters for the PUNQ-S3 model 123
6.4 Initial parameters of the PSO algorithm 124
6.5 Stopping criteria for the WCut-HDMR approach 124
6.6 Numerical settings for the WCut-HDMR approach 127
6.7 Comparison of different methods and control modes of CO_2 injection 131

7.1 Numerical settings for the AHDMR and non-AHDMR surrogates 138
7.2 Initial parameter of the PSO algorithm 145
7.3 Optimal BT values obtained from the last iteration for model 1 in Fig.7.7(a). . . 146
7.4 Optimal BT values obtained from the last iteration for model 2 in Fig.7.7(b). . . 146

LIST OF FIGURES

2.1 Global average surface temperature change . . . 5
2.2 A schematic of storage candidates in CCS . . . 6
2.3 CO_2 density with depth . . . 7
2.4 Trapping mechanisms of CO_2 during post-injection period . . . 8
2.5 A schematic of the structure/stratigraphical trapping . . . 8
2.6 A schematic of CO_2 residually trapped in pore spaces . . . 9

3.1 Sources of uncertainty related to the system design . . . 12
3.2 Comparison of optimal and robust optimal solution . . . 14

4.1 Illustration of Lagrange basis function . . . 23
4.2 Illustration of Lagrange polynomial interpolation . . . 24
4.3 Nodal and hierarchical . . . 27
4.4 Nodal basis for the interpolation of arbitrary function $f(x)$ . . . 28
4.5 Hierarchical basis for the interpolation of arbitrary function $f(x)$ . . . 29
4.6 Tensor product of the hat functions . . . 30
4.7 Sparse grid pattern of the hat functions . . . 31
4.8 Two-dimensional layout of the sparse grid points . . . 34
4.9 Hierarchical surplus . . . 35
4.10 The layout of Chebyshev-based grid points . . . 36
4.11 Numerical evaluations: sparse grid and tensor product grid . . . 38
4.12 Illustration of sparse grid interpolation for the drop-wave function . . . 39
4.13 Illustration of sparse grid interpolation for the Rastrigin function . . . 40
4.14 The rectangular synthetic model . . . 41
4.15 Well location: PUNQ-S3 model . . . 43
4.16 Potential leakage pathways of CO_2 to atmosphere in different scenarios . . . 44
4.17 Well location: Leaky PUNQ-S3 model . . . 44
4.18 Petrophysical properties of three dimensional PUNQ-S3 model . . . 47
4.19 Well configuration: PUNQ-S3 model . . . 48
4.20 Numerical evaluation: PUNQ-S3 model . . . 50
4.21 Refinement levels of the SGI surrogates . . . 50
4.22 Sparse grid pattern: PUNQ-S3 . . . 52
4.23 Numerical evaluations: non-adaptivity and adaptive sparse grid . . . 53
4.24 Pareto-optimal front and the utility function concept . . . 54
4.25 The process of building the utility surrogate . . . 55
4.26 Sparse grid surrogate under uncertainty: using a global measure . . . 56
4.27 Sparse grid surrogate under uncertainty: using a local measure . . . 57
4.28 Robust optimisation using different utility function values . . . 58
4.29 The shape of the Pareto-optimal front with the Utility function . . . 59
4.30 The worst case scenarios of the optimal strategy: PDF . . . 60
4.31 The worst case scenarios of the optimal strategy: CDF . . . 61

5.1 Concept of dominance . . . 64

5.2 Four types of Pareto-optimal front for two objectives . . . 65
5.3 Demonstration of weak and strong Pareto-optimal solutions . . . 66
5.4 Demonstration of global and local Pareto-optimal fronts . . . 67
5.5 Pareto-optimal fronts in the criterion space . . . 68
5.6 Geometrical illustration of the weighted sum and the exponential sum approaches . . . 70
5.7 Illustration of Pareto-front with sorted rank . . . 71
5.8 Illustration of crowding distance . . . 71
5.9 A flowchart of the NSGA-II procedure . . . 72
5.10 MOO under uncertainty . . . 74
5.11 MOO under two sources of uncertainty . . . 74
5.12 Pareto optimality of the rectangular model . . . 80
5.13 Pareto optimality of the PUNQ-S3 model . . . 81
5.14 The bootstrap method with 1000 resampling sets . . . 82
5.15 Histograms and probability distribution of the objective function . . . 83
5.16 A plot of skewness on the injection ratio space . . . 84
5.17 The Lilliefors test . . . 85
5.18 PDFs and CDFs of the break-even tax credit distribution of the rectangular model . . . 86
5.19 PDFs and CDFs of the break-even tax credit distribution of the PUNQ-S3 model . . . 86
5.20 PDF and probability plots of the break-even tax credit distribution of the PUNQ-S3 model . . . 86
5.21 Impact of the density functions on decision making . . . 87
5.22 Scatter plots of break-even tax and unswept volume . . . 90
5.23 Scatter plots of the mean of break-even tax and unswept volume . . . 90
5.24 Pareto-optimal fronts between the break-even tax credits and the unswept volume of the aquifer . . . 91
5.25 Pareto-optimal fronts using different cut-off variables . . . 91
5.26 A matrix of scatter plots of unswept volume, break-even tax credit, and CO_2 leakage . . . 92
5.27 Scatter plots of the mean of three criteria . . . 93
5.28 Probability density function of CO_2 leakage, BT values, and unswept volume . . . 93
5.29 Illustration of Value at Risk (VaR) for two injection strategies . . . 94
5.30 Probability density function of CO_2 leakage . . . 94
5.31 Pareto-optimal front of break-even tax credits, unswept volume of aquifer, and CO_2 leakage . . . 95
5.32 Pareto-optimal front of break-even tax credits, unswept volume of aquifer, and CO_2 leakage . . . 95
5.33 A value path method of non-dominated solutions . . . 96
5.34 Validation of Adaptive Sparse Grid Interpolation . . . 97
5.35 Validation of Adaptive Sparse Grid Interpolation using PDF . . . 97
5.36 Illustration of the Pareto-optimal front clustered by using the minimax approach . . . 99
5.37 Illustration of the Pareto-optimal front clustered by using the preference based approach . . . 99

6.1 Illustration of the Cut-HDMR approach . . . 104
6.2 Illustration of the hierarchical adaptive sparse grid high-dimensional model representation (ASGI-HDMR) . . . 106
6.3 Illustration of a surrogate built for the entire range of input space . . . 108
6.4 Demonstration of Walking Cut-HDMR method . . . 108
6.5 Convergence studies for tensor product, the RSGI, and the ASGI approaches . . . 111
6.6 The curse of dimensionality . . . 112
6.7 The relationship of normalised l_2-norm and the number of sampling points . . . 113
6.8 A convergence study of the ASGI-HDMR approach with one reference point . . . 114
6.9 A convergence study of the ASGI-HDMR approach with various reference points . . . 115
6.10 The convergence rates of HDMR for multiplicative functions . . . 117
6.11 Geometrical illustration of the subdomain region . . . 117
6.12 A layout of the sampling points for the subdomain HDMR and HDMR . . . 119
6.13 Subdomain HDMR: function values plotted versus the subdomain indices . . . 119
6.14 The function values from three functions for Subdomain HDMR studies . . . 120
6.15 The normalised error from three functions for Subdomain HDMR studies . . . 120

6.16 The normalised errors of the subdomain HDMR using only a first order expansion 121
6.17 The normalised errors of the subdomain HDMR using both first- and second-order 121
6.18 Well control configurations of two injectors for 20 input variables 123
6.19 The break-even tax credit (BT) values plotted versus the number of collocation points 125
6.20 The mass of CO_2 in various phases obtained from using 10 and 20 input variables 126
6.21 The relationship of optimal break-even tax credits and the number of collocation points 127
6.22 Optimal injection strategies obtained from the first-order WCut-HDMR and the ASGI approach . 128
6.23 The optimal injection strategies obtained from the first-order WCut-HDMR and the ASGI approach 128
6.24 The mass of CO_2 in various phases obtained from using the WCut-HDMR approach 129

7.1 A flowchart of the WCut-HDMR and WCut-AHDMR algorithms 137
7.2 Error decays of the bivariate function plotted versus the number of collocation points 140
7.3 Error decays of the trivariate function plotted versus the number of collocation points 141
7.4 Error decays of the multiplicative function and the number of collocation points 142
7.5 Error decays of the multiplicative function and the number of collocation points with various cut-offs ϵ_s 143
7.6 The break-even tax credit (BT) values plotted versus the number of collocation points from WCut-HDMR and WCut-AHDMR methods 144
7.7 The break-even tax credit (BT) values plotted versus the number of collocation points compared with the PSO 145
7.8 The break-even tax (BT) values plotted versus the number of collocation points with various alpha 147
7.9 A sensitivity analysis: first-order Sobol indices for model 1 148
7.10 A sensitivity analysis: first-order Sobol indices for model 2 148
7.11 Normalised second-order Sobol indices and the index of pairwise variables 149
7.12 The relationship of the cut-off values and the total number of effective pairwise terms for model 1 150
7.13 The relationship of the cut-off values and the total number of effective pairwise terms for model 2. 150
7.14 The mass of CO_2 in various phases: WCut-AHDMR 151

8.1 A suggested necessary condition of the mean-std approach for decision making 157
8.2 Demonstration of sensitivity analysis of Pareto-optimal front 157
8.3 Three-dimensional response surface of the mean and the standard deviation of the NPV 158
8.4 The layout of Chebyshev-based grid points with constraints 159
8.5 The parameters used in the global sensitivity analysis 161
8.6 Box plots of the main and total effects 161

A.1 Lagrange polynomial of the Runge function using different types of grids 165

CHAPTER 1

INTRODUCTION AND OBJECTIVES

The characteristics of deep saline aquifers are often poorly defined, which make it difficult to design an effective injection strategy. Even though finding the optimal injection strategy that maximise performance criteria while guaranteeing long-term storage security is difficult and challenging, it is a key factor for successful CO_2 sequestration. In the following, we will address four main challenging problems.

Firstly, the optimisation of CO_2 injection strategies requires multiple evaluations of expensive black-box functions, which is computationally intensive restricting practical applications of optimisation algorithms. Surrogate-assisted optimisation techniques could be used to reduce the computational costs. In the thesis, we propose the use of Adaptive Sparse Grid Interpolation (ASGI) to accelerate the optimisation of CO_2 injection rates. The surrogate is adaptively built with different numbers of evaluation points (simulation runs) in different dimensions to allow automatic refinement in the dimension where added resolution is needed. This technique is referred to as dimensional adaptivity and provides a good balance between accuracy of the surrogate model and the number of simulation runs to save computational costs.

Secondly, decision making for CO_2 sequestration is difficult and uncertain when the geology of the aquifer is not well-defined. In this thesis, we use a utility function to find a robust injection strategy. The utility function is approximated by the ASGI surrogate to reduce computational burden. Numerical testing of the proposed approach applied to benchmark functions and aquifer models show the efficiency of the method for the optimisation of CO_2 under geological uncertainty.

Next, there are several decision criteria for CO_2 sequestration as we not only maximise the economic performance of the project, but also maximise the mass of trapped CO_2 while simultaneously minimising the risk of CO_2 leakage. Unfortunately, these criteria are generally conflicting objectives, and often difficult to guarantee when sufficient geological data of the storage sites is not available. In the thesis, we propose multi-criterion optimisation (MOO) technique with the use of the ASGI surrogate. Additionally, instead of using the utility function described earlier, we use the MOO technique for robust optimisation.

Lastly, the ASGI surrogate is an efficient surrogate-assisted optimisation technique for a small to moderate number of input variables. However, for problems with 10s of input variables, it suffers from the curse of dimensionality — the computational time increases exponentially with the number of input variables. In the thesis, we propose a surrogate building technique for CO_2 optimisation with many input parameters. The surrogate is built based on a high-dimensional model representation to efficiently map the input-output relationship of a function by using significantly fewer function evaluations.

Generally, this thesis considers how CO_2 sequestration in deep saline formations can be operated effectively and efficiently when the geological characteristics are not known, and also when the decision criteria are conflicting. In the following, we will define the objectives of this study, followed by the organisation of the thesis.

1.1 Thesis objectives and contributions

This thesis addresses two areas of decision making under uncertainty — a single decision criterion, and multiple criteria — and one area for development of surrogate-assisted optimisation for high-dimensional problems. Specially, it introduces and applies novel surrogate techniques for optimisation under uncertainty, and develops a new optimisation algorithm to handle many input variables for application to CO_2 sequestration in deep saline aquifers.

The proposed optimisation approaches in this thesis were tested on synthetic CO_2 sequestration problems. The outcomes of the thesis are:

1. Applications of a novel surrogate-assisted optimisation technique based on adaptive sparse grid interpolation for CO_2 sequestration.
2. Development of an efficient algorithm to find optimal injection strategies for CO_2 storage in deep saline aquifers under geological uncertainty.
3. Development of robust and optimal multi-criterion decision making for CO_2 sequestration.
4. An extension of surrogate-based techniques for optimising high-dimensional problems.

The contributions of the thesis are:

1. The first application of a surrogate-assisted optimisation technique based on adaptive sparse grid interpolation to CO_2 sequestration.
2. A new economic performance criterion based on a break-even tax credit function for optimising CO_2 injection rates.
3. A new utility function of a break-even tax credit for optimising CO_2 injection rates with the consideration of model uncertainty.
4. Development of an efficient algorithm for optimising CO_2 storage in deep saline aquifers for both single and multiple objective functions under geological uncertainty.
5. The first development of a surrogate technique for high dimensional objective functions of CO_2 sequestration.
6. A new optimisation method based on the proposed surrogate technique for optimising high dimensional problems.
7. Development of a new adaptive scheme to improve the computational efficiency of the proposed optimisation method.

1.2 Organisation of the thesis

Chapter 2 begins with an overview of carbon capture and storage (CCS) as a potential technology to mitigate climate change effects induced by greenhouse gas emissions (GHGs). We review the evidence that climate change, atmospheric CO_2 concentration, and the GHGs are closely related. Atmospheric emissions of CO_2 can be reduced by injecting CO_2 into geological formations such as deep saline aquifers. Following that, we will explain why deep saline aquifers are suitable for CO_2 storage and what characteristics make them attractive. The challenge of finding good injection design for saline aquifers will be discussed, followed by the detailed descriptions of aquifer models we used in this thesis.

Uncertainty quantification (UQ) will be addressed in Chapter 3 where we discuss the main ingredients, such as types and sources of uncertainty, and explain why poorly defined characteristics of saline aquifers are treated as geological uncertainty. Later in Chapter 3, we will demonstrate how to mitigate the impact of geological uncertainty by using robust optimisation. Recent literature related to the sensitivity analysis of trapped CO_2 in aquifers to various aquifer parameters will be briefly presented, followed by the reviews of recent optimisation techniques for CO_2 sequestration.

The main thesis begins from Chapters 4 onward for decision making and optimisation techniques for CO_2 injection strategies under uncertainty, which we will present in four chapters:

1. Surrogate-assisted optimisation in Chapter 4
2. Multi-criterion optimisation in Chapter 5
3. Surrogate techniques for high-dimensional functions in Chapters 6 and 7

Chapters 4 introduces the surrogate building technique to increase computational efficiency of the optimisation procedure. This chapter mainly presents the mathematical details underlying a sparse grid surrogate approach, such as polynomial interpolation, and the types of chosen grid nodes. Later, we will present the idea of dimensional adaptivity of the sparse grid, named as adaptive sparse grid interpolation (ASGI). For the optimisation of CO_2 sequestration using the ASGI method, we first present a new break-even tax credit (BT) function, which takes into account the economic criterion of CO_2 sequestration project, as the objective function of CO_2 optimisation. Later, for robust optimisation, we discuss the use of a utility function that comprises mean and variance of the break-even tax credit function to cope with geological uncertainty. Numerical studies for both adaptive and non-adaptive sparse grid interpolation for approximating the utility function will be discussed. We will later present how optimal solutions obtained from the utility function are insensitive to the impact of the geological uncertainty of aquifer models.

Chapter 5 discusses the use of multi-objective optimisation (MOO) techniques to find robust injection strategy and trade-off solutions between conflicting criteria. In this chapter, we will demonstrate basic principles and general ideas of multi-objective optimisation, such as the concepts of domination, a Pareto-optimal set, and the dependency of decision criteria.

A short survey of the MOO approaches is given with the detailed descriptions of a Non-Dominated Sorting Algorithm (NSGA-II), which will be used for robust CO_2 optimisation presented later in two sections. For the first section, we show that the mean and variance of the BT function are two conflicting objectives, where the MOO approach can be used to find robust solutions. Later, we will discuss how different types of parametric distributions used to estimate the mean and variance could impact decision making. For the second section, in addition to the BT function, we use volumetric sweep efficiency of CO_2 flooding and the risk of CO_2 leakage from storage aquifers as additional objective functions for optimisation. Then, we present how to find the optimal injection strategies when these objectives are conflicting. Several aquifer models are used in this chapter to demonstrate the MOO approach under uncertainty.

The surrogate building techniques for optimisation of CO_2 injection strategies with many input parameters will be presented in Chapter 6. We review the mathematical principles underlying High-Dimensional Model Representation (HDMR) and several extensions of the HDMR techniques, including Cut-HDMR, truncated HDMR, and subdomain HDMR. They are fundamental concepts for our proposed surrogate building technique for optimisation. Later, we will discuss why our approach can be a potential candidate for optimisation of practical CO_2 sequestration with many injection controls, and show how to balance exploration and exploitation for finding optimal solutions. Several simple mathematical functions and aquifer models are used for demonstration in this chapter. To increase computational efficiency of the algorithm, Chapter 7 presents how a global sensitivity analysis could be used to develop an adaptive version of HDMR with applications to CO_2 sequestration.

Chapter 8 summarises the thesis by presenting the main findings of each chapter. The recommendations for future work to improve our proposed methods are discussed in the end.

CHAPTER 2

INTRODUCTION TO CARBON STORAGE IN DEEP SALINE AQUIFERS

A brief introduction of CO_2 sequestration in deep saline aquifers is given in Section 2.1, followed by general aspects of saline aquifers and trapping mechanisms in Section 2.3 and 2.4. The challenges in developing CO_2 sequestration are given in Section 2.4. For the aquifer models used in the thesis, we will provide detailed descriptions later in Chapter 4.

2.1 Climate change and carbon capture and storage

The concentrations of CO_2 in the atmosphere and anthropogenic forcing appear to be closely related to climate change. It was reported by Stocker et al [165] that an increase in global averaged temperature since 1950 is very likely due to human-related activities such as greenhouse gases (GHG) emissions. The change of averaged temperature from 1950 to 2100 is illustrated in Fig. 2.1 where four Representative Concentration Pathways (RCPs) are presented. The terms RCP2.6, RCP4.5, RCP6.0, and RCP8.5 shown here provide time-dependent projection of greenhouse gas concentrations [173] in the atmosphere, with shaded band denoting the 5 to 95% confidence intervals; these four RCPs reflex scenarios with different magnitudes of climate change; the detailed descriptions can be found in [137, 173].

From Fig. 2.1, we can see that the relationship of temperature increases and the atmospheric CO_2 concentrations, the RCPs, are closely related. Additionally, Parry [123]

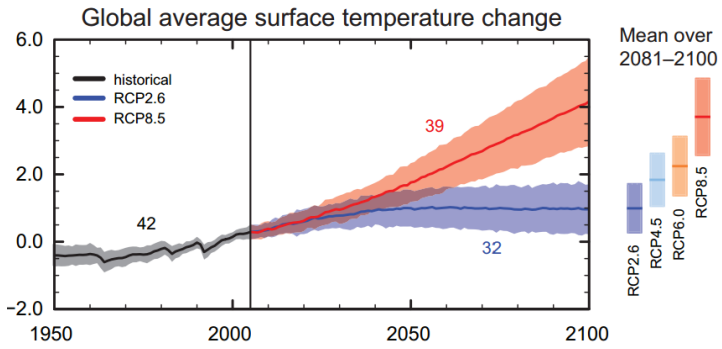


Figure 2.1: Global average surface temperature change (°C) from 1950 to 2100 is displayed using four Representative Concentration Pathways (RCP) models, with shaded band presenting the confidence intervals from 5 to 95% . The black curve presents historical data since 1950 to 2005. This image was obtained from [165].

reported that a rise in CO_2 emissions could lead the global temperature to increase in the range of 4 °C to 7 °C. Later, Bennaceur et al [11] documented in 2008 that this rise of temperature could be reached by 2050 if there are no technologies and policies to mitigate the atmospheric GHG emissions. To this end, three main human activities contributing to the GHG emissions are: transportation, utility, and industrial production.

Among human activities contributing to the GHG emissions, combustion from fossil fuels is the main contributor (about 96.6%) of CO_2 emissions to the atmosphere [43]. Renewable energy (e.g. wind, solar, hydroelectric, biomass, and geothermal energy) are alternative sources that produce low-carbon or carbon-free emissions and could replace the use of fossil fuels in future. Even though renewable and alternative energy are reliable and promising proposals to mitigate excessive emissions of the GHGs, they are long-term solutions which cannot immediately prevent the climate problems we are facing today. Carbon capture and storage (CCS) is considered one of the most promising and immediate proposal to mitigate the effect of excessive CO_2 emissions in the atmosphere. CCS is the process of capturing and sequestering CO_2 from industrial sources into geological formations, which could significantly reduce CO_2 emissions [105].

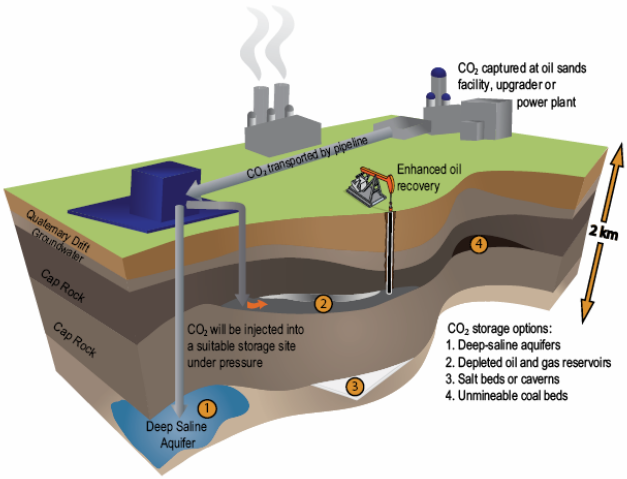


Figure 2.2: A schematic of storage candidates in CCS. Image captured from http://www.co2solutions.com [26].

Figure 2.2 shows four types of storage candidates for CCS: depleted oil and gas reservoirs, salt beds, coal beds, and deep saline aquifers. The best potential candidate for industrial scale CO_2 storage are deep saline aquifers as their worldwide capacities are the largest and estimated to be $10^3 - 10^4$ $GtCO_2$ [105]. Deep saline aquifers are therefore the focus in this study as geological formations for CO_2 sequestration. In the next section, we will describe general aspects of deep saline aquifers, followed by the fundamental trapping mechanisms for CO_2 storage.

2.2 General aspects of deep saline aquifers for CO_2 storage

Deep saline aquifers generally occur at depths deeper than a drinking water source. The formation water at this level contains high salinity such that it cannot be used. As addressed in [12], both sandstone and carbonate formations could be used to store CO_2, from which CO_2 is injected in a dense phase (supercritical state) — the state at which its properties are above both critical temperature (31 ^oC) and pressure (7.38 MPa). This means that CO_2 in the supercritical state exhibits gas-like viscosity and liquid-like density where it flows through porous rocks like gas and can dissolve material like liquid [111]. Additionally, the density of supercritical CO_2 increases with pressure, which enables CO_2 to reside in pore spaces of rocks with much less volume than its gaseous phase. The density of CO_2 with depths using a geothermal gradient 25 °C/km is shown in Fig. 2.3.

We see from Fig. 2.3 that a suitable depth for CO_2 storage is greater than 800 m as CO_2 is usually a supercritical state in this region whereas at the shallower depths CO_2 is more likely in an aqueous phase depending on the in-situ pressure and temperature conditions. However, the suitable minimum depth for CO_2 storage is between 600-1200 m and varies depending on the temperature and pressure gradients in a particular formation [7].

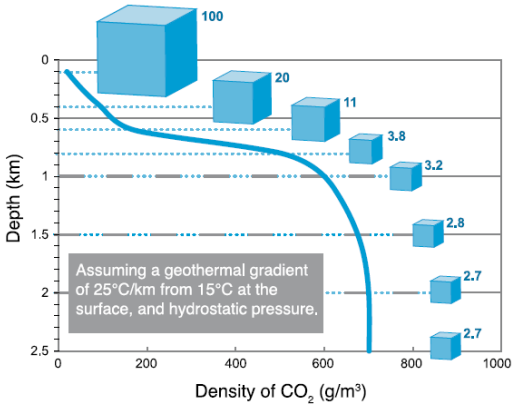


Figure 2.3: CO_2 density (g/m^3) at geothermal 25 °C/km with depth (m), obtained from [104].

2.3 Trapping mechanisms

Trapping mechanisms in deep saline aquifers are defined as a combination of physical and chemical processes that allows CO_2 to be trapped and stored in geologic formations [139, 155]. Trapping mechanisms can be categorised into four types. (1) Structure/stratigraphical trapping: trapping through impermeable rocks. (2) Residual or capillary trapping: immobile phase of CO_2 in pore spaces due to the capillary force. (3) Solubility trapping: CO_2 is dissolved into formation brine and sinks to the bottom of reservoir. (4) Mineral trapping: chemical interaction between CO_2 and brine forming solid minerals.

Figure 2.4 shows the contribution of each mechanism during the post-injection phase. Residual and solubility trapping are the main mechanisms during ten to thousands of years. We see that the contribution of the residual and dissolution trapping are within the time scale (the order of years to centuries) that we can predict underlying parameter values with some degrees of certainty [152] and they are likely to be most important in injection design. Therefore, in this study, we consider the injection design that can enhance CO_2 trapping from both residual and dissolution mechanisms. A brief review of different types of trapping mechanisms is provided in the following few paragraphs.

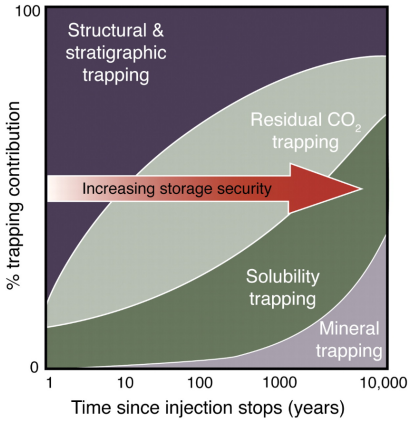


Figure 2.4: The evolution of trapping mechanisms of CO_2 during post-injection period. All mechanisms except structural trapping increase their trapping contribution over time within 10.000 years after injection [105].

Structure/Stratigraphical trapping

Structural trapping is a physical mechanism which uses an overlying impermeable (sealing) layer of rocks, known as cap rock, to prevent the buoyant CO_2 from migrating upward to the top of the formation. The mechanism is similar to what retains oil and gas in hydrocarbon reservoirs. An appropriate geometry of the sealing rock is essential for effective trapping. The structural trapping could fail to trap CO_2 if CO_2 can leak through the cap rock; this could result from the penetration of drilled or abandoned wells, fractures induced by CO_2 injection, faults, and seismic activity. For structure trapping, the risk of CO_2 leakage to the surface is high if the cap rock is not capable of holding CO_2. Figure 2.5 shows a cartoon of structural trapping with CO_2 shown in blue.

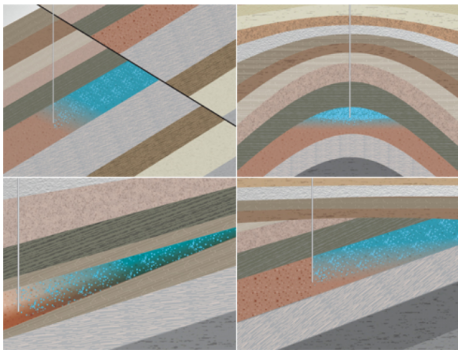
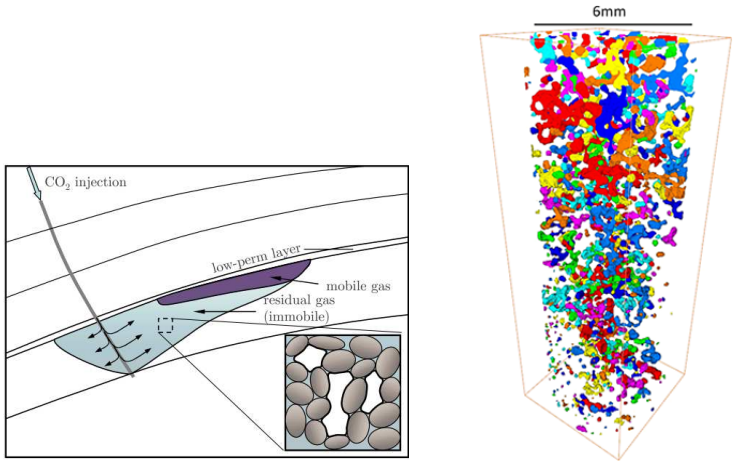


Figure 2.5: A schematic of the structure/stratigraphical trapping with CO_2 shown in blue, from [32].

Residual trapping

Residual trapping keeps CO_2 in an immobile phase in the pore spaces due to capillary forces. Residual trapping is the most important and rapid process for permanent and safe storage in the order of years to decades [41, 71, 76, 86, 116]. The trapping process occurs

when CO_2 is disconnected (snap-off mechanism) due to the capillary pressure between brine and the pore spaces. Figure 2.6(a) shows CO_2 residually trapped in pore spaces behind the flood front of the injection while Fig. 2.6(b) displays the image of CO_2 ganglia from the 3D rendering of the core flood experiment [6]. Each colour represents a unique CO_2 ganglion isolated and trapped in the capillary pores. Some authors have shown that the residual trapping mechanism could be enhanced by using an appropriate injection design [5, 133, 152].



(a) A schematic of CO_2 residually trapped in pore spaces [76]

(b) A 3D rendering of CO_2 ganglia from the core flood experiment [6]

Figure 2.6: a) A schematic of CO_2 residually trapped in pore spaces behind the flood front of the injection [76], b) A 3D rendering of the core flood experiment. Each colour represents a unique CO_2 ganglion that is isolated and residually trapped, from [6].

Dissolution trapping

In this process, CO_2 dissolves in resident brine so that the density of brine becomes slightly denser. The denser brine then sinks to the bottom of the formation and causing a convection current — dense brine sinking and less dense, CO_2-free brine rising, which allows further dissolution from the additional contact of unsaturated brine along the sinking path. The time scale required is in the order of hundreds to ten thousand years [42]. Formation pressure, brine salinity, reservoir temperature are all important parameters that can affect the dissolution process. Sweep efficiency, mobility, permeability, and anisotropy ratios also impact the rate and extent of dissolution [155].

Mineral trapping

Mineral trapping is the process of chemical interaction between formation brine and dissolved CO_2 to form carbonate minerals. The extent and rate of mineral trapping depend on mineralogy, composition of rock types, diffusion and sedimentary structures, and formation water compositions [139]. Mineral trapping is considered the most secure mechanism while being the slowest mechanism for CO_2 storage.

2.4 Challenges

Deep saline aquifers are promising candidates for CO_2 storage. However, for successful CO_2 operations, two key challenges need to be addressed. Firstly, the geology of deep saline aquifers are often poorly defined due to lack of knowledge about the formation evaluation. Thus, uncertainty quantification and injection designs that are robust under geological uncertainty need to be considered. Secondly, operation of CO_2 sequestration involves multiple objectives, where economic and environmental consequences (i.e. mass of the injected CO_2, and long-term storage security) play significant roles for designing CO_2 injection strategies. Unfortunately, these objectives are often in conflict and difficult to reconcile.

To this end, in this thesis, we address these issues and present how to design CO_2 injection strategies for deep saline storage under geological uncertainty and under multiple conflicting objectives. In the next chapter, we will present basic principles of uncertainty quantification and robust optimisation techniques for CO_2 injection design.

CHAPTER 3

UNCERTAINTY QUANTIFICATION AND ROBUST OPTIMISATION

This chapter introduces uncertainty quantification (UQ) of computer experiments using simulations. Later, we will discuss how optimisation technique could be used to account for model uncertainty — a robust optimisation method. Literature reviews of uncertainty analysis and optimisation techniques with applications to CO_2 sequestration are provided in the last section.

3.1 Uncertainty Quantification

Uncertainty can be defined as the state of lacking or having limited knowledge of information. Uncertainty quantification is the science of quantitative characterisation and reduction of impact of variability and lack of knowledge on the quantities of interest [70]. For the component of uncertainty quantification, Psuade [131] suggested the following list:

1. Identification: identifying the source of uncertainties
2. Characterisation: uncertainty can be characterised to aleatoric and epistemic uncertainties.
    - Aleatoric uncertainty (objective uncertainty or irreducible uncertainty). This type of uncertainty involves variability of physical nature or environment, which cannot be removed or reduced from the system of interest. It is usually modelled with a probabilistic framework [157] as it is not strictly due to lack of knowledge [70]. Therefore, we need a design that can handle and reduce the impact of this uncertainty on the quantity of interest.
    - Epistemic uncertainty. This uncertainty (also known as subjective or reducible uncertainty) is due to lack of knowledge, simplifying assumptions or ignoring physics in the model [157], which in principle can be reduced by using representative models or having more information through experimental data, expert opinions, observation data, etc [141]. Also, the epistemic uncertainty could significantly result in the forecast biases.

Remark : Aleatoric and epistemic uncertainties are characterised based on our existing knowledge of a system. Therefore, there exists a mixed type of uncertainty where aleatoric and epistemic uncertainties are not clearly distinct as it depends on

current knowledge, experiments, and physics embedded in the model [157]. In other words, aleatory and epistemic uncertainties are interchangeable from one to another system. For example, geological uncertainty in aquifer models is reducible uncertainty where additional information could be acquired to reduce the uncertainty; however, we can treat it under a probabilistic framework as irreducible geological uncertainty when the objective is to reduce the impact of uncertainty but not the uncertainty itself.

3. Propagation is a study of how input uncertainty can propagate to the output uncertainty where the reasonable interval of the quantity of interest needs to be defined for analysis [157]. For large-scale simulations, it is worth noting that a proxy model or a response surface may be required to obtain a reasonable predictive estimation of output interval.

4. Uncertainty analysis involves a study of the effect of uncertainty, the impact of uncertainty (risk analysis), the reliability of model, and the types of data for model calibration.

5. Uncertainty reduction: model calibration from additional data and/or experimental observations leads to uncertainty reduction. This consequently increases confidence in a model and it may improve predictive capability of the model.

3.1.1 Sources of uncertainty

We describe four sources of uncertainty related to the input/output of a system shown in Fig. 3.1, where they can be described as:

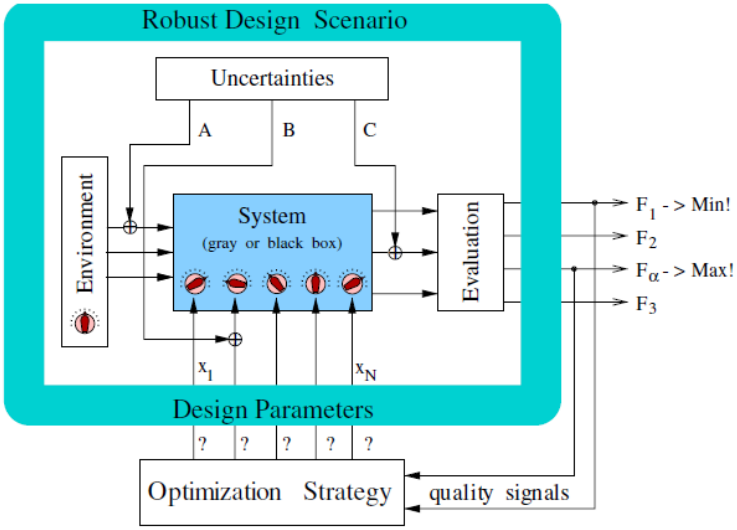


Figure 3.1: Sources of uncertainty related to the system design: (A) uncertainty related to environmental conditions; (B) uncertainty related to design variables $\mathbf{x}$; (C) uncertainty related to output; (D) uncertainty related to the constraint of input variables (not shown here), from [14].

(A) Uncertainty related to environmental conditions: this uncertainty is related to environmental conditions which we cannot control and reduce. Mathematically, if $\mathbf{f}$ and $\mathbf{x}$ represent a system output and design variables respectively, the system output with uncertain environmental conditions (ξ) is given as $\mathbf{f} = \mathbf{f}(\mathbf{x}, \xi)$.

(B) Uncertainty related design variables $\mathbf{x}$: uncertainty enters the system of the form $\mathbf{f} = \mathbf{f}(\mathbf{x} + \varepsilon)$, which relates to a certain degree of accuracy in the design variables with ε presenting the variation of input variables.

(C) Uncertainty related to system output: this uncertainty involves approximation errors associated with the use of a model (e.g. numerical dispersion of a simulation, model assumptions, simplified physics models).

(D) Feasibility uncertainty: this is more related to design spaces of input variables, rather than the output $\mathbf{f}$. This uncertainty involves the uncertain constraints of input variables; this type of uncertainty can also be characterised as either type (A) or (B).

For CO_2 sequestration applications, the system in Fig. 3.1 represents a compositional flow simulator [150] with the input ($\mathbf{x}$) and the output ($\mathbf{f}$) denoting CO_2 injection controls and a measure of performance, respectively. The geological uncertainty of aquifer models in this work represents a variability of physical nature that cannot be reduced through the design process. In order to reduce the impact of uncertainty, we use a robust optimisation approach — an optimisation technique that simultaneously mitigates the impact of uncertainty on the model output.

3.2 Robust optimisation

3.2.1 Background

Fowlkes and Creveling [49] described robust design as the design of processes that is insensitive to the effect of sources of variability by not necessarily eliminating the causes. Robust optimisation is the process of optimising an objective function where its solution is insensitive to the impact of uncertainties.

Figure 3.2 shows the difference between optimal and robust solutions when minimising the objective function. The uncertainties of the objective function are shown by the bell curve distributions on the y-axis, which are projected from the variation of the objective function (the length of the arrow). The solid curve in Fig. 3.2 represents the objective function where its uncertainty (Type C) is shown by dashed curves for the upper and lower limits. The blue arrow corresponds to variation of the objective function values at the optimal solution while the red arrow presents the variation at the robust solution. We see that the robust solution may not necessarily provide the minimum value in comparison to the optimal solution. However, it gives the best solution under the impact of uncertainties as the length of the red arrow is smaller than the blue ones. Therefore, the robust solution provides the best minimum value for both objective function and its variation from uncertainties.

In Chapters 4 and 5, we will discuss two robust optimisation methods, namely a utility function and a multi-objective approach, in detail.

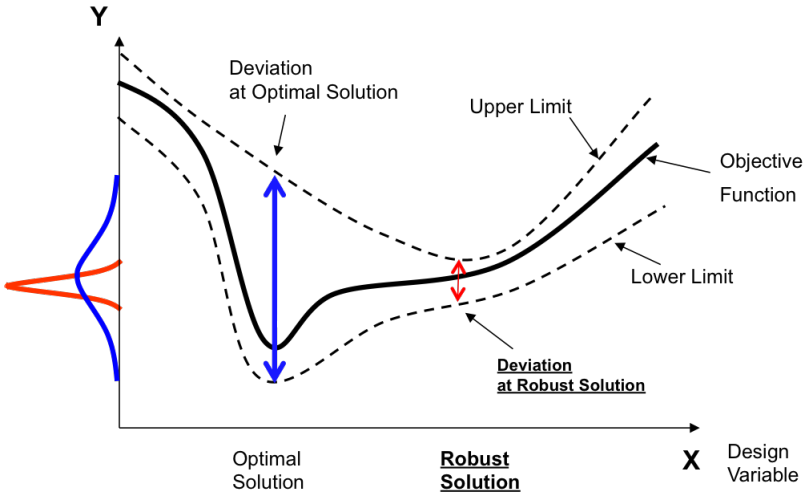


Figure 3.2: The figure shows variability in system responses from the optimal and robust solution. The solid curve represents a single deterministic function and the dashed line presents the upper and lower limit of the objective function due to uncertainty related to the system responses (type C). The robust solution is the solution that provides the best minimum value for both objective function and its variation. Figure was modified from [3].

3.3 Literature review

Several authors have assessed the key contributors to geological uncertainty and considered its impact on CO_2 storage. We present a few recent studies about geological parameters that are significant to subsurface CO_2 migration, followed by a review of optimisation techniques for CO_2 injection.

3.3.1 Sensitivity studies and uncertainty assessment related to CO_2 injection

Sensitivity analysis aims to identify the input-output relationship of a system by not necessarily quantifying uncertainty related to the input parameters and the model. However, it reveals which types of input parameters could affect the output of interest.

There are two generic types of geological parameters that are significant to CO_2 storage. Firstly, the parameters related to structural geology, i.e. the location and the shape of the reservoir and the bounding cap rock. Slope and rugosity of the top reservoir surface can have a significant impact on CO_2 migration and storage. Structure and stratigraphic studies related to CO_2 storage were explored in [4, 56]. The thickness and the permeability of a cap rock also play important roles on pressure management of the storage formation from the diffusion of pressure build-up [15].

Secondly, parameters related to petrophysical and fluid flow properties, such as porosity, permeability, residual gas saturation, capillary pressure, relative permeability hysteresis, and brine salinity, are parameters the migration of CO_2 is sensitive to [76, 89, 107, 155].

Other authors including [22, 63, 71] have documented a sensitivity analysis of CO_2 trapping to several injection designs, such as water-alternating-gas (WAG), types of injection wells, and injection rates.

Juanes et al [76] documented the impact of relative permeability based on a modified PUNQ case. By using numerical examples, they reported that relative permeability and related parameters could impact the CO_2 trapping process as follows. 1) The effect of hysteresis on CO_2 trapping: they considered the effect of relative permeability hysteresis and concluded that residual trapping occurred after the end of injection by imbibition. 2) The effect of injection rate: they showed that higher injection rate (also higher pressure) drove CO_2 into smaller pores as the capillary pressure of these smaller pores requires higher entry pressure. Thus, using a higher injection rate provides more residual trapping than a lower rate. 3) The effect of WAG and brine injection enhanced imbibition process by trapping non-wetting phase (CO_2) at the tail of the CO_2 plume. 4) The effect of BHP: despite advantages of using WAG to enhance CO_2 trapping from the imbibition process, the BHP of the injector was higher in the WAG process because the compressibility of injected water was smaller than CO_2. 5) Grid refinement; the authors suggested that using a coarse grid in the simulation model can overestimate the amount and extent of capillary trapping of CO_2.

The impact of capillary pressure and hysteresis was also documented by Mo and Akervoll [107] where they considered the impact of injection rate, reservoir layering, capillary pressure, and hysteresis on CO_2 distribution. They concluded that permeability was the main contributor that affects CO_2 migration, distribution, and pressure build-up.

Bryant et al [22] studied the effect of buoyancy-dominated flow under geological uncertainty where permeability, anisotropy, and dip angle of the aquifer model were considered. Without spatially correlated permeability, the CO_2 plume exhibited a fingering flow pattern with more effective trapping apparent in the correlated permeability case. Based on the effects of capillary and anisotropy, they observed that capillary pressure and anisotropy mainly controlled the CO_2 movement as high capillary entry pressure could prevent CO_2 movement vertically, leading to a vertical pressure barrier and smoothing out the fingering effect. The authors concluded that heterogeneity of the aquifer formation controlled the movement of the CO_2 plume, rather than the buoyancy-driven flow.

Ide et al [71] investigated the impact of gravitational, viscous, and capillary effects on the amount and rate of CO_2 residual trapping. Among other results, injection schemes, such as horizontal and vertical injection wells, and water alternate gas (WAG) were conducted using numerical simulation. They reported the following: 1) Greater immobilised CO_2 was observed when the flow was dominated by the viscous effects; however, the rate at which CO_2 became immobilised was greater in the case of gravity-dominated flow. The greater the lateral extension of the CO_2 plume, the higher the residual saturation. 2) Capillary pressure and the inclination of the saline aquifer could increase CO_2 residual trapping. 3) Injection rate and well placement could affect the amount and the rate of trapped CO_2. The authors reported that a vertical brine injector could enhance trapping capacity greater than a horizontal well if there was more space for CO_2 to migrate in the lateral direction.

Other reservoir properties, such as temperature, brine salinity, horizontal permeability,

residual gas saturation, dipping angle, reservoir pressure, cell block size, injection strategies, well completion, completion placement, and hysteresis effects, were analysed by Sifuentes et al [155] where the sensitivity of trapping mechanisms to reservoir parameters were investigated simultaneously through an experimental design approach. The reservoir model was based on Kezin in Germany while the fluid properties were obtained from Frio, Texas. The main findings were as follows: 1) Heterogeneity of reservoir increased solubility and residual trapping. 2) Imbibition including hysteresis contributed to residual trapping and needed to be included in the model. 3) Field strategies, such as well placements and well controls, were essential strategies to increase trapped CO_2.

Later, Lengler et al [89] documented the impact of geophysical heterogeneity on CO_2 distribution and breakthrough time. Spatial geostatistics of permeability, such as anisotropy ratios, correlation lengths, standard deviation, and upscaling techniques for permeability (i.e. harmonic and geometric average) were used to study the effect of CO_2 distributions. Among other results, they reported that: 1) Anisotropy ratio and correlation lengths of permeability led to a preferred flow path which resulted in a higher probability of early CO_2 breakthrough compared to the homogeneous case. 2) A small correlation length resulted in an increase in heterogeneity (spatial variability), which led to a slower mean arrival time of CO_2. 3) An increase in standard deviation (σ_{log_k}) showed a greater degree of heterogeneity in the storage formation. 4) The use of upscaling for permeability did not affect the CO_2 distribution and injection pressure.

For uncertainty analysis related to geological uncertainty, Senel and Chugunov [151] considered how uncertainty could be reduced when additional geophysical information becomes available. The authors showed three different levels of information in the model, and compared each scenario with the probability distribution (P10 and P90) of the mobile CO_2, the dissolved and residual trapping of CO_2, and the area flooded by the CO_2 plume. Their results presented the reduction of uncertainty when more information was acquired, as we expect.

Economic aspects related to geological uncertainty were documented by Heath et al [63]. The authors presented the relationship of injection cost per tonne of injected CO_2 with regard to geological heterogeneity. The authors proposed the use of a transfer function that relates heterogeneity to the uncertainty of injection cost and used a simple analytical function to quantify the uncertainty of injection cost caused by geological uncertainty of the model. They concluded that cost uncertainty is highly sensitive to geological uncertainty. This result also highlighted the impact of geological uncertainty to the economic designs of CO_2 injection.

In summary, most of the studies about uncertainty in literature showed that the prediction of CO_2 distribution and injection design is sensitive to geological uncertainty. Based on near-well properties, conceptual models and geological studies of surface analogues, it is difficult or impossible to know which geological realisation is more realistic. Therefore, it is essential to design robust injection strategies with regard to geological uncertainty. In the next section, we will review how optimisation techniques could be applied to enhance CO_2 trapping in aquifers.

3.3.2 A survey of CO_2 sequestration management using optimisation approaches

For oil and gas reservoirs, resources are often managed using production optimisation techniques to maximise performance criteria of a given field where some parameters such as well locations, well rates, types and trajectories of wells are manipulated under an optimisation framework. As reported in [19, 27, 29, 172, 174], optimisation approaches have been successfully applied in many oil and gas fields in recent decades, but these techniques have just become increasingly attractive in the past ten years for CO_2 sequestration, e.g. [24, 72, 73, 87, 109, 110, 152]. In the following, we will provide a brief literature review of CO_2 storage optimisation.

Kumar and Bryant [87] studied the effect of injection intervals on residual trapping. A semi-analytical algorithm was used for optimising the CO_2 completion depths. The perforation lengths and rates of vertical and horizontal wells were determined to maximise the amount of trapped CO_2 in a simple aquifer model. For a vertical well, they reported that completing the well in the bottom layer was the best strategy to enhance CO_2 trapping (both residual and dissolution) as it increased the travel distances between the injected CO_2 and the top-seal, and also increased the contact area between CO_2 and formation brine. However, this depends on the petrophysical properties of the formation near the injection interval and the injection rates used. For a horizontal well study, they observed that well length could impact trapping efficiency as CO_2 preferably travelled in a vertical direction rather than the lateral direction of the horizontal wells for a given injection rate, and hence it reduced overall trapping efficiency.

Qi et al [133] considered injecting brine together with CO_2 followed by chase brine using model data from a North Sea aquifer to increase storage security and capacity. Numerical studies based on streamline simulation were carried out using a one-dimensional and a synthetic three-dimensional model. They showed that using brine and chase brine with CO_2 injection reduced the mobility contrast between injected and displaced fluid, resulting in a uniform sweep pattern of CO_2. With respect to a fractional flow of injected CO_2, injecting a mixture of CO_2/brine followed by chase brine produced rapid trapping than injecting only the mixture of CO_2/brine since the mobility of chase brine was lower (and favourable) and moved rapidly to trap mobile CO_2. They concluded that the best injection strategy was to inject at a fractional flow of CO_2 between 85% to 100% followed by the injection of chase brine.

Later, Nghiem et al [109, 110] studied how brine injection scenarios could increase residual and solubility trapping. They maximised trapped CO_2 saturation by varying the depth of brine injection, brine injection rate, and brine injection period for both low and high permeability regions. Trapping efficiency index (TEI), shown below as a function of time, was used as the objective function in this study. Rather than mathematical optimisation, the authors chose to select variables from the pre-defined arbitrary finite set, and compared with

other selected variables. Their results therefore strictly depended on the chosen variables.

$$\text{TEI(t)} = \frac{\text{Total mass of } CO_2 \text{ trapped as residual gas(t)}}{\text{Total mass of } CO_2 \text{ injection(t)}} + \frac{\text{Total mass of } CO_2 \text{ dissolved in brine(t)}}{\text{Total mass of } CO_2 \text{ injection(t)}} \tag{3.1}$$

Based on their synthetic homogeneous permeability results, the authors reported the following: 1) The total amount of CO_2 trapping in high permeability zones was insensitive to brine injection; however, trapped CO_2 could be enhanced by injecting brine in the lower permeability region. 2) Brine injection near the CO_2 injection depth in the lower permeability region contributed to greater residual trapping, while injecting brine further from the depth of CO_2 injection favoured dissolution trapping in high permeability regions.

The idea of brine injection [110, 133] was developed further by Cameron and Durlofsky [24] under an optimisation framework. They considered the location of injectors and associated well configurations as well as the time and volumes of CO_2 injection to minimise the mobile phase of CO_2 in storage aquifers. Brine cycling and re-injecting were also considered: they reported that injection and re-injection of brine had an influential effect on trapping mechanisms and helped to minimise the mobile phase.

Similar to Cameron and Durlofsky [24], Goda and Sato [57] studied only locations and rates of injectors in relation to both immobile and dissolved phases of CO_2 in a realistic field case. The authors used a global optimisation technique for placing injectors in the Nagaoka pilot field, sandstone formation in Japan. Three vertical wells associated with injection rates were considered in this work to minimise a proportion of mobile CO_2, explicitly given as $f = 1 - \text{TEI(t)}$. From the study, they showed that optimising the well location of injectors and the injection rates could improve residual and dissolution trapping of CO_2.

As well as minimising the mobile phase of CO_2, Shamshiri and Jafarpour [152] designed injection strategies with two different objective functions: sweep efficiency, and total mass trapped in the reservoir, and studied the effect of geological uncertainty on the control strategies. Also, they proposed techniques to reduce the risk of CO_2 leakage by manipulating the injection rate. The main findings of this study were: 1) Maximising the mass of trapped CO_2 provided better trapping capacity than the optimisation of sweep efficiency; their hypothesis was that the optimisation of trapped CO_2 targeted directly to a higher storage region (higher porosity) whereas the sweep efficiency technique provided a uniform distribution and did not target to the high porosity region. 2) Using optimisation over an ensemble of geological realisations provided better overall performance than using a single realization for optimisation. 3) The objective function could be further modified to minimise CO_2 saturation in risky zones such as active faults and abandoned wells, which showed that it could be done under the optimisation framework.

Jahangiri and Zhang [72] introduced a co-optimisation method which maximised oil recovery and CO_2 storage simultaneously. The revenue from oil production and the amount of cumulative CO_2 stored were maximised by manipulating the BHP of producers. They studied various injection strategies, such as continuous CO_2 injection, water-alternating-gas (WAG), and gas injection after waterflood with miscible and immiscible cases. They reported the following: 1) Using the WAG process provided the greatest NPV value than

using pure CO_2 injection in immiscible conditions for a given tax credit at \$40 dollar per tonne of CO_2 stored. 2) The revenues from oil production dominated the carbon credits from CO_2 injection in miscible conditions. Therefore, the benefit of CO_2-EOR was insensitive to carbon credits gained from CO_2 sequestration under miscible conditions. In contrast, it was sensitive to the carbon credits obtained in the immiscible flood case. 3) Optimisation results for both miscible and immiscible cases provided significant improvement of both NPV and cumulative oil production, in comparison to the base case.

Later, Jahangiri and Zhang [73] extended their previous work to incorporate an ensemble-based technique (EnOpt) in an optimisation algorithm, which allowed them to incorporate dynamic injection of CO_2 and brine. Three scenarios were tested by comparing the optimisation results with equally distributed rates of CO_2 injection. They considered the following: 1) Pure CO_2 injection constrained by the total of CO_2 injection rate for both miscible and immiscible cases. 2) CO_2 and brine injection constrained by the total liquid injection in the immiscible case. 3) CO_2 and brine injection constrained by total CO_2 injection in the immiscible case. In the first case, there were two findings. Firstly, they observed that the optimal CO_2 injection rate is a pulse-shaped, which explains why the interruption of continuous rates could dominate the fluid flow process behind the flood front and help to improve sweep efficiency. Secondly, to delay CO_2 breakthrough, a higher injection rate was used in the lower permeability region while the lower rate was used in the higher permeability region in order to achieve a uniform sweep pattern. In the second and third cases, the authors applied a level set function to switch the injection phase from injecting CO_2 to brine or vice versa. They reported that injecting brine and CO_2 simultaneously provided better mobility control of the slug brine and CO_2 than injecting only pure CO_2. However, the authors did not consider hysteresis effect. Thus, this practice is no doubt leading to a poor performance of the total amount of the stored CO_2 compared to the base case and the pure CO_2 injection case.

Mathematical optimisation methods such as gradient-based [72, 73], direct search [24], and global optimisation algorithms [57] have been used in the previous studies, where the authors showed that the optimal CO_2 operating strategies could increase storage capacities and, in some cases, simultaneously reduce the risk of CO_2 leakage. However, the choice of optimisation methods usually depends on the problems and the objective of optimisation. Even though the classical gradient-based optimisation methods such as steepest descent, conjugate gradient [66], and quasi-newton methods are powerful and effective algorithms, they may not be cost-effective methods to find a global optimal solution in this thesis. As function evaluations of CO_2 flow equations are computationally expensive, finding the global solutions from the gradient-based approach may requires multiple starting points and forward runs. Though several global optimisation algorithms are also computationally expensive, they can be coupled with a surrogate approach to obtain a global (a better local) optimal solution. Additionally, the characteristics of objective function are often not known a priori and, in most cases, the gradient information is not available for the gradient-based algorithms. In contrast, we propose the use of a surrogate model to assist the optimisation process because 1) a surrogate model can provide information about the objective function characteristics (the nature of input-output relationship, nonlinearity,

nonconvexity, multimodal, etc.). This can be use for global sensitivity analysis [160] and uncertainty assessment. 2) It can be used to assist global optimisation algorithms to improve computational efficiency of the optimisation process. In the next chapter, we will discuss the use of a surrogate to assist optimisation process in application to CO_2 sequestration.

CHAPTER 4

SURROGATE-ASSISTED OPTIMISATION

4.1 Introduction

This chapter presents mathematical descriptions for building an approximate model for CO_2 injection in saline aquifers based on few simulation runs. This approximate model will be used to increase the efficiency of the optimisation algorithms used to find optimal CO_2 injection rates. The chapter begins with terminologies and definitions of polynomial interpolation and error analysis of the interpolation. Later, different forms of basis functions used for univariate interpolation are presented, followed by a tensor product approach for multivariate interpolation cases. Last, we present a sparse grid interpolation approach and its dimensional adaptivity, which will be used to assist decision making for CO_2 sequestration.

4.1.1 A surrogate model

In many engineering applications, it is often difficult to obtain the input-output relationship (IO) of the system of interest when the outcome of the system is too expensive to obtain numerically or experimentally. Therefore, a surrogate function, a function used to approximate the IO of a computationally expensive system, can be used. This approximate function is sometimes referred to as a proxy, a response surface, a meta-model, an emulator, or a surrogate model; these terms have been used interchangeably in different engineering fields. The main principle of the surrogate model is to mimic the behaviour of the model output by constructing a cheap-to-compute function that approximates a time-consuming or an expensive-to-compute function. For robust optimisation of CO_2 storage problems, the optimisation procedure usually requires many objective function evaluations, each of which corresponds to a full forward run of a compositional simulator that is computationally expensive. To reduce the computational burden, a surrogate model is used to accelerate the solution process for the expensive optimisation problems.

In recent years the surrogate approach has been successfully applied to engineering problems by several researchers, e.g. [47, 84, 153]. The underlying principle of the surrogate model originates from approximation theory from which it is characterised into two main areas: 1) a selection of basis functions for the purpose of interpolation; and 2) a sampling strategy technique for building the interpolating function. In this study, we focus on the use of polynomial basis functions and a Smolyak [159] sampling strategy for building the surrogate model.

4.1.2 Polynomial interpolation

Polynomials can be used as a surrogate model to mimic the IO of the system by using the interpolation of a given data set. There are many types of polynomials that can be used but we present polynomials that are related to this study, i.e. Lagrange polynomials and a Barycentric Lagrange representation, in the following.

Let x be an input variable and $f(x)$ be a continuous function that represents the IO of the system — the objective function of CO_2 sequestration in this study, and let $p(x)$ be an interpolating polynomial function that approximates $f(x)$. Then the polynomial form of the 1-dimensional polynomial function is:

$$f(x) \approx p(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \ldots + a_{n-1} x^{n-1} \tag{4.1}$$

where $a_0, a_1, \ldots, a_{n-1}$ present coefficients of the polynomial function passing through n data points represented as $x_1, x_2 \ldots, x_n$. If j presents a data point, the coefficients can be determined from the following condition:

$$f(x_j) = p(x_j) \text{ , for } \forall j \in \{1, \ldots, n\} \tag{4.2}$$

The condition above is a system of linear equations where the coefficients $a_0, a_1, \ldots, a_{n-1}$ are determined from the solution of the linear system.

$$\begin{pmatrix} 1 & x_1 & \cdots & x_1^{n-2} & x_1^{n-1} \\ 1 & x_2 & \cdots & x_2^{n-2} & x_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots & \\ 1 & x_n & \cdots & x_n^{n-2} & x_n^{n-1} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-1} \end{pmatrix} = \begin{pmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{pmatrix} \tag{4.3}$$

The matrix on the left-hand side of the equation is referred to as the Vandermonde matrix [106]. The linear solution is unique if the Vandermonde matrix is non-singular where the determinant given below is non-zero.

$$\det(V) = \prod_{1 \leq i < k \leq n} (x_k - x_i) \tag{4.4}$$

where V presents the Vandermonde matrix. If x_j for $\forall j \in \{1, \ldots, n\}$ are distinct, the determinant of the matrix V is non-zero and hence the solution of Eq. 4.3 is unique. However, for many data points, the size of the Vandermonde matrix becomes large such that solving the linear equations can be a cumbersome task. A large matrix can also suffer from ill-conditioning meaning that a small variation in the matrix causes considerable variation of the solution. To avoid this, we introduce an alternative polynomial function based on the Lagrange basis, which does not require matrix inversion for finding coefficients of the polynomial.

4.1.3 Lagrange polynomial interpolation

The 1-dimensional polynomial function $p(x)$ of degree $n-1$ can be expressed in Lagrange form as:

$$p(x) = \sum_{i=1}^{n} \ell_i(x) f(x_i) \tag{4.5}$$

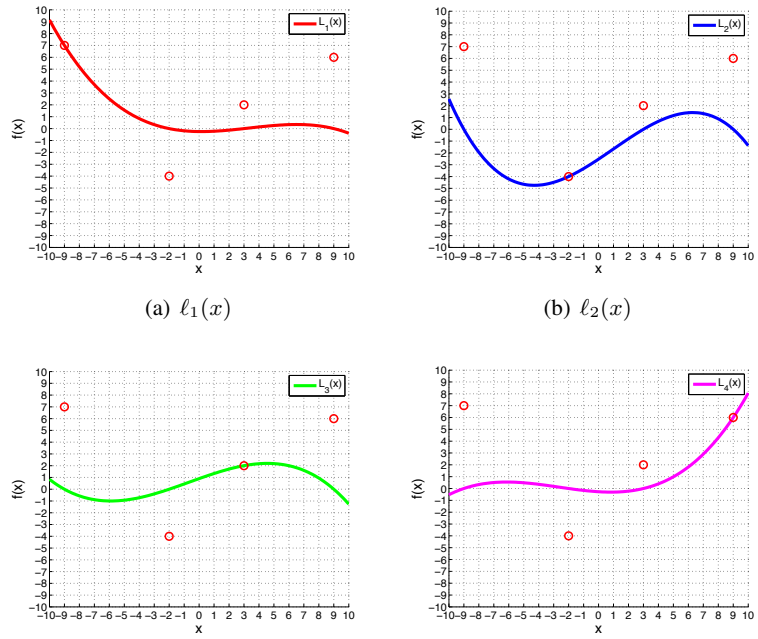
where the Lagrange basis $\ell_i(x)$ is:

$$\ell_i(x) = \prod_{\substack{j=1 \\ j \neq i}}^{n} \frac{x - x_j}{x_i - x_j}, \qquad \text{for } i = 1, 2, \ldots, n \tag{4.6}$$

which satisfies the following condition.

$$\ell_i(x_j) = \delta_{ij} \ , \ \ 1 \leq i, j \leq n \tag{4.7}$$

where δ_{ij} is the Kronecker delta such that $\ell_i(x_j) = 1$ for $i = j$, and $\ell_i(x_j) = 0$ for $i \neq j$. With this condition, the Lagrange polynomial $p(x)$ is exact at given points $\{x_i\}_1^n$. Additionally, the coefficients from the Lagrange representation do not require solving a linear system of equations (Eq. 4.3) since they are calculated directly from the data points. The Lagrange polynomial is illustrated in Fig. 4.1 for four data points (x_1, x_2, x_3, x_4 from left to right) where the circle markers represent the function $f(x_i)$ on the y-axis and the corresponding point x_i in the x-axis. We see from Fig. 4.1 that the Lagrange basis function $\ell_i(x)$ passes through the function value $f(x_i)$ and the corresponding point x_i. The Lagrange polynomial, the black curve in Fig. 4.2, is obtained from the sum of each Lagrange basis curve weighted by its corresponding function value.



(a) $\ell_1(x)$ (b) $\ell_2(x)$

(c) $\ell_3(x)$ (d) $\ell_4(x)$

Figure 4.1: Illustration of Lagrange basis function for four different data points. The circle markers denote the function $f(x_j)$ on the y-axis and the corresponding point x_j in the x-axis.

Using Lagrange polynomials is a straightforward method to represent the polynomials; however, there are some drawbacks in terms of computational efficiency, especially when there are many data points. These drawbacks are:

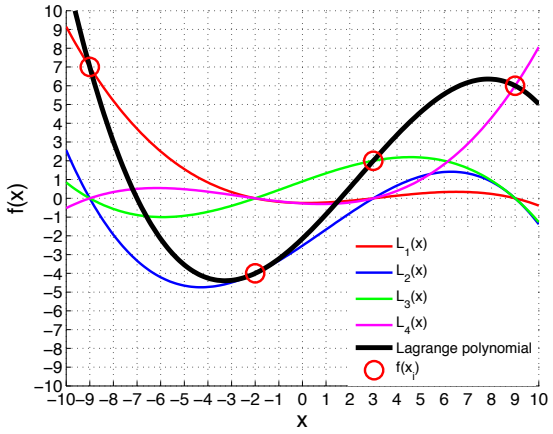


Figure 4.2: Illustration of Lagrange polynomial interpolation for four given data points, x_1, x_2, x_3, x_4 from left to right. The circle markers denote the function $f(x_i)$ on the y-axis and the corresponding point x_i on the x-axis.

1. Each evaluation of the Lagrange basis function $\ell_i(x)$ requires $\mathcal{O}(n)$ operations. The $\mathcal{O}(n)$ denotes an expression bounded by cn where $c > 0$ and n approaches ∞. Then, evaluating a point x in $p(x)$ requires $\mathcal{O}(n^2)$ operations from the summation and multiplication.

2. The Lagrange basis function is dependent on data points $\{x\}_i^n$. Therefore, having a new set of data, for example x_{n+1} and $f(x_{n+1})$, requires a new calculation of Lagrange basis functions.

The Lagrange form can be arranged in a Barycentric polynomial representation [147] to mitigate the drawbacks described above to enhance computational efficiency of the Lagrange polynomial representation.

4.1.4 Barycentric Lagrange interpolation

The main idea behind the Barycentric representation [147] is to make the Lagrange basis function independent from any interpolated point x such that the polynomials can be evaluated with $\mathcal{O}(n)$ operations. The Lagrange basis $\ell_i(x)$ is rewritten as:

$$\ell_i(x) = \overbrace{\prod_{j=1}^{n}(x - x_j)}^{\ell(x)} \frac{1}{x - x_i} \overbrace{\frac{1}{\prod\limits_{\substack{j=1 \\ j\neq i}}^{n}(x_i - x_j)}}^{w_i} \tag{4.8}$$

where $\ell(x)$ is a nodal basis function defined as

$$\ell(x) = \prod_{j=1}^{n}(x - x_j) \tag{4.9}$$

The barycentric weight in the following is not dependent on any data point x_i; the weights are calculated at once and can be reused many times during the interpolation process.

$$w_i = \frac{1}{\prod\limits_{\substack{j=1 \\ j \neq i}}^{n} (x_i - x_j)} \tag{4.10}$$

Then, the Barycentric representation of Lagrange interpolation is expressed [13, 62] as

$$p(x) = \ell(x) \sum_{i=1}^{n} \frac{w_i}{x - x_i} f(x_i) \tag{4.11}$$

Equation 4.11 is the first form of Barycentric representation. It should be noted that this Barycentric form is not defined for $x = x_i$; however, the function value of any given point x_i is obtained from the condition (Eq. 4.2) that $f(x_i) = p(x_i)$. The Barycentric representation can be simplified to the second form by making the polynomial be exact for the constant function $f(x) = 1$, as shown in the following.

$$1 = \ell(x) \sum_{i=1}^{n} \frac{w_i}{x - x_i} \tag{4.12}$$

$$\ell(x) = \frac{1}{\sum\limits_{i=1}^{n} \frac{w_i}{x - x_i}} \tag{4.13}$$

Substituting the nodal basis function $\ell(x)$ from Eq. 4.13 into Eq. 4.11 yields the second form of Barycentric Lagrange polynomial.

$$p(x) = \frac{\sum\limits_{i=1}^{n} \frac{w_i}{x - x_i} f(x_i)}{\sum\limits_{i=1}^{n} \frac{w_i}{x - x_i}} \tag{4.14}$$

The improved formulae of the Barycentric representation have several advantages in comparison to the Lagrange form as:

1. Even though the Barycentric form requires $\mathcal{O}(n^2)$ operations to calculate w_i, the main advantage is that each evaluation of any interpolating point x only needs $\mathcal{O}(n)$ operations given that the Barycentric weights (Eq. 4.10) are known. This makes function evaluations using the Barycentric form more efficient.

2. The second form of the Barycentric representation allows a common term c in both nominator and denominator terms to cancel out each other. Additionally, Higham [67] showed that the second form is a stable form of the Barycentric Lagrange polynomial. In this study, we use the second form of the Barycentric representation for the surrogate.

4.2 Basis functions in one-dimensional interpolation

We first define the terminology that will be used later to describe a sparse grid interpolation approach. One-dimensional piecewise functions with equidistant points are used in this section to illustrate two types of basis functions: nodal and hierarchical basis functions. Then,

the terminology of univariate interpolation will be extended to multivariate interpolation in the next section.

4.2.1 Definitions

We use the same notation as described in [51, 80, 126] in this section. Let $l \in \mathbb{N}$ be a level of interpolation, and let X_l be a set of grid nodes in the interval $[0, 1]$, defined as:

$$X_l := \{x_{l,0}, x_{l,1}, \ldots, x_{l,2^l}\} = \{x_{l,j_l}\}_{j_l=0}^{j_l=2^l} \tag{4.15}$$

where the equispaced points X_l have a mesh size $h_l := 2^{-l}$ so that $x_{l,j_l} = j_l h_l$ where j_l presents the index, $0 \leq j_l \leq 2^l$ where $j_l \in \mathbb{N}$. We consider the following piecewise linear hat function for a basis function.

$$\phi(x) = \max(1 - |x|, 0) \tag{4.16}$$

4.2.2 Nodal basis function

According to Eq. 4.16, we define nodal basis functions in the following.

$$\phi_{l,j}(x) := \phi(2^l x - j) \tag{4.17}$$

From Eq. 4.17, the basis function $\phi_{l,j}(x)$ is equal to 1 at point $x_{l,j}$ and zeros at other points. Then, we can define an associated space V_l of the nodal basis functions.

$$V_l := \text{span}\{\phi_{l,j_l} : 0 \leq j_l \leq 2^l\} \tag{4.18}$$

4.2.3 Hierarchical basis functions

The basic idea behind hierarchical basis functions is that the structure of grid points at a higher level of interpolation is arranged in a way that the grid points from the lower level are used. In other words, the grid points at the lower level is a subset of the higher one, $X_l \subset X_{l+1}$. We define a hierarchical subspace $\phi(x)$ as:

$$W_l := \text{span}\{\phi_{l,j_l} : j_l \in B_l\} \tag{4.19}$$

where a set of index B_l is given below:

$$B_l := \left\{ j_l \in \mathbb{N} \;\middle|\; \begin{array}{ll} j_l = 1, \ldots, 2^l - 1, \text{ if } l > 0 & \text{if } l > 0 \text{ and } j_l \text{ is odd} \\ j_l = 0, 1 & \text{if } l = 0 \end{array} \right. \tag{4.20}$$

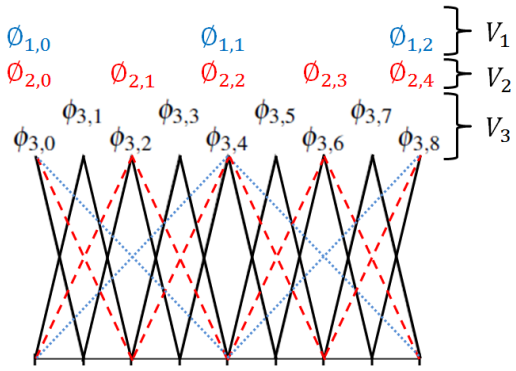
Then we can write the hierarchical subspace W_l in terms of V_l as follows:

$$W_l := V_l \backslash V_{l-1} \tag{4.21}$$

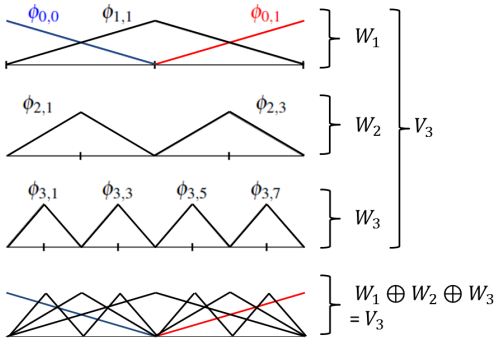
The hierarchical subspace W_l for a level l is the difference of grid nodes between V_l and V_{l-1} and $W_0 = V_0$. The space of basis functions V_l can be written as a direct sum of the hierarchical subspaces in the following form.

$$V_l := \bigoplus_{l'=0}^{l'=l} W_{l'} = V_{l-1} \bigoplus W_l \tag{4.22}$$

where the operator $\bigoplus$ denotes a direct sum of the hierarchical subspaces from each level. A comparison of the nodal basis and the hierarchical basis for 3 levels of interpolation is shown in Fig. 4.3. The nodal basis function in Fig. 4.3(a) is displayed for level 1 (the dotted line), level 2 (the dashed line) and level 3 (the solid line); we see that some points at the higher level are the same as the points in the lower level. For example, the peak of the basis $\phi_{2,0}, \phi_{2,2}$, and $\phi_{2,4}$ have the same location as the basis $\phi_{1,0}, \phi_{1,1}$, and $\phi_{1,2}$. This means that using a nodal basis for 3 levels requires 9 evaluating points, but only the points $\{\phi_{3,1}, \phi_{3,3}, \phi_{3,5}, \phi_{3,7}\}$ provide new information. In contrast, a hierarchical basis function in Fig. 4.3(b) is a direct sum of the hierarchical subspace W_l so that the higher level always reuses points from the lower one. Thus, V_3 is a sum of $W_1 \bigoplus W_2 \bigoplus W_3$, which gives the same points as the nodal bases in Fig. 4.3(a) but only additional 4 evaluation points from W_3 is needed because $V_3 = V_2 \bigoplus W_3$. The underlying principle of the sparse grid interpolation is based on the hierarchical basis function.



(a) Nodal basis for V_3



(b) Hierarchical basis for V_3

Figure 4.3: Nodal basis (a) and hierarchical basis (b) using the hat function (Eq. 4.16) for 3 levels of interpolation.

Figure 4.4 illustrates the interpolation of a function $f(x)$ by using nodal basis functions as shown in Fig. 4.3(a). On the left figure of Fig. 4.4, the function $f(x)$ (the solid curve) and the interpolating polynomial (the dashed curve) is displayed, while the nodal basis functions used to build the interpolation is shown on the right figure. We see that the interpolating polynomial $p(x)$ is the sum of a function value weighted by the corresponding nodal basis function (the hat function in this case).

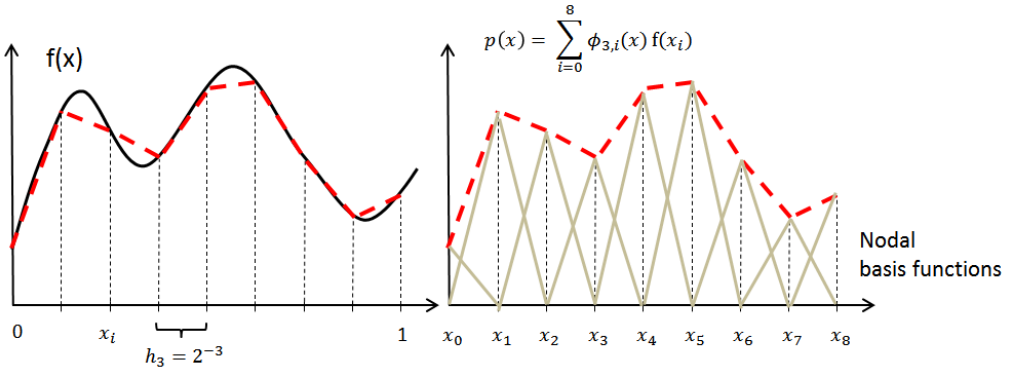


Figure 4.4: Illustration of nodal basis functions for the interpolation of an arbitrary function $f(x)$. The left figure presents the function $f(x)$ (the solid curve) and the interpolating polynomial (the dashed curve), while the right figure displays the interpolating polynomial built from the nodal basis function.

For the interpolation of the same function $f(x)$ using the hierarchical basis, Fig. 4.5 displays the interpolating polynomial (the dashed curve) built from different levels of the hierarchical basis. As can be seen on the left figure, the interpolating polynomial is built by adding the basis functions at the higher level to the basis of the lower one. Thus, the interpolating polynomial is obtained by the sum of hierarchical basis weighted by coefficients $w_{l,j}$ for the level l and index j. The coefficients $w_{l,j}$, referred to as hierarchical surpluses, are the difference between the function value $f(x)$ and the basis function at a point $x_{l,j}$. A set of points $x_{l,j}$ used to evaluate the hierarchical surpluses $w_{l,j}$ is defined as:

$$X_{\triangle}^{l} := \{x_{l,j} \in X_l \backslash X_{l-1} : j_l \in B_l\} \tag{4.23}$$

where $X_l \backslash X_{l-1}$ denotes the set X_l that excludes the set X_{l-1}. This set implies that only points at the current level are used to determine the hierarchical surpluses. Hence, using the hierarchical basis requires fewer function evaluations than using the nodal basis function for building the interpolant. On the right figure of Fig. 4.5, we present different basis functions the height of which is the hierarchical surplus $w_{l,j}$. It should be noted that the value of the hierarchical surplus can be either negative or positive. Later in this chapter, we will present the use of hierarchical surpluses for dimensional adaptivity of the sparse grid interpolation approach.

4.3 Tensor product grid and multivariate interpolation

The definitions of univariate interpolation given in Section 4.2 can be extended to the multivariate case from which the level l and mesh size h_l are expressed in d-dimensional multi-indices as $\underline{l} := (l_1, \ldots, l_n)$ and $\underline{h} := (h_{l_1}, \ldots, h_{l_d})$, respectively. For multivariate interpolation, the set of indices in the one-dimensional case (Eq. 4.20) becomes the set of multi-indices $B_{\underline{l}}$ where $\underline{j} \in B_{\underline{l}}$; $\underline{j} := (j_1, \ldots, j_d)$; and $j_t = 0, 1, \ldots, 2^{l_t}$. The basis functions $\phi_{l,j}$

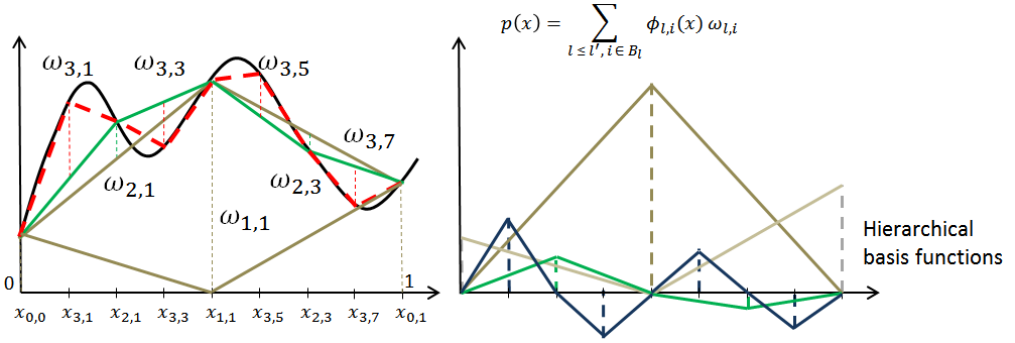


Figure 4.5: Illustration of hierarchical basis functions for the interpolation of an arbitrary function $f(x)$. The left figure presents the function $f(x)$ (the solid curve), the interpolating polynomial (the dashed curve), and the hierarchical basis functions; the difference between $f(x)$ and the basis function is denoted by $w_{l,j}$ for level l and index j. The figure on the right presents different levels of hierarchical basis functions used to build the interpolating polynomial.

in d-dimension are expressed in a tensor product form:

$$\phi_{\underline{l},\underline{j}} := \prod_{t=1}^{d} \phi_{l_t,j_t}(x_t) \tag{4.24}$$

from which the subspace $W_{\underline{l}}$ of multivariate piecewise functions for level $\underline{l}$ is given in the following:

$$W_{\underline{l}} := \text{span}\{\phi_{\underline{l},\underline{j}} : \underline{j} \in B_{\underline{l}}\} \tag{4.25}$$

We recall from Eq. 4.21 that W_l is the hierarchical subspace in one-dimension, which is expressed in d-dimensions as [51]:

$$W_{\underline{l}} := V_{\underline{l}} \setminus \bigoplus_{t=1}^{d} V_{\underline{l}-\underline{e}_t} \tag{4.26}$$

where $\underline{e}_t$ is the t-unit vector; for example, if $d = 2$ and $\underline{l} = (l_1 = 3, l_2 = 2)$, then $W_{\underline{l}} = V_{(3,2)} \setminus (V_{(2,2)} \bigoplus V_{(3,1)})$ since the unit vectors $\underline{e}_1 = (1, 0)$ and $\underline{e}_2 = (0, 1)$. As shown previously in Eq. 4.22 for one-dimension, the direct sum of the hierarchical subspace $W_{\underline{l}}$ is the space V_n, defined in the following:

$$V_n := \bigoplus_{l_1=0}^{n} \cdots \bigoplus_{l_d=0}^{n} W_{\underline{l}} = \bigoplus_{|\underline{l}|_\infty \le n} W_{\underline{l}} \tag{4.27}$$

where n is the level of interpolation and $|\underline{l}|_\infty$ is the L_∞-norm of $\underline{l}$, defined as:

$$|\underline{l}|_\infty := \max_{1 \le t \le d} l_t \tag{4.28}$$

The space V_n given in Eq. 4.27 is a foundation of a tensor product or a full grid approach. We use the hat function (Eq. 4.20) with equispaced points (Eq. 4.15) without the boundary to illustrate the 2-dimensional tensor product by showing each level of the hierarchical subspace $W_{(l_1,l_2)}$. A layout of points for the level 3 of tensor product interpolation is shown in Fig. 4.6 where each sub-figure $W_{(l_1,l_2)}$ is displayed with corresponding points. The result

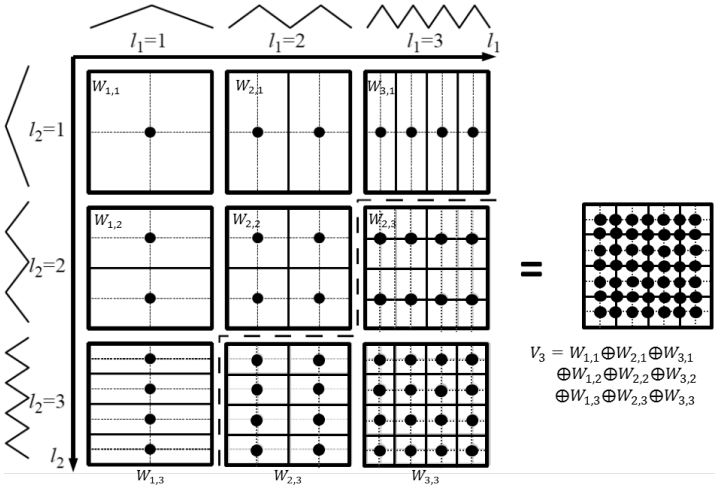


Figure 4.6: Illustration of the tensor product; each sub-figure displays the layout of points with the hierarchical subspace $W_{(l_1,l_2)}$ of the hat basis functions from which the level l is in an ascending order from left to right and from top to bottom. The result of the tensor product is shown on the right of the figure where it is a direct sum of the subspace W_{l_1,l_2}.

of the tensor product is shown in the figure on the right. The tensor product are obtained from a direct sum of W_{l_1,l_2}, which is a combination of all sub-figures in Fig. 4.6.

It should be noted that we use Chebyshev-based grid points for building the interpolant since this grid pattern mitigates the effect of Runge phenomenon [142] (more detailed discussion about error analysis of Lagrange polynomial interpolation is given in appendix A). However, we see in Fig. 4.6 that the tensor product approach requires a large number of evaluation points from a direct sum of each level. Therefore, building the interpolant from the tensor product approach can be computationally expensive. For this reason, Smolyak [159] proposed a combination strategy in such a way that the error of the interpolant built from this method is close to the error from the tensor product approach but with significantly fewer points. The Smolyak algorithm is the principle behind a sparse grid interpolation approach.

4.4 Sparse grid

The underlying principle of the sparse grid method originates from the Smolyak algorithm [159] which is used to extend univariate interpolation formulas to the multivariate interpolation by choosing a set of points in a lower order product grid. In other words, a sparse grid approach is a linear combination of lower-order tensor products, which results in fewer grid sampling points than the full tensor product interpolation. Sparse grid methods were first introduced for numerical integration and interpolation problems in [9, 54] and were further developed by several authors [51, 55, 58, 80, 126].

In this section, we first present the sparse grid approach in terms of the hierarchical subspace $W_{\underline{l}}$ and the space V_n, followed by the description of the Smolyak algorithm and hierarchical surpluses. Later, we demonstrate the algorithm with examples. Last, the description

of Chebyshev-based grid points and dimensional adaptivity of the sparse grid method are presented.

4.4.1 Hierarchical basis functions

Let us define the L_1-norm as $|\underline{l}| := \sum_{t=1}^{d} l_t$, then the sparse grid is an arrangement of the grid such that the space V_n is defined in the following form:

$$V_n := \bigoplus_{|\underline{l}| \leq n+d-1} W_{\underline{l}} \quad \text{for } n \geq 1 \ , \ l_t \geq \ 0 \tag{4.29}$$

where n denotes the level of sparse grid interpolation, and the condition $|\underline{l}| \leq n + d - 1$ provides a combination rule such that only the points in the upper-left corner triangle of the full tensor product are used. This combination of the hierarchical subspace $W_{\underline{l}}$ shown in Fig. 4.7 results in the sparse grid patten.

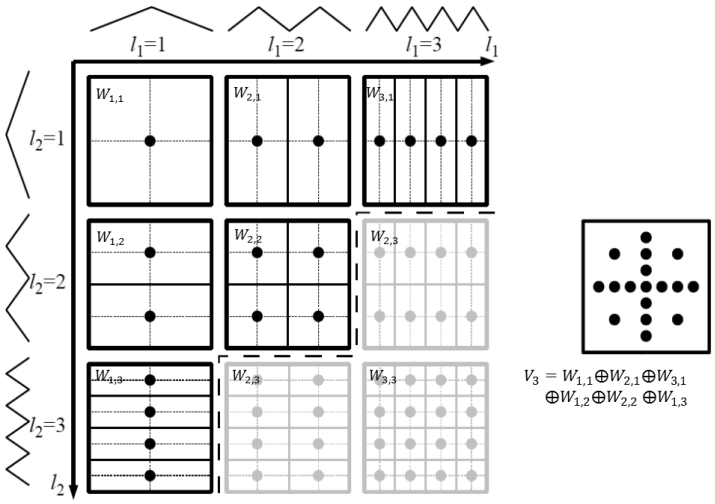


Figure 4.7: Illustration of the sparse grid approach; each sub-figure displays the layout of points with the hierarchical subspace $W_{(l_1,l_2)}$ of the hat basis functions from which the level l is in an ascending order from left to right and from top to bottom. The result of the sparse grid is shown on the right of the figure where it is a direct sum of the subspace W_{l_1,l_2} using the condition $|\underline{l}| \leq n + d - 1$, from [126].

We recall that the approximation of smooth functions $f : [0, 1]^d \to \mathbb{R}$ with a finite number of grid nodes X^l for one-dimension is expressed in Lagrange form as:

$$\mathcal{Q}^l(f) = \sum_{x_l \in X^l} \phi_{x_l} \cdot f(x_l) \tag{4.30}$$

where $l \in \mathbb{N}$ is the level of interpolation; $\mathcal{Q}^l(f)$ is the interpolant of $f(\cdot)$; the set of support grids in one dimension is $X^l = \{x_l \mid x_l \in [0, 1]\}$; the basis functions ϕ_{x^l} follow Lagrange characteristic (Eq. 4.7). For univariate hierarchical functions, we define $\triangle^l(f) \in W_l$ for the difference between the interpolant at subsequent levels of interpolation.

$$\triangle^l(f) := \mathcal{Q}^l(f) - \mathcal{Q}^{l-1}(f) \tag{4.31}$$

Since the interpolant $\mathcal{Q}^l$ is built based on the previous level $\mathcal{Q}^{l-1}$, the interpolant $\mathcal{Q}^l(\mathcal{Q}^{l-1}(f))$ is an exact interpolant $\mathcal{Q}^{l-1}(f)$ [80]. Then, we can write the following relation.

$$\mathcal{Q}^{l-1}(f) = \mathcal{Q}^l(\mathcal{Q}^{l-1}(f)) \tag{4.32}$$

Substitute Eqs. 4.30 and 4.32 to Eq. 4.31, yields the following $\triangle^l(f)$ as:

$$\begin{aligned} \triangle^l(f) &= \sum_{x_l \in X^l} \phi_{x_l} \cdot f(x_l) - \sum_{x_l \in X^l} \phi_{x_l} \cdot \mathcal{Q}^{l-1}(f)(x_l) \\ &= \sum_{x_l \in X^l} \phi_{x_l} [f(x_l) - \mathcal{Q}^{l-1}(f)(x_l)] \end{aligned} \tag{4.33}$$

where the points $x_l \in X^l$ are hierarchical points which satisfy Eq. 4.23 and we know that $f(x_{l,j}) - \mathcal{Q}^{l-1}(f)(x_{l,j}) = 0$ for $\forall x_{l,j} \in X^{l-1}$. If we define $m^l_\triangle$ as the cardinality of the set $X^l_\triangle$, then the term $\triangle^l(f)$ is given below.

$$\triangle^l(f) = \sum_{j=1}^{m^l_\triangle} \phi_{l,j} \underbrace{[f(x_{l,j}) - \mathcal{Q}^{l-1}(f)(x_{l,j})]}_{w^l_j} \tag{4.34}$$

where w^l_j is a hierarchical surplus which is the difference between function value $f(x)$ and the interpolant $Q^{l-1}(x)$ as demonstrated graphically in Fig. 4.5. For multivariate cases, the tensor product of Eq. 4.34 yields

$$\begin{aligned} (\triangle^{l_1} \otimes \cdots \otimes \triangle^{l_d})(f) &= \sum_{\underline{j}} \phi^{\underline{l}}_{\underline{j}} \cdot w^{\underline{l}}_{\underline{j}} \\ \text{where} \quad \phi^{\underline{l}}_{\underline{j}} &= (\phi_{l_1,j_1} \otimes \cdots \otimes \phi_{l_d,j_d}) \\ w^{\underline{l}}_{\underline{j}} &= f(x_{l_1,j_1}, \ldots, x_{l_d,j_d}) - \mathcal{Q}^{l-1}(x_{l_1,j_1}, \ldots, x_{l_d,j_d}) \end{aligned} \tag{4.35}$$

where $\phi^{\underline{l}}_{\underline{j}}$ and $w^{\underline{l}}_{\underline{j}}$ are basis functions and hierarchical surpluses for multivariate cases. This equation will be used for the Smolyak algorithm in the next section.

4.4.2 Smolyak algorithm

Smolyak proposed a sparse tensor product that reduces the number of support grids while preserving the asymptotic error decay of the full grid approach [51, 126, 157]. There are two approaches: 1) superposition of low-order tensor products, referred to as the combination technique [60], and 2) superposition of hierarchical subspaces $W_{\underline{l}}$. In this work, we use the later approach which utilises the subspace $W_{\underline{l}}$ to formulate the sparse grid [54, 80]. Given a Smolyak parameter q and the number of dimensions d, the Smolyak formulation [80] is given by

$$\begin{aligned} \mathcal{Q}_{q,d}(f) &= \sum_{|\underline{l}| \leq q} (\triangle^{l_1} \otimes \cdots \otimes \triangle^{l_d})(f) \\ &= \sum_{|\underline{l}| \leq q-1} (\triangle^{l_1} \otimes \cdots \otimes \triangle^{l_d})(f) + \sum_{|\underline{l}| = q} (\triangle^{l_1} \otimes \cdots \otimes \triangle^{l_d})(f) \\ &= \mathcal{Q}_{q-1,d}(f) + \sum_{|\underline{l}| = q} \sum_{\underline{j}} \phi^{\underline{l}}_{\underline{j}} \cdot w^{\underline{l}}_{\underline{j}} \end{aligned} \tag{4.36}$$

where $\mathcal{Q}_{q,d}(f) \in V_{(q+1-d)}$ denotes the interpolation of function $f(\cdot)$ and $\mathcal{Q}_{d-1,d}(f) = 0$; the Smolyak parameter q satisfies the conditions, that are $q \geq d$ and $q = n + d - 1$, $q \in \mathbb{N}$. It should be noted that only the hierarchical surpluses $w_{\underline{j}}^{\underline{l}}$ of the last term in Eq. 4.36 is calculated for the current level l of the interpolation. Thus, the hierarchical surplus $w_{\underline{j}}^{\underline{l}}$ is used to indicate the quality of the interpolation at the current level l. Also, it can be used to decide whether the next level of interpolation is necessary. In addition to that, many authors [55, 58, 80, 126] have successfully used the hierarchical surpluses to detect a degree of non-linearity of the function from which it can be extended for dimensional adaptivity to enhance the cost-benefit of building the surrogate. We will discuss the dimensional adaptivity in the next section.

4.4.3 Hierarchical surplus

The hierarchical surplus $w_{\underline{j}}^{\underline{l}}$ provides significant information on which dimensions need to be further refined based on a pre-defined threshold value. In this work, we use the absolute and the relative form of error measure to globally control the error of the interpolation. As suggested in [80], the absolute $\varepsilon_{w_q}^{abs}$ and the relative error measures $\varepsilon_{w_q}^{rel}$ are obtained from the hierarchical surpluses. The error measures of both forms are defined in the following:

$$\varepsilon_{w_q}^{abs} = \max_{|\underline{l}|=q,\, \underline{j}} \{|w_{q,\underline{j}}^{\underline{l}}|\} \tag{4.37}$$

$$\varepsilon_{w_q}^{rel} = \frac{\max\limits_{|\underline{l}|=q,\underline{j}} \{|w_{\underline{j}}^{\underline{l}}|\}}{\max\limits_{|\underline{l}|\leq q,\, \underline{j}} \{f(x_{\underline{j}}^{\underline{l}})\} - \min\limits_{|\underline{l}|\leq q,\, \underline{j}} \{f(x_{\underline{j}}^{\underline{l}})\}} \tag{4.38}$$

where $x_{\underline{j}}^{\underline{l}} = (x_{j_1}^{l_1}, \ldots, x_{j_d}^{l_d})$ and $\underline{j}$ presents multi-index $(j_1, \ldots, j_d)$, $j_k = 1, 2, \ldots, m_{\triangle}^{l_k}$ and $k = (1, \ldots, d)$; the Smolyak parameter is denoted by q. Detailed descriptions of hierarchical surpluses can be found in [55, 64, 80, 81].

4.4.4 Illustration of sparse grid Interpolation

In this section, we first illustrate the Smolyak algorithm (Eq. 4.36) by using the piecewise hat function for two dimensions ($d = 2$) with equidistant grid nodes. Next, the hierarchical surplus used in the sparse grid approach is demonstrated in one dimension using the parabola function.

4.4.4.1 Smolyak algorithm

Figure 4.8 shows the sparse grid layout of the Smolyak interpolant $\mathcal{Q}_{q=3,d=2}(f)$ of the function f for interpolation level $q = 3$; the Smolyak grid set is the set that satisfies $|\underline{l}| \leq q$ where $|\underline{l}| = l_1 + l_2 + \ldots l_d$ and $l_d \in \mathbb{N}$ is the level of interpolation used for d^{th} dimension. In this example, the sparse grid set needs to satisfy $l_1 + l_2 \leq 3$; it is therefore the union of the set: $\{l_1 = 1, l_2 = 1\}, \{l_1 = 2, l_2 = 1\}, \{l_1 = 1, l_2 = 2\}$ to satisfy $l_1 + l_2 \leq 3$. In other words, the sparse grid interpolation points are the members of the set $\{(|\underline{l}| = 2) \cup (|\underline{l}| = 3)\}$. The Smolyak sampling points and the construction of a product rule in the example above

are illustrated in Fig. 4.8. It should be noted that we use the non-equidistant interpolation points x^l (Eq. 4.39) in this thesis instead of the grid points shown in the example.

We show the sparse grid points obtained from the union of the univariate interpolation shown in Fig. 4.8(a), where the red circles represent collocation points of the univariate interpolation. In Fig. 4.8(b), we illustrate the grid layout of the product rule and the sparse grid; the grey shaded area denotes the selection of grid points ($A \cup B \cup D$) introduced by Smolyak, while the product rule is the union of all intermediate levels ($A \cup B \cup C \cup D$). If $f(x) = x + 1$ and $f(y) = y + 1$, the product rule is written as $f(x, y) = f(x)f(y) = (x+1)(y+1)$. The structure of the sparse grid requires fewer supporting points than the full tensor product grid for equivalent errors evaluated from the function f, leading to a lower computational cost.

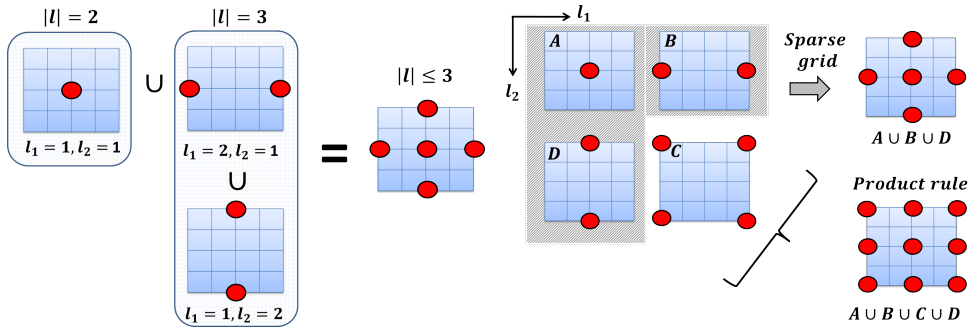


Figure 4.8: Two-dimensional layout of the sparse grid points: a) The sparse grid layout for the interpolant $\mathcal{Q}_{q=3,d=2}(f)$ where points shown are a member of the set $\{(|\underline{l}| = 2) \cup (|\underline{l}| = 3)\}$ that satisfies $|\underline{l}| \leq 3$: b) A comparison of the product grid and sparse grid sampling points. The grey shaded area denotes the selection of grid points ($A \cup B \cup D$) of the sparse grid, while the product grid are the grid points that follow $A \cup B \cup C \cup D$.

4.4.4.2 Hierarchical surplus

Figure 4.9 displays hierarchical surpluses w_j^l for level l of grid location j. As stated earlier, the hierarchical surplus is the difference between $f(x)$ and the interpolant $\mathcal{Q}^{l-1}(f(x))$. Thus we can express the hierarchical surplus as $w_j^l = f(x_j^l) - \mathcal{Q}^{l-1}(x_j^l)$. The parabola function in Fig 4.9 is the function $f(\cdot)$ for interpolation, and the shape of the grid points x_j^l denotes different levels of interpolation. For a continuous function, the greater the level of interpolation, the smaller the hierarchical surplus, as can be seen in Fig. 4.9 that $w_1^2 < w_1^1$.

4.4.5 Chebyshev-based grid

Even though several grid patterns, such as Clenshaw-Curtis and Newton-Cotes points, can be used with the sparse grid approach, the Chebyshev-based grid (CGL) is chosen in this work for two reasons: 1) the Lebesgue value of CGL is close to the optimal value (see Eq. A.10 in appendix A), and 2) its non-equidistant property reduces oscillation of the polynomial at the edges of the interval, referred to as the Runge phenomenon [142]. Additionally, the Chebyshev-based grid points is also recommended in [9, 80] for sparse grid

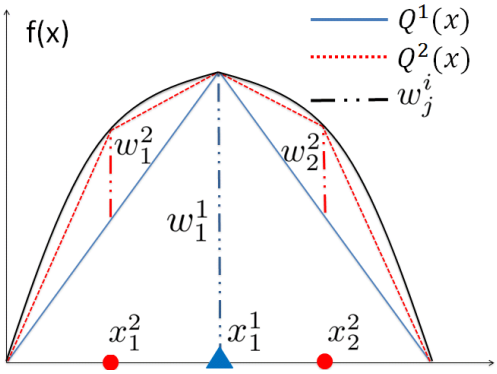


Figure 4.9: Illustration of the hierarchical surplus of the interpolant of level 1 and 2 in one dimension. The solid blue curve and the red dashed curve represent the approximate function from the sparse grid points, x_j^l at level 1 and 2, respectively. Hierarchical surplus w_j^l is the error measure between two consecutive models at the grid points of current level l. For example, the w_1^1 at the point x_1^1 is evaluated at the level 1, while w_1^2 and w_2^2 are the hierarchical surplus at level 2.

interpolation. Given the set of one-dimension grids $X^l = \{x_1^l, \ldots, x_{m_l}^l\}, l \in \mathbb{N}, x^l \in [0,1]$ and $X^{l-1} \subset X^l$, we define the set of grid points in the following:

$$
\begin{aligned}
m_l &= \begin{cases} 1 & \text{if } \; l = 1, \\ 2^{l-1}+1 & \text{if } \; l > 1, \end{cases} \\
x_j^l &= \begin{cases} (-\cos \; (\pi(j-1)/(m_l-1)) \; +1 \;)/2 & \\ \qquad \text{for } j = 1, \ldots, m_i \quad \text{if} \quad m_l > 1, \\ 0.5 \qquad \text{for } j = 1 \qquad\quad \text{if} \quad m_l = 1, \end{cases}
\end{aligned}
\tag{4.39}
$$

where m_l denotes the number of grid points for the interpolation level l; x_j^l represents the position of the grid points at j^{th} position with the level l. Figure 4.10 demonstrates the layout of Chebyshev-based grid points in 2 and 3 dimensions. The corresponding Smolyak parameter q is denoted on top of each sub-figure where the level of interpolation from 2 to 5 is shown from left to right.

4.4.6 Dimensional adaptivity

For a regular sparse grid, all dimensions are treated equally and no particular order of dimensional sampling is preferred [80]. However, for many functions, the explicit form is often unknown from which the degree of non-linearity in one dimension might be greater than the other. Therefore, it is efficient to fill more sampling points in the dimension that exhibits a greater degree of the non-linearity, which leads to dimensional adaptivity that allows automatic sampling along dimensions. Several authors [55, 64, 80, 81] have successfully implemented dimensional adaptivity for sparse grid interpolation. In this work, we use the method proposed in [80] where the hierarchical surplus is used as an error indicator for dimensional adaptivity of the surrogate.

For the adaptive sparse grid approach, the Smolyak algorithm given in Eq. 4.36 can be

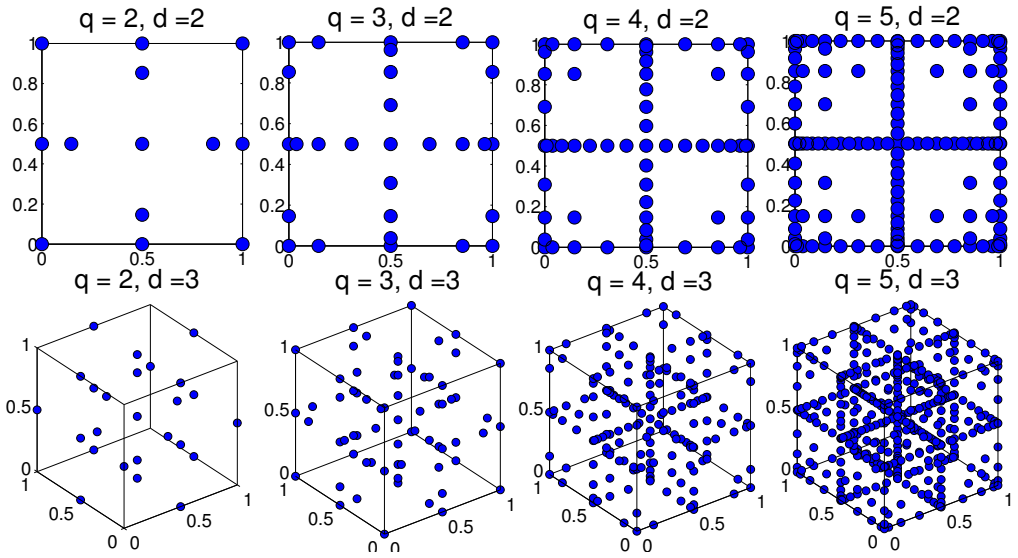


Figure 4.10: The layout of Chebyshev-based grid points with corresponding Smolyak parameters q in 2 and 3 dimensions, as shown in the top and bottom row. From left to right, the level of interpolation q is from 2 to 5.

generalised in the following form [157]:

$$\mathcal{Q}_{q,d}(f) \quad = \quad \sum_{\underline{\ell}' \in \mathbb{I}(q)} (\triangle^{l_1} \otimes \cdots \otimes \triangle^{l_d})(f)$$

where the multi-index set $\mathbb{I}(\ell)$ is a set that determines the structure of the sparse grid for dimensional adaptivity; this set is given in [157] as follows:

$$\mathbb{I}(q) = \{\underline{\ell}' \in \mathbb{N}^n \mid \underline{\ell}' \cdot \mathbf{a} = \sum_{i=1}^{d} a_i \ell_i \leq q\} \tag{4.40}$$

where $\mathbf{a}$ is a vector of weights that controls the interpolation level l_i in the i^{th}-dimension of the sparse grid. Each weight a_i is determined based on the hierarchical surplus such that the approximation error of the interpolation is significantly reduced. By using the multi-index set $\mathbb{I}(\ell)$, the interpolation level l of the sparse grid becomes anisotropic, meaning that the grid points are placed in each dimension differently. However, if the vector $\mathbf{a}$ is a vector of all ones, the set $\mathbb{I}(\ell)$ is the multi-index for the regular sparse grid without dimensional adaptivity. The applications of adaptive sparse grid interpolation will be discussed in the next chapter.

4.4.7 Implementation and software

Several toolboxes of the sparse grid are freely available under the GUI license from the DAKOTA project [33] and Oak Ridge National Laboratories [169]. MATLAB implementation is available from the Universität Stuttgart [80, 81, 170] from which we have used their library for the sparse grid algorithm in this thesis.

4.5 Applications

In this section, we present the use of the Sparse Grid Interpolation (SGI) technique for building a surrogate of the objective function for CO_2 storage under geological uncertainty. This chapter focuses on three main parts: 1) numerical evaluations of the SGI technique to justify the quality of the SGI surrogate by using several mathematical test functions and a reservoir PUNQ-S3 model; 2) the use of dimensional adaptivity for the SGI surrogate; 3) surrogate building techniques that deal with the geological uncertainty of deep saline aquifers.

4.5.1 Numerical evaluations of mathematical functions

Numerical evaluations are performed to demonstrate the error convergence of the Sparse Grid Interpolation (SGI) approach in comparison to the product grid method. Four different benchmark functions were chosen to represent different characteristics of the objective function in CO_2 storage problems.

Three test functions ($f_1(\mathbf{x}), f_2(\mathbf{x}), f_3(\mathbf{x})$) were originally designed for numerical integration, see [115], and also were used to demonstrate sparse grid interpolation in high dimensions by Barthelmann et al [9]. The $f_4(\mathbf{x})$ function obtained from [16] was used to simulate a highly multi-modal landscape with many local maximum and one global maximum. Let functions $f(\mathbf{x})$ be $f : X \to \mathbb{R}$ where X is defined on the unit-hypercube $[0, 1]^{10}$, then the test functions are given below:

$$\text{Oscillatory:} \quad f_1(\mathbf{x}) = \cos(2\pi r_i + \sum_{i=1}^{d} c_i x_i) \tag{4.41}$$

$$\text{Gaussian:} \quad f_2(\mathbf{x}) = \exp(-\sum_{i=1}^{d} c_i^2 \cdot (x_i - r_i)^2) \tag{4.42}$$

$$\text{Continuous:} \quad f_3(\mathbf{x}) = \exp(-\sum_{i=1}^{d} c_i \cdot |x_i - r_i|) \tag{4.43}$$

$$f_4(\mathbf{x}) \begin{cases} = -(x_1^2 + 2x_2^2 - 0.3\cos(3\pi x_1) \\ \quad -0.4\cos(4\pi x_2) + 0.7), \text{for } d = 2 \\ = -\sum_{j=1}^{d-2}\sum_{i=1}^{2}(x_i^2 + 2x_{i+j}^2 - 0.3\cos(3\pi x_i) \\ \quad -0.4\cos(4\pi x_{i+j}) + 0.7), \text{for } d > 2, \\ \quad d \in \{4, 6, \ldots, 2n;\ n \in \mathbb{N}\} \end{cases} \tag{4.44}$$

where $\mathbf{x} := \{x_1, \ldots, x_d\}$ and d is equal to 4 in this example; c_i and r_i are random parameters chosen from the range [0,1]. We built the surrogate of explicit functions, Eqs. 4.41 – 4.44, and compare interpolation errors between the SGI and the full grid (tensor product grid) surrogate.

The reference response surface constructed from the explicit function are shown in the top row in Fig. 4.11, while the response surface obtained from the SGI approach is shown in the middle row. The response surfaces in Fig. 4.11 show only two dimensions (x_1 and x_2) that cut through the fixed coordinates x_3 and x_4 at zero from the total of four dimensions.

As can be seen in Fig. 4.11, the response surfaces from the SGI technique are similar to the reference surfaces. For numerical evaluations, the difference in the function values between the explicit function and the surrogate model is used as an error measure. For this example we first draw $1,000$ samples from a uniform distribution covering the domain X, and then l_∞-norm was used for the maximum error measure.

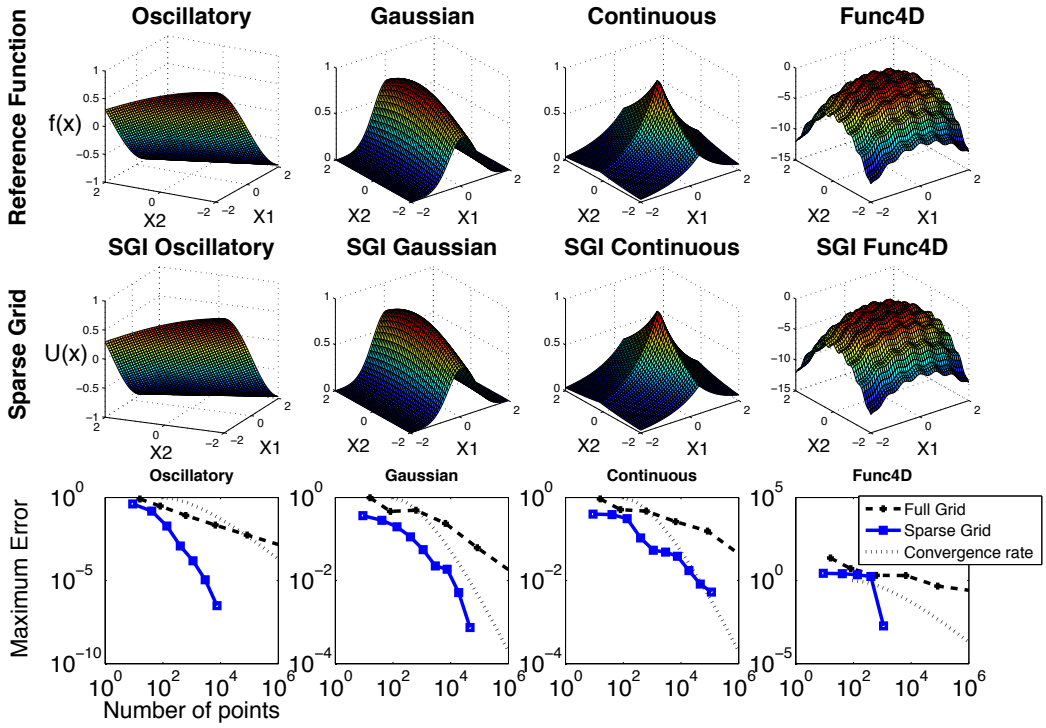


Figure 4.11: Top row: reference surrogates $f(x)$ from left to right obtained from the functions in Eqs. 4.41 – 4.44. Middle row: each surrogate model $U(x)$ is obtained from Sparse Grid Interpolation (SGI). Bottom row: the relationship between maximum error and the number of points used to build the surrogate. The figure shows only two dimensions (x_1 and x_2) that cut through the fixed coordinates x_3 and x_4 at zero from the total of four dimensions.

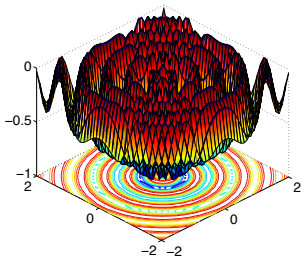
The bottom row of Fig. 4.11 shows the plot of maximum error and the number of points used to build the surrogate. The error decays are plotted versus the number of collocation points where we see that increasing the number of points improves accuracy of interpolation for both techniques. According to Fig. 4.11, the SGI (the solid curve with square markers) requires significantly fewer points than the full grid approach (the dashed curve) for the same maximum error for all four functions. It could be explained that the SGI selects only the points that are important to the objective function and discards less important points, resulting in a significant reduction of evaluation points required to achieve the same level of accuracy of full grid approach.

To illustrate the SGI surrogate for different levels of interpolation, we used drop-wave [38] and Rastrigin [128] functions as explicit functions to compare the quality of the SGI surrogates. The following 2-dimensional drop-wave function is evaluated over the range

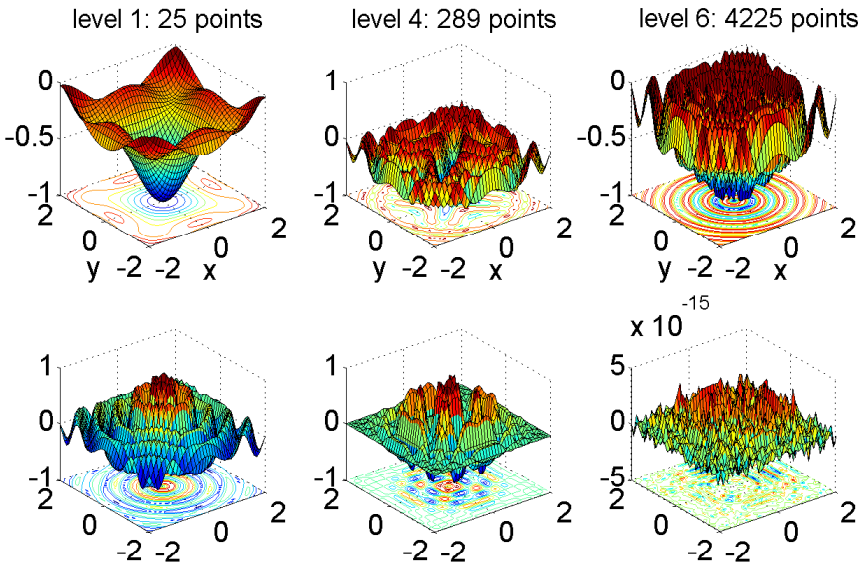
$[-2, 2]^2$.

$$f(\mathbf{x}) = -\frac{1 + \cos\left(12\sqrt{x_1^2 + x_2^2}\right)}{0.5(x_1^2 + x_2^2) + 2} \tag{4.45}$$

Figure 4.12(a) displays the response surface of the drop-wave function. The top row in Fig. 4.12(b) presents the SGI surrogates using different interpolation levels, while the bottom row represents the response surface of the error obtained from the difference between the reference function and the SGI surrogate. Each column in Fig. 4.12(b) presents the levels 1, 4, and 6 with the corresponding number of points used to construct the surrogate. Comparing Figs. 4.12(a) and 4.12(b) shows that the SGI surrogate using the higher level of interpolation is better than using the lower level since the SGI surrogate with the high interpolation level is in good agreement with the reference surface. As can be seen, the SGI surrogate using 6 levels is very similar to the reference surface as the error surface values over the range $[-2, 2]^2$ are approximately identical to zero.



(a) Drop-wave function [38]



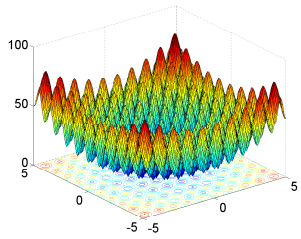
(b)

Figure 4.12: Illustration of the sparse grid interpolation for the drop-wave function: a) the response surface of the drop-wave function; b) the SGI surrogates with different interpolation levels are shown at the top row, while the error between the drop-wave function and the associated SGI surrogate is displayed along the bottom row.

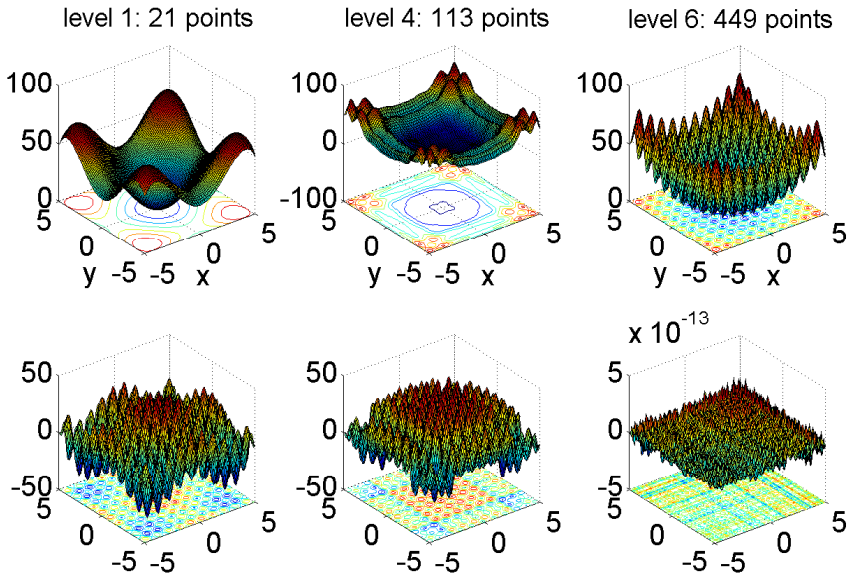
As well as the study of the drop-wave function in Fig. 4.12, we used the Rastrigin function, which contains highly multimodal with multiple local minima, for a similar study of the SGI surrogate for different interpolation levels. The following 2-dimensional Rastrigin function over the range $[-5, 5]^2$ is given below:

$$f(\mathbf{x}) = 10n + \sum_{i=1}^{n} \left[x_i^2 - 10\cos(2\pi x_i) \right] \tag{4.46}$$

Figure 4.13 shows the response surface of the Rastrigin function and the SGI surrogates for different levels of interpolation. Similar results of the drop-wave function in Fig. 4.12 are also observed for the Rastrigin function, which implies that the SGI surrogate using the higher interpolation level is better than the surrogate using the lower level.



(a) Rastrigin function [128]



(b)

Figure 4.13: Illustration of the sparse grid interpolation for the Rastrigin function: a) the response surface of the Rastrigin function; b) the SGI surrogates with different interpolation levels are shown at the top row, while the error between the Rastrigin function and the associated SGI surrogate is displayed along the bottom row.

We have seen from both Figs. 4.12 and 4.13 that the quality of the SGI surrogate can be improved through the level of interpolation. However, it should be noted that building the surrogate with a high interpolation level may require more points, which is a time-

consuming process when building the surrogate of computationally expensive functions. Thus, the use of dimensional adaptivity may be used to allocate more points to significant locations, and consequently reduces the computational cost. Dimensional adaptivity and several numerical evaluations performed on a response surface relevant to the CO_2 optimisation problem will be discussed in the next section for synthetic deep saline aquifer models.

4.5.2 Deep saline storage aquifers

This section first describes the deep saline aquifer models that will be used to demonstrate the application of optimisation methods presented in this thesis. Next, we introduce economic criteria for CO_2 sequestration. Then, we present numerical evaluations of the SGI surrogate, followed by a discussion of dimensional adaptivity. Later, we consider the use of surrogate-assisted optimisation under uncertainty. Last, we introduce a utility function to find robust injection strategies where two surrogate building techniques under uncertainty are presented.

4.5.2.1 Model descriptions

There are three models used in this thesis: an idealised rectangular model, the PUNQ-S3 model [45], and the leaky PUNQ-S3 model. The details of these three models are presented, followed by a description of geological uncertainty and well configurations.

Rectangular synthetic model. The aquifer model has $17 \times 17 \times 50$ grid blocks where the dimension of each grid block is $50\ m \times 50\ m \times 5\ m$. The top formation is located at a depth of 1400 m with an initial pressure of 17.4 MPa and the temperature is constant at 32.2 °C. Permeability and porosity distributions are assumed to be uncorrelated and independent, which was generated from the geostatistical model described later in Section 4.5.2.2. The commercial compositional simulator (Eclipse E300) with the CO2STORE option [150] was used to simulate subsurface CO_2 migration. Aquifer and fluid properties in the model are described in Tables 4.1 and 4.2. We ignore any hysteresis in capillary pressure and relative permeability. Well locations are shown in Fig. 4.14 whereas all wells were perforated at the most-bottom layer with well control configurations given in Section 4.5.2.3.

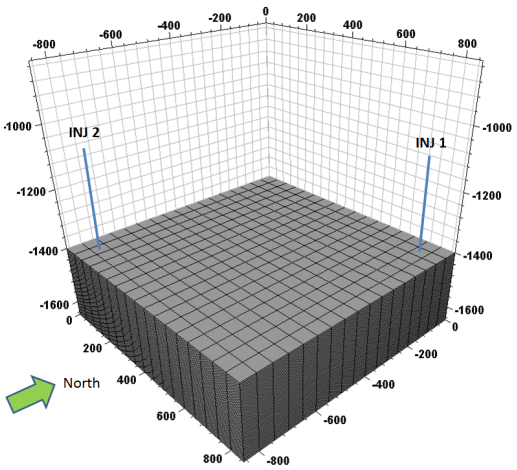


Figure 4.14: A three-dimensional rectangular synthetic model is shown with injector 1 and 2.

Table 4.1: Description of the rectangular model and fluid properties

Description	Parameter values
Number of grid blocks	$17 \times 17 \times 50$
Number of layers	50
Grid block size	$50m \times 50m \times 5m$
Rock compressibility (P_{Ref} =13.7 GPa)	0.725 GPa^{-1}
Top depth of the aquifer	$1400\ m$
Initial aquifer pressure (at 1400 m)	17.4 MPa
Aquifer temperature	$32.2\ ^{\circ}C$
Residual water saturation	0.30
Maximum residual gas saturation	0.30
Fluid and solid components of saline aquifer	CO_2, H_2O, NaCl, $CaCl_2$

Table 4.2: Total mole fraction and diffusion coefficient for initial conditions of the rectangular model.

Descriptions at initial aquifer conditions	H_2O	CO_2	NaCl	$CaCl_2$
Total mole phase fraction	0.9109	0.0	0.0741	0.015
Water diffusion coefficient (m^2/s)	1.16×10^{-9}		1.16×10^{-9}	
Gas diffusion coefficient (m^2/s)	1.16×10^{-8}		-	

Remark : The rectangular model will be used in Chapter 5 for multi-criterion decision making.

PUNQ-S3 model. A synthetic reservoir model, PUNQ-S3, is the standard model used in this thesis. Detailed descriptions of the PUNQ-S3 model can be found in [45, 132] which was modified to represent a storage aquifer by Juanes et al [76]. PUNQ-S3 is a sandstone formation for all five layers. The relative permeability curves were taken from Juanes et al [76] with the Killough hysteresis model [79] applied to the non-wetting phase (CO_2). The Eclipse simulator for the PUNQ-S3 model was used to simulate the flow of two phases: a CO_2-rich phase and a H_2O-rich phase with two components CO_2 and H_2O. Other details of the aquifer model are summarised in Table 4.3.

The model is discretised into $19 \times 28 \times 5$ grid blocks with the top formation located at a depth of 2340 m and with an average thickness of 15 m. The initial pressure is 90 bars with a fixed temperature of 32.2 °C. There are two injectors completed on the bottom most layers of the aquifer, shown in Fig. 4.15. To represent an unconfined aquifer, we set pressure boundaries by assigning a pore volume of $1,000$ times the model size at the boundary. The rest of the model parameters are summarised in Table 4.3.

Remark : The PUNQ-3 model is a standard model in this thesis and will be used throughout all chapters, unless otherwise stated.

Leaky PUNQ-S3 model. The risks of CO_2 leakage from storage aquifers to the atmosphere are characterised for global and local risks. The global risk of leakage is the scale at which leaking CO_2 can significantly impact the global climate. However, the local risk is the releases of CO_2 such that it contaminates potable water sources, ecosystems, human health and communities [104]. For the leaky PUNQ-S3, we modelled the local risk scenar-

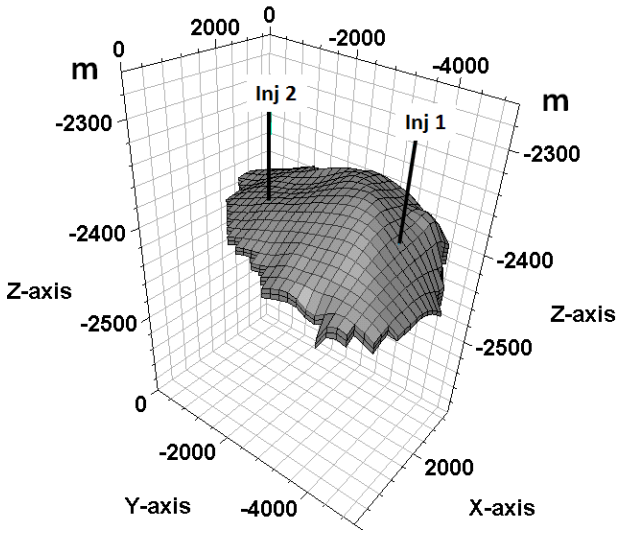


Figure 4.15: The location of injectors in the three dimensional PUNQ-S3 model. Injector 1 is located to the north while injector 2 is located to the south of the model.

Table 4.3: Description of the PUNQ-S3 model and fluid properties

Description	Parameter values	
Number of grid blocks	$19 \times 28 \times 5$	
Number of layers	5	
Grid block size	$180m \times 180m \times 3m$	
Rock compressibility	0.5 GPa^{-1}	
Top depth of the aquifer	$2340\ m$	
Initial aquifer pressure (at 2355 m)	9 MPa	
Aquifer temperature	$32.2\ ^{\circ}C$	
Fluid components of saline aquifer	CO_2, H_2O	
Residual water saturation	0.3	
Maximum residual gas saturation	0.3	
Total mole phase fraction	H_2O	CO_2
(at initial aquifer conditions)	1.0	0.0
Water diffusion coefficient (m^2/s)	1.16×10^{-9}	1.16×10^{-9}
Gas diffusion coefficient (m^2/s)	1.16×10^{-8}	1.16×10^{-8}

ios when CO_2 could leak to the atmosphere. The potential leakage pathways were addressed by Metz et al [104] in Fig. 4.16. Particularly, Gasda et al [52] reported that many potential CO_2 storage sites in the United States and in Canada (Alberta basin) have the potential of CO_2 leakage from existing and abandoned wells. The leakage of CO_2 from abandoned wells can be one of the most probable escape pathways [112–114]. Thus, this study mainly focuses on the risk of CO_2 migration through an abandoned well. The optimisation approach that considers the risk of CO_2 leakage will be shown later in Chapter 5 using the leaky PUNQ-S3 model.

For the leaky PUNQ-S3, the model parameters are the same as the PUNQ-S3 model, except that we added an abandoned well, shown in Fig 4.17, for CO_2 to escape to the atmosphere. The leakage of CO_2 could be modelled by a permeable medium where vertical permeability is significantly greater than the formation [39] since this represents the flow through cement material around the well. Alternatively, CO_2 produced from the abandoned

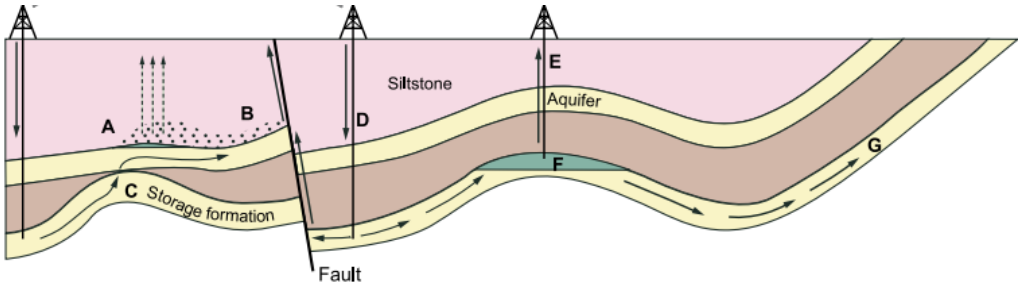
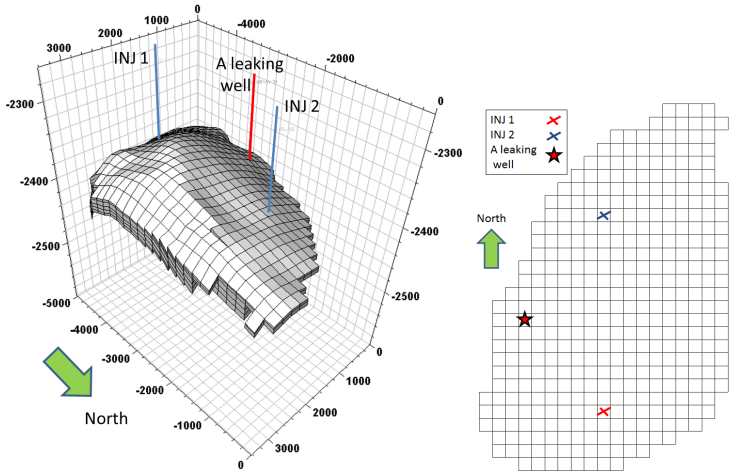


Figure 4.16: Potential leakage pathways of CO_2 to atmosphere in different scenarios as described by [104]. A: CO_2 pressure exceeds capillary pressure and seeps through formation above. B: CO_2 leaks through open faults or fractures. C: CO_2 leaks through open or fracture in the caprock. D: Injected CO_2 migrates directly through faults due to high injection pressure. E: CO_2 leaks via abandoned wells. F: CO_2 is dissolved in natural flow and transport outside the storage. G: Dissolved CO_2 migration, from [104].

well could also be used as a proxy for the CO_2 leakage [152]. In this study, we used the later approach where the produced CO_2 from the well represents the leakage from the storage aquifer.

For modelling the risk of CO_2 leakage through wells, Oldenburg et al [117] reported that the risk comprised the probability of intersecting CO_2 and wells, of CO_2 flowing in a conduit to atmosphere, and of the impact of CO_2 on the leaking area at the surface. To this end, for the leaky PUNQ-S3 model, the bottom-hole pressure (BHP) of the leaky well was used at the initial reservoir pressure with a transmissibility multiplier of the connecting grid boxes around the well of 0.05; this reduces the rate at which the CO_2 column intersecting the abandoned well would escape to the atmosphere.



(a) Three-dimensioanl model (b) Two-dmesional model

Figure 4.17: a) A three-dimensional leaky PUNQ-S3 model shown with injectors 1 and 2 and a leaky well. b) Two-dimensional model with well locations.

Remark : The Leaky PUNQ-3 model will be used in Chapter 5 for multi-criterion decision making.

4.5.2.2 Geological uncertainty

Horizontal permeability and porosity fields were generated independently; each field was spatially correlated and generated geostatistically from Gaussian random fields using the following Gaussian covariance function [31, 119].

$$C_y(h) = \sigma^2 \exp\left[-3\left(\frac{|h|}{a}\right)^2\right] \quad (4.47)$$

$$\mathbf{C}_y = \mathbf{L}\mathbf{L}^T \quad (4.48)$$

where the Gaussian covariance function $C_y(h)$ contains lag distances between two arbitrary points h and correlation length a; $\mathbf{C}_y$ is a matrix of covariance for various lag distances, which is written in a lower triangular matrix $\mathbf{L}$ using Cholesky decomposition so that we can generate a vector of random fields y with permeability and porosity mean (μ) as:

$$y = \mu + \mathbf{L}z \quad (4.49)$$

where μ is a vector containing the mean of model parameters (e.g. a priori mean of permeability and porosity distribution) and z is a vector of independent identically distributed (iid) random variables with zero mean and variance equal to 1. Table 4.4 gives the geostatistical parameters used in Eqs. 4.47 and 4.49 to generate 29 and 20 realisations for the rectangular and the PUNQ-S3 models, respectively. The detailed descriptions for generating geological uncertainty for the rectangular and the PUNQ-S3 geological models are given in the following.

Table 4.4: Geostatistical parameters used to generate Gaussian random fields for the rectangular and the PUNQ-S3 models

		Horizontal permeability (k_h)	Porosity	Anisotropy ratios
Rectangular Model	mean** (μ)	5	0.25	0.1, 0.3, 0.5, 0.7
	Corr* (a)	8	8	-
	Variance (σ^2)	0.36	0.04	-
PUNQ-S3 Model	mean** (μ)	100, 200, 300, 500	0.15, 0.2, 0.1, 0.25	0.5, 0.25, 0.65, 0.1
	Corr* (a)	12	15	-
	Variance (σ^2)	0.16	0.0144	-

*Correlation length based on the number of grid boxes in the geological model,**Permeability in mD

Geological uncertainty for the rectangular model. We first generated two-dimensional permeability and porosity fields, each for 29 realisations. Then, one of the permeability and porosity fields is used for all 50 layers — each layer has the same properties. Therefore, we obtained 29 three-dimensional models from the 29 two-dimensional permeability and porosity fields. The anisotropy ratio ($\frac{k_v}{k_h}$) is randomly sampled from a finite set of $0.1,\ 0.3,\ 0.5$, and 0.7 for each three-dimensional model.

Geological uncertainty for the PUNQ-S3 and leaky PUNQ-S3 models. We used a greater range of geological models by varying the mean of permeability and porosity values, rather than using a static value as in the rectangular case. It is worth noting that geological properties of the PUNQ-S3 model are the same as the leaky PUNQ-S3 model and we will refer only to the PUNQ-S3 model in this section. To this end, we chose 20 model realisations where each layer comprises different permeability and porosity fields. The samples of the geological realisations of permeability and porosity are shown in Figs. 4.18(a) and 4.18(b), respectively; the top to the bottom layers of the PUNQ-S3 model are shown from left to right, while each row represents a sample of model realisations; the location of injectors are shown only in the upper-left corner figure using the × symbol. The anisotropy ratio ($\frac{k_v}{k_h}$) is randomly sampled for each model realisation from a finite set of $0.1,\ 0.3,\ 0.5$, and 0.7.

4.5.2.3 Well control configurations

Two injectors were constrained by bottom hole pressure (BHP) and operated by injection rate controls. The well locations are shown in Fig. 4.14 for the rectangular model, in Fig. 4.15 for the PUNQ-S3 model, and in Fig. 4.17 for the leaky PUNQ-S3 model. We consider two injection periods, each of 15 years. The total injection and post-injection (monitoring) period is 30 and 50 years, respectively. In this work, we manipulate the injection ratio (x_1, x_2) of each injector instead of injection rate. The injection ratios of the injector 1 at the first and the second injection periods, x_1 and x_2, are shown in Fig. 4.19. The injection ratio (x_1) is obtained from the injection rate of a particular well divided by the total supplied rate. Thus, over the same injection period, the sum of the injection ratios of CO_2 in each well is unity at all time $(x_1 + (1 - x_1) = 1)$; x_2 is obtained in a similar way to x_1. The detailed well configuration parameters are listed in Table 4.5.

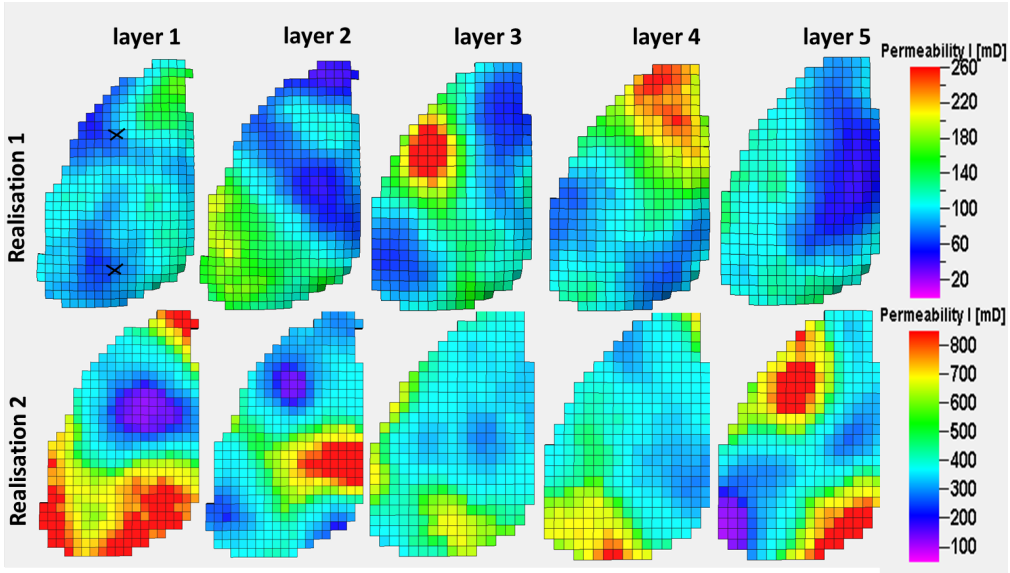
Table 4.5: Well configuration parameters for both the rectangular and the PUNQ-S3 model

Description	Parameter values
Total supplied rate (Sm^3/day)	2.5×10^6
Minimum injection rate (Sm^3/day)	0
Maximum injection rate (Sm^3/day)	2.5×10^6
Maximum BHP (bars)	600 (PUNQ-S3 model)
	400 (rectangular model)
Injection period	30 years
Post-injection period	50 years
Time period per control step	15 years

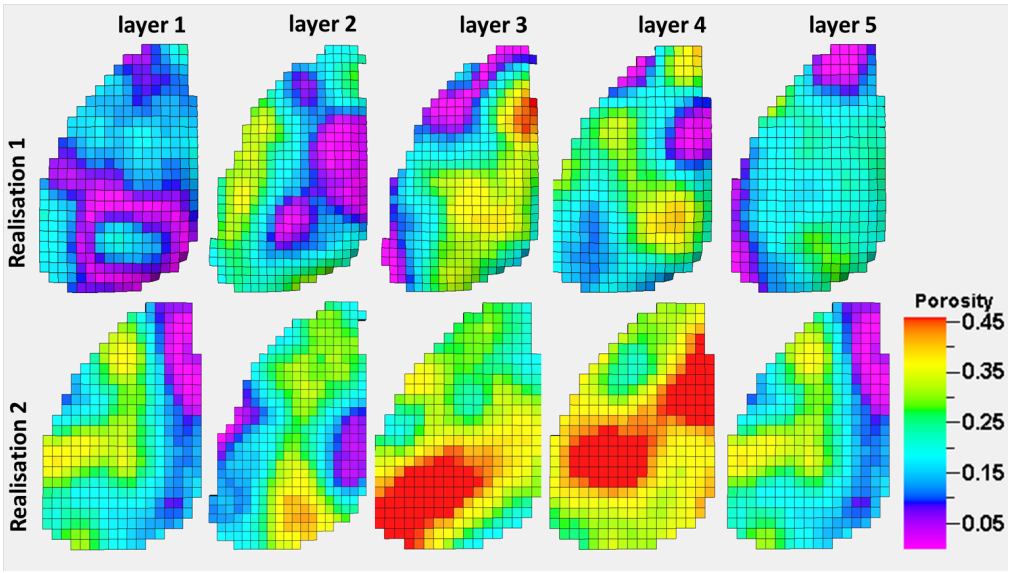
Remark : We used the well configurations above for all models in all chapters except Chapter 6 where we used a greater number of control variables.

4.5.2.4 Economic performance criteria

In the oil and gas industry, the usual economic performance criterion used for decision-making is Net Present Value (NPV), the sum of net benefit and cost cash flows over a period of time discounted to its present value. In other words, NPV provides the present value of



(a) Permeability



(b) Porosity

Figure 4.18: Petrophysical properties of three dimensional PUNQ-S3 model with five layers: (a) and (b) show permeability and porosity distributions for the top to bottom layers (*left to right*), respectively; each row of figure (a) and (b) presents a sample of 20 model realisations; injection wells are shown only in upper-left corner figure using the × symbol.

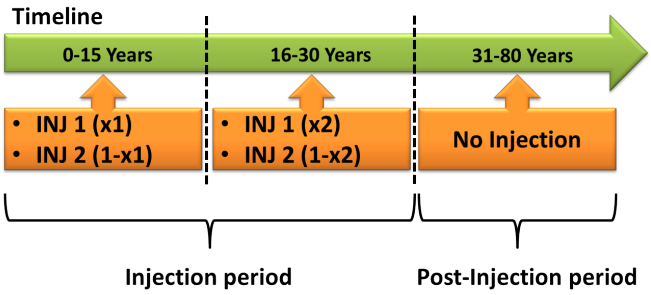


Figure 4.19: Well control configurations of two injectors: the injection ratios of the injector 1 at the first and the second injection periods are denoted by x_1 and x_2, respectively.

an investment based on the expected net benefit in the future of that investment. For oil and gas production optimisation, NPV is often maximised by varying production strategies. However, for CO_2 sequestration, the economic criteria are slightly different from the NPV due to the fact that there is currently no explicit revenue from sequestering CO_2 into deep saline aquifers.

Several alternative performance criteria from other disciplines could be applied to CO_2 sequestration. For example, we could minimise the negative associated costs of CO_2 injection (e.g. the capital and operational costs of sequestration project), or maximising the mass of CO_2 injection [110], or minimising the negative impact from releasing CO_2 to the atmosphere. One could also maximise the contact area of CO_2 plume to the formation brine by allocating CO_2 injection [152]. While these objectives are technically useful, they are currently not suitable for go/no-go decision making as they lack judgement and assessment of the economic value of CO_2 sequestration. For this reason, we could maximise the potential economic benefit received from the government if we can reduce CO_2 emissions. This economic benefit could be in the form of reducing tax bills or the government incentives in terms of tax credits.

In this chapter, we model the profitability of CO_2 sequestration based on the assumption that tax credits are given for the amount of CO_2 trapped in the subsurface. As the explicit form of environmental tax credits are currently not available, we therefore formulate the minimum tax credits that should be received in order to balance the cost of an investment. This idea is similar to the concept of internal rate of return (IRR) [61] where we slightly modify to the tax credit required to break even — the point where the NPV of the project is zero, described in the following:

$$\sum_{t=1}^{T} \frac{C_{in}(t) - C_{out}(t)}{(1+r)^t} - C_o = 0 \tag{4.50}$$

where $C_{in}(t)$ and $C_{out}(t)$ are cash inflow and outflow at time t; C_0 is a capital investment and t represents time steps which can be described on a daily, monthly or annual basis; T is the total duration of the project and the term $C_{in}(t)$ is the product of the tax credit and the amount of trapped CO_2. Given that $C_{in}(t) =$ tax credit $\times$ trapped CO_2, we can rearrange

the tax credit to one side of Eq. 4.50 and set C_0 to zero, giving

$$\text{tax credit} = \frac{\sum_{t=1}^{T} \left(C_{out}(t) \times (1+r)^{-t}\right)}{\sum_{t=1}^{T} \left(\text{trapped } \mathrm{CO_2} \times (1+r)^{-t}\right)} \tag{4.51}$$

The tax credit in Eq. 4.51 is actually the break-even tax credit where we formulate the following function as the economic performance criterion for CO_2 sequestration.

$$BT(\mathbf{x},\mathbf{y}) = \frac{\sum_{k=1}^{N_t} \left(c\Delta q_i(\mathbf{x},\mathbf{y})(1+r_T)^{\frac{-k\Delta t}{T}}\right)}{\sum_{k=1}^{N_t} \left(\left[\Delta D(\mathbf{x},\mathbf{y}) + \Delta R(\mathbf{x},\mathbf{y})\right]\beta(k)(1+r_T)^{\frac{-k\Delta t}{T}}\right)} \tag{4.52}$$

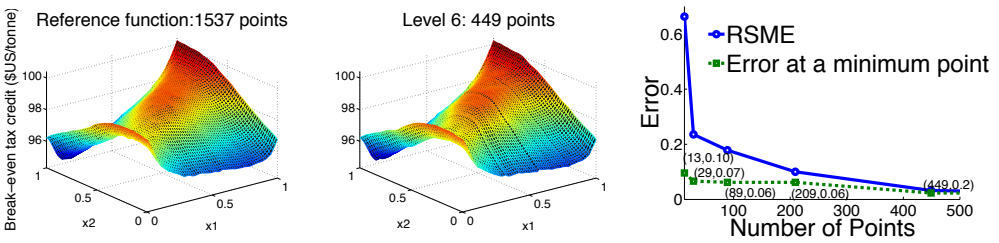
where $BT(\mathbf{x},\mathbf{y})$ presents the break-even tax credit function with $\mathbf{x}$ representing a vector corresponding to the control settings imposed on the wells and $\mathbf{y}$ presenting the spatial state variables in the geological model such as permeability, porosity, and anisotropy ratio; k and N_t represent time step and total time step; Δt is the time interval (days) per time step k; T is the time period (days) used to calculate the discount term r_T; r_T is the discount rate during the time period T; c is the cost associated with CO_2 injection and compression (set here to be 60 US$/tonne); $\Delta q_i(\mathbf{x},\mathbf{y})$, $\Delta D(\mathbf{x},\mathbf{y})$, $\Delta R(\mathbf{x},\mathbf{y})$ are the amount of injected CO_2, the mass of CO_2 dissolved in brine, and the mass of capillary-trapped CO_2 in tonnes during the time Δt, respectively. The future cash flow is usually discounted to the present value. However, the mass of CO_2 trapping in a post-injection period is rewarded by adding an arbitrary exponential function $\beta(k) = 1 + e^{(k-N_t)}$ to cancel out the discount term and emphasise geological security of long-term CO_2 storage. In this chapter, the BT function is minimised by allocating the injection rates of each well where the control settings and the well configurations for CO_2 optimisation are referred to Section 4.5.2.3.

4.5.2.5 Numerical evaluations of the sparse grid interpolation on the PUNQ-S3 model

In this section, we present the SGI surrogate of the BT function in comparison to a reference surface constructed from very fine scale grid points. Figures 4.20(a) and 4.20(b) show the response surface of the break-even tax credit (Eq. 4.52) corresponding to various injection rates (x_1, x_2). The reference surface shown in Fig. 4.20(a) is obtained by a regular sparse grid pattern of $1,537$ points and the surface shown in Fig. 4.20(b) is built by the SGI approach using 449 function evaluations.

As can be seen, the SGI surface is qualitatively similar to the reference surface with a maximum error tolerance of 0.02%. The error tolerance was evaluated by two methods: the difference in the function values between the reference surface and the SGI surface using a root mean square error (RMSE), and a relative error at the optimal solution (a minimum point of the response surface). The relative error was calculated as $(\frac{\mathcal{Q}_f(x_i)-\mathcal{Q}(x_j)}{\mathcal{Q}_f(x_i)}) \times 100$, where $\mathcal{Q}_f(\cdot)$ denotes the interpolant built from a very fine grid pattern (1,537 points) to approximately represent the surface of the reference function; $\mathcal{Q}(\cdot)$ represents the interpolant, and x_i, x_j present the minimum point of $\mathcal{Q}_f(\cdot)$ and $\mathcal{Q}(\cdot)$, respectively. Figure 4.20(c) shows the error convergence with the levels of interpolation (from 1 to 6). The RMSE and the relative error at the optimal solution were evaluated at the product grid pattern of

10,201 points. It is clear that the RMSE decreases significantly with an increase in the interpolation level; however, the errors at the optimal point are relatively small when the interpolation level is greater than 2, which may imply the potential of the SGI surrogate in identifying optimal solutions using a limited number of interpolation levels.



(a) (b) (c)

Figure 4.20: a) The response surface obtained from 1,537 evaluation points is used as a reference surface. b) The response surface obtained from the SGI approach using the 6^{th} level (449 evaluation points) interpolation with an error tolerance of 0.02%. c) The relative error and the root mean square error (RMSE) are plotted against the number of points required for building the surrogate. The relative error, shown by the dashed curve, was calculated from the minimum point of the reference and the SGI surface.

The SGI surrogates at different levels of interpolation (level 1 to 6) are shown in Fig. 4.21 where the corresponding number of points used to construct the SGI surrogate is presented at the top of each figure. Comparing Fig 4.20(a) and 4.21, it can be seen that the higher the level of refinement, the greater the accuracy of the surrogate to the reference surface.

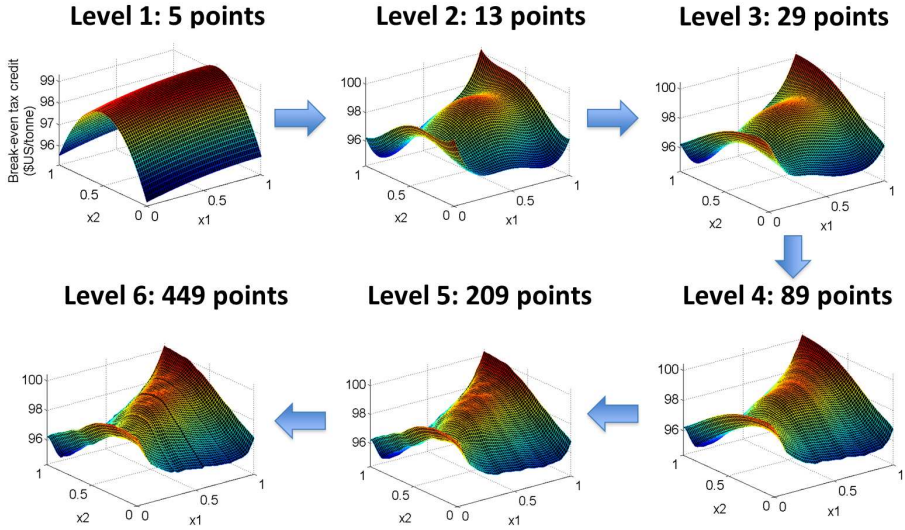


Figure 4.21: Illustration of the SGI surrogates at different levels of interpolation; the level of interpolation and its corresponding number of points used to build the surrogate is shown at the top of each figure.

4.5.2.6 The sparse grid surrogate with dimensional adaptivity

In this section, we present the layouts of sampling points used to build the adaptive sparse grid interpolation (ASGI) and the SGI surrogates, followed by numerical comparisons of the ASGI and the SGI surrogates of the break-even tax credit (BT) function using the PUNQ-S3 model. In this example, we built 20 surrogate models of the break-even tax credit in one-to-one correspondence with 20 geological realisations of the PUNQ-S3 model. Each corresponding surrogate is the response surface of the break-even tax credit from various injection rates (x_1, x_2). The results of this comparison are shown in Figs. 4.22 and 4.23, where 4 models out of 20 evaluated models were selected for illustrative purposes. It is noteworthy that these selected 4 cases represent different patterns of objective function appearing in the entire set of the evaluated models.

In Fig. 4.22, the top row displays the response surface of the break-even tax credit from various injection rates (x_1, x_2). The middle and the bottom rows present the layout of the sampling points (the dot markers) obtained from the SGI and the ASGI approach, respectively. The number of points used to build the surrogate model is shown at the top of each figure. For the ASGI, a relative hierarchical tolerance error at the threshold value 2.4×10^{-2} is used. The results show that the ASGI requires fewer sampling points than the SGI approach as it adaptively enriches the important dimension while constructing the response surface. This can be seen in the model 1 of the figure that the ASGI approach distributes more points in the dimension x_1 than in the x_2, while the ASGI in the model 4 locates more points along the x_2 dimension than x_1.

Figure 4.23 displays the maximum error of the response surface versus the number of interpolation points. The error measure (l_∞-norm) was used to evaluate the difference between the response surface values (ASGI or SGI) and solutions obtained from the reference surface using a product grid pattern of 10,201 points. For all the models shown in Fig. 4.23, the ASGI approach requires fewer points than the SGI method to achieve the same maximum error, which means that the ASGI approach is more efficient than the SGI method.

The results presented in Figs. 4.22 and 4.23 generally imply the significance of dimensional adaptivity for building the surrogate model. We therefore use the ASGI approach to construct the surrogate model to find robust injection strategies for CO_2 sequestration.

4.5.2.7 Surrogate models under geological uncertainty

In this section, we first introduce a utility function for finding an injection strategy whose BT value is insensitive to geological uncertainty. Next, we present numerical evaluations of the SGI surrogate of a utility function where two techniques of surrogate building are presented, followed by the role of utility parameters to the optimal solutions. Last, the impact of optimal and robust injection strategies on the mass of trapped CO_2 are shown.

A utility function of the economic criteria under geological uncertainty. A fixed set of controls will provide various objective function values for different geological realisations. One common approach to design a robust injection strategy is to consider the average of objective function values over different geological models. However, this approach often

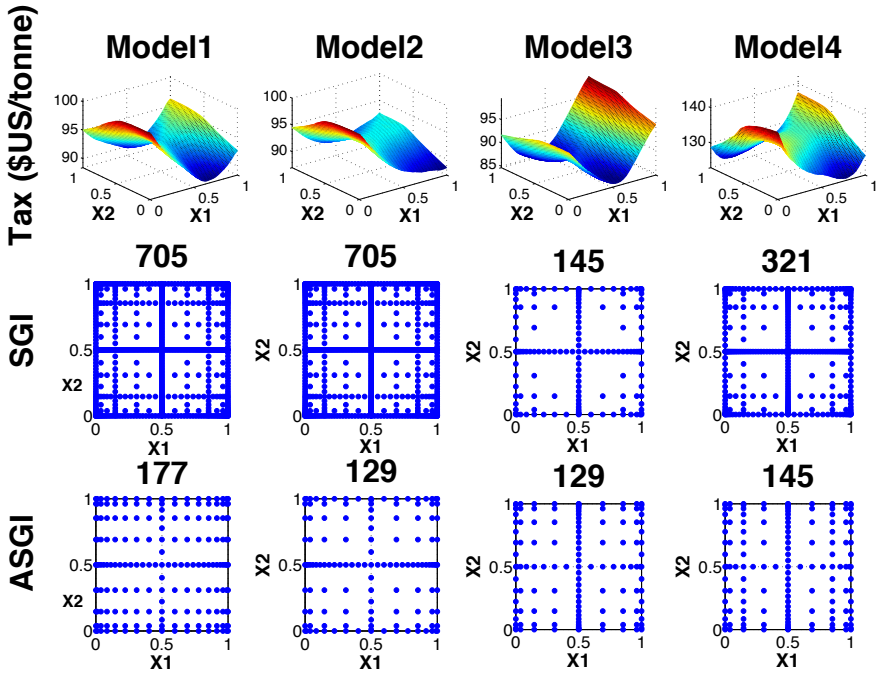


Figure 4.22: Top row: four response surfaces of the objective function (the break-even tax credit) used as the reference surface. Each response surface corresponds to a single geological model. Middle and bottom rows: a layout of the sparse grid in two dimensions of the SGI and the ASGI techniques. The number of grid points associated with the grid layout is shown at the top of each figure.

neglects the variability of the objective functions about the mean (i.e. the variance). In this study, we consider both mean and variance of the objective functions as a new objective function, which is referred to as a utility function, to find a robust injection strategy.

Several authors [28, 108, 175] have successfully used the mean and variance in defining their objective functions for robust optimisation of several practical problems such as mechanical design applications [28, 175] and large-scale power capacity problems [108]; this approach is often referred to as a mean-variance method. In this work, we use the mean-variance method for the economic criteria (over a set of geological realisations) to define a robust objective function. The average and the variance of the BT function (Eq. 4.52) can be simultaneously minimised through a multi-objective optimisation framework or through the use of a utility function. In this chapter, we will first present the utility function whereas multi-objective optimisation for the mean and the variance will be presented later in the next chapter.

The utility function in this work was modified from Alexander and Baptista [1] for the

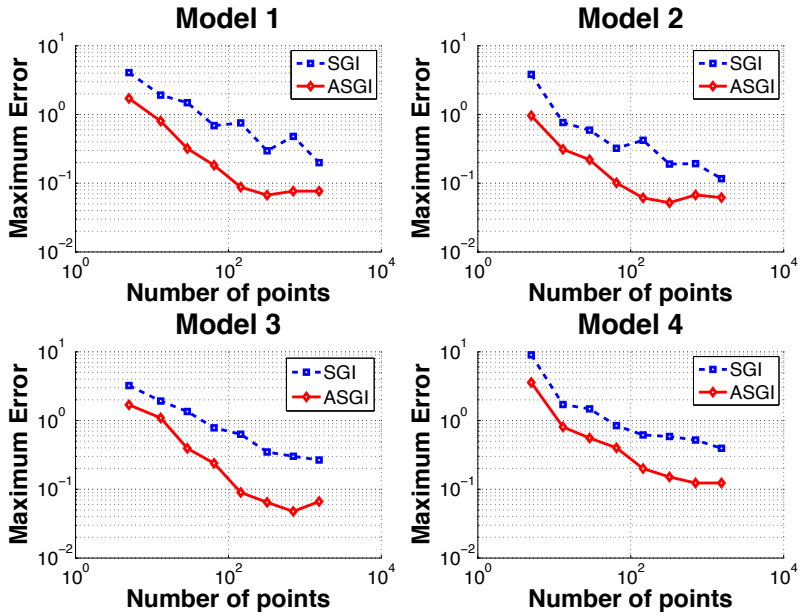


Figure 4.23: The absolute maximum error is plotted versus the number of points used to build the surrogate where the maximum error is evaluated from the difference between the reference model and the surrogate built by regular sparse grid (SGI) (the square markers curves) and dimensional adaptivity sparse grid (ASGI) (the diamond marker curves). Models 1-4 are the same models as in Fig. 4.22.

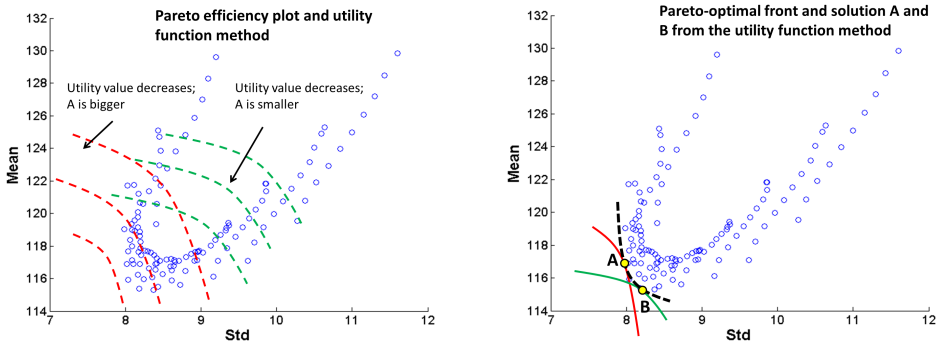
mean-variance method of the BT function. The utility function $\mathcal{U}$ is formulated as:

$$\begin{aligned}
\text{Minimise:} \quad & \mathcal{U}(\mathrm{E}_{BT}(\mathbf{x}), \sigma^2_{BT}(\mathbf{x})) = \mathrm{E}_{BT}(\mathbf{x}) + \frac{1}{2} A \sigma^2_{BT}(\mathbf{x}) \\
\text{Subject to} \quad & z_i(\mathbf{x}, \mathbf{y}) = 0 \qquad i = 1, 2, \ldots, N_z \\
& g_i(\mathbf{x}) = 0 \qquad i = 1, 2, \ldots, N_g \\
& h_i(\mathbf{x}) \leq 0 \qquad i = 1, 2, \ldots, N_h \\
\text{where} \quad & \mathrm{E}_{BT}(\mathbf{x}) = \mathrm{E}[\widetilde{BT}(\mathbf{x}, \tilde{\mathbf{y}})] \\
& \sigma^2_{BT}(\mathbf{x}) = \mathrm{E}[\widetilde{BT}^2(\mathbf{x}, \tilde{\mathbf{y}})] - \mathrm{E}^2_{BT}(\mathbf{x}) \\
& \forall (\mathrm{E}_{BT}, \sigma^2_{BT}) \in \mathbb{R} \times \mathbb{R}, \ A > 0
\end{aligned} \tag{4.53}$$

where $z_i(\mathbf{x}, \mathbf{y})$ denotes the governing equations of CO_2 in each grid cell for the total of N_z cells as a function of control setting $\mathbf{x}$ and state variable $\mathbf{y}$ defined earlier; $g_i(\mathbf{x})$ and $h_i(\mathbf{x})$ represent linear (or non-linear) equality and inequality constraints of control variable $\mathbf{x}$, respectively; N_g and N_h denote the total numbers of equality and inequality constraints, respectively; $\tilde{\mathbf{y}}$ presents different geological realisations; the mean and variance over multiple geological realisations are defined by $\mathrm{E}_{BT}(\mathbf{x})$ and $\sigma^2_{BT}(\mathbf{x})$, respectively. The utility parameter A is a risk coefficient. A risk coefficient set to 0 indicates a risk-neutral attitude. In contrast, a risk coefficient set to a non-zero value denotes a risk-averse ($A > 0$) or risk-prone ($A < 0$) attitude. Since the utility function is formulated based on an individual's attitude towards risks, the preference of risk should be specified a priori. If it is not known a priori, bi-objective optimisation with mean and variance could be applied, or the risk co-

efficient could be varied within a single objective optimisation framework to obtain a set of robust solutions.

In Fig. 4.24(a), we present how utility functions can be used to find a robust solution in a mean and standard deviation space. We see that there are two utility functions with two different utility parameters, as shown with the red and green dashed curves; the blue dots in the figure represent the data in the mean-variance space. The utility function is formulated to ensure that it is a monotonically increasing function where the negative gradient directs to the Pareto-optimal front, a line containing robust solutions. A detailed description about the Pareto-optimal front is given later in the next chapter. In this example, we minimise the utility function such that it is tangent to the Pareto-optimal front (the dashed curve) as shown in Fig. 4.24(b) where the red and green curves present the utility functions using two different utility parameters. From the tangent of the Pareto-optimal front, points A and B are robust solutions.



(a) (b)

Figure 4.24: a) Illustration of the Pareto-optimal front and the utility function; the shape of the utility function is governed by the utility parameter (A) in Eq. 4.53. There are two utility functions with different utility parameters, as shown by the red and green curves; the blue dots in the figure represent the data in mean-variance space. b) The robust solutions A and B are obtained from the tangent of the utility functions and the Pareto-optimal front.

In this thesis, we use the particle swarm optimiser (PSO) [78, 130], a global and stochastic optimisation method inspired by swarm behaviour (e.g. flocks of birds and school of fish), with the surrogate model to find robust and optimal solution. The PSO algorithm is similar to genetic algorithm (GA) in the sense that they both use the initial generation of random population and the use of information sharing among their members, e.g. PSO uses the velocities and locations of its members to improve a solution while GA uses cross-over and mutation. However, in terms of computational efficiency, PSO provides a better convergence rate than GA [101, 120, 122]. Because of its efficiency, PSO has been applied in several applications [129]. It should, however, be noted that gradient-based algorithms such as conjugate gradient [66] and quasi-newton methods can also be used with the surrogate model, but we will discuss later in chapters 6 and 7 that using PSO with surrogate model can be more effective in terms of balancing exploration and exploitation of optimisation search methods, especially for high-dimensional functions.

Numerical evaluations of sparse grid interpolation under uncertainty. Geological models were generated with multiple realisations to represent the uncertainty about the geological properties. The mean and variance of the BT function (Eq. 4.52) from multiple geological models were used in the utility function to find a robust solution. Solving for the robust solution is usually expensive due to the computational costs associated with multiple function evaluations. However, this computational burden can be reduced through the use of the surrogate model built by the ASGI technique. In this study, we present two approaches to building the surrogate of the utility function: to construct the surrogate associated with each geological model independently (referred to as the one-to-one technique), or to build only the surrogate of the mean and standard deviation from multiple geological models (referred to as the mean-std technique). Figure 4.25 illustrates the two techniques (one-to-one and mean-std approach) for constructing the surrogate of the utility function. The quality of these two techniques will be evaluated by two approaches: a spatial error measure, and the error measured at the optimum point.

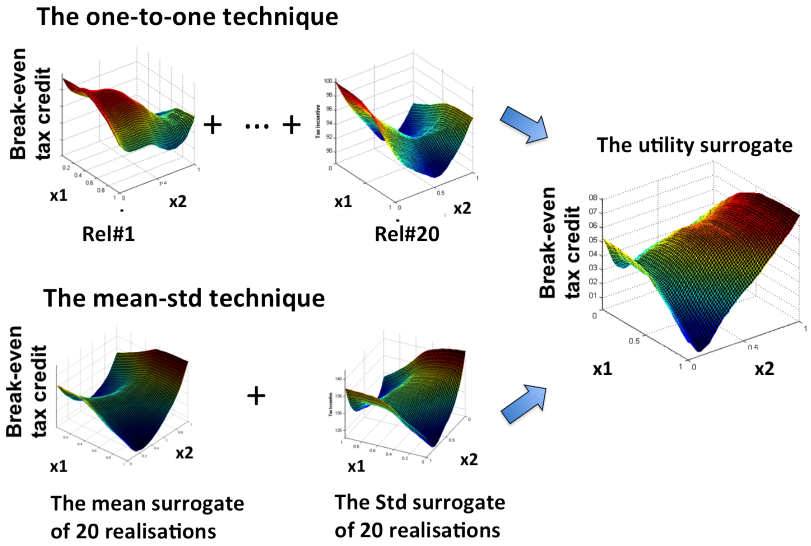


Figure 4.25: Response surfaces shown here are the response of the BT function where the injection rate ratios can be adjusted ($x1, x2$). The figure illustrates two techniques for building the surrogate of the utility function from the ASGI approach; the one-to-one approach on the top row is to build the surrogate for each geological model independently, while the mean-std approach on the bottom row is to build the surrogate of the mean and standard deviation.

Firstly, the spatial error is calculated from the difference in the function values between the reference surface and the surrogate. We use the l_∞-norm as the error measure to evaluate the quality of the surface using $1,000$ points drawn from a uniform distribution over the surrogate domain.

Figure 4.26 presents the error decays of different surrogate models (i.e. the mean, the standard deviation, and the utility function) built using the one-to-one and the mean-std approach. With equivalent maximum errors for all sub-figures, the mean-std technique (the solid curves with circle markers in Fig. 4.26) requires significantly fewer function evalua-

tions than the one-to-one technique (the curves with square markers). In addition, the sparse grid approach (either the one-to-one or the mean-std techniques) still requires fewer grid points than the full grid approach with the same maximum error, which is consistent with the results in Fig. 4.11. Figure 4.26 indicates that the mean-std approach is recommended for building the surrogate of the utility function in the presence of uncertainty.

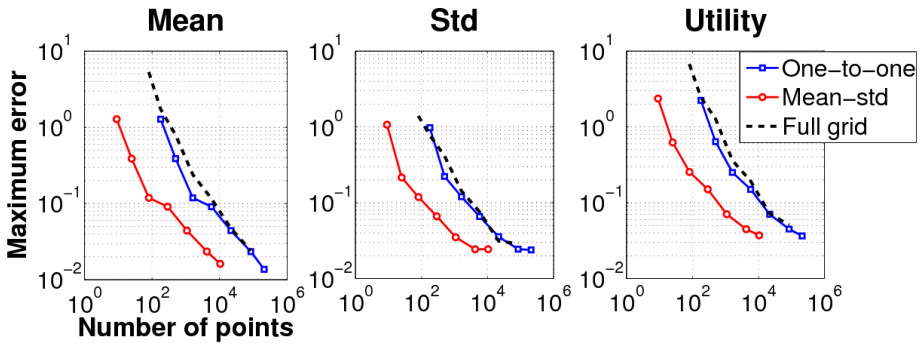


Figure 4.26: The relationship between the maximum error and the number of grid points of the surrogates of the mean, standard deviation and the utility function are illustrated. The solid curve with circle markers represents the surrogate built by the ASGI using the mean-std technique, while the curve with square markers presents the surrogate built by the ASGI using the one-to-one technique. The dashed curve denotes the surrogate constructed from the full grid approach using the one-to-one technique.

Secondly, instead of evaluating $1,000$ spatial sampling points throughout the surrogate, the difference in the function values between the reference surface and the surrogate of the utility function at the minimum point (the minimum point were obtained from the reference surface) was measured to evaluate the deviation of the optimum solution from the surrogate. In this study, the optimal point at the reference surface was obtained from the particle swarm optimiser (PSO) [78, 130].

Figure 4.27 shows the relationship of the error at the optimal solution (the minimum point) versus the number of points required to build the surrogate. The mean-std technique still requires fewer points than the one-to-one approach for the same error measure. To investigate the errors for different global optimal solutions, we changed the shape of the utility response surface by varying the utility parameter (A) in Eq. 4.53 to 0, 0.01, 0.05, 0.1, 1 and 5, as shown in Fig. 4.27. For all utility parameters, we see that the error at the optimal point of the mean-std approach (the curve with circle markers) is less than the one-to-one approach for the same number of points. The results of the mean-std approach using the error measured at the optimum point also corresponds with the results using the spatial error measure in Fig. 4.26.

The observed superior performance of the mean-std technique over the one-to-one approach (as shown in Figs. 4.26 and 4.27) can be attributed to the smoothness of the response surfaces obtained from the mean-std approach in comparison to the surfaces obtained by the one-to-one approach. This smoothness has a great impact on the interpolation errors as spectral convergence can be achieved using Chebyshev points on smooth surfaces [80].

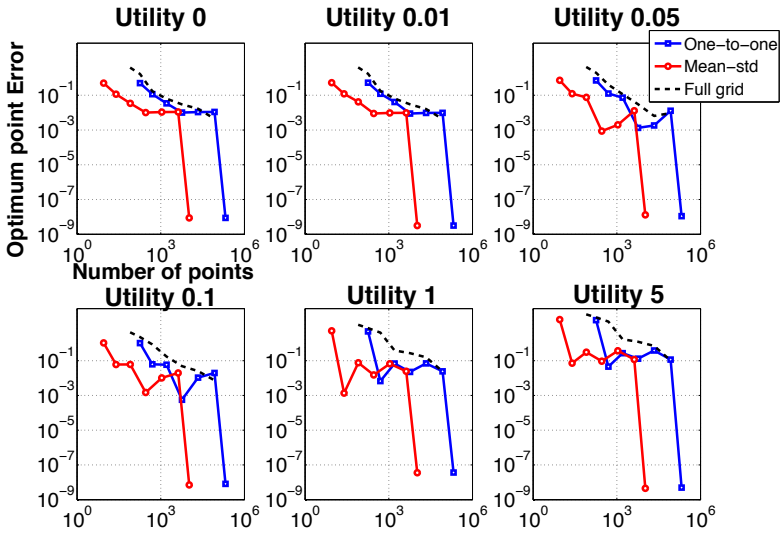


Figure 4.27: The relationship between the error evaluated at the optimal point and the number of grid points required to build the surrogate. The curves with circle markers present the mean-std technique, while the curves with square markers represent the one-to-one technique. Each sub-figure displays the utility parameter (A) used in the utility function.

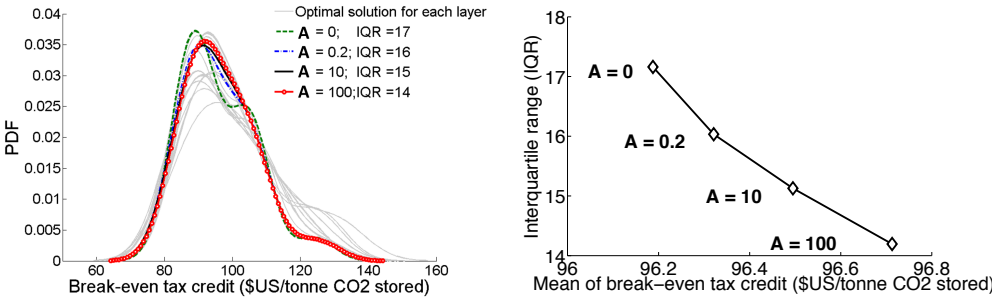
4.5.2.8 Optimal and robust optimisation

The optimal CO_2 injection strategy for one geological realisation may not necessarily be optimal for other geological realisations, and can be a very poor strategy in other realisations. It is quite common to have inaccurate and limited knowledge of geological formations; therefore, we use a strategy that takes into account the range of possible outcomes to find a robust solution.

We demonstrate the optimal solution and the robust solution in Fig. 4.28(a) where a Gaussian probability density function (PDF), estimated from a kernel density function [17] with 100 artificially generated samples, is assumed for the distribution of BT under geological uncertainty. Figure 4.28(a) shows the distribution of the BT values obtained by applying the optimal injection strategy (optimisation based on a single geological realisation) to 20 realisations, shown with the grey thin curves. The figure also shows the distributions of the BT values using the robust strategies with different utility parameters. The associated interquartile range (IQR), defined as the difference between the 75^{th} and 25^{th} percentiles, are used to measure statistical dispersion of the BT values. The dashed curve in Fig. 4.28(a) ($A = 0$) presents the BT distribution obtained from the robust strategy using the utility function without the variance term — a risk-neutral preference. Thus its BT distribution has the greatest interquartile range in comparison to the other BT distributions. In contrast, having a greater value of utility parameter results in a smaller interquartile range because the solution obtained from the utility function tends to minimise the variance term. In addition, we see from Fig. 4.28(a) that the right tail areas of the PDFs obtained from the optimal strategies (thin grey curves) are greater than the area obtained from the robust strategies. This implies the non-robustness behaviour of the optimal strategies, which means that the

optimal strategies applied to some geological realisations can result in poor BT values.

The relationship of the mean and IQR is displayed in Fig. 4.28(b) where a trade-off between the mean and the dispersion (IQR) of the BT is presented. We see that the lower the mean value is, the greater the IQR, or vice versa. Since the utility parameter is a coefficient of the dispersion term in the utility function (Eq. 4.53), the greater the utility parameter, the smaller the dispersion of the BT values around its mean. The trade-off between the mean and the IQR depends on the attitude to risk defined through the utility parameter by decision makers.

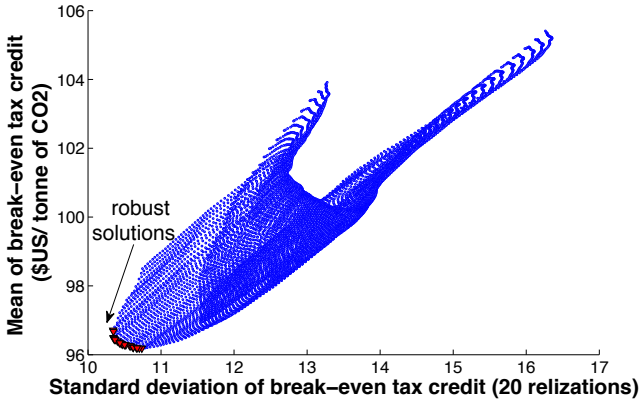


(a) (b)

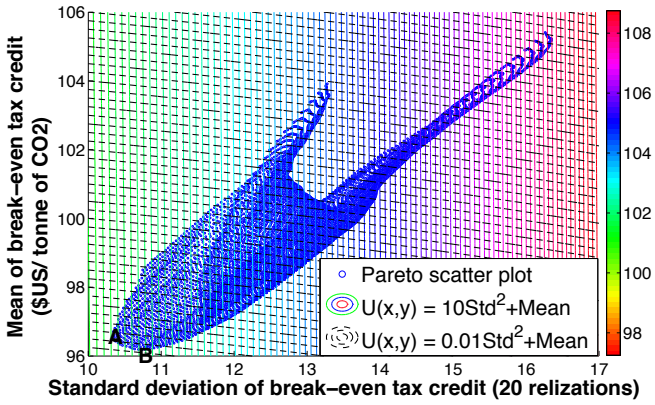
Figure 4.28: a) The probability density function (PDF) of the break-even tax credits (BT) obtained from the robust and optimal injection strategies where the BT distributions were obtained from each injection scenario applying to 20 geological realisations; the thin curves (grey) present the distribution from the optimal injection rates at each geological realisation, while the other curves display the BT distributions obtained from the robust injection strategy at different utility parameters; the dashed curve, the dot-dashed curve, the thick solid curve, and the circle-marker curve dictate the utility parameters at 0, 0.2, 10, and 100, respectively: b) A trade-off between the mean and the dispersion (IQR) of the BT at different robust strategies with the utility parameters A = 0, 0.2, 10, and 100.

To demonstrate the role of utility parameters, Fig. 4.29(a) displays a scatter plot of the mean and standard deviation of the BT using 20 geological realisations; each data point corresponds with the mean and standard deviation obtained from individual injection strategies. The triangle markers at the lower-left corner of the figure present the optimal points as they either provide the minimum value of the mean or the standard deviation. A set of the optimal points is referred to as a Pareto-optimal set from which any injection strategy that corresponds to a point in this set is a robust strategy. In this study, we minimise the utility function to find a solution in the Pareto-optimal set. We present a geometric representation of the utility function in Fig. 4.29(b) where the tangent of the utility function and the Pareto-optimal set is the robust solution. In Fig. 4.29(b), two utility functions with different utility parameters (20 and 0.02) are shown by the coloured and the black-dashed curves. We see that the utility functions monotonically decrease towards the origin. Thus minimising the utility functions with utility parameters 20 and 0.02 results in the solution **A** and **B**, respectively. As can be seen, **A** has a lower standard deviation value than **B** because the utility parameter is the weight of the variance term in the utility function; using a high weight value results in a lower standard deviation.

From the shape of the Pareto-optimal set in Fig. 4.29, two extreme BT values can be observed. Firstly, the minimum of the mean value regardless of the standard deviation is around 96.2 \$US/tonne of CO_2, which implies the scenario of a risk-neutral design (the best scenario we can achieve in terms of average BT). Secondly, the minimum standard deviation of approximately around 10.3 \$US/tonne of CO_2 has the worst mean value at 96.8 \$US/tonne of CO_2 in comparison to other solutions in the Pareto-optimal set. Unless additional geological information is available, the standard deviation of the BT can never be less than 10.3 \$US/tonne of CO_2 in the PUNQ-S3 model irrespective of the injection strategies.



(a) a scatter plot



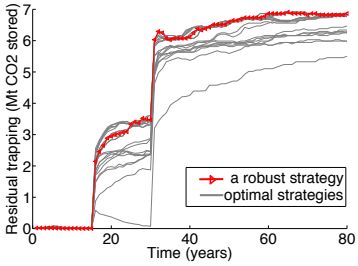
(b) a scatter plot with utility functions

Figure 4.29: a) A scatter plot of the mean and standard deviation of the break-even tax credit with the red markers denoting the robust solutions. The blue data points were evaluated from the SGI surrogate from different injection strategies. b) The scatter plot in a) is shown over the contour of two utility functions from Eq. 4.53 where the utility parameter is 20 (the colour lines) and 0.02 (the dash black lines). The colour bar on the right represents the range of utility values.

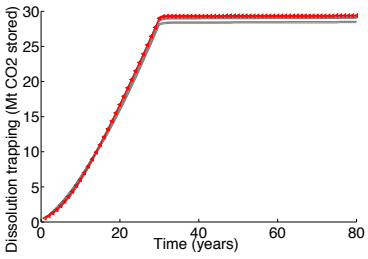
Figures 4.30 and 4.31 present the impact of the optimal strategy on the BT when the uncertainty of the model is not taken into account. In Fig. 4.30, the worst case scenarios for CO_2 in mobile phase, residual and dissolution trapping are illustrated with the correspond-

ing optimal and robust injection strategies; the grey curves represent the optimal injection strategies, each of which is the optimal injection rates obtained from a single geological realisation. The dashed curve corresponds to a robust strategy over 20 geological realisations. We set the utility parameter at 0.5 in the utility function for the robust strategy; the worst case scenario is the realisation that has the highest BT value. We see in Figs. 4.30(a) and 4.30(b) that the worst case of the optimal strategy obtained from a single realisation (grey curves) could lead to poor CO_2 trapping in the presence of uncertainty. In contrast, the robust strategy (the dashed curve) is insensitive to the variability of uncertainties and hence the worst case does not affect the performance criterion of the robust strategy. A similar result is observed in Fig. 4.30(c) where CO_2 in mobile phase is minimised.

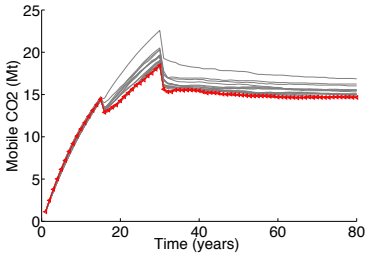
Figure 4.31 presents cumulative distribution functions (CDFs) of CO_2 in mobile phase, residual and dissolution trapping over 20 geological realisations. The CDF is assumed to follow a Gaussian distribution, estimated from a kernel density function with 100 artificially generated samples. The other notations in Fig. 4.31 is the same as in Fig. 4.30. The robust strategy (the dashed curves) in Fig. 4.31 provides satisfactory results in terms of maximising the mean while minimising the standard deviation of the residual and dissolution trapped CO_2 in comparison to the optimal strategy (the grey curves). Since the temporal scale of dissolution trapping could last for 10 to 100 years during the post-injection period [111], it could be one of the reasons why dissolution trapping in Figs. 4.30 and 4.31 shows small variability in this study, which is limited to 50 years for the post-injection period. In addition, the abrupt changes of mobile phase and residual trapping at 15 and 30 years are due to the changes of injection rates of CO_2.



(a) Residual trapping

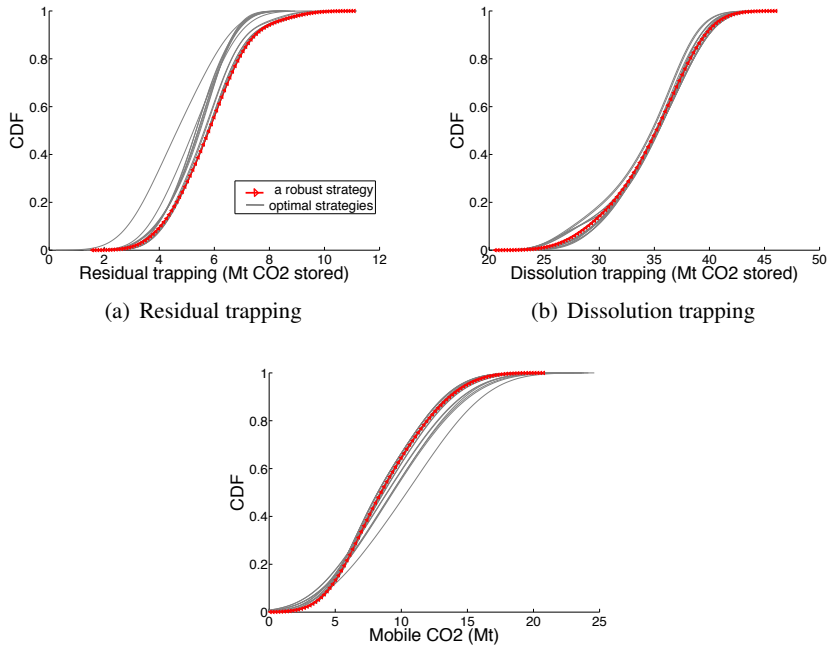


(b) Dissolution trapping



(c) Mobile phase of CO_2

Figure 4.30: The worst case scenarios of the optimal strategy obtained from a single realisation (the grey curves) and the robust strategy (the dashed curves) are shown during the injection and post-injection period in various forms of CO_2; a) CO_2 residual trapped (Mt), b) CO_2 dissolution trapped (Mt), and c) the mobile phase (Mt).



(a) Residual trapping

(b) Dissolution trapping

(c) Mobile phase of CO_2

Figure 4.31: The cumulative distribution (CDFs) of the optimal strategy obtained from a single realisation (the grey curves) and the robust strategy (the dashed curve) are shown during the injection and post-injection period in various forms of CO_2; a) CO_2 residual trapped (Mt), b) CO_2 dissolution trapped (Mt), and c) the mobile phase (Mt).

4.6 Summary

We developed an optimisation technique by using the Adaptive Sparse Grid Interpolation (ASGI) technique with the use of the utility function to find robust CO_2 injection strategies under uncertain formation characteristics, such as permeability, porosity, and anisotropy ratios. Our results are summarised as follows:

1. We proposed a new objective function that considers economic performance of CO_2 sequestration. The break-even tax credits (BT) is inversely proportional to the mass of trapped CO_2 in the subsurface. Additionally, it represents the minimum tax credit gained from the government such that the NPV of the sequestration project is positive. Hence, the BT objective function was minimised to maximise the economic performance of the project.

2. To cope with geological uncertainty, a utility function was formulated using both the mean and the variance of the BT function for CO_2 optimisation under uncertainty. We built the surrogate using both adaptive and non-adaptive sparse grid interpolation approaches to approximate the utility function, which consequently reduced the computational cost associated with the optimisation process.

3. Sparse Grid Interpolation (SGI) requires fewer function evaluations than the full grid interpolation approach with the same error for the mathematical test functions. The

SGI selects only the grid points that contribute the most to building the surrogate and discards less important grid points. This advantage is achieved through the selection of the grid points at different hierarchical levels in each dimension. Due to the minimal grid points required, the SGI is suitable for CO_2 sequestration and appears to be a promising surrogate technique for other situations that would normally require repeated time-consuming flow simulation problems.

4. Adaptive Sparse Grid Interpolation (ASGI) distributes the layout of the grid points in each dimension differently depending on the importance of the dimension. The grid points are more refined intentionally in the significant dimension with respect to the pre-defined threshold value. As a result, ASGI requires fewer function evaluations in comparison to SGI. The results suggested using dimensional adaptivity for building the surrogate.

5. Two techniques for building the utility surrogate were considered in this study: the one-to-one approach, and the mean-std approach. We evaluated the quality of the surrogate with the reference model by using two error criteria: the maximum spatial error and the error at the optimum point. The results revealed that the mean-std approach requires fewer grid points than the one-to-one approach at the same error for the two criteria. The mean-std approach therefore is a better technique for building the utility surrogate with multiple model realisations.

6. Based on the PUNQ-S3 model, we showed that the robust solution obtained from the utility function relies on the preference of the decision makers and their attitude towards risk. The worst case scenarios were used to compare CO_2 in mobile phase as well as trapped as a residual phase and by dissolution from the robust and optimal injection strategy. The results revealed that the worst case of the optimal strategy could lead to poor CO_2 sequestration economic performance in contrast to the robust strategy.

CHAPTER 5

MULTI-OBJECTIVE OPTIMISATION

5.1 Introduction

This chapter introduces the fundamental principles of multi-objective optimisation (MOO), followed by a brief review of several MOO techniques commonly applied in engineering applications. Next, we describe the MOO algorithm used in this thesis, followed by the discussions of MOO problems under uncertainty and the use of adaptive sparse grid interpolation for the MOO algorithm. Then we end the chapter with the applications to CO_2 sequestration.

5.2 Multi-objective optimisation (MOO)

Multi-objective optimisation (MOO) deals with problems when more than one decision criterion is considered. In contrast to single-objective optimisation, we aim to find optimal solutions for multiple decision criteria. Let M, $\mathcal{F}_x$, and $f_m(\mathbf{x})$ be the number of objectives, the criterion domain, and the objective function at m^{th} criterion, respectively. If we define $f_m(\mathbf{x}) : \mathcal{S}_x \rightarrow \mathcal{F}_x$, then the MOO under a minimisation framework is given as:

$$\begin{aligned} \text{Minimise:} \quad & f(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_M(\mathbf{x})] \\ \text{where} \quad & \mathbf{x} \in \mathcal{S}_x,\ f_m \in \mathcal{F}_x,\ m \in \mathbb{Z}^+ \\ & \mathcal{F}_x = \{f(\mathbf{x}) | f(\mathbf{x}) \in \mathbb{R}^M, \mathbf{x} \in \mathcal{S}_x, \} \\ & \mathcal{S}_x = \{x | x \in \mathbb{R}^n, h(\mathbf{x}) = 0, g(\mathbf{x}) \geq 0\} \end{aligned} \tag{5.1}$$

where $\mathbf{x}$ is a vector of n factors: $\mathbf{x} = (x_1, \ldots, x_n)$ with a domain $\mathcal{S}_x$; $h(\mathbf{x})$ and $g(\mathbf{x})$ represent linear (or non-linear) equality and inequality constraints, respectively. $\mathbb{Z}^+$ denotes a set of positive integers. Later, we shall refer to a feasible set on $\mathcal{F}_x$ as an attained set even though feasibility and attainability have slightly different meanings. Marler and Arora [98] showed that every point in the design space $\mathcal{S}_x$ maps to a point in the criterion space $\mathcal{F}_x$ but not every point in $\mathcal{F}_x$ maps to a point in $\mathcal{S}_x$. Hence, attainability implies that a point in the criterion space $\mathcal{F}_x$ corresponds to a point in $\mathcal{S}_x$, whereas feasibility of the criterion space is the space such that the constrain conditions are not violated. In this thesis, we consider the space $\mathcal{F}_x$ such that it is feasible and attainable.

As described by Deb and Gupta [35], the goal of MOO is to search for solutions in $\mathcal{F}_x$ such that they satisfy two or more criteria: an optimal trade-off between different objectives and the diversity of the obtained optimal solutions. These criteria will be discussed in the following subsection where we first describe the concept of domination, weak and strong Pareto-optimal set, and dependency of the criteria.

5.2.1 Concept of dominance and a Pareto-optimal set

In MOO, the solutions $f(\mathbf{x})$ in criterion space $\mathcal{F}_x$ are compared among other solutions; the solutions that dominate the others are grouped to a non-dominated set where each member does not dominate the other members. Figure 5.1 illustrates the concept of dominance

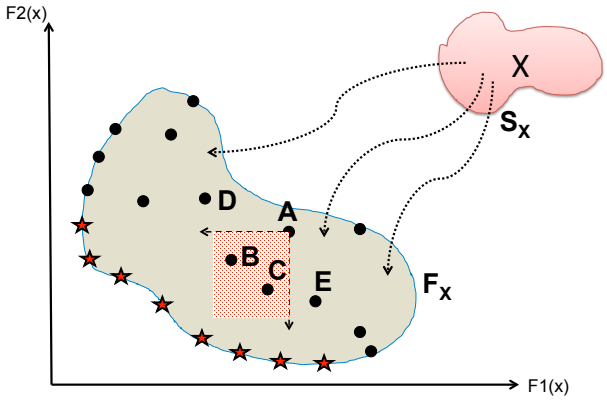


Figure 5.1: An illustration of multi-objective optimisation (MOO); the dashed arrow lines present a domain mapping from the variable domain $\mathcal{S}_x$ to the criterion domain $\mathcal{F}_x$. The black dots represent the data in the domain $\mathcal{F}_x$, each of which is compared to find a non-dominated solution. The stars-shape marks denote the non-dominated set or the Pareto-optimal set. The shaded area near the solution A demonstrates that any solution within this area dominate the solution A.

for a minimisation problem (both objective functions are minimised); the dots represent solutions in $\mathcal{F}_x$ that are mapped from variables $\mathbf{x}$ in the design space $\mathcal{S}_x$. Any solution in the shaded area, e.g. B and C, dominates the solution A whereas the solutions outside this area, e.g. D and E, do not dominate the solution A. If the members of the non-dominated set are obtained from the entire criterion space, this set is a Pareto-optimal set. We see from Fig. 5.1 that the star-shape markers are the solutions in the Pareto-optimal set as no other solutions outside this set in $\mathcal{S}_x$ can dominate the member in this set. Hence, the optimal solution in the MOO is a set of solutions rather than a single solution as in the single-objective optimisation (SOO). Dominance concept is an essential concept for MOO. The mathematical expression of the dominance concept and the related terms are given by Deb [34] and in the following.

Given the operator $f_i(\mathbf{x}) \lhd f_j(\mathbf{x})$ expresses that the objective i is better than the objective j and the operator $f_i(\mathbf{x}) \rhd f_j(\mathbf{x})$ means that the the objective i is worse that the objective j while $f_i(\mathbf{x}) \not\rhd f_j(\mathbf{x})$ means that the objective i is no worse than the objective j. If all objectives are minimised, the symbol '$\lhd$' can be simply replaced by '$<$'. The concept of dominance is given by [34] below.

Definition 1: *A solution* $\mathbf{x_1}$ *is said to dominate a solution* $\mathbf{x_2}$ *(Or mathematically,* ($\mathbf{x_1} \preceq \mathbf{x_2}$) *if and only if*

(1) *The solution* $\mathbf{x_1}$ *is no worse than the solution* $\mathbf{x_2}$ *for all objectives:*
$f_i(\mathbf{x_1}) \not\rhd f_i(\mathbf{x_2})$ *for* $\forall i \in \{1, \ldots, M\}$

(2) *The solution* $\mathbf{x_1}$ *is strictly better than the solution* $\mathbf{x_2}$ *at least one objective:*

$f_j(\mathbf{x_1}) \lhd f_j(\mathbf{x_2})$ *for* $\exists j \in \{1, \ldots, M\}$

Based on the dominance concept from the Definition 1, Fig. 5.2 displays four types of Pareto-optimal front for two objectives. The 'Min–Max' shown in the figure represents the minimisation of the objective 1 (x-axis) while maximising the objective 2 (y-axis). However, in our study, we are only interested in minimisation of all objectives from which the Pareto-optimal front is at the lower left corner.

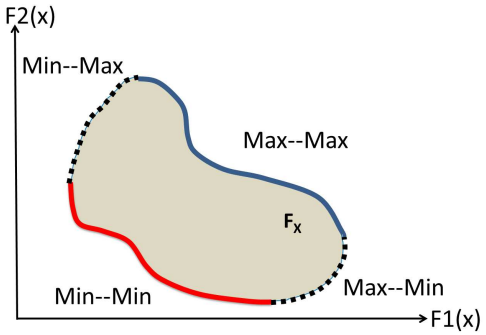


Figure 5.2: Four types of Pareto-optimal front for two objectives. The *Min–Min* represents the minimisation of both objective 1 (x-axis) and 2 (y-axis), where the red line shows the Pareto front. While the dotted line near the *Min–Max* represents the Pareto front for minimising the objective 1 while maximising the objective 2. The other types of Pareto front follow the same logic here.

5.2.2 Concept of weak and strong dominance

The concept of dominance can be further characterised by strong and week dominance. Definition 1 is referred to as the weak dominance, whereas strong dominance is expressed by Deb [34] in the following:

Definition 2 : *A solution* $\mathbf{x_1}$ *is said to strongly dominate a solution* $\mathbf{x_2}$ *(Or mathematically,* $(\mathbf{x_1} \prec \mathbf{x_2})$ *if and only if*

(1) *The solution* $\mathbf{x_1}$ *is strictly better than the solution* $\mathbf{x_2}$ *for all objectives:*

$f_i(\mathbf{x_1}) \lhd f_i(\mathbf{x_2})$ *for* $\forall i \in \{1, \ldots, M\}$

A solution denoted by $\mathbf{x^s}$ is a strong Pareto optimum if and only if there is no $\mathbf{x} \in \mathcal{S}_x$ such that $f_i(\mathbf{x}) \leq f_i(\mathbf{x^s})$ for all $i \in \{1, \ldots, M\}$ which at least one objective satisfies strict inequality $f_i(\mathbf{x}) < f_i(\mathbf{x^s})$. However, a solution $\mathbf{x^w}$ is a weak Pareto optimum if and only if there is no $\mathbf{x} \in \mathcal{S}_x$ such that $f_i(\mathbf{x}) < f_i(\mathbf{x^w})$ for all objectives. It is worth noting that the strong Pareto optimum $\mathbf{x^s}$ also satisfies a weakly dominance, but not for vice versa [35].

Figure 5.3 displays the weak and the strong Pareto-optimal solutions as the square and the triangle markers, respectively. We see that points **a**, **b**, **c**, **d** and **e** are weak Pareto optima $\mathbf{x^w}$ as there is no $\mathbf{x}$ such that $f_i(\mathbf{x})$ is less than $f_i(\mathbf{x^w})$ for all objectives; however, only the points **b**, **d** and **e** are strong Pareto optima $\mathbf{x^s}$ because there are no $\mathbf{x}$ such that $f_i(\mathbf{x})$ is less

than or equal to $f_i(\mathbf{x}^{\mathbf{s}})$ for all i. We see that **a** is not a strong Pareto optimum because there is a solution **b** such that $f_i(\mathbf{b}) \leq f_i(\mathbf{a})$ for $i = 1, 2$; the same logic is applied to point **c** where there are points **b** and **d** such that $f_i(\mathbf{b}) \leq f_i(\mathbf{c})$ and $f_i(\mathbf{d}) \leq f_i(\mathbf{c})$ for $i = 1, 2$.

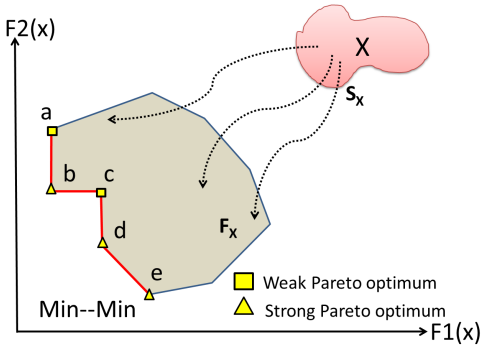


Figure 5.3: Demonstration of weak and strong Pareto-optimal solutions: points **a**, **b**, **c**, **d**, and **e** are all weak Pareto optima while only the points **b**, **d**, and **e** are strong Pareto optima.

5.2.3 Concept of efficiency and efficient frontier

In addition to the concept of dominance, it is noteworthy that efficiency is another concept of dominance in that it refers to the solution in the design spaces $\mathcal{S}_x$ rather than the criterion space $\mathcal{F}_x$. As described by Marler and Arora [98], the efficiency and efficient frontier are given by:

Definition 3 : Efficient and inefficient: *A point, $\mathbf{x}^* \in \mathcal{S}_x$, is efficient if and only if there does not exist a point, $\mathbf{x} \in \mathcal{S}_x$, such that $f_i(\mathbf{x}) \leq f_i(\mathbf{x}^*)$ for all objectives, with at least one objective satisfies $f_i(\mathbf{x}) < f_i(\mathbf{x}^*)$. Otherwise, $\mathbf{x}^*$ is inefficient.*

Definition 4 : *A set of solutions $\mathbf{x} \in \mathcal{S}_x$ that satisfies Definition 3 is an efficient frontier.*

5.2.4 Globally and locally Pareto-optimal front

A set of solutions is said to be a globally Pareto-optimal set if the Pareto-optimality conditions are satisfied for entire design space $\mathcal{S}_x$. However, if the conditions of Pareto optimality are satisfied around the feasible neighbourhood π of $\mathbf{x}$, it is said to be a local Pareto optimum. For example, let $R(f(\mathbf{x}^{\mathbf{w}}))$ be a neighbourhood around $f(\mathbf{x}^{\mathbf{w}})$ in $\mathcal{F}_x$ which is mapped from the neighbourhood $R_\pi(\mathbf{x}^{\mathbf{w}})$ with a radius $\pi > 0$ in $\mathcal{S}_x$; the solution is said to be a locally weak Pareto optimum for the given radius π if and only if there is no $\mathbf{x} \in \{\mathcal{S}_x \cap R_\pi(\mathbf{x}^{\mathbf{w}})\}$ such that $f(\mathbf{x}) < f(\mathbf{x}^{\mathbf{w}})$ for all objectives. We demonstrate the globally and locally Pareto-optimal fronts in Fig. 5.4 where it should be noted that the region $R(f(\mathbf{x}^{\mathbf{w}}))$ may not be continuous when it is mapped from the continuous $R_\pi(\mathbf{x}^{\mathbf{w}})$. Later in this chapter, we will discuss the evolutionary algorithm that can find the globally Pareto-optimal set.

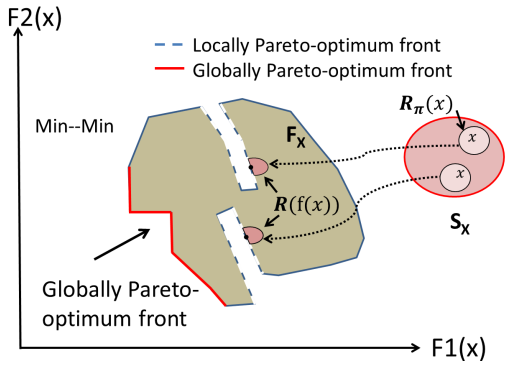


Figure 5.4: Demonstration of global and local Pareto-optimal fronts. The circle presents $R_\pi(\mathbf{x})$ as the feasible neighbourhood of the local Pareto optimum and the half sphere represents $R(f(\mathbf{x}))$ as the neighbourhood in the criterion space of the locally optimum solution.

5.2.5 Dependency of criteria

The principle of MOO techniques may not be strictly applied to optimisation problems with multiple objectives. Since most MOO techniques are developed to assist decision makers to find an optimal trade-off among objectives, some problems with multiple objectives that do not exhibit the trade-off between objectives may not be suitable for the MOO. This depends on the way each objective is optimised, and is not limited to the conflicting (or non-conflicting) behaviour of the objectives. For instance, if two objective functions are expected to be minimised, the conflicting nature of the objective functions should exist such that it is possible to obtain the optimum trade-off relationship from both criteria. However, if two objective functions are non-conflictive, there would be only one optimal solution in this case. Hence, it could be more efficient to use SOO techniques by combining multiple objectives to a single objective for optimisation. It should however be noted that the presence of negative (or positive) dependence of different criteria and the way in which each objective function is optimised should be considered thoroughly so that the MOO technique can be applied efficiently.

Figure 5.5 illustrates this point by presenting the criterion space $\mathcal{F}_x$ of bi-objective optimisation. Similar to Fig. 5.1, the Pareto-optimal fronts and four possible ways of optimising two objective functions are shown. As can be seen in the figure, using SOO could be sufficient for the Min-Min and Max-Min optimisations since there is only one solution (the star shaped marker). However, MOO techniques could be used to obtain the Pareto-optimal front for the Min-Max problem because there is a trade-off relationship between both objective functions, which has a positive dependence (one objective function increases or decrease with the other). Additionally, if both objective functions are maximised, the MOO technique could also be applied to obtain the Pareto-optimal set even though the negative dependence of the criteria is present. The other aspects and detailed discussions of MOO can be found in Deb [34], whereas we only highlight some important concepts of MOO techniques in this thesis.

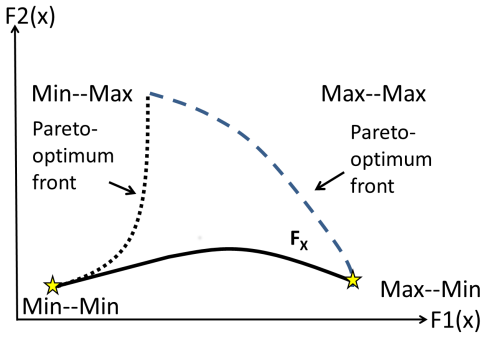


Figure 5.5: Demonstration of Pareto-optimal fronts in the criterion space $\mathcal{F}_x$ of bi-objective optimisation. The Pareto-optimal front of the positive-dependence criteria is shown at the upper-left corner, while the upper-right corner presents the Pareto-optimal front of the negative-dependence criteria. The optimal solutions from the Min-Min and Max-Min optimisation are denoted by the star shaped markers.

5.3 A survey of multi-objective optimisation techniques

MOO techniques can be characterised based on the articulation of preference towards the criteria from decision makers. The preference and optimisation procedure are used to define different techniques for solving MOO problems. Three main articulations of preference are: 1) Prior preference: the decision maker indicates his preference prior to the optimisation procedure. 2) Posterior preference: the techniques that help decision makers to chose the solution after optimisation process. 3) No preference: there is no sufficient information for decision maker to make a preference. In addition to this, one might consider an interactive method when the preference is made real time or in each iteration of the optimisation algorithm. Detailed descriptions are given in [98]: we provide a brief review in the following.

5.3.1 Prior articulation of preference methods

An a priori preference technique is an approach to reducing the degrees of freedom associated with solving MOO problems. Since the optimisation complexity of MOO problems increases with the number of objectives, the prior preference technique reduces the degree of freedom of the problem by determining or ranking the relative importance of each criterion before performing the optimisation process. This can be done by assigning the relative weight associated with each objective to form a preference function or a utility function; this function transforms multiple objective functions to a single objective function such that the techniques used for SOO can be applied.

As described in [34, 98], the prior articulation methods are: weighted-global criteria method, weighted sum method, lexicographic method, min-max method, exponential weighted sum, weighted product method, and goal programming method. However, it should be noted that the underlying principle of these techniques is a preference-based approach from different utility functions. In this thesis, we will highlight only common approaches.

5.3.1.1 Weighted sum approach

The weighted sum method is commonly used to solve MOO problems. It has been applied to many fields, one of which is the cost functions or the Net Present Values (NPV). For example, the NPV function commonly used in oil and gas industries is simply a utility function of oil/gas production and water production/injection, which are grouped by the weight factors as a single function. In general, the weighted sum method is given in the following form:

$$U = \sum_{i=1}^{m} w_i f_i(\mathbf{x}) \tag{5.2}$$

$$\text{where: } \sum_{i=1}^{m} w_i = 1 \ , \text{ and } w_i > 0$$

where U denotes the utility function; w_i is a weight factor for each objective i. It should however be noted that the weights are not necessarily a positive number, but the positive number of all the weight factors provides a sufficient condition for Pareto optimality [35].

5.3.1.2 Weighted exponential sum method

This method is similar to the weighted sum approach but with an exponential parameter p.

$$U = \sum_{i=1}^{m} w_i [f_i(\mathbf{x})]^p \ , \ f_i(\mathbf{x}) > 0 \quad \text{for } \forall i \tag{5.3}$$

$$U = \sum_{i=1}^{m} [w_i f_i(\mathbf{x})]^p \ , \ f_i(\mathbf{x}) > 0 \quad \text{for } \forall i \tag{5.4}$$

The parameter p can be fixed or varied to increase the effectiveness of the method in order to capture the solutions in the Pareto-optimal set. Figure 5.6 illustrates the weighted sum and exponential sum approaches with the Pareto optimum solution obtained from the tangent of the curves to the Pareto-optimal front. Generally, the parameter p is used to change the shape of the utility function in the criterion space, which results in different Pareto-optimal solutions.

5.3.1.3 Value function method

This method enables users a flexibility to specify his/her own preference towards each objective by providing a mathematical function $U : \mathbb{R}^M \rightarrow \mathbb{R}$. The root of the utility-based approach stems from decision theory, which has been intensively applied to the area of economics. This method is formulated as:

$$\begin{aligned} &\text{Maximise } \ U(f(\mathbf{x})) \\ &\text{subjected to } \ \mathbf{x} \in \mathcal{S}_x \end{aligned} \tag{5.5}$$

In order to use Eq. 5.5 to solve MOO problems, Rosenthal [140] showed that the function must be monotonically decreasing such that the concavity of the function is held such that the maximisation of Eq. 5.5 can be used to obtain the Pareto-optimal solution.

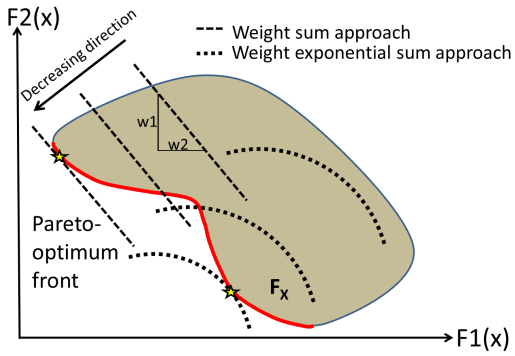


Figure 5.6: Geometrical illustration of the weighted sum and the exponential sum approaches. The dashed curves represent the weighted sum approach decreasing towards the Pareto-optimal front, while the dotted curves present the exponential weight approaches. The weight factors w_1 and w_2 are the slope of the curves.

It is worth noting that priori preference methods aim to combine multiple objectives to a single objective function where the SOO technique is used to solve for a single optimum solution. However, it can be used to find a set of Pareto-optimal solutions by varying multiple model parameters, such as the utility parameter described in Chapter 4, and the exponential parameter p in Eqs. 5.3 and 5.4.

5.3.2 Posterior articulation of preference methods

The aim of a posterior articulation method is to provide a set of Pareto-optimal solutions so that decision makers can make their preference over the optimal choices. These techniques may not be efficient if the computational cost associated with each function evaluation is large since the efforts spent on the search for the set of solutions may not be feasible in the case that we only need one solution. However, posterior preference-based methods enable the users to select the optimal solution from the set of solutions by not necessarily specifying the relative importance of the objective a priori. Theses methods including physical programming, normal boundary intersection (NBI), and the normal constraint (NC) method can be found in Deb [34]. In this study, we use an evolutionary algorithm based on a non-dominated sorting procedure to obtain the set of Pareto optimum solutions.

5.3.2.1 Non-dominated sorting genetic algorithm

Most multi-objective evolutionary algorithms (MOEAs) use ranking schemes for sorting non-dominated solutions. However, the main difficulties of MOEAs used to be their computational complexity and lack of keeping useful solutions during an evolution process. An elitist multi-objective evolutionary optimisation was developed to overcome this challenge, the technique of which includes a Non-dominated Sorting Algorithm (NSGA-II) [36], a Pareto-archived evolution strategy (PAES) [82], and a strength-Pareto EA (SPEA) [179]. The non-dominated sorting procedure based on a genetic algorithm (GA) was first proposed by Srinivas and Deb [164] and was further developed by Deb et al [36] to improve computational efficiency and diversity of Pareto-optimal solutions. Deb et al [36] showed that the

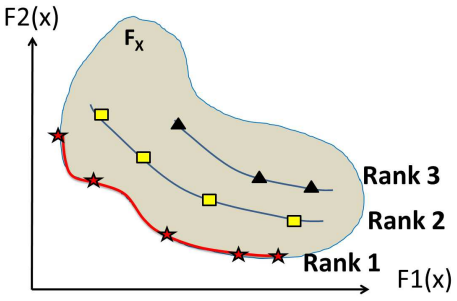


Figure 5.7: Illustration of Pareto-front with sorted rank for minimisation of two objectives.

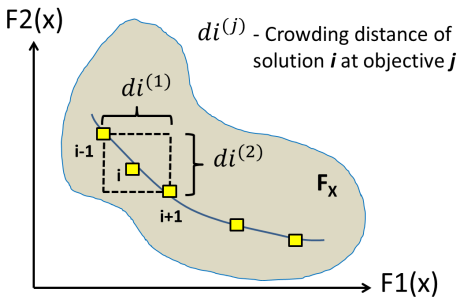


Figure 5.8: Illustration of crowding distance of solution i for bi-objective problems.

NSGA-II outperforms the SPEA and PAES approaches on several test functions. In the past ten years, the NSGA-II has therefore been applied to many engineering applications, such as biomedical engineering [65], economics and finance [168], and civil engineering [176], among others. For this reason, we will use the NSGA-II algorithm, proposed in [36, 164], for CO_2 sequestration applications in this study.

The detailed description of the NSGA-II can be found in [34, 36]: here we highlight important aspects of the algorithm. The NSGA-II has two main features: 1) Fast non-dominated sorting. 2) Crowding distance and elitism; the first feature is a density estimator to emphasise diversity of the solutions while the second allows the best solutions to be carried over to the next generation.

Fast non-dominated sorting approach : Non-dominated solutions are sorted based on two principles: 1) rank classification of non-dominated solutions, and 2) a domination counter, which assigns an individual solution the number of solutions that dominates this individual. Non-dominated solutions obtained from the whole population are grouped as a non-dominated set for the first rank where a domination counter is set to zero. Thereafter, the domination counters of other solutions excluding the set of the first rank are reduced by one. The solutions that have the counter equal to zero are grouped to the second rank. The procedure is repeated until different ranks of non-dominated solutions are obtained. An illustration of different ranks of non-dominated solutions for a minimisation problem is shown in Fig. 5.7.

Diversity preservation and elitism concept : A density estimator using a crowding distance indicates a relative Euclidean distance of an individual solution to its neighbourhood in the criterion space. A larger average crowding distance indicates a greater individual diversity in the population. The crowding distance can be advantageously used to compare the solutions with the same rank of the non-dominated set. An illustration of crowding distance is shown in Fig. 5.8. The crowding distance can also be used in a tournament selection for generating a population of children using a GA procedure. To preserve elitism, Deb [34] defined a crowd tournament operator as stated in the following.

Definition 5 : *In a tournament selection, a solution i is preferred to a solution j if one of the following is true:*
1. If a solution i has a lower rank of the non-dominated set than a solution j.
2. If a solution i has the same rank of the non-dominated set but the solution i has a larger crowding distance than a solution j.

A flowchart of the NSGA-II procedure is shown in Fig. 5.9, where P_t and Q_t denote parent and child populations of size N. It is noteworthy that sorting the rank in the elitism process is done for N individuals since only N parents of the total 2N will be selected as the next generation of a parent population. Hence, sorting the Pareto rank and selecting N individuals are a simultaneous process. In this thesis, we use 20 as a population, unless otherwise stated.

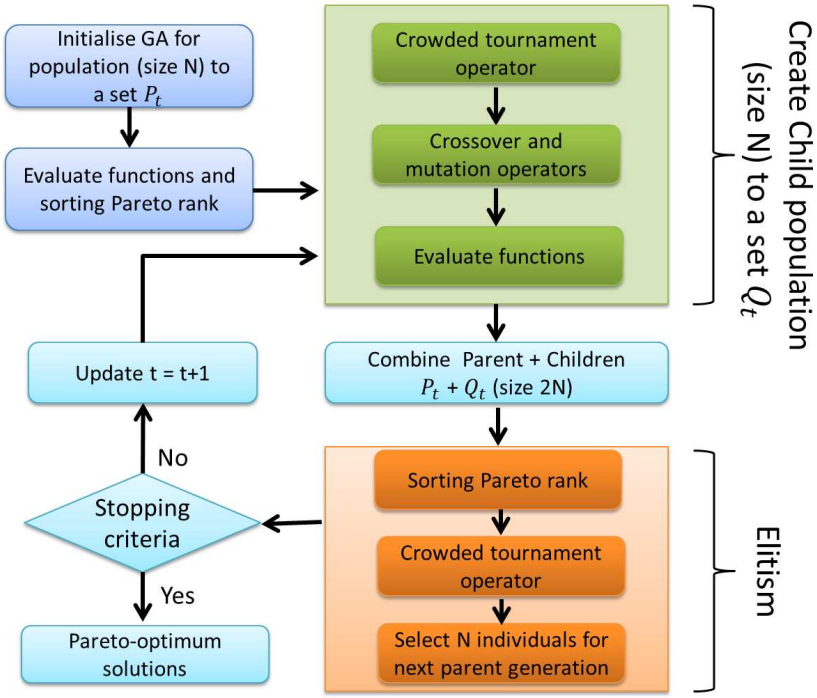


Figure 5.9: A flowchart of NSGA-II procedure.

5.3.3 No articulation of preference method

As described previously, this method is suitable when the designer has no preference to the design. The method of no-preference is similar to the prior articulation of preference where model parameters such as the exponential parameter p and the weight factor $\mathbf{w}$ are set to one or typically excluded so that each objective is equivalently important.

5.3.4 Discussion

1. The priori articulation of preference saves computer time since there are fewer degrees of freedom in the search space of SOO problems $f(\mathbf{x}) : \mathbb{R}^{\mathbb{M}} \rightarrow \mathbb{R}$. Also, this

method is recommended in applications where only one optimal solution is required since finding a set of Pareto-optimal solutions will be wasted and may not be nesccesary. The example in the previous chapter demonstrates the utility function applied to robust optimisation. For that application, the risk preference is determined a priori and the trade-off between optimum solutions is determined by varying the utility parameter.

2. The posterior articulation of preference technique is suitable when the relative importance of the objectives is not known. Hence, providing a set of optimum solutions allows decision makers to state their preference after the optimisation procedure. In addition, a set of Pareto-optimal solutions reveals the trade-off among objectives, which can be useful for decision making by multiple decision makers. However, finding a set of Pareto-optimal solutions is generally more computationally intensive than solving problems for a single solution. We overcome this challenge by using the surrogate-assisted approach to reduce this computational burden. Therefore, the cost of computation is relatively cheap in comparison to the cost of building the surrogate.

3. For the posterior preference-based method, users make a decision based on a set of Pareto-optimal points. The selection of these solutions would be difficult if there are large numbers of objectives involved. Even though most techniques are capable of solving many objective functions, it should be noted that the number of objectives should be small such that the selection process of the optimum points is efficient. Later in the chapter, we will present methods that allow us to efficiently select solutions from a pool of Pareto-optimal solutions.

5.4 Multi-criterion optimisation under uncertainty

The main goal of MOO under uncertainty is to find a Pareto-optimal front that is insensitive to the sources of variability. These variabilities, such as uncertainty of input variable, and uncertainty embedded in the model, are two main factors in a robust design. Figure 5.10 illustrates a relationship between the input variables in $\mathcal{S}_x$ and their corresponding objectives $f(\mathbf{x})$ in $\mathcal{F}_x$. A smaller variability (small circle) of $f(\mathbf{x})$ in the criterion space is said to be more robust since the objective $f_i(\mathbf{x})$ is insensitive to the uncertainty. If variance is the measure of variability, we see in the figure that the solution x_2 is more robust than the solution x_1 because its probability density function (PDF) in $\mathcal{F}_x$ has a smaller variance.

In this thesis, we review different methods that handle MOO under two sources of uncertainty: the uncertainty related design variables and the model uncertainty. Even though these sources of uncertainty can both be embedded in a system, we present each technique separately. Two sources of uncertainty are demonstrated in Fig. 5.11 where we first show that the uncertainty of input variables $\tilde{\mathbf{x}}$, denoted by the local neighbourhood around $\mathbf{x}$, propagates through a certain model to the criterion space, and consequently causes the variability of the objectives in $\mathcal{F}_x$. As well as the input uncertainty, any certain input variables entering an uncertain model to the criterion space also causes a variability of the objective $f(\mathbf{x})$. The techniques handling these uncertainties are described in the following sections.

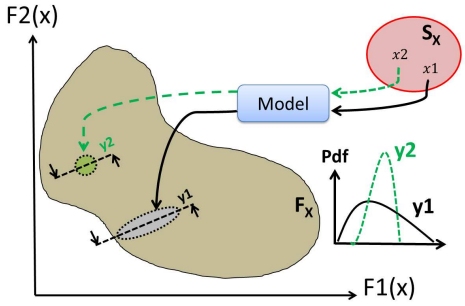


Figure 5.10: An illustration of MOO under uncertainty.

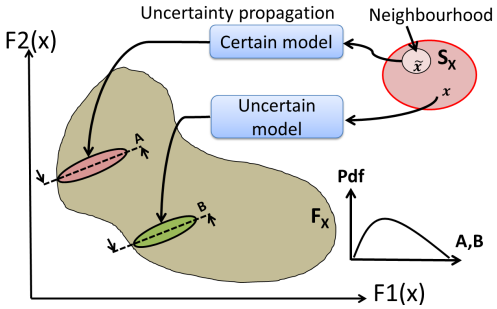


Figure 5.11: An illustration of MOO under two sources of uncertainty.

5.4.1 MOO under input uncertainty

Deb and Gupta [35] described two methods for handling input uncertainty as follows: 1) an effective mean method, and 2) a constrained MOO technique. Given that $R_\pi(\mathbf{x})$ presents the neighbourhood around $\mathbf{x}$ at radius π and $f_i^e(\mathbf{x})$ denotes the effective mean of the objective i, this method is similar to Eq. 5.1 by replacing $f_i^e(\mathbf{x})$ for all objectives. The effective mean is defined:

$$f_i^e(\mathbf{x}) = \frac{1}{|R_\pi(\mathbf{x})|} \int_{\mathbf{z} \in R_\pi(\mathbf{x})} f_i(\mathbf{z}) \mathrm{d}\mathbf{z} \tag{5.6}$$

where $|R_\pi(\mathbf{x})|$ denotes the hypervolume of the neighbourhood around $\mathbf{x}$. This equation is basically an average of the objective function $f_i(\mathbf{x})$ around its input neighbourhood. It, however, should be noted that the robust solution may be sensitive to the chosen radius π, and the local continuity mapping from neighbourhood $\mathbf{x}$ to criterion space should exist.

A constraint imposed on MOO can be used to handle input uncertainty as described in [35]:

$$\begin{aligned}
&\text{Minimise:} && f(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_M(\mathbf{x})] \\
&\text{subject to:} && (1)\ \frac{\|f^p(\mathbf{x}) - f(\mathbf{x})\|}{\|f(\mathbf{x})\|} \leq \eta && (5.7) \\
&\text{or} && (2)\ \|f^p(\mathbf{x}) - f(\mathbf{x})\| \leq \eta && (5.8) \\
&\text{or} && (3)\ \frac{1}{M} \sum_{j=1}^{M} \frac{f_j^p(\mathbf{x}) - f_j(\mathbf{x})}{f_j(\mathbf{x})} \leq \eta && (5.9) \\
&\text{or} && (4)\ \left\| f_j^p(\mathbf{x}) - f_j(\mathbf{x}) \right\| \leq \eta_j && (5.10)
\end{aligned}$$

where the constraints Eqs. 5.7 – 5.10 are defined in various forms of measure with η presenting a degree of robustness; the operator $||\cdot||$ is a user defined norm measure. $f^p(\mathbf{x})$ represents a locally perturbed function (or the worst objective function) within its neighbourhood $R(f(\mathbf{x}))$. The other variables (e.g. $M, f, \ldots$) were described previously in Eq. 5.1.

5.4.2 MOO under model uncertainty

As stated previously, an incomplete description of a model can lead to model uncertainty. To cope with model uncertainty, multiple realisations of the model are generated to represent

the model output in a probabilistic fashion. This model uncertainty is common in CO_2 storage applications since most characteristics of deep saline formations are poorly defined.

In this thesis, we therefore focus on model uncertainty techniques, such as probabilistic density estimation, and parameter estimation. Probabilistic density estimation is used to describe the characteristics of the outcome population in criterion space where the parameter of the distribution is used to describe the main features of the distribution. The uncertainty region in the criterion space of the Pareto-optimal front can be obtained from the statistics of the density function, such as the variance, and the IQR. In contrast, the techniques in Eqs. 5.6 – 5.10 describe the statistics of the outcome $f(\mathbf{x})$ from the local neighbourhood $R_\pi(\mathbf{x})$ of the input variable. In the following subsection, we will provide a brief description of probabilistic density functions and parameter estimation that are used to deal with the model uncertainty.

5.4.2.1 Probabilistic density estimation

We chose four parametric density functions to describe the probability of the objective function arising from the model uncertainty. The density functions include a Gaussian function, a location-scale t-distribution function, a generalised extreme value function, and a mixture function. These density functions given below will be demonstrated later for CO_2 sequestration.

Gaussian density function : A normal distribution is commonly used in many engineering applications. Given that μ and σ denote a measure of central tendency and variability respectively, the Gaussian function is:

$$f(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \tag{5.11}$$

Location-scale t-distribution function : This function is modified for translation and rescaling of a standard student's t-distribution. The student's t-distribution is suitable for small samples (with a number of samples, N $\leq$30), and approaches to the normal distribution for large sample sizes. The shape of the student's t-distribution is a bell shape with a higher (or thicker) tail distribution than the normal distribution. The tail weight is adjusted through the degree of freedom ν, which can be useful if the tail probability is an area of interest e.g. risk and reliability accessment. If parameter ν denotes the shape of the distribution (or degree of freedom) and $\Gamma(x)$ is a gamma function, the location-scale t-distribution is given by:

$$f(x|\mu,\sigma,\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\sigma\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}\left[1+\frac{(\frac{x-\mu}{\sigma})}{\nu}\right] \tag{5.12}$$

$$\text{where} \quad \sigma > 0,\ \nu > 0, \text{ and} \tag{5.13}$$

$$\Gamma(x) = (x-1)! \quad \text{for } x \in \mathbb{Z}^+$$

$$\Gamma(\frac{1}{2}+x) = \frac{1\cdot 3\cdot 5\cdots(2x-1)}{2^x}\sqrt{\pi} \quad \text{for } x \in \mathbb{Z}^+$$

Generalised extreme value (Gev) function : This function combines three extreme value distributions, which are Gumbel, Frećhet, and Weibull distributions, into a single generalised function. The shape parameter k is used to define different types of extreme value distribution. The Gev function is given below:

$$f(x|\mu,\sigma,k) = (\frac{1}{\sigma})\text{exp}\left[-\left(1+k\frac{(x-\mu)}{\sigma}\right)^{-\frac{1}{k}}\right]\left(1+k\frac{(x-\mu)}{\sigma}\right)^{-1-\frac{1}{k}}, \quad (5.14)$$

$$\text{for } k \neq 0 \text{ and } 1+k\frac{(x-\mu)}{\sigma} > 0$$

$$f(x|\mu,\sigma,k=0) = (\frac{1}{\sigma})\text{exp}\left[-\text{exp}\left(-\frac{(x-\mu)}{\sigma}\right) - \frac{(x-\mu)}{\sigma}\right], \text{ for } k=0 \quad (5.15)$$

where $k > 0$ presents the Frećhet distribution, and $k < 0$ represents the Weibull distribution, while Eq. 5.15 with $k = 0$ is the Gumbel function.

Mixture model function : This is a mixture function that ties each density function together with predefined weights. The formula of the mixture model function is given by:

$$f(x|\boldsymbol{\lambda},\boldsymbol{\Theta}) = \sum_{k=1}^{K}\lambda_k f_k(x|\Theta_k) \quad (5.16)$$

$$\text{subject to:} \quad \sum_{k=1}^{K}\lambda_k = 1, \quad \text{and} \quad 0 < \lambda_k < 1$$

where K represents the number of component distributions, and λ_k defines a weight factor associated with its distribution function with $\lambda_k \in \boldsymbol{\lambda}$; $f_k(x|\Theta_k)$ denotes a distribution function with its associated parameter Θ_k; the parameter $\boldsymbol{\Theta}$ is the vector of parameter Θ_k, $\Theta_k \in \boldsymbol{\Theta}$. We express the first and second moment of the mixture distribution as:

$$\mu_M = \sum_{i=k}^{K}\lambda_k\mu_k \quad (5.17)$$

$$\sigma_M^2 = \sum_{k=1}^{K}\lambda_k\sigma_k^2 + \sum_{1\leq k<r\leq K}\lambda_k\lambda_r(\mu_k-\mu_r)^2 \quad (5.18)$$

It should be noted that if K is set to 1, the mixture model becomes a single parametric distribution, and if K is equal to the number of samples, the mixture function is a kernel density function. In the next section, we will present a method for estimating parameters of the density function.

5.4.2.2 Parameter estimation : maximum likelihood estimation

In frequentist statistics, the maximum likelihood estimation (MLE) is a method for finding the parameters of a statistical distribution function corresponding to the maximum probability value given the observed samples [157]. However, it should be noted that the parameters are not the most likely parameters from the given random samples; the likelihood function is not the probability density function (PDF) of the parameter, but a function of the PDF. The description of the MLE is provided in the following.

Let x_i be independent and identically distributed random variables drawn from the PDF with a vector of parameters $\mathbf{\Theta}$ presenting as $f(x_i|\mathbf{\Theta})$, and let n denote the number of observed samples, then we can write a likelihood function as follows:

$$L(\Theta_j|x) = \Pi_{i=1}^{n} f(x_i|\Theta_j) \tag{5.19}$$

The MLE of parameter Θ_j, where $\Theta_j \in \mathbf{\Theta}$, is obtained by maximising this likelihood function with given observed variables x.

$$\Theta_j^{\text{MLE}} = \underset{\Theta_j \in \Theta}{\arg\max}\, L(\Theta_j|x) \tag{5.20}$$

In this study, we used the MLE approach obtained from the MATLAB® [100] command *mle* to approximate the mean and variance of the parametric distribution in Eqs. 5.11 – 5.16. It is worth noting that the mle algorithm from MATLAB® is a gradient based optimisation approach, where there is no guarantee that the solution will converge to the global maxima. For this reason, we set the maximum number of iterations to 12,000 with a tolerance error for both input variables x and the likelihood function at 10^{-6}.

5.4.2.3 Mathematical formulation

For the MOO under model uncertainty in this thesis, we will optimise a vector of statistical measures instead of a vector of objective functions as in Eq. 5.1. The formulation of MOO under model uncertainty is written in the following form:

$$\begin{aligned} \text{Minimise:} \quad & f(\mathbf{x}) = [\Theta_{f_1}^{j_1}, \Theta_{f_2}^{j_2}, \ldots, \Theta_{f_M}^{j_M}] \\ \text{where} \quad & \mathbf{x} \in \mathcal{S}_x,\ \Theta_{f_m}^{j_m} \in \mathcal{F}_x,\ m \in \mathbb{Z}^+ \end{aligned} \tag{5.21}$$

where $\Theta_{f_m}^{j_m}$ denotes a scalar parameter of the density function f_m and the index j_m represents types of parameters such as mean, variance, and shape parameters in the parameter set $\mathbf{\Theta}$ with $\Theta^{j_m} \in \mathbf{\Theta}$.

5.5 Surrogate-assisted multi-criterion optimisation

The objective functions $f(\mathbf{x})$, evaluated from a reservoir simulator, are approximated from the sparse grid surrogate $\mathcal{Q}_{q,d}(f)$ with q and d representing the Smolyak parameter and the problem dimension respectively. For brevity, we define $\mathcal{Q}_{q,d}(f)(\mathbf{x})$ to be equivalent to $\mathcal{Q}_q(\mathbf{x})$, then the MOO formula in Eq. 5.1 can be written in the following form:

$$\begin{aligned} \text{Minimise:} \quad & \zeta(\mathbf{x}) = [\mathcal{Q}_{q^1}(\mathbf{x}), \mathcal{Q}_{q^2}(\mathbf{x}), \ldots, \mathcal{Q}_{q^M}(\mathbf{x})] \\ \text{where} \quad & \mathbf{x} \in \mathcal{S}_x,\ \mathcal{Q}_{q^m} \in \mathcal{F}_x,\ m \in \mathbb{Z}^+, q^m \in Q \end{aligned} \tag{5.22}$$

where $\mathcal{Q}_{q^m}$ is the surrogate of the m^{th} objective function; this surrogate was built separately for each objective function; therefore, the Smolyak parameters q^m may be different. Different Smolyak parameters imply that each objective function have different degrees of non-linearity as the individual surrogate has its own dimensional adaptivity.

For the surrogate used in MOO under uncertainty, let $\widetilde{y}$ be the model uncertainty and let x be independent variables, then we define:

$$\widetilde{f}(\mathbf{x}) := f(\mathbf{x}, \widetilde{y}) \tag{5.23}$$

Let $\mu^{(i)}$ be the i^{th}-statistical moment of the objective function distribution $\widetilde{f}(\cdot)$, from which we define:

$$\widehat{\mathcal{Q}}^{j}_{q,i}(\mathbf{x}) \approx \mu^{(i)}(\widetilde{f}_j(\mathbf{x})) \tag{5.24}$$

where j denotes the index of the objective function. Then, we define the MOO under uncertainty using the ASGI surrogate as follows:

$$\begin{aligned} \text{Minimise:} \quad & \zeta(\mathbf{x}) = [\widehat{\mathcal{Q}}^{1}_{q,i}, \widehat{\mathcal{Q}}^{2}_{q,i}, \ldots, \widehat{\mathcal{Q}}^{M}_{q,i}] \\ \text{where} \quad & \mathbf{x} \in \mathcal{S}_x,\ \widehat{\mathcal{Q}}^{m}_{q,i} \in \mathcal{F}_x,\ m \in \mathbb{Z}^{+} \end{aligned} \tag{5.25}$$

where the Smolyak parameter q and the index of the statistical moment i for each objective function may be different.

5.6 Applications

In this section, we present MOO techniques for CO_2 sequestration in deep saline aquifers under geological uncertainty. This chapter focuses on two main parts: (i) a robust optimisation technique using the MOO approach for a single objective function, and (ii) a robust multi-criterion optimisation approach. Later, the results for CO_2 sequestration using the rectangular and the PUNQ-S3 models are shown.

5.6.1 Robust optimisation based on the MOO approach

In Chapter 4, a utility function approach using the mean and variance of the objective function was formulated for the robust optimisation. Instead of setting the relative weights of the two components of the utility functions a priori, a multi-objective optimisation explores the Pareto-optimal front for the complete range of weights. This section presents the use of the MOO technique for a single objective function to obtain a set of robust solutions.

To cope with model uncertainty, we focus on the choice of parametric distributions, such as Gaussian and non-Gaussian functions, and their parameters. The objective function in this study is presented first, followed by parameter estimation of the Gaussian density function. Later, both Gaussian and non-Gaussian parametric distributions are compared. Last, the impact of different types of distributions on decision making is presented.

5.6.1.1 Objective function

We slightly modified the break-even tax credit (BT) function given in Section 4.5.2.4 for CO_2 optimisation in this chapter. The BT function we used is:

$$BT(\mathbf{x}, \mathbf{y}) = \frac{\sum_{k=1}^{N_t} \left(c \Delta q_i^k(\mathbf{x}, \mathbf{y}) (1 + r_T)^{\frac{-k\Delta t}{T}} \right)}{\sum_{k=1}^{N_t} \left(\left[\Delta D^k(\mathbf{x}, \mathbf{y}) + \Delta R^k(\mathbf{x}, \mathbf{y}) \right] \right)} \tag{5.26}$$

where $\mathbf{x}$ is a vector associated with injection rates of the wells and $\mathbf{y}$ represents the geophysical properties; the other parameters are described in Eq. 4.52. The future cash flow is discounted to the present value for the cost of injection. However, we add a term that cancels out the discount term for the trapped CO_2 to reward long-term security of the storage. Hence, there is no discount term in the denominator.

The optimal solution is achieved through the minimisation of the BT function by allocating CO_2 injection rates at each step of each well. The objective function is formulated under a bi-objective optimisation framework.

$$\text{Minimise:} \qquad f(\mathbf{x}) = [\mu_{BT}^{(1)}(\mathbf{x}), \mu_{BT}^{(2)}(\mathbf{x})] \tag{5.27}$$

where $f(\mathbf{x})$ is a vector of two objectives, and $\mu_{BT}^{(i)}$ represents the i^{th} statistical measures of the BT function. Unless otherwise stated, $\mu_{BT}^{(1)}$ and $\mu_{BT}^{(2)}$ denote the mean and the variance of the BT function. Unlike the utility function, the mean and the variance in Eq. 5.27 are formulated in a separate objective function such that the NSGA-II, an algorithm for solving multi-objective optimisation, can be used to obtain the Pareto-optimal front for a trade-off between the mean and the variance.

5.6.1.2 Parameter estimation of a Gaussian density function

In this section, a Gaussian density function (Eq. 5.11) is assumed to describe the probability of the BT value in the presence of geological uncertainty for a given injection strategy. We consider both point and interval estimators, such as mean, interval mean, and dispersion index, of the Gaussian density function as multiple objective functions for the MOO. Then, the NSGA-II is used to find a trade-off between these criteria for a set of robust solutions. Later in the section, we will quantify the robustness of the chosen estimators by using a bootstrapping method.

Different types of statistical measures. We present a robust optimisation method for CO_2 sequestration using the rectangular and the PUNQ-S3 models. For the rectangular model, the robust solutions are obtained from the NSGA-II of the mean, the interval mean, and the dispersion index of the BT functions. The mean is estimated from the random BT values; the interval estimator of the mean is obtained from the 95% confidence interval. The dispersion index, referred to as variance-to-mean ratio (VMR), represents the dispersion of the BT value per unit mean. Additionally, these estimators are approximated from 29 realisations of the BT function, where the samples are assumed to be independent and identically distributed (iid) samples.

Figure 5.12(a) presents the mean, the interval mean, and the dispersion index for different injection strategies, while Figs. 5.12(b) – 5.12(d) displays a cross-section of Fig. 5.12(a). There are 10,201 scattered data points (the blue circles) in the criterion space mapped from injection strategies in the design space. We use a spline surrogate [166] of the BT function to obtain the scatter plot. We see in Fig 5.12(a) that the circles are clustered; they do not spread throughout the criterion space. For example, the cross-section plots in Fig. 5.12(b) clearly shows that the mean and the interval mean are weakly conflicting as it does not

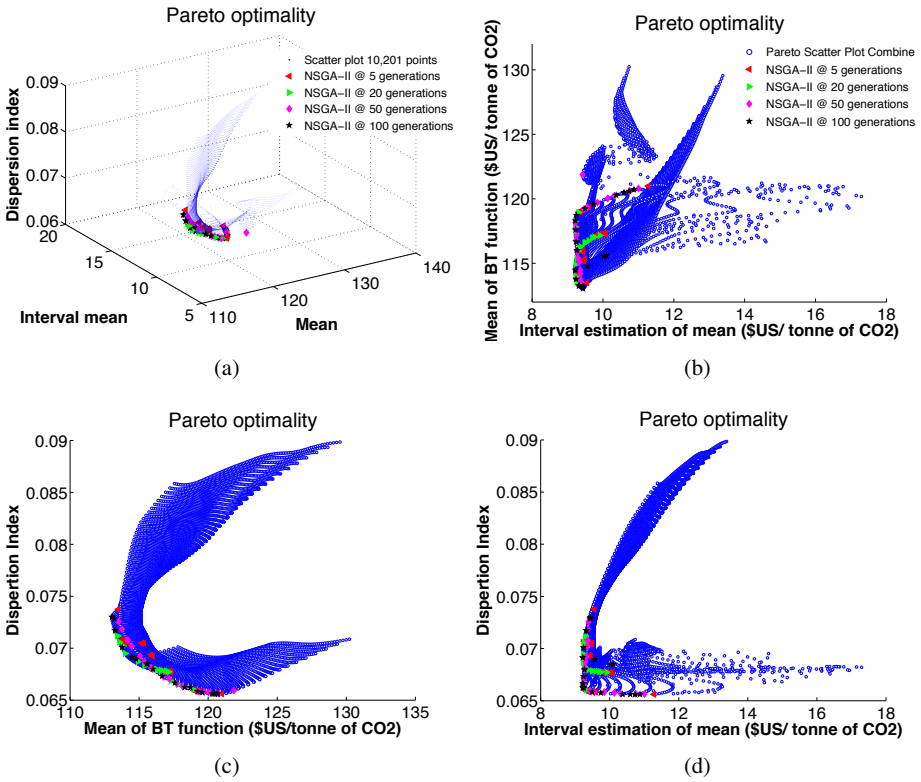
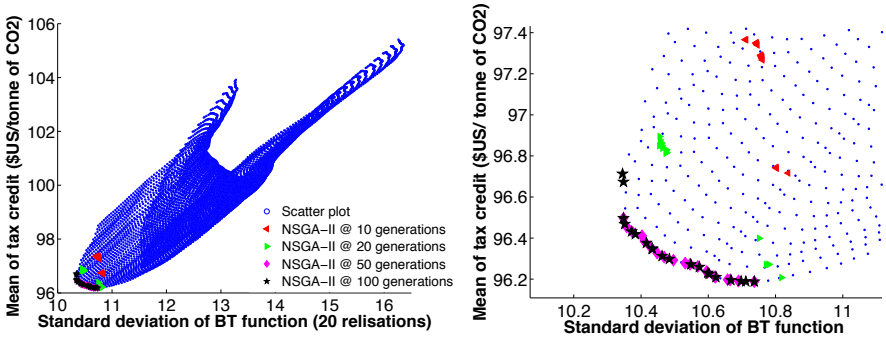


Figure 5.12: The relationship between statistical measures of the break-even tax credit function, using mean, interval mean, and dispersion index, is shown in a), while b)-d) display cross sections of these statistical measures.

exhibit a trade-off relationship where only a single optimal solution exists. It also implies that using MOO for the mean and interval mean may not be efficient. The dependency of the interval mean and the dispersion index is also observed in Fig. 5.12(d). However, in Fig. 5.12(c), the dispersion index and the mean are conflicting, which implies that there exists the optimum trade-off which can be captured by the NSGA-II. In addition, we see that the NSGA-II at different generations, shown by different marker types, capture the entire Pareto-optimal front quite well. As expected, the greater number of the NSGA-II generations results in a better set of solutions in terms of its diversity and closeness to the Pareto-optimal front.

For the PUNQ-S3 model, Fig. 5.13 presents mean and standard deviation of the BT function; the mean and standard deviation were obtained from 20 ASGI surrogates of the BT function. Each ASGI surrogate corresponds to a particular geological model. Therefore, we built 20 ASGI surrogates for 20 geological realisations of the PUNQ-S3 model. The blue circles shown in the figure represent the data in criterion space mapped from 10,201 injection strategies in the design space. The close-up view near the Pareto-optimal front is shown in Fig. 5.13(b) where the front is captured by the NSGA-II using 50 and 100 generations. We see that using 100 generations provides a better result in terms of diversity of the solutions in comparison to the result from 50 generations. This also corresponds to the results in Fig. 5.12 where using the greater number of generations is better. As expected,

the solutions of the evolutionary algorithm are improved when the number of generations is increased, due to the fact that parent populations always pass all good genes to their children. However, using the NSGA-II with the greater number of generations requires more computational resources.



(a) (b)

Figure 5.13: a) Criterion space of the mean and standard deviation of the PUNQ-S3 model with different NSGA-II generations; b) a close-up view of the Pareto-optimal front.

The examples of both rectangular and PUNQ-S3 models show that: 1) The first two moments of the Gaussian distribution such as the mean and the variance are conflicting objectives, which is in contrast to the interval mean and other measures. 2) The NSGA-II can be used to obtain robust solutions (the trade-off between the mean and the variance) where the number of NSGA-II generations should be selected based on the complexity of the problem and computational budget. It is worth noting that the computational cost in this study is dominated by the cost of surrogate construction rather than the NSGA-II.

Bootstrapping for dispersion measures. The accuracy of statistical estimators can be accessed using a bootstrapping method [40] where different estimators are compared based on their distributions. An unbiased estimator which has the minimum dispersion is preferred over other estimators having a greater dispersion. We use the dispersion measure from the estimator distribution to access the accuracy of individual estimators. For a measure of central tendency, Chernick [30] recommended using the median instead of the mean if the population distribution is unknown, where the population median can be consistently obtained from the samples. Bootstrapping could also be used to estimate the mean but the bootstrapping mean is the same as the sample mean [30]. For a measure of dispersion, we compare variance, interquartile range (IQR), and 90% confidence interval (CI) in this study, where the IQR is the difference between P75 and P25, as well as the 90% CI defined as the difference between P95 and P5.

The bootstrapping method is based on a resampling strategy from its own sampling set. The estimator distributions are shown in Fig. 5.14, each of which is obtained from the bootstrapping with 1,000 sampling sets. An estimator with the minimum standard deviation or the IQR is preferred over other estimators. In Fig. 5.14, the standard deviation and the IQR of the variance estimator are 45.77 and 63, which are the highest values in comparison to the others. The IQR estimator has the lowest standard deviation and interquartile range.

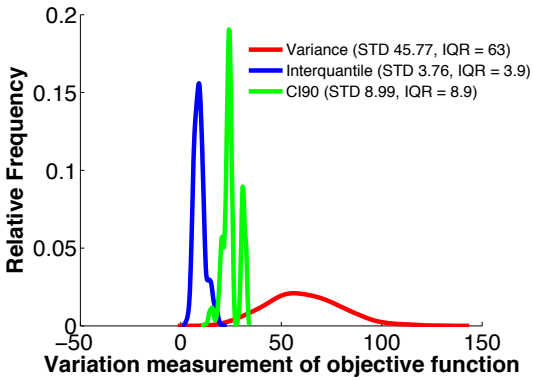


Figure 5.14: The distribution of estimators resampled from the bootstrapping with 1000 sets are shown. Blue indicates interquartile (P75-P95) distribution; red represents the distribution of the variance and the confidence interval at 90% is displayed in green.

Therefore, the IQR estimator is recommended for the variation estimator of the BT function. The numerical values of this comparison are provided in Table. 5.1.

Table 5.1: Dispersion estimator using bootstrapping method with 1,000 resampling sets

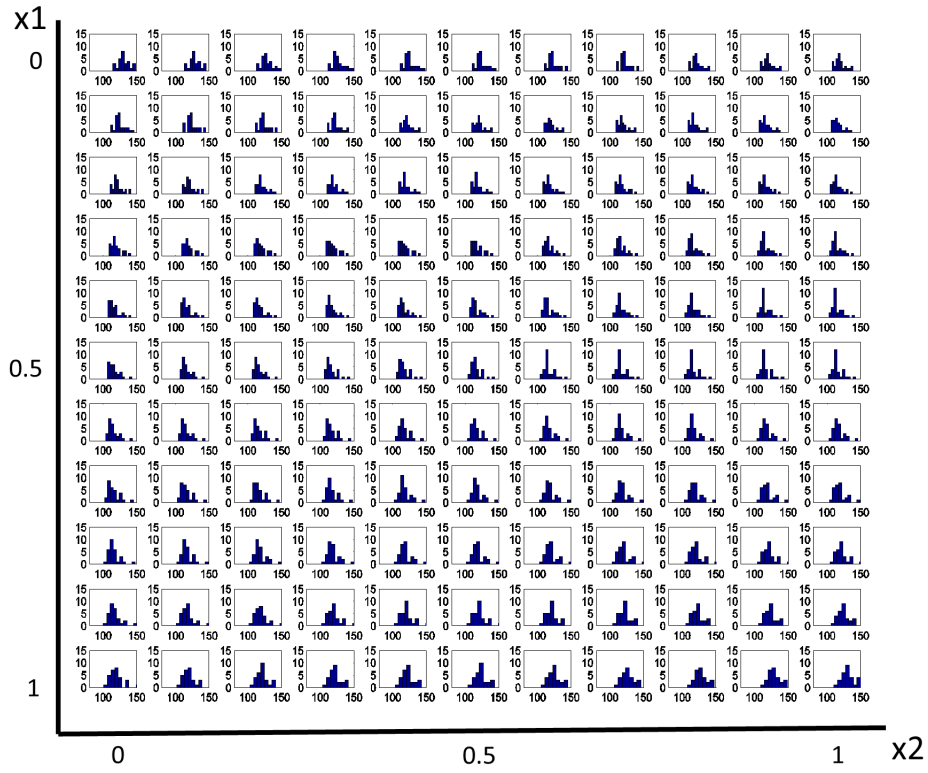
Estimator	Bootstrapping STD	Bootstrapping IQR
Variance	45.77	63
Interquartile range (IQR)	**3.76**	**3.9**
Confidence interval at 90%	8.99	8.9

5.6.1.3 Parameter estimation of the non-Gaussian density function

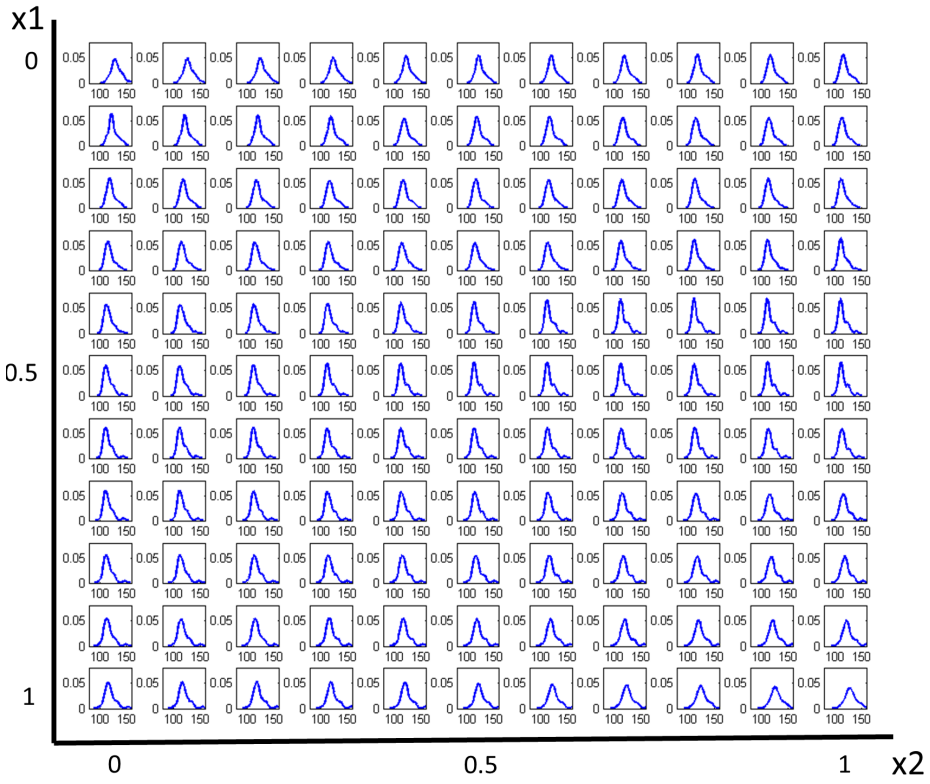
In the previous section, we made an assumption that the random samples follow a Gaussian probability distribution from which the statistical measures were obtained. However, in this section, we first validate the Gaussian assumption, followed by a discussion of alternative density functions. Then, the impact of the chosen density functions on decision making will be presented last.

Validation of Gaussianity. There are three common techniques to access characteristics of populations from which observed independent variables are drawn. These techniques are 1) graphical methods, such as histogram, quantile plot; 2) numerical methods based on moments of distributions such as the skewness or the other shape parameters; and 3) hypothesis normality tests, such as the Kolmogorov-Smirnov (K-S) [156] and Lilliefors [95] tests.

For the graphical method, we use a histogram plot to show the BT distributions of the rectangular model. Figure 5.15 shows a matrix plot of histograms and probability distributions from 29 model realisations of each control variable (i.e. x_1 and x_2); the probability density function is estimated from the kernel density function for visualisation. We see that most of the histograms in Fig. 5.15(a) and the PDFs in Fig. 5.15(b) do not exhibit symmetric distributions as they tend to have positively-skewed distributions (a right-tailed distribution).



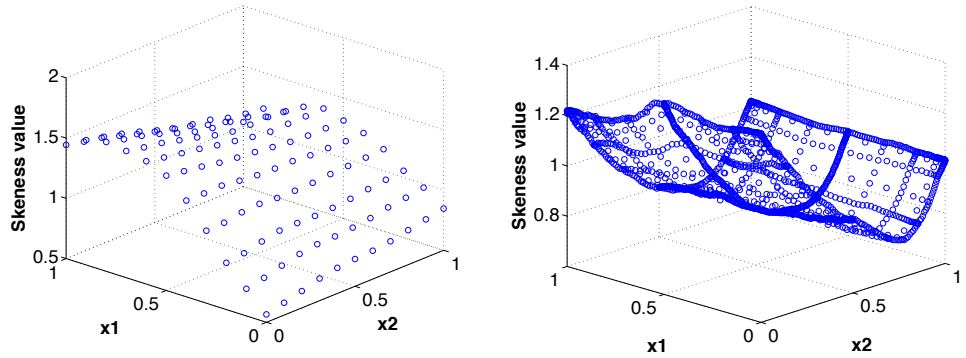
(a) Histograms plot



(b) Probability distribution

Figure 5.15: A matrix plot of histograms and the probability distributions of the objective function for 29 realizations of the rectangular model. x_1 is the control ratio of well 1 in the first period while x_2 is the ratio of the control of well 1 in the second period. Detailed well control configurations are defined in Section 4.5.2.3.

For the skewness test, we plot the skewness values over the domain of injection ratios as shown in Figs. 5.16(a) and 5.16(b) for the rectangular and the PUNQ-S3 model, respectively. We see that the skewness values are all positive for both models, which implies that the BT distribution for a particular injection ratio is asymmetric and skewed to the right (as shown in Fig. 5.15). Hence, the parametric distribution with a shape parameter can be used to describe the characteristics of the population of the BT function.



(a) The rectangular model

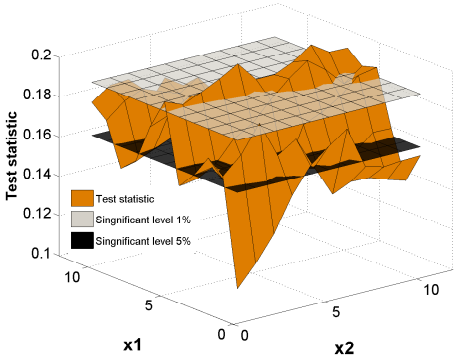
(b) The PUNQ model

Figure 5.16: A plot of skewness values over the injection ratio space: a) the rectangular model with 29 model realisations, b) the PUNQ-S3 model with 20 realisations.

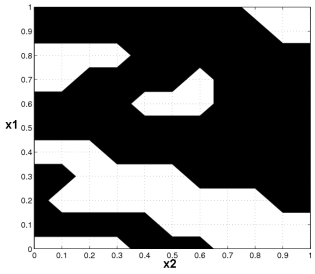
For the normality test, we choose the Lilliefors test [95] with a null-hypothesis being that the hypothesised data is drawn from a family of normal distributions. The Lilliefors test is similar to the K-S test [156] in that the maximum distance between the empirical and the hypothesised distribution is used for the test. If the maximum distance exceeds the critical value at a specified confidence level, the null-hypothesis will be rejected. Unlike the K-S test, the parameters of hypothesised distributions in the Lilliefors test are based on the empirical samples, rather than the predefined parameters in the K-S test.

Figure 5.17 presents the Lilliefors test for 121 injection strategies of the rectangular model with 29 realisations, where Fig. 5.17(a) displays the plot of the test statistics with two cutting planes presenting the critical value at significance level 0.05 and 0.01. The cross-sections of the test statistics with significance levels 0.05 and 0.01 are shown in Figs. 5.17(b) and 5.17(c), where black indicates that the test statistics exceeds the critical value and consequently implies that the data are highly unlikely to be drawn from the family of normal distributions. More than 50% and 10% of the injection strategies do not follow a normal distribution in Fig. 5.17(b) and Fig. 5.17(c), respectively. Hence, we conclude with 95% confidence that the BT distribution does not follow the normal distribution and the parameters of which such as mean and variance may not well describe the statistics of the BT population for a given injection strategy. However, it should be noted that this hypothesis testing is based on a relatively small sample size.

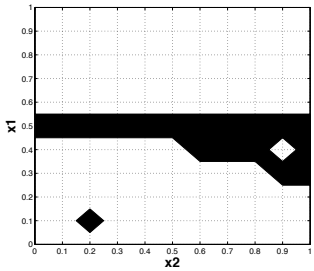
For this reason, we will use other density functions such as location-scale t-distribution and generalised extreme value (Gev) functions that exhibit the shape parameters to capture the BT value distribution. We first compare the histogram of the BT distribution with differ-



(a) Lilliefors test



(b) Significance level 0.05

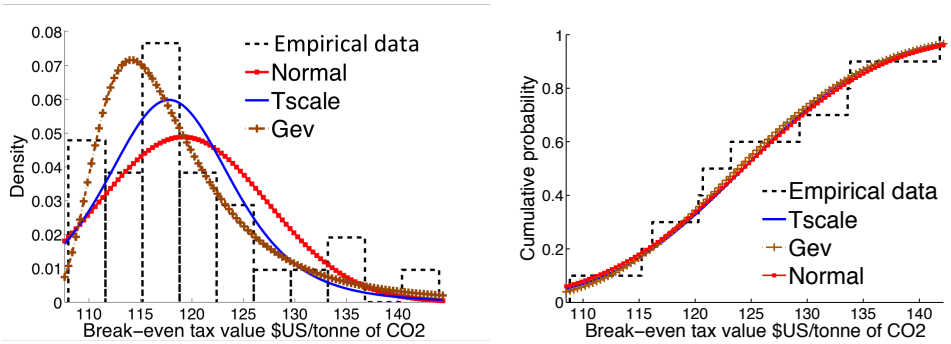


(c) Significance level 0.01

Figure 5.17: The Lilliefors test for 121 injection strategies of the rectangular model with 29 realisations is shown in (a) where the cutting planes represent the critical values of the significance level at 0.01 and 0.05. Cross-sections of the cutting planes are shown in (b) and (c) with black denoting the rejected test statistics whose value exceeds the critical value.

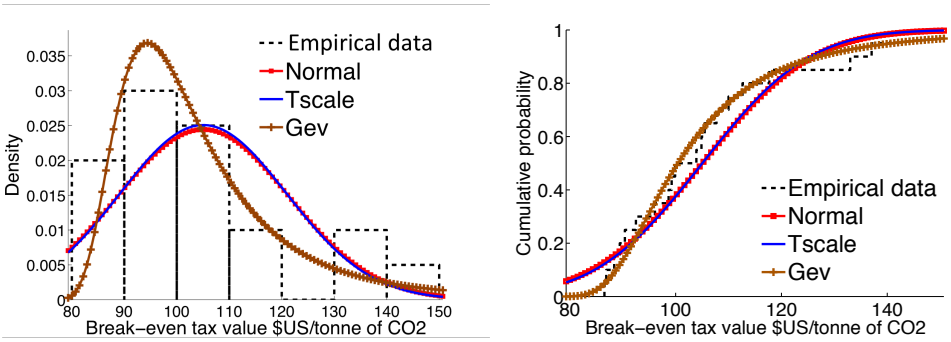
ent parametric density functions in Figs. 5.18 and 5.19 for the rectangular and the PUNQ-S3 model, respectively. We display both PDFs and CDFs of BT values obtained from random geological models (29 for the rectangular model and 20 for the PUNQ-S3 model) with the injection ratios x_1 and x_2 set to 0.3 and 0.6 for the rectangular case, and 0.5 and 0.5 for the PUNQ-S3 case. It shows in both Figs. 5.18(a) and 5.18(b) that the GeV can capture the left side of the PDF while there are no significant differences among the CDFs. However, for the PUNQ-S3 model in Fig. 5.19, we see that the Gev is slightly better than the other two functions for matching both the PDF and CDF.

Since the sample size is relatively small in this example, we drew a new set of 1,000 realisations of the PUNQ-S3 model using the same injection strategy in Fig. 5.19. The PDFs and a probability plot are shown in Fig. 5.20. We see from the PDFs in Fig. 5.20(a) that there is no significant difference between the density functions; however, the Gev function tends to represent the data best in the range of 80 – 90 \$US/tonne of CO_2. For the probability plot shown in Fig. 5.20(b), the Gaussian probability scale is used on the y-axis so that the data from the Gaussian distribution would exhibit a linear line. The figure shows that the Gev function represents the data reasonably well in the range of 80 – 90 and 115 – 125 \$US/tonne of CO_2. It should be noted that the location-scale t-distribution (Tscale) approaches to the Gaussian distribution for a large sample size. Thus, the Gaussian function overlies the



(a) PDFs (b) CDFs

Figure 5.18: PDFs and CDFs of the break-even tax distribution of the rectangular model with the injection ratio x_1 and x_2 set at 0.3 and 0.6, respectively.

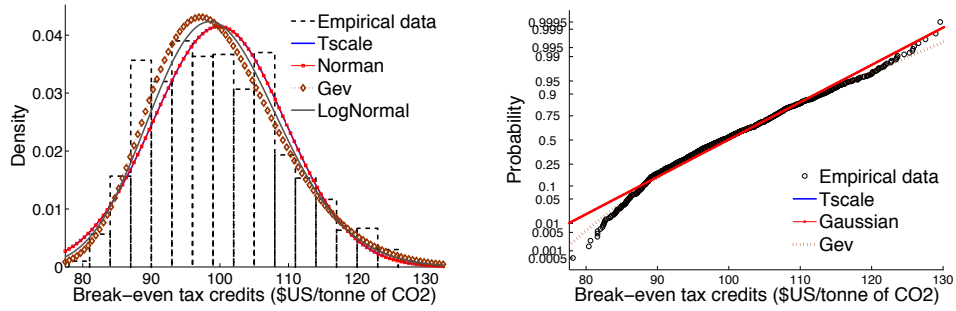


(a) PDFs (b) CDFs

Figure 5.19: PDFs and CDFs of the break-even tax distribution of the PUNQ-S3 model with the injection ratio x_1 and x_2 set at 0.5 and 0.5, respectively.

Tscale in this example.

Since these graphical methods are based on the individual injection strategy, it might be difficult to infer this observation to all injection strategies. Hence, in the next section, we will make use of a mixture model to combine various density functions where the impact of different density functions on decision making will be discussed.



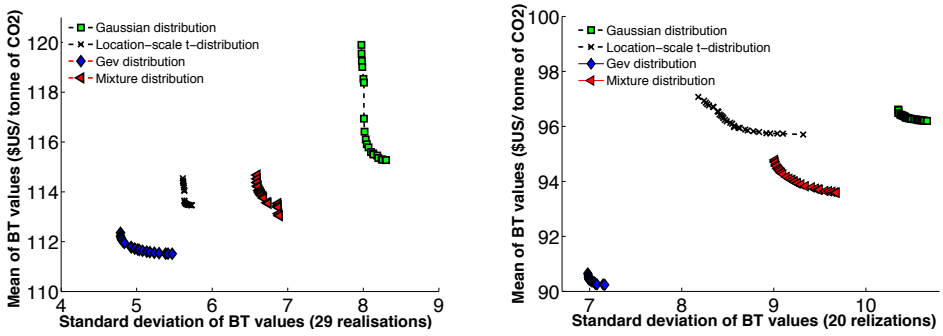
(a) PDF (b) Probability plot

Figure 5.20: PDF and probability plot of the break-even tax credit distribution of the PUNQ-S3 model using 1,000 samples

Impact of density functions on decision making In addition to the Gaussian distribution, we use the location-scale t-distribution, the Gev, and the mixture distribution function to compare a set of robust solutions obtained from the mean and the standard deviation of different density functions. The mixture model is used to combine the Gaussian, the location-scale t-distribution and the Gev function to a single density function with the statistics obtained from the component distributions. Even though we could include non-parametric estimation or the use of polynomial basis functions in this study to approximate the statistical properties of the distribution for a robust solution, it is not within the scope of the chapter.

We display the Pareto-optimal fronts obtained from the NSGA-II (for 1,000 iterations with 20 populations) using various types of density functions for the rectangular and the PUNQ-S3 models as shown in Figs. 5.21(a) and 5.21(b), respectively. Since we aim to minimise the BT function, the injection strategy providing the minimum mean and standard deviation of the BT value is preferred to the strategy with the higher mean and standard deviation. Then, it can be seen from Fig. 5.21 that decision making based on the Pareto-optimal front of the Gaussian function for both the rectangular and the PUNQ-S3 models underestimates economic performance since the BT values are higher (worse) than the other functions. On the other hand, this will be overestimated if the Gev function is used for making a decision since the Pareto-optimal front is at the lower-left corner of the figure.

In this example, even though we show that types of density functions could impact the decision making, it is very difficult to know which functions should be used to capture the characteristic of the BT distribution. Therefore, the mixture model, shown by the triangle-marker curve, is a compromise choice of the parametric distributions that combines different density functions for robust decision making. However, it should be noted that using the mixture model requires more computational resources than using individual parametric distributions.



(a) Pareto front: Synthetic model

(b) Pareto front: PUNQ-S3 model

Figure 5.21: Pareto-optimal fronts obtained from the NSGA-II using various types of the density functions such as the Gaussian, the location-scale t-distribution, the Gev, and the mixture distribution functions are shown for the rectangular (a) and the PUNQ-S3(b) model.

5.6.2 Robust multi-criterion optimisation

In this section, we use the NSGA-II to solve multiple engineering criteria for CO_2 sequestration. We first describe mathematical models of CO_2 injection criteria, followed by a discussion of the different objectives used for decision making. Later in this chapter, we will present an approach to selecting the solution from the Pareto-optimal front for efficient decision making.

5.6.2.1 Mathematical models of CO_2 injection criteria

We consider economic efficiency, CO_2 sweep efficiency, and the risks of CO_2 leakage from the storage sites as three conflicting criteria for CO_2 sequestration management. For the economic efficiency, we use the BT function as described in Eq. 5.26 to represent the profitability of investment of the project. The other two criteria are described in the following subsection.

Volumetric sweep efficiency. Two important trapping mechanisms for CO_2 storage in deep saline aquifers are dissolution and capillary trapping from which the mass of trapped CO_2 can be enhanced by increasing the contact area between CO_2 and formation brine, as discussed in Section 3.3.2. Hence, we consider volumetric sweep efficiency as another criterion for CO_2 injection. The mathematical model of volumetric sweep efficiency is described below:

$$\text{Maximise:} \qquad \mathrm{S}(\mathbf{x},\mathbf{y}) = \left[1 - \frac{\sum_{j=1}^{N_G} \delta_j(\mathbf{x},\mathbf{y}) v_j(\mathbf{y})}{\sum_{j=1}^{N_G} v_j(\mathbf{y})}\right] \times 100 \tag{5.28}$$

$$\text{where} \qquad \delta_j(\mathbf{x},\mathbf{y}) = \begin{cases} 0 & \text{for} \quad Sw_j(\mathbf{x},\mathbf{y}) \leq C_{Sw} \\ 1 & \text{for} \quad Sw_j(\mathbf{x},\mathbf{y}) > C_{Sw} \end{cases}$$

where $\mathrm{S}(\mathbf{x},\mathbf{y})$ represents the sweep efficiency in percent with $\mathbf{x}$ and $\mathbf{y}$ presenting injection strategies and geophysical properties respectively; Sw_j and v_j represent the water saturation and reservoir pore volume of grid cell j, and N_G denotes the total grid cells in the model. The indicator δ_j takes the values 0 or 1 depending on the condition of CO_2 in grid cell j by checking the water saturation with the decision criterion C_{Sw} within the cell. For example, any cell j that has Sw_j greater than C_{Sw} is considered an inaccessible cell to CO_2 and consequently it is treated as an unswept cell. It should be noted that the choice of decision criterion C_{Sw} can affect the swept volume calculation as the criterion C_{Sw} implicitly decides the boundary of the swept area. For instance, the approximate swept volume could be underestimated if a smaller C_{Sw} value is chosen. In this study, unless otherwise stated, we set the decision variable C_{Sw} at 0.9 and the volumetric sweep efficiency calculation is evaluated at the end of injection period.

It is more convenient to minimise Eq. 5.28 under the MOO framework. Therefore, the sweep efficiency can be written explicitly as the volume of unswept CO_2 as follows:

$$\text{Minimise:} \qquad \mathrm{S}_u(\mathbf{x},\mathbf{y}) = \frac{\sum_{j=1}^{N_G} \delta_j(\mathbf{x},\mathbf{y}) v_j(\mathbf{y})}{\sum_{j=1}^{N_G} v_j(\mathbf{y})} \times 100 \tag{5.29}$$

CO_2 leakage. For the CO_2 leakage, we consider the situation of CO_2 leaking an abandoned well to the atmosphere. The mass of CO_2 produced from the leaky wells is used as a proxy of CO_2 leakage from aquifers. The detailed description of this leaky pathway from the well was presented previously in Chapter 4. The objective function for CO_2 passing through the abandoned well is formulated by the percentage of the mass of CO_2 leakage to the total mass of CO_2 injection. Minimisation of CO_2 leakage is given in the following:

$$\text{Minimise:} \qquad \mathrm{L}(\mathbf{x},\mathbf{y}) = \frac{\sum_{i=1}^{N_w} \sum_{k=1}^{N_t} \Delta q_p^{i,k}(\mathbf{x},\mathbf{y})}{\sum_{k=1}^{N_t} \Delta q_i^k} \times 100 \tag{5.30}$$

where Δq_i^k presents the total mass of CO_2 injection per iterative injection step k with the total injection N_t for 30 years; $\Delta q_p^{i,k}$ denotes the mass of CO_2 leakage per injection step while N_w presents the number of the potential leakage wells in the model. In practice it might be possible to get tax penalties due to the leakage of CO_2. For this reason, a single utility function capturing both BT function and the penalty terms could be formulated to address the tax penalty issue. Even though it is possible and relatively easy to include the penalty terms in the utility function, it is not within a scope of this work since solving MOO problems is the main objective of this chapter.

5.6.2.2 PUNQ-S3 model

In this section, we use the NSGA-II for the BT (Eq. 5.26) and the unswept volume of aquifer (Eq. 5.29) to find the Pareto-optimal solutions. A matrix of 20 scatter plots for both criteria is shown in Fig. 5.22 where each plot, obtained from a geological realisation of the PUNQ-S3 model, shows the criterion space of 1,537 injection strategies. The unswept volume criterion C_{Sw} was set to 0.9 for all plots. We see that most of the scatter plots in Fig. 5.22 show a compromise relationship between the break-even tax credit and the unswept volume of aquifer. For different C_{Sw} values, Fig. 5.23 displays the mean of both criteria for various C_{Sw} values at 0.8, 0.85, 0.9 and 0.95. In Figs 5.22 and 5.23, we see that there is a trade-off relationship between both criteria, and hence the NSGA-II can be used to find the Pareto-optimal front.

Figure 5.24 displays the Pareto-optimal front of the mean of the two criteria where each Pareto-optimal front is obtained from the NSGA-II using different generations. We expect the Pareto-optimal front to be as close as possible to the origin since both objectives are minimised. We see that the greater the number of GA generations, the closer the Pareto-optimal front to the origin. However, there is a slight difference between the front from 100

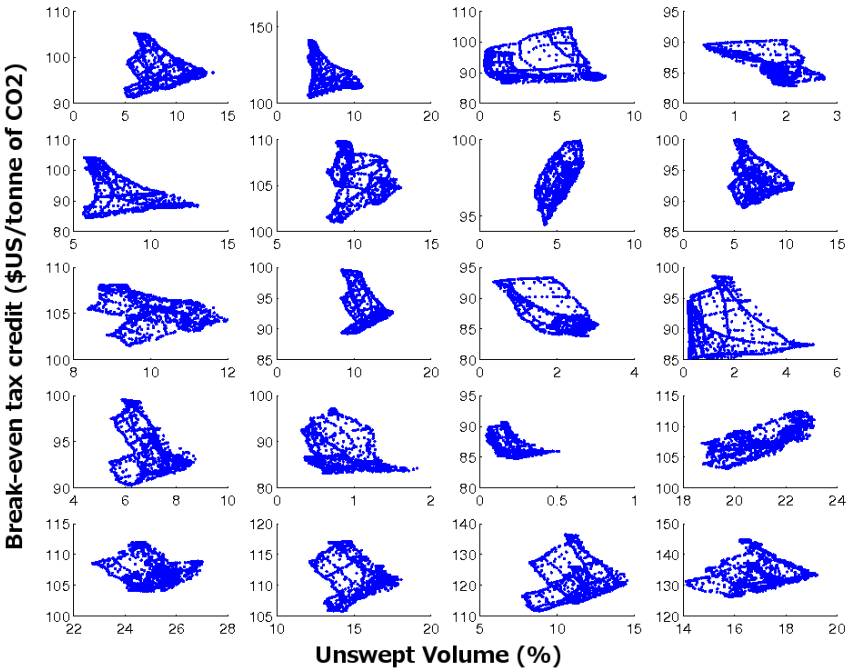


Figure 5.22: A matrix of scatter plots between break-even tax and unswept volume of aquifer from CO_2 injection for 20 realisations of the PUNQ-S3 model.

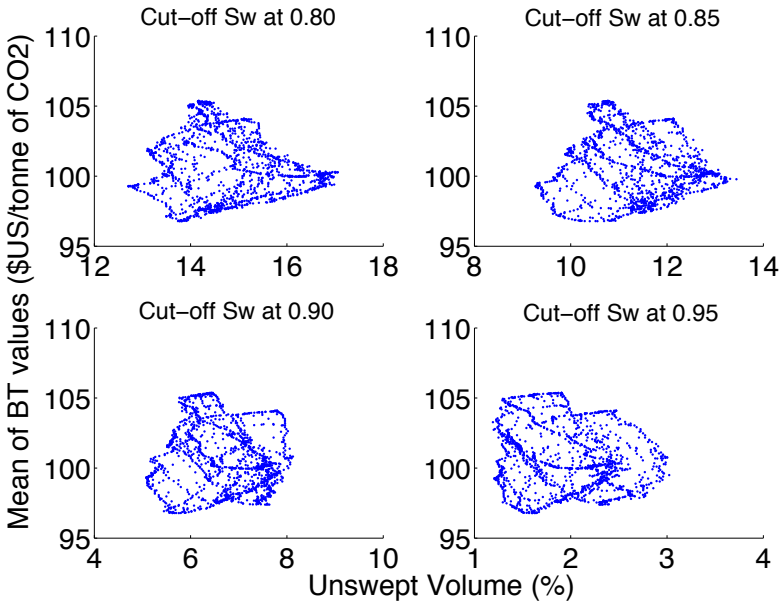


Figure 5.23: Scatter plots of the mean of break-even tax credit and the unswept volume with different C_{Sw} values .i.e. 0.80, 0.85, 0.9, and 0.95.

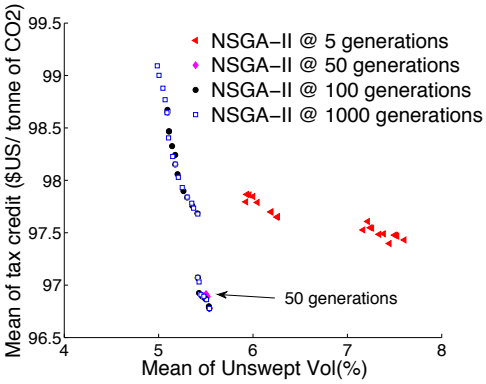


Figure 5.24: Pareto-optimal fronts between the mean of the break-even tax credits and the mean of the unswept volume of aquifer are shown. The Pareto-optimal fronts are obtained from the NSGA-II using 20 populations with different GA generations.

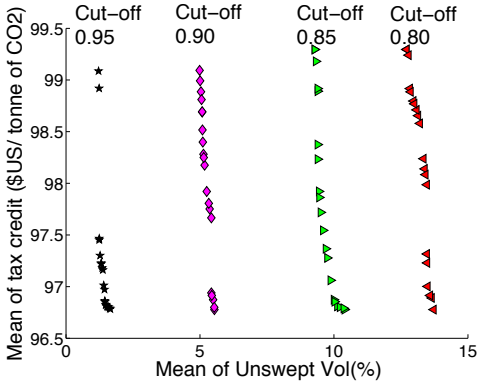


Figure 5.25: Pareto-optimal front between the mean of the break-even tax credits and the mean of the unswept volume of aquifer are shown. The Pareto-optimal fronts are obtained from the NSGA-II using 20 populations with different C_{Sw} values.

(the circles) and 1000 (the squares) generations. Therefore, form now on, we will use the NSGA-II with 100 generations as a norm, unless otherwise stated.

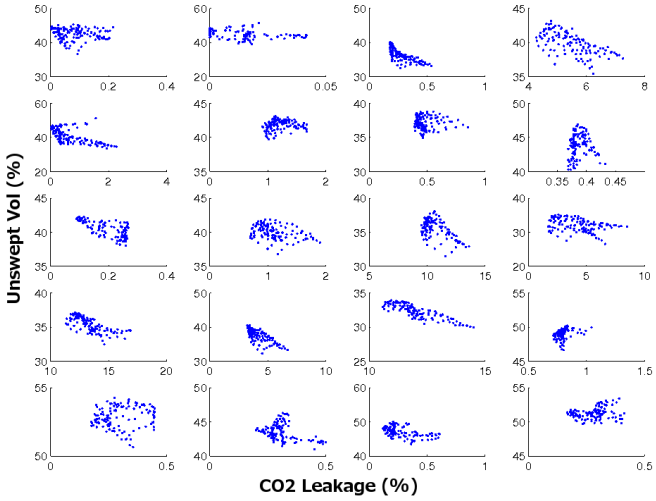
In Fig. 5.25, the Pareto-optimal fronts obtained from the NSGA-II for different C_{Sw} values are shown. As expected, the value of C_{Sw} can affect the calculation of the sweep efficiency. This can be seen from the figure that the mean of unswept aquifer volume varies by 14% and the Pareto-optimal front using the C_{Sw} value at 0.95 shows the least percentage of the unswept volume in comparison to the others.

5.6.2.3 Leaky PUNQ-S3 model

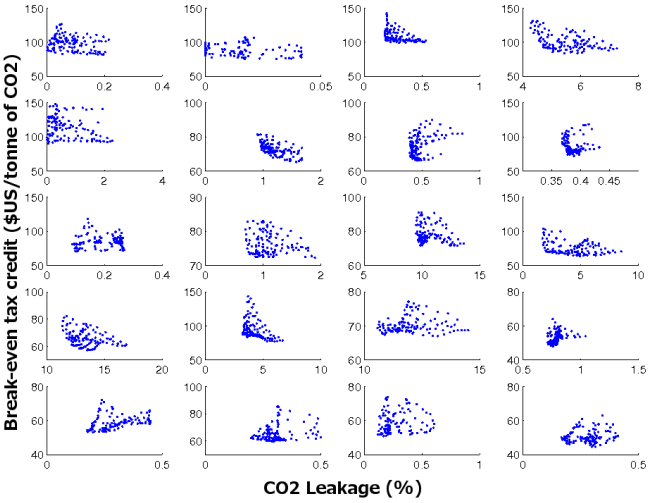
We use the leaky PUNQ-S3 model to demonstrate CO_2 leakage from the aquifer as described in Chapter 2. All three criteria are considered in this study and the NSGA-II is used to find a set of robust solutions.

Dependency of decision criteria. Matrices of scatter plots in Fig. 5.26 are used to reveal dependency of the criteria; each matrix plot displays 20 realisations of the leaky PUNQ-S3 model, and each scatter plot demonstrates the criterion space mapped from 121 injection strategies. Most of the plots in Fig. 5.26 imply the existence of Pareto-optimal solutions of the criteria. This can be seen from the lower-left corner of each plot that it exhibits a set of solutions rather than a single solution, as described previously in Fig. 5.5.

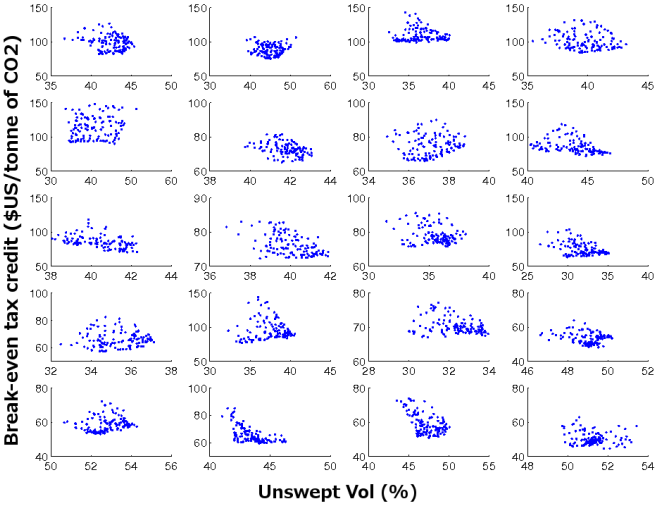
The scatter plots using the mean of the criteria presented in Fig. 5.26 are shown in Fig. 5.27 where the trade-off relationship between the mean of the criteria (shown by the solid curve) is clearly seen. Additionally, Fig. 5.27(b) exhibits the obvious trade-off relationship in the case of using the mean of the criteria even though this trade-off is not clearly seen in Fig. 5.26(b). Hence, based on the results shown in Figs 5.26 and 5.27, the NSGA-II can be used to optimise the multiple objectives of the leaky PUNQ-S3 model.



(a)



(b)



(c)

Figure 5.26: A matrix of scatter plots of unswept volume, break-even tax credit, and CO_2 leakage; each scatter plot represents the criterion space from 121 injection strategies evaluated on the ASGI surrogate. Each matrix plot presents 20 model realisations of the leaky PUNQ-S3 model.

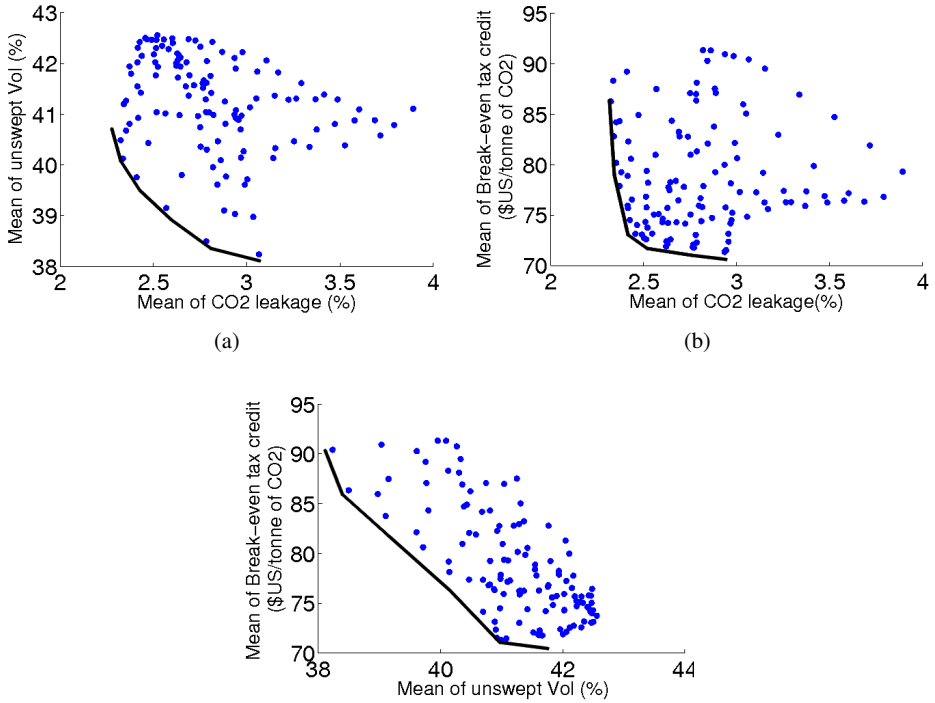


Figure 5.27: Scatter plots of the mean of three criteria with the solid curve representing the trade-off relationship (the Pareto-optimal front).

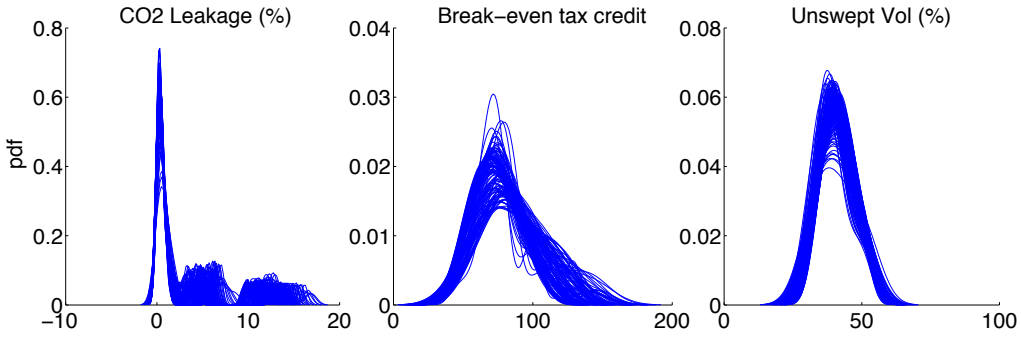


Figure 5.28: Probabilistic density function of CO_2 leakage, BT values, and unswept volume.

Statistical measures of decision criteria. As described in Eq. 5.21, we use statistical measures of the objective function rather than the objective function values to handle model uncertainty. g The probability density function (PDF) of each objective function from 20 model realisations is shown in Fig. 5.28 where each PDF, estimated from the Gaussian kernel density estimator [17], corresponds to individual injection strategies. Fig. 5.28 presents PDFs of BT values, unswept volume of aquifer, and CO_2 leakage, each plot for 121 injection strategies (121 PDFs). We see that the PDFs of the BT and the unswept volume of the aquifer are similar to a normal distribution, but the PDFs of the BT exhibit a positively skewed distribution. However, the positive skewness is more obvious for the PDFs of the CO_2 leakage; this indicates that the data above the mean is spread out more than the data

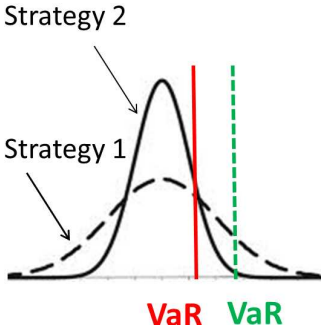


Figure 5.29: Illustration of Value at Risk (VaR) for two injection strategies. Green indicates the VaR of the strategy 1 (dashed curve); red represents the VaR of the strategy 2 (solid curve). For risk mitigation, we prefer the injection strategy 2 to the strategy 1 in this example.

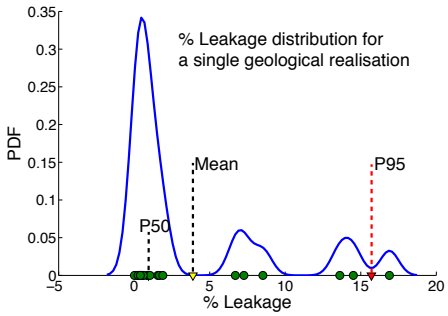


Figure 5.30: Probability density function of the mass of CO_2 leakage from a single geological realisation is shown with different statistical measures such as mean, P50, and P95; the green dots are the samples we used to estimate the PDF. We will use P95 as a risk measurement for CO_2 leakage.

below the mean. The PDFs of the CO_2 leakage are crucial for risk assessments where the extreme value having low probability might have a very large impact on the outcome (the black swan event [167]), in this case that is the leakage of CO_2.

For that reason, we specify a Value at Risk (VaR) [103] for a risk measurement with significance level 0.05 for CO_2 leakage. This VaR generally means the maximum value that we can predict for the mass of CO_2 leakage with probability less than 5% that it will exceed this value. Let $F_Y(\cdot)$ be the CDF of variable y and let $\alpha \in (0, 1)$ be the confidence level, then the mathematical formulation of VaR is defined as:

$$VaR_\alpha(Y) = \inf\{z|F_Y(z) \geq \alpha\} = \inf\{z|Pr(Y \leq z) \geq \alpha\} \tag{5.31}$$

The definition of VaR is basically the percentile of the loss distribution [85]. We illustrate the VaR in Fig. 5.29 from which we prefer the injection strategy 2 to the strategy 1 for minimising the risk. Hence, for minimising the risk of CO_2 leakage, we will minimise the term $VaR_\alpha(Y|\mathbf{x})$, where Y represents the distribution of CO_2 leakage mass and $\mathbf{x}$ denotes the injection strategy. For the distribution of CO_2 leakage, Fig. 5.30 presents different statistical measures such as mean, P50, and P95 with the circle dots displaying the sample of CO_2 leakage from a single realisation. In this plot, the kernel density estimator was used to obtain the distribution function. It should be noted that the VaR with significance level 0.05 is simply the 95th percentile (P95) and it will be used for the risk measurement of CO_2 leakage in this study, while the P50 will be used for a statistical measure of both BT and unswept volume objectives.

Non-dominated solutions. The statistical measures described previously are used to handle uncertainty of the objection functions from which the NSGA-II is applied to find the Pareto-optimal front. Figure 5.31 shows the front of three criteria using the NSGA-II with 500 populations; we used 500 populations to illustrate the front of the non-dominated solutions. However, using 500 populations may not be necessary in practice since the small

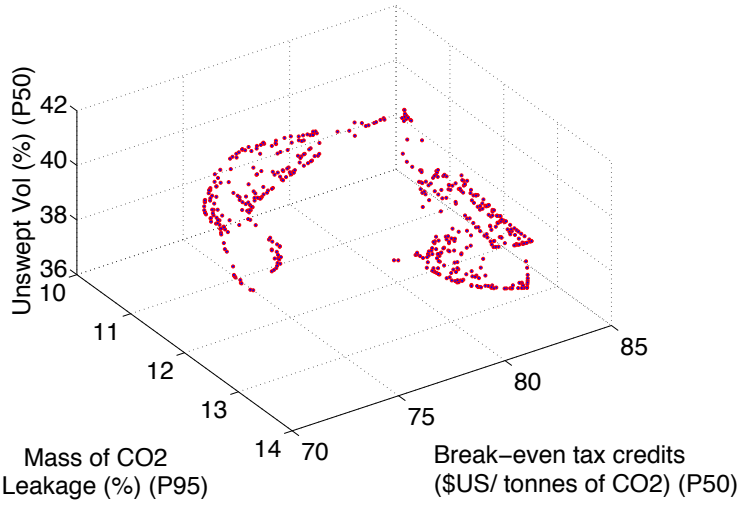
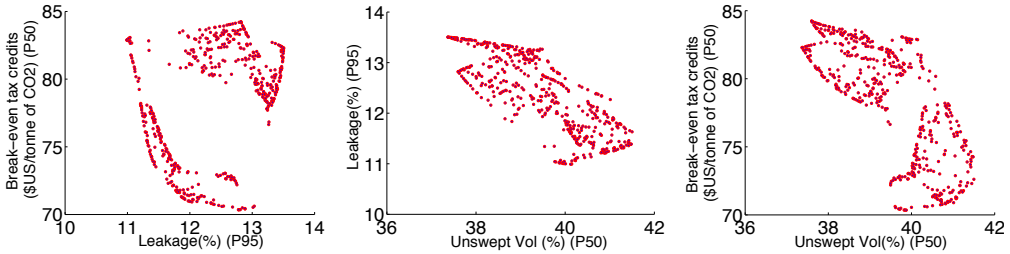


Figure 5.31: Pareto-optimal front of break-even tax credits, unswept volume of aquifer, and CO_2 leakage using NSGA-II with 500 populations.



(a) BT and CO_2 Leakage (%) (b) CO_2 Leakage (%) and unswept volume (%) (c) BT and unswept volume (%)

Figure 5.32: Pareto-optimal front of break-even tax credits, unswept volume of aquifer, and CO_2 leakage using NSGA-II with 500 populations.

number of populations (decisions) is more efficient when making a decision. The cross-sections of the Pareto-optimal front in Fig. 5.31 are displayed in Fig. 5.32. We see from both Figs. 5.31 and 5.32 that the curve of the front is nonconvex. Though the real Pareto-optimal front might be different from the one obtained from the NSGA-II, based on several test functions having nonconvex Pareto-optimal front in [34], the NSGA-II can efficiently capture the front close to the true one. For this reason, we can infer that the front in Fig. 5.31 is nonconvex, where using the prior articulation approach such as the utility function or the weighted sum technique would be problematic to obtain the set of non-dominated solutions even though the preference or the weights are varied for different values. Therefore, not only does this example illustrate the trade-off relationship among three objectives, but it also emphasises the benefit of using the NSGA-II for MOO of CO_2 sequestration.

The trade-off relationship among three criteria in Fig. 5.31 can be visualised by using a value path method [53] shown in Fig. 5.33. The objective function values from the non-

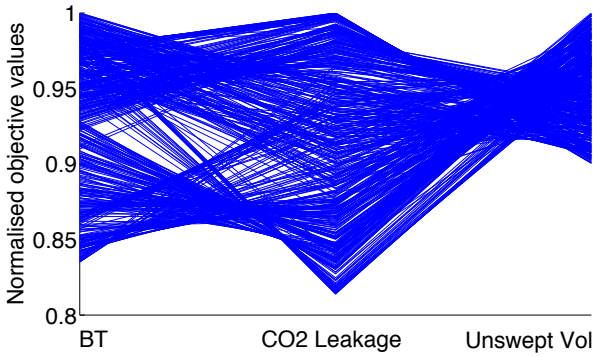


Figure 5.33: A value path method of non-dominated solutions.

dominated solutions are normalised using the following formulation.

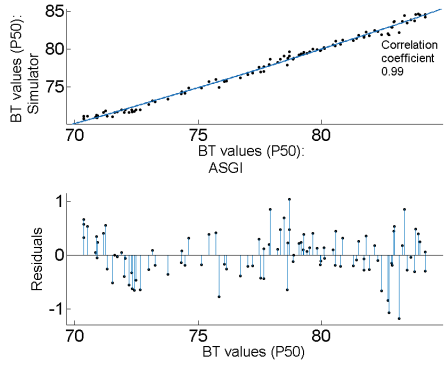
$$f_{\mathrm{N}}^{j} = \frac{f^{j} - f_{\mathrm{min}}^{j}}{f_{\mathrm{max}}^{j} - f_{\mathrm{min}}^{j}} \tag{5.32}$$

where f_{N}^{j} denotes the normalised objective function value with j representing the objective index. As shown in Fig. 5.33, each curve represents a non-dominated solution from a total of 500 solutions. It can be seen that the zigzag curves reveal the conflicting criteria.

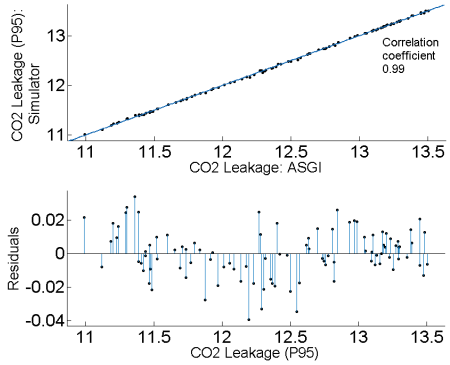
Validation of adaptive sparse grid interpolation. In this chapter, we have used the ASGI surrogate to assist the MOO process to obtain the Pareto-optimal front. To validate the quality of the ASGI surrogate, all non-dominated solutions obtained from the ASGI were evaluated directly with the data from simulator. The cross-plot between the data obtained from the surrogate and the simulator indicates the quality of the surrogate from the degree of linear dependence of the plot — we used the correlation coefficient.

Figure 5.34 presents the cross plots and the residual error plots of three criteria together with correlation coefficients. We see that the objective values of the non-dominated solutions from the ASGI surrogate correspond very well with the data from the simulator as the coefficient value of the plot is close to unity. It, however, should be noted that each plot shown in Fig. 5.34 are statistical measures of the objective values from 20 model realisations. Therefore, we validate the surrogate of the statistical data rather than a single surrogate obtained from each model realisation.

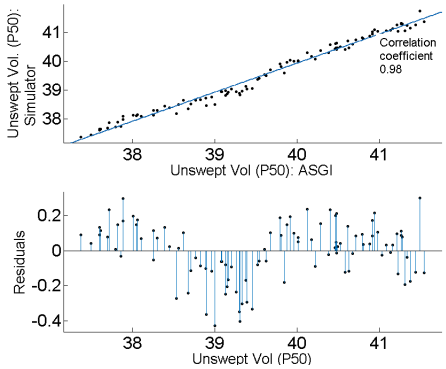
In Fig. 5.35, the objective values of 500 non-dominated solutions are displayed in frequency plots, where the data obtained from the surrogate and the simulator are shown. We see that the data from the surrogate corresponds with the data from the simulator for three criteria, which are also in line with the correlation coefficients in Fig. 5.34. The results in both Figs. 5.34 and 5.35 indicate that the ASGI approach is reliable and can be used to assist MOO for CO_2 sequestration. In the next section, we will discuss how to select the non-dominated solutions from the Pareto-optimal front for effective decision making.



(a) Break-even tax credit (P50)

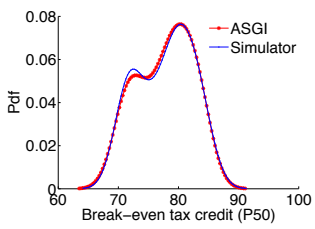


(b) CO_2 Leakage % (P95)

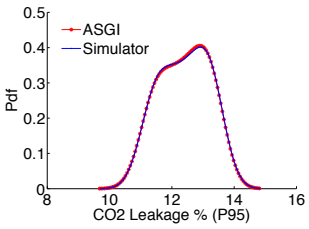


(c) Unswept Vol. % (P50)

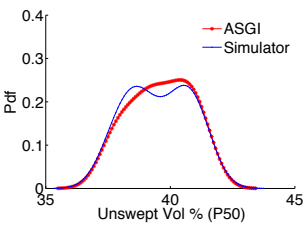
Figure 5.34: The data obtained from the surrogate are validated with the data from the simulator with correlation coefficients shown at each criteria a) ,b) and c). Cross-plots of the data are presented in the top figure with the bottom one representing the residual of the data.



(a) Break-even tax credit (P50)



(b) CO_2 Leakage % (P95)



(c) Unswept Vol. % (P50)

Figure 5.35: Frequency plots of the data obtained from the surrogate and the simulator are displayed with three criteria: (a) Break-even tax credit (P50), (b) CO_2 Leakage % (P95), and (c) Unswept Vol. % (P50).

A guideline to Pareto-optimal selection. As described in Section 5.3.4, one drawback of the posterior articulation approach for MOO is that the number of objective functions should be as small as possible such that the decision can be made effectively. Additionally, the number of non-dominated solutions should also be small. For this reason, we present a technique to assist a decision maker to identify and select alternative solutions effectively.

Although numerous techniques have been proposed in many disciplines, it is difficult to justify which method is the best since it is a subjective topic and usually depends on the preference of a decision maker (DM). However, a discussion of various techniques can be found in Kahraman [77]. In this study, we will discuss two approaches which base the Pareto-optimal selection on 1) the solutions that are robust under model uncertainty, and 2) the preference-driven biases of DM [102].

For a selection of the robust solutions, we make use of a k-means clustering method [97] to classify the non-dominated solutions into a predefined number of groups based on the distance measures. This k-means technique divides the data into k groups by minimising the sum of the distance measures of each group, explicitly given in the following:

$$\arg\min_{\mathbf{X}} \sum_{i=1}^{k} \sum_{\mathbf{x}_j \in \mathbf{X}_i} ||\mathbf{x}_j - \boldsymbol{\mu}_i||^2 \tag{5.33}$$

where μ_i is the centroid of group i. We used the algorithm *kmeans* in MATLAB® for the k-means clustering method to cluster the solutions into different groups so that we could find the representative solution of the group. This solution could be a centroid point or a mean of the entire group; however, we applied the minimax approach from statistical decision theory [171] to obtain the robust solution in comparison to the other solutions within the same group. Let $f(x, y)$ be an objective function with x and y presenting injection strategies and the model realisations, respectively. The principle of the minimax approach is to find x that minimises the maximum loss function (or minimise the worst case scenario), given as:

$$\sup_{y \in Y} f(\mathbf{x}^*, y) = \inf_{\mathbf{x} \in X} \sup_{\mathbf{y} \in Y} f(\mathbf{x}, \mathbf{y}) \tag{5.34}$$

where x^* is a minimax with respect to the function $f(x, y)$ which is the BT function in this study, and the term $\sup_{y \in Y} f(x, y)$ is approximated by using the Value at Risk (VaR) term from Eq. 5.31. Figure 5.36 displays the Pareto-optimal front of the BT values and the CO_2 leakage. We see that the solutions are clustered into three groups obtained from the k-means approach, where the star markers represent the minimax solutions. By using the minimax approach, we can reduce 100 non-dominated solutions into three solutions, providing an effective method for decision making.

For the preference-driven bias approach, the DM is required to specify his/her preference in the criterion space where the Pareto-optimal solutions that strictly dominate the preference points will be selected. This method is used to reduce the total number of solutions into a smaller group from which the k-means clustering method or the minimax approach can then be applied to each group.

Let $\text{Pf}(x, y)$ be the preference points with x presenting the BT values and y for CO_2 leakage. Figure. 5.37(a) demonstrates how data is partitioned using the preference points.

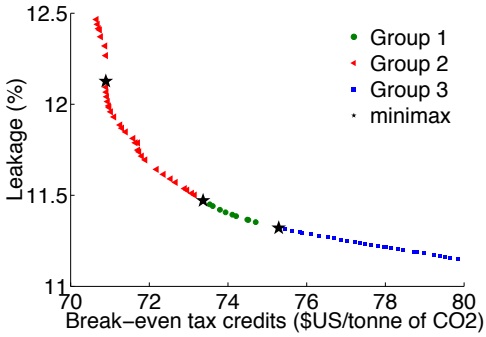


Figure 5.36: Illustration of the Pareto-optimal front clustered into three groups by k-means clustering method where the star marker represent the minimax solution for each group.

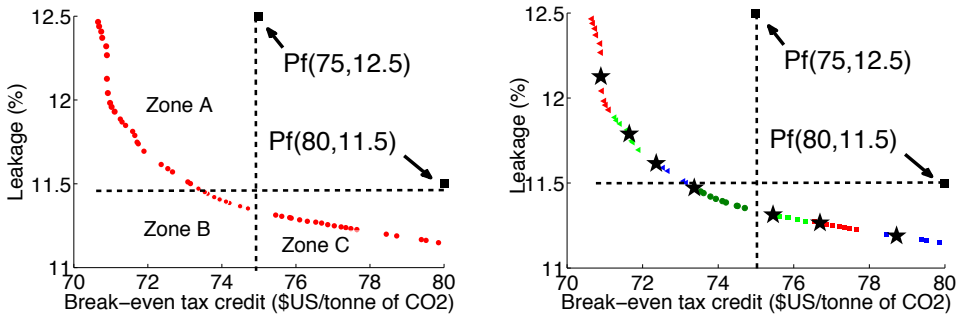


Figure 5.37: a) Illustration of the Pareto-optimal front clustered by the preference points; the solutions in zone B dominates both preference points whereas zone A and C dominate only the point Pf(x_1, y_1) and Pf(x_2, y_2), respectively. b) k-means clustering is used to partition the solutions in to three zones where the minimax approach is applied to each zone to obtain a representative solution (the star marker).

We see that zone B strictly dominates both preference points whereas zones A and C strictly dominate only the point Pf$(75, 12.5)$ and Pf$(80, 11.5)$, respectively. Based on each zone, the k-means clustering approach is applied to partition the data into three groups (for zone A and C) and one group for zone B as shown in Fig. 5.37(b). Then, the minimax approach is used in each group to obtain representative solutions as shown by the star markers. We see that the proposed method using the preference points can reduce 100 solutions to only 7 solutions in this example.

It should be noted that this method can be used in a higher-dimensional criterion space as well, but it might be difficult for the DM to make the preference choices. Additionally, if the DM knows his/her preferences a priori, the bias of the DM preference can be implemented in the NSGA-II to get the Pareto-optimal solutions only in the zone that strictly dominates the preference points.

5.7 Summary

This chapter has focused on solving multiple-objective optimisation for CO_2 sequestration under geological uncertainty. In the first section, we considered mean, interval mean, and dispersion index of the BT function as multiple objectives for finding a set of robust solutions. In the second section, we considered multiple engineering criteria, such as break-even tax credits (BT), volumetric CO_2 sweep efficiency, and the risk of CO_2 leakage, for the practical optimisation of CO_2 sequestration. We summarise this chapter in the following:

1. Based on the rectangular and the PUNQ-S3 models, the trade-off relationship between the mean and the variance of the BT was obtained by using NSGA-II. The results indicated that the mean and the variance could be used under the MOO framework to find robust solutions, as an alternative approach to the utility function method (Chapter 4). Later, we showed that statistical measures could affect the decision making process. For this reason, the bootstrapping method was used to compare different dispersion measures, such as variance, IQR, and 90% CI of the BT function from the rectangular model. It was shown that the distribution of IQR had the smallest dispersion in comparison to the others. Therefore, it is a recommended dispersion measure for the robust optimisation of CO_2 sequestration.

2. For MOO under uncertainty, the choice of parametric distribution functions could impact decision making. Since the distribution function describing the uncertain objective function is generally not known a priori, we used the mixture model to combine three density functions into a single function in this study. Then, the statistical measure based on the mixture distribution was used for robust decision making. However, using the mixture distribution requires more computational resources than using individual parametric distributions.

3. We considered the economic, the long-term security of CO_2 storages, and the volumetric sweep efficiency of CO_2 as the decision criteria for the optimisation of CO_2 sequestration. For the PUNQ-S3 model, we showed that the BT and the unswept volume of aquifer are conflicting objectives. For the leaky PUNQ-S3 model, we used statistical measures of the objective functions to cope with geological uncertainty and we showed that the Pareto-optimal front was efficiently obtained by the NSGA-II approach. The results from both models revealed the potential benefit of using NSGA-II for CO_2 sequestration as the optimal solutions from the non-convex Pareto-front can be accurately captured.

4. Adaptive sparse grid interpolation (ASGI) is a promising surrogate approach to assist MOO. We showed that the non-dominated solutions from the ASGI surrogate correspond to the ones obtained from the simulator very well.

5. We introduced k-means clustering, the minimax method, and the use of preference-driven bias of DM to assist the decision making process by reducing large numbers of non-dominated solutions into a smaller and manageable size, providing an effective tool for decision makers to identify and select alternative solutions.

CHAPTER 6

HIGH-DIMENSIONAL MODEL REPRESENTATION

6.1 Introduction

The chapter presents approaches to building a surrogate model for high-dimensional functions. High-Dimensional Model Representation (HDMR) is introduced first in Section 6.2, followed by several related representation techniques such as ANOVA-HDMR in Section 6.3 and Cut-HDMR in Section 6.4. The truncation form of these HDMRs is discussed in Section 6.5, which becomes a foundation of the surrogate used in the rest of this thesis. Next, Section 6.6 presents an integration of the Adaptive Sparse Grid Interpolation (ASGI) approach and the HDMR for efficient surrogate construction. In the last section, we develop this approach for solving optimisation with many input variables.

6.2 High-dimensional model representation techniques

The idea of representing multivariate functions stems from the Kolmogorov superposition theorem [83] where Kolmogorov proved that every multivariate functions can be written as sums and compositions of monovariate functions. This theorem was further refined and developed by [2, 90, 135, 136, 160] for High-Dimensional Model Representation (HDMR). HDMR can be used as an approximate model to efficiently map the input-output relationship (IO) of the function in a hierarchical structure of correlation effects among the input variables. The HDMR formulation [135, 136] is provided in the following form:

$$\begin{aligned} f(\mathbf{x}) = f_0 + \sum_{i=1}^{d} f_i(x_i) + \sum_{1\le i<j\le d} f_{ij}(x_i, x_j) + \\ + \ldots + \sum_{1\le \ldots \le d} f_{i_1,i_2,\ldots,i_m}(x_{i_1}, x_{i_2}, \ldots, x_{i_m}) + \\ + \ldots + f_{1,2,\ldots,d}(x_1, x_2, \ldots, x_n) \end{aligned} \tag{6.1}$$

where $\mathbf{x}$ is a vector of $[x_1, x_2, \ldots, x_n]$ and the constant f_0 function (the zeroth order of the HDMR expansion) is the mean of the response function $f(\mathbf{x})$ or the constant value evaluated at the reference point. The HDMR expansion is a finite sum of component functions $f_i(x_i), f_{ij}(x_i, x_j), \ldots, f_{1,2,\ldots,d}(x_1, x_2, \ldots, x_n)$; the component functions are not unique and can be defined in different ways depending on the measure criteria used. The univariate functions $f_i(x_i)$ (the first-order HDMR) present the response output associated

independently with the variable x_i. The second-order component functions $f_{ij}(x_i, x_j)$ are associated with the joint variables x_i and x_j to account for the interaction of bivariate inputs on the output response. The last term in the expansion is the residual d-order term to ensure that the HDMR expansion provides an exact representation for $f(\mathbf{x})$. Unlike the polynomial expansion, the HDMR requires only a finite number of terms to have an exact representation.

Given a set $\mathbf{u} \subseteq \mathcal{D}$ where $\mathcal{D} = \{1, \ldots, d\}$ denotes coordinate indices (i) of the input variables x_i. If $\#\mathbf{u}$ presents the cardinality of the set $\mathbf{u}$, then $f_{\mathbf{u}}$ represents $\#\mathbf{u}$-dimensional component functions. For example, if $\mathbf{u} = \{1, 5, 7\}$, then $f_{\mathbf{u}} = f_{1,5,7}(x_1, x_5, x_7)$. The generalised HDMR formulation [135, 136] can be written in the compact notation as expressed by Griebel and Holtz [59], Rabitz and Alis [135] as follows:

$$f(\mathbf{x}) = \sum_{\mathbf{u} \subseteq \mathcal{D}} f_{\mathbf{u}}(\mathbf{x}_{\mathbf{u}}) \tag{6.2}$$

The HDMR component functions in Eq. 6.2 can be derived in the optimisation framework by minimising the objective function below:

$$\begin{aligned} &\underset{f_{i,\ldots,d}}{\arg\min} \int_{\Omega^{\mathrm{d}}} \rho(\mathbf{x})[f(\mathbf{x}) - \sum_{\mathbf{u} \subseteq \mathcal{D}} f_{\mathbf{u}}(\mathbf{x}_{\mathbf{u}})]^2 \mathrm{d}\mu(\mathbf{x}) \\ &\text{subject to:} \quad \int_{\Omega^{\mathrm{d}}} f_{\mathbf{u}}(\mathbf{x}_{\mathbf{u}}) f_{\mathbf{u}'}(\mathbf{x}_{\mathbf{u}'}) \mathrm{d}\mu(\mathbf{x}) = 0 \end{aligned} \tag{6.3}$$

where Ω^d presents the entire domain of variables $\mathbf{x}$ in $[0, 1]^d$ and d is the total number of dimensions; the density $\rho(\mathbf{x})$ is uniform on $[0, 1]^d$ and equal to unity; $f(\mathbf{x})$ is the function to be approximated whereas the component functions of HDMR structure are $f_{\mathbf{u}}(\mathbf{x}_{\mathbf{u}})$ as described in Eq. 6.2; the term $\mathrm{d}\mu(\mathbf{x})$ denotes a measure used for integration. The constraint of the optimisation in Eq. 6.3 determines the orthogonality of the component functions at different dimensions where the term $f_{\mathbf{u}'}(\mathbf{x}_{\mathbf{u}'})$ is a component function with a set $\mathbf{u}'$ defined as $\mathbf{u}' \subseteq \mathcal{D}$ and $\mathbf{u}' \neq \mathbf{u}$.

The component functions $f_{\mathbf{u}}(\mathbf{x}_{\mathbf{u}})$ can be derived from an optimisation framework in Eq. 6.3 from which they are not unique and can be derived in several forms, such as ANOVA-HDMR [135, 160], a random sampling HDMR (RS-HDMR) [91, 92], and Cut-HDMR [135]. This depends on the term $\mathrm{d}\mu(\mathbf{x})$, a measure criterion used to define the component functions. In this work, we have used both ANOVA-HDMR and Cut-HDMR for CO_2 optimisation, the description of which is given in the following section.

6.3 ANOVA high-dimensional model representation

HDMR can be derived to ANOVA-HDMR if the Lebesgue measure is employed on Ω such that $\mathrm{d}\mu(\mathbf{x}) = \prod_{i=1}^{d} dx_i$. ANOVA-HDMR is commonly used in statistics to evaluate the output uncertainty (e.g. analysis of variance (ANOVA) [149]) and to quantity the influential parameters to the model output.

As described in [135, 161], the component functions of the ANOVA-HDMR are given by:

$$\begin{aligned} f_0 &= \int_{\Omega^d} f(\mathbf{x}) \mathrm{d}\mu(\mathbf{x}) \\ f_i(x_i) &= \int_{\Omega^{d-1}} f(\mathbf{x}) \mathrm{d}\mu(\mathbf{x})_{\sim i} - f_0 \\ f_{ij}(x_i, x_j) &= \int_{\Omega^{d-2}} f(\mathbf{x}) \mathrm{d}\mu(\mathbf{x})_{\sim\{ij\}} - f_i(x_i) - f_j(x_j) - f_0 \end{aligned} \tag{6.4}$$

The constant term f_0 is the average over the entire domain $\Omega^d = [0,1]^d$. The symbol $\sim\{i\}$ denotes the exclusive indices of the component i in the set. Therefore, the term $\mathrm{d}\mu(\mathbf{x})_{\sim\{i\}}$ excludes the term dx_i and becomes $\prod_{j=1,j\neq i}^{d} dx_j$; the term $\sim\{ij\}$ follows the same logic. One application of ANOVA-HDMR is a global sensitivity analysis, which captures the significant input uncertainty that attributes to the output uncertainty [23, 92, 146]. However, evaluating the numerical integration of Eq. 6.4 can be computationally expensive especially for high-dimensional problems. For this reason, a Monte Carlo approach or orthonormal basis functions could be employed to approximate the multivariate quadrature; this method is referred to as Random Sampling (RS-HDMR) [91], which relies on random samples or the basis functions. Alternatively, instead of finding the average function over the input domain, the component functions can be derived from a reference point (referred to as Cut-HDMR), which attempts to mitigate the computational cost from multivariate quadrature.

6.4 Cut high-dimensional model representation

In contrast to solving multivariate quadrature of the component functions in the ANOVA-HDMR, a reference or an anchored point is used to approximate the component function instead. Cut-High-Dimensional Model Representation (Cut-HDMR) is derived from the optimisation framework in Eq. 6.3 by employing a Dirac measure such that $\mathrm{d}\mu(\mathbf{x^c}) = \prod_{i=1}^{d} \delta\left(x_i - x_i^c\right) dx_i$ at the cut-point x^c and consequently the multivariate quadrature becomes the function evaluated at the cut-point, as $\int f(\mathbf{x}) \mathrm{d}\mu(\mathbf{x^c}) = f(\mathbf{x^c})$. The component functions of the Cut-HDMR have the following terms:

$$\begin{aligned} f_0 &= f(\mathbf{x^c}) \\ f_i(x_i) &= f(\mathbf{x}) \mid_{\mathbf{x}=\mathbf{x}^c \backslash x_i} - f_0 \\ f_{ij}(x_i, x_j) &= f(\mathbf{x}) \mid_{\mathbf{x}=\mathbf{x}^c \backslash \{x_i, x_j\}} - f_i(x_i) - f_j(x_j) - f_0 \end{aligned} \tag{6.5}$$

The notation $\mathbf{x} = \mathbf{x^c} \backslash x_i$ denotes that $\mathbf{x}$ except x_i is set equal to the cut-point or anchored point $\mathbf{x}^c$. For example, if $\mathbf{x} = (x_1, \ldots, x_n)$, then $\mathbf{x}^c \backslash x_i = (x_1^c, \ldots, x_{i-1}^c, x_i, x_{i+1}^c, \ldots, x_n^c)$. The subtraction of the lower-order terms from the component functions (e.g. f_0 is subtracted from the $f(\mathbf{x}) \mid_{\mathbf{x}=\mathbf{x}^c \backslash x_i}$ components; the f_0 and $f_i(x_i)$ are subtracted from the $f(\mathbf{x}) \mid_{\mathbf{x}=\mathbf{x}^c \backslash \{x_i, x_j\}}$ components) ensures that only the contribution of the particular order is present. The component functions in Eq. 6.5 are defined along the cut lines or planes or hyper-volumes passing through the cut-point $\mathbf{x}^c$; therefore, the function values are always exact along these hyper-volumes through the cut-point. In contrast, the function values at other locations are approximate by projecting to the hyper-volumes.

Figure 6.1 illustrates the Cut-HDMR with the cut-point $\mathbf{x}_c$ to approximate a 2-dimensional saddle function (the brown surface). We see that the tangent plane is the function f_0 (Eq. 6.5) evaluated at the cut-point $\mathbf{x}_c$ whereas the dashed curves shown on the cutting planes represent the functions $f(\mathbf{x}) \mid_{\mathbf{x}=\mathbf{x}^c \setminus x_i}$, a projection from any point $\mathbf{x}$ to the cutting plane. The function values on the cutting planes subtracted by the function value of the tangent plane is basically the first-order component function $f_i(x_i)$ in Eq. 6.5.

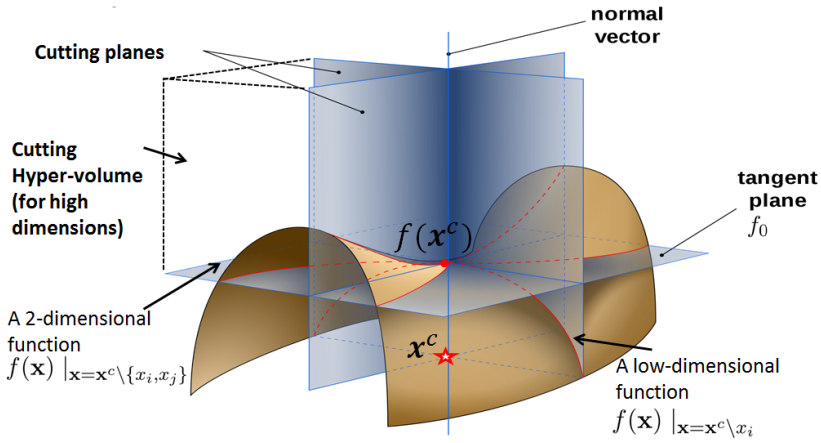


Figure 6.1: Illustration of the Cut-HDMR approach with the cut-point $\mathbf{x}_c$. The saddle surface is a 2-dimensional function where the tangent plane is the function f_0 (Eq. 6.5) evaluated at the cut-point whereas the dashed curves shown on the cutting planes represent the functions $f(\mathbf{x}) \mid_{\mathbf{x}=\mathbf{x}^c \setminus x_i}$, a projection from any point $\mathbf{x}$ to the plane. For a higher-dimensional function ($d > 2$), we also show a 2-dimensional function the domain of which is projected to the cut-point (x_i, x_j). This figure was modified from [143].

Given the component functions $f_{\mathbf{u}}(\mathbf{x_u}) = f_{\mathbf{u}}^{cut}(\mathbf{x_u})$ in Eq. 6.2, we can write the Cut-HDMR expansion in a more compact form [59] below:

$$f_{\mathbf{u}}^{cut}(\mathbf{x_u}) = f(\mathbf{x}) \mid_{\mathbf{x}=\mathbf{x}^c \setminus \mathbf{x_u}} - \sum_{\mathbf{v} \subset \mathbf{u}} f_{\mathbf{v}}^{cut}(\mathbf{x_v}) \tag{6.6}$$

where $\mathbf{u}$ is described previously and $\mathbf{v}$ is a set of lower-order indices to represent the recursive subtraction from $f(\mathbf{x}) \mid_{\mathbf{x}=\mathbf{x}^c \setminus \mathbf{x_u}}$.

6.5 Truncated high-dimensional model representation

In most physical problems, high-order interactions among input variables, especially in a high-dimensional space, may not necessarily have an impact on the output response and the relatively low-order correlations of input variables can be used to represent the model in high-dimensional problems. As addressed in [148], the structure of the HDMR is analogous to the many-body expansion employed in the molecular physics to represent energy at the atomic level where only the two-body terms (second-order interactions) dominate the higher-body terms. As well as other practical applications, it has been observed that the higher-order terms in the HDMR expansion are less significant [25, 90].

This idea leads to a truncated HDMR formulation which has only lower-order terms (the second order is suggested in some applications). We will show later that the truncation

of the HDMR is a reasonable assumption and can be used to approximate functions with highly multiplicative structure (e.g. $f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i)$) when the domain of input variables is small. The formulation of the truncated HDMR at the second order is expressed as:

$$f(\mathbf{x}) \approx f_0 + \sum_{i=1}^{d} f_i(x_i) + \sum_{1 \leq i < j \leq d} f_{ij}(x_i, x_j) \tag{6.7}$$

When truncated HDMR (Eq. 6.7) is employed with the Cut-HDMR, the choice of the cut-point x^c could have a large impact on the approximation. We refer to [50, 177] for strategies to select the cut-point; one suggestion is to use the mean of the uniform sampling of the input variables for the cut-point so that $f(x^c = \hat{x}) \approx \hat{f}$, where $\hat{x}$ and $\hat{f}$ represent the mean of x and the function $f(\cdot)$, respectively. Alternatively, multiple cut-points could be employed by placing multiple local HDMR for each cut-point; this technique is referred to as Multicut-HDMR [8, 93]. However, inspired by the idea of the Multicut-HDMR and the nested partition method by Shi and Olafsson [154], we have adapted this technique to accelerate the optimisation process by iteratively moving the cut-point close to the global optimum, the method of which is referred to as Walking Cut-HDMR and will be described later in this chapter.

6.6 Hierarchical adaptive sparse grid using high-dimensional model representation

The idea is to integrate the Adaptive Sparse Grid Interpolation (ASGI) with the HDMR expansion, to build a surrogate for high-dimensional models. We will refer to this method as the ASGI-HDMR method. The sparse grid method was presented in Chapter 4 where we showed that the ASGI is an efficient surrogate technique to approximate the input-output relationship of the objective function.

The underlying principle of hierarchical adaptive sparse grid with HDMR (ASGI-HDMR) is the use of the ASGI technique to approximate the component functions of the Cut-HDMR in Eq. 6.6. Therefore, the lower-component function in the HDMR structure is given by the approximation of the ASGI approach. It, however, should be noted that the HDMR referred to in this work is the Cut-HDMR and the term ASGI-HDMR is actually the hierarchical ASGI using the Cut-HDMR method, unless otherwise stated. Figure 6.2 illustrates the concept of the ASGI-HDMR by denoting $\mathcal{Q}_{q,d}(f)$ for the surrogate obtained from the ASGI technique. The surrogate of the function $f(\mathbf{x})$ is the summation of multiple component functions, each of which is approximated by the ASGI surrogate.

We recall from Chapter 4 that $f(\cdot) \in C^{\alpha}[0,1]$ denotes a smooth objective function and l_j presents the level of interpolation along jth dimension where $|\underline{l}| = l_1 + \cdots + l_d$ for $\underline{l} \in \mathbb{N}^d$. Then the d-dimensional sparse grid interpolation given in Eq. 4.36 is:

$$\mathcal{Q}_{q,d}(f) = \mathcal{Q}_{q-1,d}(f) + \sum_{|\underline{l}|=q} \sum_{\underline{j}} \phi_{\underline{j}}^{\underline{l}} \cdot w_{\underline{j}}^{\underline{l}} \tag{6.8}$$

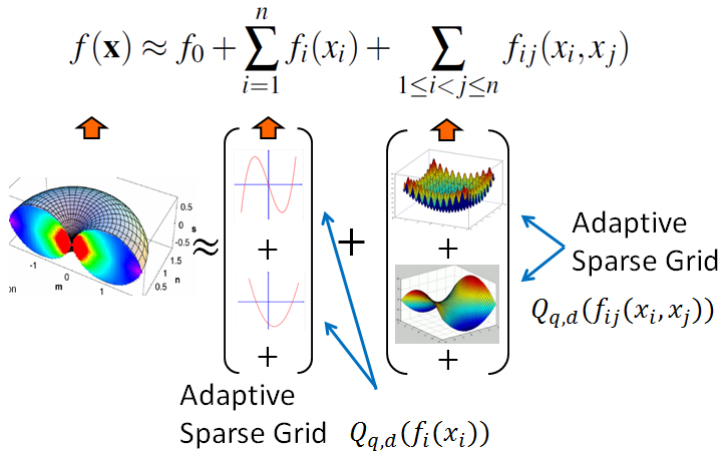


Figure 6.2: Illustration of the hierarchical adaptive sparse grid high-dimensional model representation (ASGI-HDMR); $\mathcal{Q}_{q,d}(f)$ represents the surrogate of the component functions.

where $\phi_{\underline{j}}^{\underline{l}}$ represents basis functions with interpolation level $\underline{l}$, $\underline{l} = (l_1, \ldots, l_d)$, $\underline{j}$ denotes a set of multi-indices $(j_{l_1}, \ldots, j_{l_n})$ where j_{l_d} presents the set of collocation points that do not belong to the previous interpolation level $(l-1)$ at d-dimension and $w_{\underline{j}}^{\underline{l}}$ is the hierarchical surplus used for dimensional adaptivity.

Based on the ASGI structure in Eq. 6.8 and the HDMR formulation in Eq. 6.2, we can write the ASGI-HDMR with the cut-point in the following form:

$$f(\mathbf{x}_\Omega) = \sum_{\mathbf{u} \subseteq \mathcal{D}} \mathcal{Q}_{q,d=\#\mathbf{u}}(f_{\mathbf{u}}^{cut}(\mathbf{x}_\mathbf{u})) \tag{6.9}$$

where $\mathbf{x}_\Omega$ denotes $\forall \mathbf{x} \in \Omega$ with the domain $\Omega = [0,1]^d$; the other parameters were described previously. Based on the component functions in Eq. 6.6, the recursive sum of the component function surrogates in Eq. 6.9 is explicitly expressed below:

$$\mathcal{Q}_{q_j,d=\#\mathbf{u}}(f_{\mathbf{u}}^{cut}(\mathbf{x}_\mathbf{u})) = \mathcal{Q}_{q_j,d=\#\mathbf{u}}(f(\mathbf{x}) \mid_{\mathbf{x}=\mathbf{x}^c \setminus \mathbf{x}_\mathbf{u}} - \sum_{\mathbf{v} \subset \mathbf{u}} \mathcal{Q}_{q_k,d=\#\mathbf{v}}(f_{\mathbf{v}}^{cut}(\mathbf{x}_\mathbf{v}))) \tag{6.10}$$

where $f(\cdot)$ presents the objective function calculated from a numerical flow simulator; the Smolyak numbers q_j and q_k are independent as they are defined individually based on the hierarchical surplus at different dimensional functions. The ASGI-HDMR approach is developed further for a surrogate-assisted optimisation method in the next section.

6.7 Surrogate building techniques for optimisation

An optimisation procedure in many practical problems usually requires a large number of objective function evaluations, which is computationally expensive and might be impossible to solve in some applications. In order to reduce the computational costs, we develop the use of ASGI-HDMR to accelerate the optimisation process in this study. Generally, there are three main approaches for building the surrogate. (1) A surrogate is built globally to approximate the objective function for the entire range of input variables. The term 'global' in this case means the global range of input variables; this also includes multiple

subregion surrogates built to approximate the entire input domain (e.g. Li et al [93], Zhao and Xue [178]. (2) The surrogate is built locally in a subregion of the objective function. (3) A hybrid method where the surrogate is first built globally and then refined locally; this approach includes, for example, an expected improvement algorithm, see e.g. [37, 75], and our proposed algorithm described later in this chapter. However, the implementation of these techniques usually depends on the specific applications.

In the field of uncertainty quantification, building a global surrogate could reveal valid information about the global properties of the model such as the non-linear uncertainty structure embedded in the model [157]. On the other hand, for most optimisation problems, the aim is to guide the search algorithm to the optimal solution iteratively by not necessarily building the surrogate for the entire domain.

In this study, we propose an optimisation technique that can balance exploration and exploitation of the search algorithm. The global characteristics of the objective function are approximate by the low-order polynomial before the higher-order surrogate is built locally. Having the surrogate built locally near the optimum solution is the principle underlying Walking Cut-HDMR.

6.7.1 A walking Cut-HDMR technique

We first demonstrate a surrogate building technique that is used to approximate the objective function for the entire range of input variables. Later, we will present our proposed surrogate built locally to approximate the objective function.

In Fig. 6.3, the surrogate function built globally is shown, where the contour of the function and the global optimum solution are displayed with the coloured curves and the star-shape marker, respectively. Collocation points used for building the surrogate are presented by the blue dots. We divide the input domain by region A and B. As can be seen, the region A (shaded area) does not contain the optimal point. Therefore, building the surrogate in this region would be unnecessary because we are only interested in the global optimum point. In contrast, more sampling points should be allocated to the region B to construct the surrogate near the optimum solution. Motivated by this idea and the nested partition method [154], we build a surrogate that iteratively shrinks the domain while recentering the surrogate to the optimum solution. We denote this technique as Walking Cut-HDMR (WCut-HDMR).

The WCut-HDMR process involves four iterative steps. (1) Exploration: a low-order surrogate model is built for the entire input domain using the ASGI technique. (2) Preparation using the particle swarm optimisation (PSO) solver [78] on the ASGI surrogate to find the best optimal solution, which will be used as a cut-point for the ASGI-HDMR expansion in the next step. The criteria for the shrinking domain are calculated in this process. (3) Zooming: the ASGI-HDMR surrogate is built with the shrunk domain using a cut-point obtained from the previous step. (4) Walking: the PSO solver is employed on the surrogate in step (3) to find a new cut-point. The domain shrinking criteria are updated. Steps (3) and (4) are repeated until the stopping criteria are satisfied. The stopping criteria are specified based on 1) the difference in the optimal value is below a specified threshold, 2)

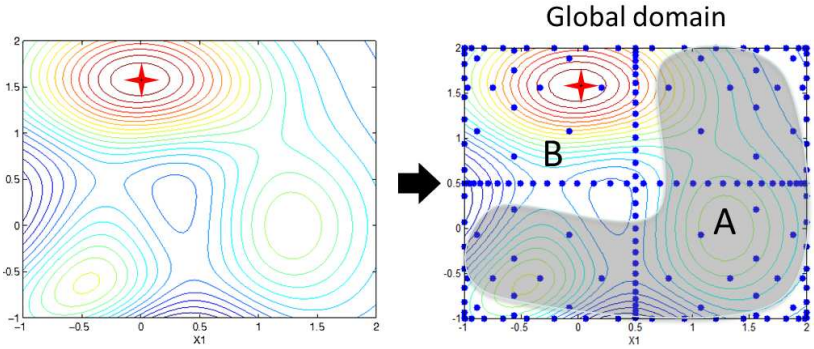


Figure 6.3: Illustration of a surrogate built for the entire range of input space. The contour of the function and the global optimum solution are shown with the coloured curves and the star-shape marker, respectively. Collocation points for building the surrogate are presented by the blue dots. In contrast to the region B, the shaded region A does not contain the optimal point.

the Euclidean distance of the optimal input variables in each iteration is below a specified threshold, 3) the hyper-volume domain of the surrogate, and 4) the number of iterations.

Figure 6.4 illustrates the algorithm by showing the function contour with collocation points (blue dots) for building the surrogate. The iterative steps of the algorithm are shown in the figure from left to right. The triangle marker, represented as the local optimal point, is the cut-point of the ASGI-HDMR expansion in the next iteration. The detailed procedure is provided in Table 6.1.

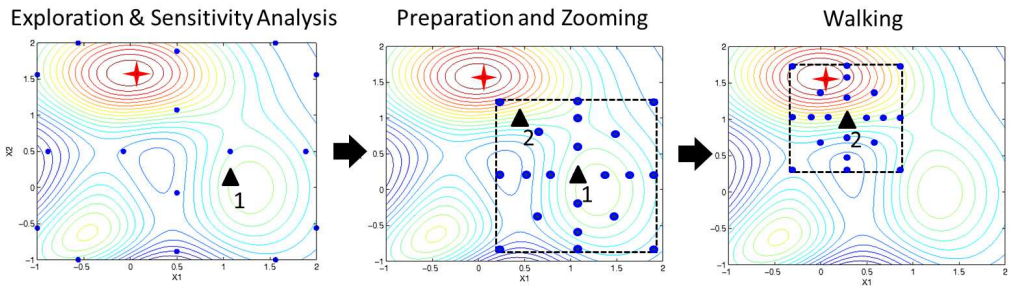


Figure 6.4: A Walking Cut-HDMR approach is illustrated at different steps: 1) exploration, 2) preparation, 3) zooming, and 4) walking. The triangle markers denote the optimal point obtained from the PSO optimisation; the optimal point will then be used as the cut-point for the next iteration.

In Table 6.1, step 3 of the preparation procedure is to check whether the optimum value from the surrogate $\mathcal{Q}(\mathbf{x}_j^*)$ is lower than the function values evaluated from the collocation points. This procedure can be done without additional cost since the function values of these collocation points are obtained when building the surrogate. On the other hand, if there exists one collocation point which has a lower function value than $\mathcal{Q}(\mathbf{x}_j^*)$, then this point will be used as the optimal point for the current surface and the cut-point for the next iteration. In step 5, the π-notation is used to iteratively update the domain shrinking criteria α_j. If π is equal to 1, then the α-parameter will be a constant value, and hence the updated

domain is given in the following:

$$\Omega_{j+1} = \alpha\Omega_j, \qquad \alpha \in [0,1] \tag{6.11}$$

where the α-alpha is an isotropic parameter that reduces the range of input variables in all dimensions. Therefore, it is a ratio of the hyper-volume of the updated domain (Ω_{j+1}) in relation to the previous one (Ω_j).

Table 6.1: A detailed procedure of the Walking Cut-HDMR method

Exploration:	Step 1	Set initial iteration $j = 1$, input domain Ω_0, and domain criteria α_0 equal to 1; build the surrogate from the ASGI technique for the entire input space; keep all collocation points obtained from the ASGI in the set X^j_{col} where $\mathbf{x}^j_{col} \in X^j_{col}$ and $X^j_{col} \subset X^{j+1}_{col}$; define $\mathcal{F}(\mathbf{x}^j_{col})$ and $\mathcal{Q}(\mathbf{x}^j_{col})$ for the function value evaluated at the collocation point $\mathbf{x}^j_{col}$ using numerical flow simulator and the surrogate, respectively.
Preparation:	Step 2	Particle swarm optimiser (PSO) [78] is employed on the ASGI surrogate to find the optimal solution $\mathbf{x}^*_j$ and the optimal value $\mathcal{Q}(\mathbf{x}^*_j)$
	Step 3	Check if $\mathcal{Q}(\mathbf{x}^*_j) <$ Min$\{\mathcal{F}(\mathbf{x}) : \forall \mathbf{x} \in X^j_{col}\}$, evaluate $\mathcal{F}(\mathbf{x}^*_j)$ and set $\mathcal{Q}(\mathbf{x}^*_j) = \mathcal{F}(\mathbf{x}^*_j)$ else, find $\mathcal{F}(\mathbf{x}^{j,*}_{col}) =$ Min$\{\mathcal{F}(\mathbf{x}) : \forall \mathbf{x} \in X^j_{col}\}$ set $\mathbf{x}^*_j = \mathbf{x}^{j,*}_{col}$ and $\mathcal{Q}(\mathbf{x}^*_j) = \mathcal{F}(\mathbf{x}^{j,*}_{col})$ end
	Step 4	Define optimal point $\mathbf{x}^*_j$ as a cut-point $\mathbf{x}^c_j$ for building the Cut-HDMR
	Step 5	The criterion for the domain shrinking is calculated by setting $\alpha_j = \pi\alpha_{j-1}$ where π is a predefined value. The updated domain becomes $\Omega_j = \alpha_j\Omega_{j-1}$
Zooming:	Step 6	The ASGI-HDMR is built at the cut-point $\mathbf{x}^c_j$ for the new domain Ω_j
Walking:	Step 7	Update $j = j + 1$; Step 2-7 are repeated until the stopping criteria are satisfied.

Based on Eq. 6.9, we use the Cut-HDMR expansion with the optimisation procedure proposed in Table 6.1 to formulate the WCut-HDMR method,

$$\underset{\mathbf{x}\in\Omega_j}{\arg\min} \sum_{\mathbf{u}\subseteq\mathcal{D}} \mathcal{Q}_{q,d=\#\mathbf{u}}(f^{cut}_{\mathbf{u}}(\mathbf{x}_{\mathbf{u}})) \tag{6.12}$$

where j represents an iterative step of the optimisation process; Ω_j is the input domain of iteration j, whose properties are given by $\Omega_{j+1} \subset \Omega_j$ and $\Omega_j \subseteq \Omega$; the other parameters were defined previously in Eq. 6.9.

We see that the WCut-HDMR method uses both global and local (more refined) surrogates to search for optimal solutions, where the trade-off between exploration and exploitation phases of the optimisation process is balanced by the α-parameter (Eq. 6.11). The idea behind using a low order surrogate for the global search is already available, e.g. the efficient global optimization (EGO) algorithm [75]. The EGO algorithm first builds a surrogate with a simple but practical space-filling experimental design. Then it uses this surrogate to estimate the next sampling point based on the maximum expected improvement conditioned to the previously evaluated points. The EGO algorithm has been shown to be an effective method (see [46, 48]) as it considers both uncertainty of the surrogate and

the prediction of the next sampling point. However, for high-dimensional objective functions, the EGO can suffer from the curse of dimensionality when using the space-filling method to construct the response surface. This is in contrast to our proposed algorithm that utilises dimensional adaptivity to reduce the number of sampling points. In addition, our proposed WCut-HDMR method can utilise the hierarchical surpluses to individually probe the linear/non-linear relationship of objective functions for each dimension, which information can be used to improve surrogate quality.

6.8 Applications : hierarchical adaptive sparse grid using high-dimensional model representation

The section presents the applications of the ASGI-HDMR approach in comparison to the ASGI technique for high-dimensional problems. Simple mathematical functions are used for convergence studies and to demonstrate several characteristics of HDMR expansion, from which we discuss the curse of dimensionality, the order of HDMR, and subdomain HDMR. Last, numerical evaluations of the WCut-HDMR approach for CO_2 optimisation using the PUNQ-S3 model are presented.

6.8.1 Convergence studies using simple mathematical test functions

In Fig. 6.5, we compare convergence rates of different benchmark functions using three surrogate building techniques: the tensor product grid, the regular sparse grid (RSGI), and the adaptive sparse grid (ASGI). The reference response surfaces constructed from the explicit functions (Eqs. 4.41 – 4.44) are displayed at the top row of Fig. 6.5, while the convergence rates of the tensor product grid approach (the dashed curves with circle markers), the RSGI (the blue curves with square markers), and the ASGI (the red curves with triangle markers) are shown at the bottom row. As can be seen in Fig. 6.5, the adaptive sparse grid generally provides better convergence rates. This implies a significant reduction of evaluation points required to achieve the same level of accuracy of the other approaches (i.e. the tensor product grid and the regular sparse grid). The detailed description of ASGI technique given earlier in Section 4.4.6 showed that the ASGI selects only the points that are important to the objective function and discards less significant points. The advantages of using the ASGI for CO_2 sequestration applications was presented in Chapter 4. However, for problems with a large number of input variables, the ASGI technique could suffer from the curse of dimensionality.

6.8.2 The curse of dimensionality

The curse of dimensionality refers to the exponential growth of the domain volume with an increase of the number of dimensions, which results in an exponential growth in the required sampling points to accurately represent a function on that domain [10]. Even though the ASGI is an efficient surrogate technique, it can sometimes suffer from the curse of dimensionality.

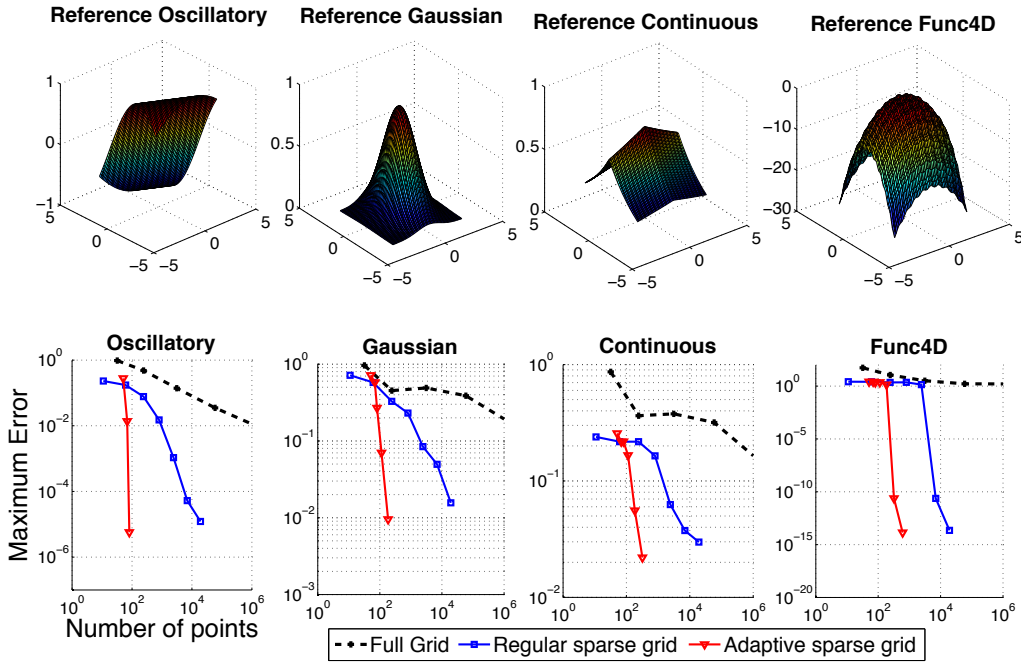
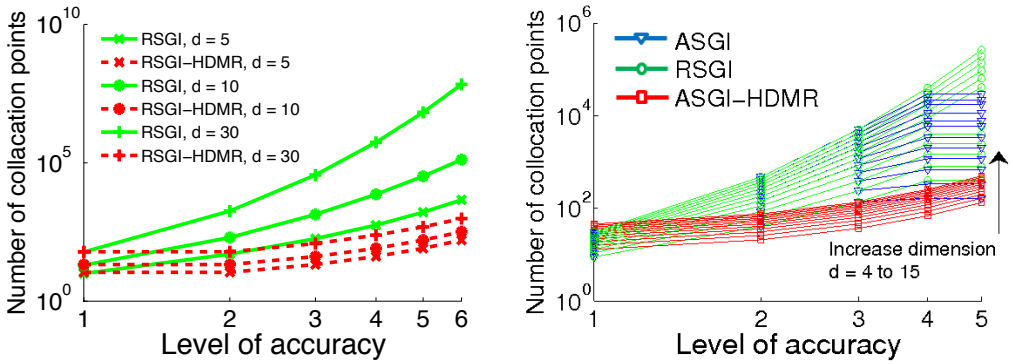


Figure 6.5: Sparse grid with dimensional adaptivity (ASGI), with non-dimensional adaptivity (RSGI), and the product grid approach are shown by the red, blue, and black curves, respectively. At the same maximum error (y-axis), the ASGI requires significantly fewer sampling points than the other methods (RSGI and the product grid).

We illustrate the curse of dimensionality in Fig. 6.6(a) by showing the relationship between the number of collocation points and the level of interpolation for different problem dimensions ($d = 5, 10, 30$) with two different surrogate building techniques: the regular (non-dimensional adaptivity) sparse grid interpolation (RSGI), and the RSGI with the HDMR structure. It should be noted that the higher the level of interpolation, the better the approximation of the smooth function. As can be seen in Fig. 6.6(a), the number of collocation points of the sparse grid (the green curves) increases exponentially with dimension. However, the number of collocation points of the RSGI with the HDMR technique (RSGI-HDMR) increases at a much slower rate than the RSGI; this implies that the RSGI-HDMR requires a substantially smaller number of collocation points for higher dimension in comparison to the RSGI method.

In Fig. 6.6(b), we use a Gaussian function (Eq. 4.42) to compare the RSGI, dimensional adaptivity (ASGI), and the ASGI coupled with the HDMR expansion (ASGI-HDMR). The Gaussian function was tested for different dimensions from d = 4 to 15 increasing monotonically as shown. We see in Fig. 6.6(b) that the curves from the ASGI technique requires fewer points than the RSGI approach for a higher level of interpolation. However, for higher dimensions it suffers from the curse of dimensionality — requiring a large number of collocation points. In contrast, for higher dimensions and higher levels of interpolation, the ASGI-HDMR outperforms the ASGI. For this reason, we will use the ASGI-HDMR technique to assist the optimisation process in this study.



(a) (b)

Figure 6.6: The relationship between the number of collocation points and the level of interpolation for different problem dimensions is shown. (a) The solid curves show Regular Sparse Grid Interpolation (RSGI) and the dashed curves present the RSGI with HDMR for different dimensions ($d = 5, 10, 30$). (b) The green curves show the RSGI approach, while the blue curves display the Adaptive Sparse Grid (ASGI) technique. The ASGI with HDMR technique is shown with the red curves. The dimension of the Gaussian function (Eq. 4.42) shown in (b) increases monotonically upward from d = 4 to 15.

6.8.3 Mathematical function structures

We present numerical evaluations on two different functional structures: additive and multiplicative functions by setting the threshold of the hierarchical surpluses for dimensional adaptivity (the $\varepsilon_{w_q}^{abs}$ and $\varepsilon_{w_q}^{rel}$ from Eqs. 4.37 and 4.38) to 10^{-6} in all case studies, unless otherwise stated. Based on the HDMR structure in Eq. 6.1, any purely additive functions can be represented exactly by the first-order HDMR so the second- or higher-order terms may not be necessary. However, in a multiplicative structure, additional higher component functions are required for an exact representation. Hence, building the HDMR for multiplicative functions could be more computationally expensive than additive functions. We will demonstrate this in the following examples.

6.8.3.1 Example 1: additive structure function

We first consider a function with additive structure over the range $[0, 1]^d$ for $d = 10$ dimensions as follows:

$$f(\mathbf{x}^j) = \sum_{i=1}^{d} \tan(x_i) \tag{6.13}$$

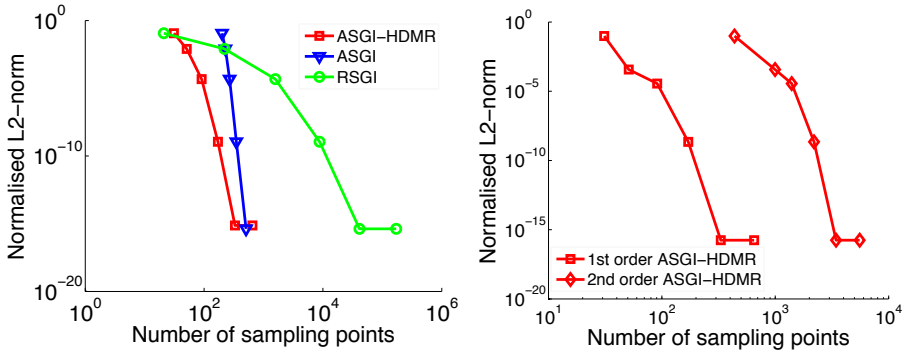
where $\mathbf{x^j} = (x_1, \ldots, x_d)$ is the input vector of the j^{th} sampling set. In this convergence study, the following normalised l_2-norm is used as an error measure.

$$\varepsilon_{L_2} = \frac{\sqrt{\sum_{j=1}^{N_s} (f_{exact}(\mathbf{x}^j) - f_{approx}(\mathbf{x}^j))^2}}{\sqrt{\sum_{j=1}^{N_s} f_{exact}(\mathbf{x}^j)^2}} \tag{6.14}$$

where N_s is a set of uniformly distributed random points within the domain; the number of samples used in this study is 10,000 points. The terms f_{exact} and f_{approx} present the test function (Eq. 6.13) and the surrogate built from the ASGI-HDMR, respectively.

Orders of ASGI-HDMR expansion. We present the convergence rates of the ASGI-HDMR, in comparison to the ASGI and the RSGI, in Fig. 6.7 by using the additive function in Eq. 6.13. Since HDMR is a hierarchical structure of correlation effects from different input variables, its expansion provides an exact representation for functions with additive structure. Hence, building the surrogate from the ASGI-HDMR would be more efficient than the RSGI and ASGI. As expected, Fig. 6.7 shows that the ASGI-HDMR has a better convergence rate than the other approaches for the same error measure because building the surrogate of component functions $f_i(x)$ is more efficient than constructing the surrogate of the original function $f(x)$ which have much higher dimensions than the component functions.

In Fig. 6.7(b), we show that the first-order HDMR ($\#\mathbf{u}$ = 1 from Eq. 6.2) is simply the sum of univariate functions and adding the second-order terms to the first-order ASGI-HDMR does not improve the accuracy, but increases the collocation points. It can be seen that the diamond-marker curve is shifted to the right-hand side of the square-marker curve without improving the accuracy of the interpolation. As expected, all higher-order terms (e.g. $f_{ij}, f_{ijk}, \ldots$) in this example are equal to zero and do not contribute to the approximation of the function.



(a) (b)

Figure 6.7: The relationship of normalised l_2-norm and the number of sampling points: a) The convergence rates of RSGI (the green curve), ASGI (the blue curve), and ASGI-HDMR (the red curve) are shown. The test function of Eq. 6.13 was evaluated using 10 input variables. We present only the first-order ASGI-HDMR. b) Both first- and the second-order ASGI-HDMR are shown with the square and the diamond markers, respectively.

The convergence rates of the ASGI-HDMR for additive functions with 10 dimensions using Eqs. 6.13 and 4.44 are shown in Figs. 6.8(a) and 6.8(b), respectively. In the figure, three different orders of the truncated ASGI-HDMR using the same reference point $\mathbf{x_c} = (0.5, \ldots, 0.5)$ are shown. The explicit forms of the ASGI-HDMR expansion for different

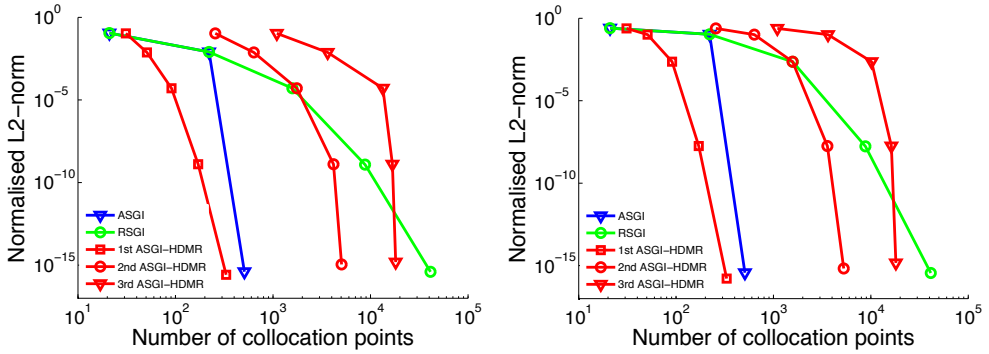
orders are given below.

$$f_1(\mathbf{x}) = f_0 + \sum_{i=1}^{d} f_i(x_i) \tag{6.15}$$

$$f_2(\mathbf{x}) = f_0 + \sum_{i=1}^{d} f_i(x_i) + \sum_{1\le i<j\le d} f_{ij}(x_i, x_j) \tag{6.16}$$

$$f_3(\mathbf{x}) = f_0 + \sum_{i=1}^{d} f_i(x_i) + \sum_{1\le i<j\le d} f_{ij}(x_i, x_j) + \sum_{1\le i<j<k\le d} f_{ijk}(x_i, x_j, x_k) \tag{6.17}$$

where the functions f_1, f_2, and f_3 are the ASGI-HDMR expansion up to the order 1, 2, and 3, respectively. The descriptions of the component functions f_i, f_{ij}, and $f_{i,j,k}$ were given in Eq. 6.1. We see from Fig. 6.8 that the second- and the third-order ASGI-HDMR do not contribute to the approximation of the function since the error decays from both methods are not better than the first-order while requiring more collocation points. However, the error decays of the first order from both test functions in Figs. 6.8(a) and 6.8(b) are in line with the result in Fig. 6.7, which confirms that the high-order component functions do not contribute to the model response for additive functions.

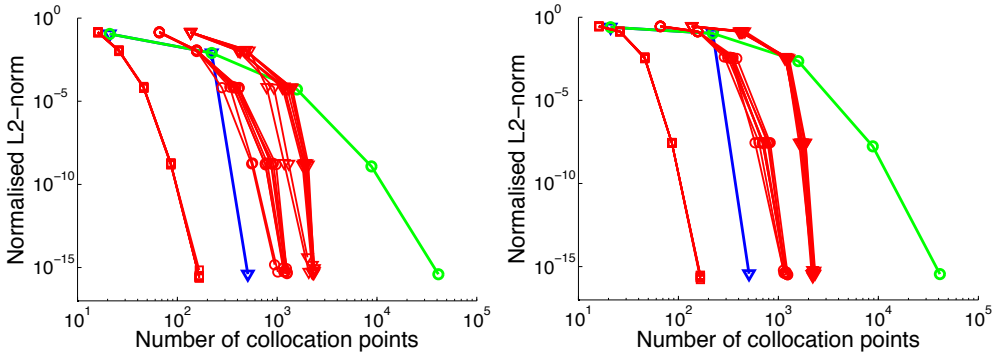


(a) Funciton from Eq. 6.13 (b) Funciton from Eq. 4.44

Figure 6.8: The convergence rates of the approximate function from the ASGI (the blue curves), the RSGI (the green curves), and the ASGI-HDMR (the red curves) approaches. The 10-dimensional test functions in a) and b) are from Eq. 6.13 and Eq. 4.44.

Reference points of the ASGI-HDMR expansion. As discussed previously in Section 6.5, the choice of reference points could impact the approximation of the function if the truncated Cut-HDMR is employed. In this section, Fig. 6.9 presents convergence rates of additive functions (Eqs. 6.13 and 4.44) with various reference points randomly sampled from the uniform distribution for the total of 9 points. The ASGI-HDMR curve in the figure corresponds to each reference point, resulting in 9 curves for each order of the ASGI-HDMR. We see from Fig. 6.9 that the error curves of both second- and third-order ASGI-HDMR (the red-circle and the red-triangle markers) are sensitive to the reference points, while the error curves of the first order show a relatively smaller magnitude of variation. This may imply that the reference points have a smaller effect on the approximation using the first-order ASGI-HDMR in comparison to the higher-order ones. In addition, the convergence

rate of the first order is still better that the ASGI and the RSGI approaches for both functions, which corresponds to the results in Fig. 6.8. It should be noted that the sampling set $\mathbf{x}^j$ used in the normalised l_2-norm in Fig. 6.9 is different from the one used in Fig. 6.8.



(a) Funciton from Eq. 6.13 (b) Function from Eq. 4.44

Figure 6.9: The convergence rates of the approximate function from the ASGI (the blue curves), the RSGI (the green curves), and the ASGI-HDMR (the red curves) are shown with various reference points randomly sampled from the uniform distribution for the total of 9 points. The ASGI-HDMR curve in the figure corresponds to each reference point, resulting in 9 curves for each order of the ASGI-HDMR. The 10-dimensional test functions in a) and b) are from Eq. 6.13 and Eq. 4.44.

6.8.3.2 Example 2: multiplicative structure function

For the function with multiplicative structure, we consider the following function over the range $[0, 1]^d$ for $d = 10$ dimensions in the following form:

$$f(\mathbf{x}^j) = (\sum_{i=1}^{d} x_i)^m \tag{6.18}$$

where the variable $\mathbf{x}^j$ defined previously was drawn from a uniform distribution for $1,000$ samples; m is a positive number and indicates the level of variable correlations. For example, if m = 1, it is a purely additive function since it is the sum of univariate variables, whereas m = 2 or higher guarantees that the bivariate or higher correlated interactions exist. This can be illustrated further when $m = 2$ as $(\sum_{i=1}^{2} x_i)^2 = x_1^2 + 2x_1x_2 + x_2^2$. Hence, the second-order HDMR (with univariate and bivariate component functions) would be sufficient for this representation because we can write $f_{HDMR} = f_0 + f_1(x_1) + f_2(x_2) + f_{1,2}(x_1, x_2)$ where the component functions are $f_0 = 0$, $f_1(x_1) = x_1^2$, $f_2(x_2) = x_2^2$, and the residual term $f_{1,2}(x_1, x_2)$ which is equal to $2x_1x_2$. For the convergence study, we used the normalised error measure (Eq. 6.14) on the multiplicative function.

Figure 6.10(a) presents the normalised l_2-norm plotted versus the order of ASGI-HDMR. Each curve represents different multiplicative functions by varying m. As can be seen, there is no improvement in the convergence once the order of ASGI-HDMR is greater than m. As expected, the higher component functions (greater than m) are not necessary for approximating the function since the component terms are all zeros in this case. A similar

convergence rate for the function with 6 and 10 dimensions for $m = 2$ are shown by the square-marker and the diamond-marker curves. The number of dimensions does not affect the approximation as long as both first- and second-order component functions are included in the expansion. As expected, the ASGI-HDMR expansion represents the interaction effects among input variables and its approximation does not depend on the number of dimensions.

In Fig. 6.10(b), we report the convergence rates of the ASGI-HDMR for two levels of interpolation; the level is indicated by the index number shown in the figure. For smooth functions, we recall that the error of polynomial interpolants is zero when the interpolation level approaches infinity. Therefore, the higher the index number, the greater the accuracy of the interpolation. For each interpolation level, the error convergences plotted versus the number of collocation points were obtained by varying the order of ASGI-HDMR expansion from 1 to 5. In other words, the order j was varied from 1 to 5 where j represents the ASGI-HDMR expansion truncated at the $(j+1)^{th}$-dimensional component functions.

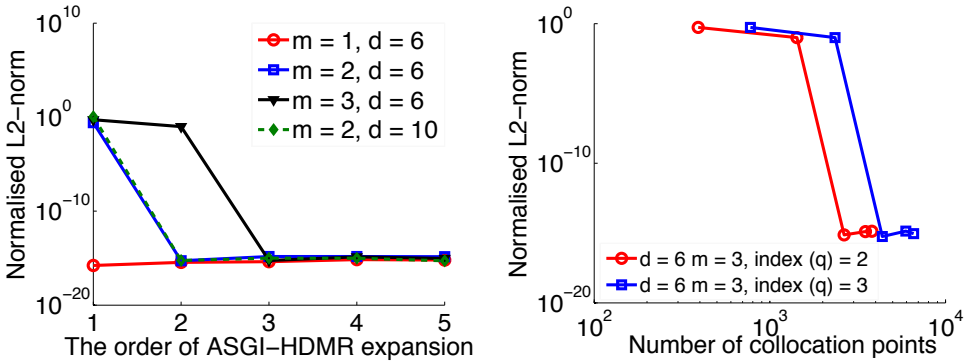
As can be seen from Fig. 6.10(b), increasing the level of interpolation (from index 2 to 3) does not improve the error convergence. This may imply that using the ASGI with the interpolation level $q = 2$ in this example is sufficient to approximate component functions in the ASGI-HDMR expansion. In this example, we also see that using the index number $q = 3$ does not contribute to the approximation of the component functions, and consequently does not improve the accuracy of the ASGI-HDMR. Additionally, the normalised error converges at the third-order ASGI-HDMR for the multiplicative function with $m = 3$. Therefore, in comparison to the level of interpolation, we can infer that the accuracy of the approximation in this example is mainly dominated by the ASGI-HDMR order, rather than the interpolation level. For that reason, using a high-order ASGI-HDMR is necessary to represent highly multiplicative functions — we shall discuss this in more detail later.

However, in most practical applications, the explicit functions or the level of multiplicative structure (m values) are not known a priori. The lower-order ASGI-HDMR can thus be used to approximate high-order multiplicative functions by dividing the domain into a set of smaller sub-domains [121]. This method, referred to as Subdomain HDMR, will be discussed in the next section. The general idea of the Subdomain HDMR leads to the development of our proposed method described in Section 6.7.

6.8.4 Subdomain high-dimensional model representation

Motivated by finite element methods, the idea behind subdomain HDMR was proposed by Özay and Demiralp [121] where the main geometries are divided into smaller sub-geometries. This enables a much lower-order HDMR to efficiently approximate multivariate functions. In this section, we will present numerical studies of the subdomain HDMR using both first- and second-order ASGI-HDMR.

We first describe the subdomain region by denoting η_i as a positive integer for the number of partitioned cells along dimension i; K_i denotes a set of indices for each cell along i-dimension where $K_i = \{1, \ldots, \eta_i\}$ and $k_i \in K_i$. Given a_i and b_i present the range of the domain in i-dimension, we can write the subdomain $\lambda_{k_1,\ldots,k_n}$ with the range l_{i,k_i}



(a) (b)

Figure 6.10: The convergence rates of HDMR for multiplicative functions: a) HDMR at different m-levels of multiplicative functions are shown. b) The error plotted versus number of collocation points at different interpolation indexes used for building a surrogate of the component functions.

along i-dimension in the following form:

$$
\begin{aligned}
& l_{i,k_i} \equiv [\psi^-_{i,k_i}, \psi^+_{i,k_i}], \quad l_{i,1} = l_{i,2} = \cdots = l_{i,\eta_i} \\
& \psi^-_{i,k_i} = a_i + \frac{b_i - a_i}{\eta_i}(k_i - 1), \quad \psi^+_{i,k_i} = \psi^-_{i,k_i} + \frac{b_i - a_i}{\eta_i} \\
& \lambda_{k_1,\ldots,k_n} = \prod_{i=1}^{n} l_{i,k_i} = l_{1,k_1} \times \ldots \times l_{n,k_n}
\end{aligned} \tag{6.19}
$$

The geometrical illustration of Eq. 6.19 is shown in Fig. 6.11. The shaded area $\lambda_{k_1=2,k_2=3}$ demonstrates the subdomain region; the union of all subdomains is equal to the entire domain Ω.

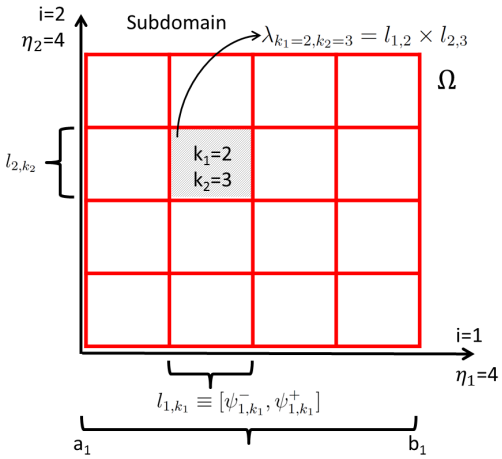


Figure 6.11: Geometrical illustration of the subdomain region from Eq. 6.19

We consider a domain $\Omega = [0, 5]^d$ for three multiplicative functions, two of which are obtained from Eq. 6.18 using m at 3 and 10. The third is the following trivariate function

modified from Hu et al [69]:

$$f(\mathbf{x}) = \sum_{i=1}^{d} x_i^2 - \sum_{i=1}^{d-2} x_i^3 x_{i+1}^3 x_{i+2}^3 \tag{6.20}$$

where d is the dimension of the problem which is 3 in this example.

In this section, we first present the subdomain HDMR with $\eta_i = 2$ for all dimensions in the first example, followed by a comparison of both $\eta_i = 2$ and $\eta_i = 4$ for all dimension. Hence, for three-dimensional functions, the total number of subdomains is $2^3 = 8$ for $\eta_i = 2$, and $4^3 = 64$ for $\eta_i = 4$.

6.8.4.1 Numerical studies of subdomain HDMR using 8 subregions

To evaluate the quality of the surrogate built by the subdomain HDMR, we used two measure criteria: 1) the value of the function, and 2) the Euclidean distance using the normalised l_2-norm (Eq. 6.14).

For the first measure, we drew a sample point from a uniform distribution in each subdomain region. The sampling point was used to evaluate the functions to justify the quality of two surrogate building approaches: the subdomain HDMR and the HDMR. Figure 6.12 demonstrates an example of the sampling point locations for the HDMR (left) and the subdomain HDMR (right). However, these points are not the sample points used in this test (each point was generated randomly within its subdomain).

Each subdomain region is indexed from 1 to N cycled from left to right and top to bottom (the number shown in Fig. 6.12); N denotes the total number of the subdomain regions ($N = \prod_{i=1}^{d} \eta_i$). The indices were used to project the function values from the subdomain region into one dimension for a comparison purpose. Figure 6.13 presents the function values of three functions plotted versus the subdomain indices. The indices shown in the figure are sorted by the explicit function value in a descending order. The subdomain HDMR and the HDMR are both displayed in the figure by using the same sorted indices obtained from the explicit function. For the same sample points, we expect the better surrogate to have function values approximately close to the values from the explicit function. In other words, we prefer that the curve in Fig. 6.13 overlies the curve from the explicit function (the red curve) exactly.

In Fig. 6.13, the function values of the subdomain HDMR (the square-marker curves) are closer to the values from the explicit function curve than the values from the HDMR (the triangular-marker curves) for the three test functions. We only used the first-order for both HDMR and subdomain HDMR in Fig. 6.13. In other words, both HDMR and subdomain HDMR are the univariate functions that approximate the multivariate functions.

In Fig. 6.14, both first- and second-order of the HDMR and the subdomain HDMR are displayed. It can be seen that the subdomain HDMR curve is much closer to the explicit function curve than the HDMR curve. Additionally, by comparing the result to the first order expansion in Fig. 6.13, we see that the subdomain HDMR and the HDMR using second-order expansion provide much better approximations. This is because using a higher-order HDMR is preferred to the lower-order one for approximating multiplicative functions, as

shown previously in Example 2. However, using only high-order expansions may not be necessary when a domain is discretised to smaller subdomains. As can be seen from both Figs. 6.13 and 6.14, the low-order HDMR expansion can be efficiently used to approximate highly multiplicative functions (i.e. $m = 3, 10$).

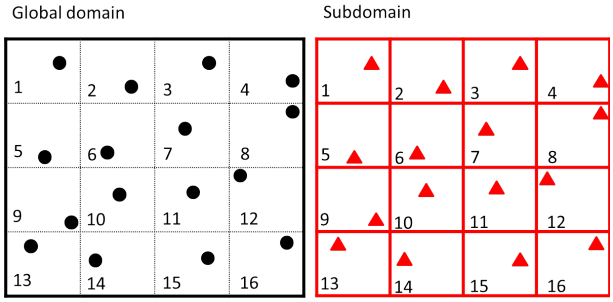
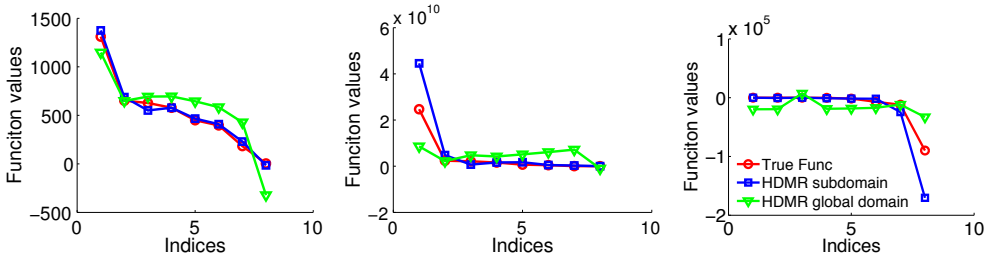


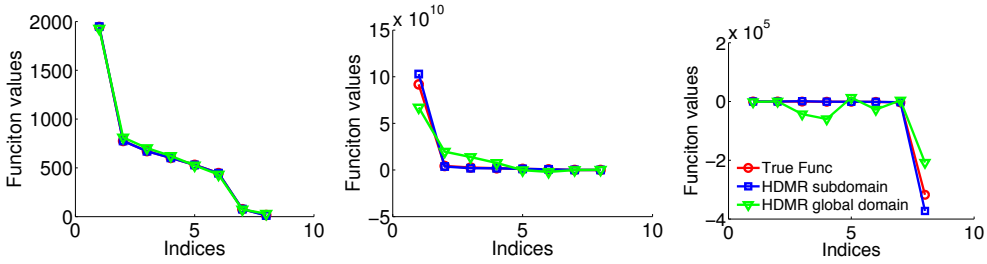
Figure 6.12: An example of the sampling point locations for the HDMR (left) and the subdomain HDMR (right) are shown. It should be noted that the locations of the points shown here are not the ones we used in this test as each point was generated randomly within its subdomain.



(a) Eq. 6.18 with m=3 (b) Eq. 6.18 with m=10 (c) Eq. 6.20

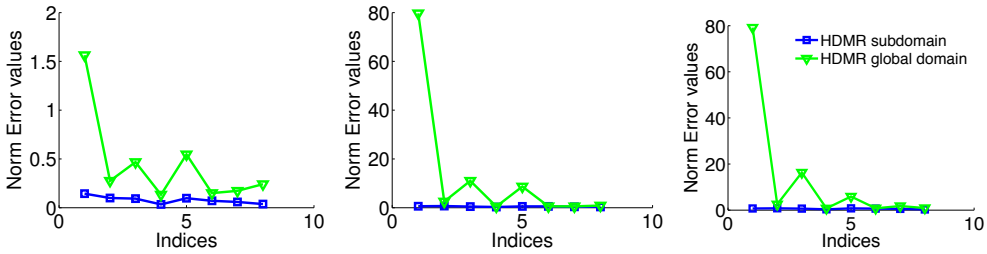
Figure 6.13: Function values of three different functions are plotted versus the subdomain indices; the curves with circle, square, and triangular markers represent the explicit function, the subdomain HDMR and the HDMR, respectively. Only the first-order expansion is shown in this figure.

For the second measure, the normalised l_2-norm described in Eq. 6.14 was used with 50 sampling points in each subdomain region to evaluate the quality of the surrogate. Figure 6.15 represents the normalised l_2-norm plotted versus the subdomain indices for the subdomain HDMR and the HDMR approaches; only the first-order expansion is used in this plot. As can be seen, the normalised norm from the subdomain HDMR is closer to zero than the HDMR. The results in Fig. 6.15 corresponds to the results shown in Figs. 6.13 and 6.14. Hence, it is evident that the subdomain HDMR provides better approximation for high multiplicative functions than the HDMR without the subdomain approach. A similar observation was also reported in [121].



(a) Eq. 6.18 with m=3 (b) Eq. 6.18 with m=10 (c) Eq. 6.20

Figure 6.14: Function values from three functions are plotted versus the subdomain indices; the curves with circle, square, and triangular markers represent the explicit function, the subdomain HDMR and the HDMR, respectively. The second-order expansion is shown in this figure.



(a) Eq. 6.18 with m=3 (b) Eq. 6.18 with m=3 (c) Eq. 6.20

Figure 6.15: The normalised error of the first-order HDMR using the subdomain and the global domain are shown at different multiplicative functions in (a), (b), and (c).

6.8.4.2 Numerical studies of subdomain HDMR using 8 and 64 subregions

Instead of using $N = 8$ in the previous example, we reduce the size of the subdomain and increase the total number of subdomains to $N = 64$ in order to compare the effect of the subdomain sizes. Only the normalised l_2-norm is used in this example.

The normalised norm plotted versus the indices of subdomain elements are shown in Figs. 6.16 and 6.17 for both first- and second-order expansions, respectively; the blue curves present the subdomain HDMR for $N = 8$ whereas the red presents the subdomain HDMR for $N = 64$. We discuss the following results from both figures below:

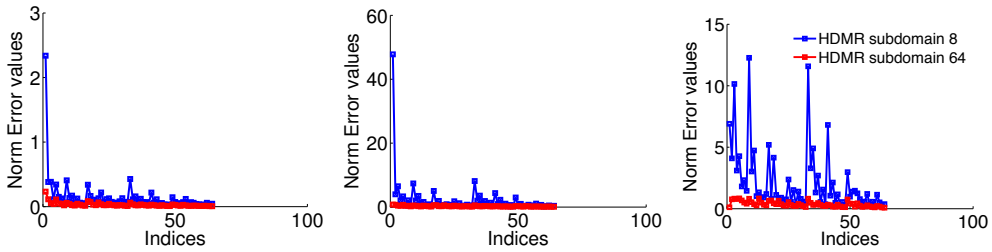
1. Both Figs. 6.16 and 6.17 show that the subdomain HDMR for $N = 64$ has much lower normalised norm values than the ones with $N = 8$ for all functions. In other words, the normalised errors from the red curves are smaller than the one from the blue curves for all functions. The results indicate that the number of subdomain elements is substantial to the approximation quality as can be seen by several magnitude decreases in the normalised norm when the total number of the elements increases from $N = 8$ to $N = 64$. However, using a greater number of the subdomain elements is more computationally demanding.

2. Comparing Figs. 6.16 and 6.17 shows that the subdomain HDMR for $N = 8$ using both first- and second-order expansion is better than the one using only first-

order expansion for all functions. In other words, we compare the blue curves in Figs. 6.16 and 6.17. The result indicates that the subdomain HDMR using high-order expansion provides better approximation than the one using only first-order for the same number of domain size.

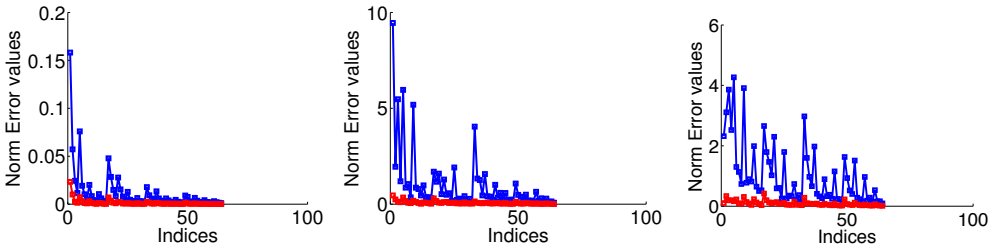
3. The subdomain HDMR using first-order expansion and with the number of elements $N = 64$ is approximately equivalent to the one using both first- and second-order with the number of elements $N = 8$ for the function in Fig. 6.16(a). In other words, the red curve in Fig. 6.16(a) is compared to the blue one in Fig. 6.17(a). The choices of using the expansion order and the number of subdomain elements can be investigated further and could be included for future study. However, in this example, the subdomain HDMR can be used to approximate the multiplicative functions with a much lower-order expansion.

For the reasons above, we adapt the idea of subdomain HDMR to a moving subdomain HDMR to find a global optimum, which will be discussed in the next section with reservoir examples.



(a) Eq. 6.18 with m=3 (b) Eq. 6.18 with m=10 (c) Eq. 6.20

Figure 6.16: The normalised errors of the subdomain HDMR using only a first order expansion are shown for different multiplicative functions in (a), (b), and (c) for the total number of elements $N = 8$ (blue curves) and $N = 64$ (red curves).



(a) Eq. 6.18 with m=3 (b) Eq. 6.18 with m=10 (c) Eq. 6.20

Figure 6.17: The normalised errors of the subdomain HDMR using both first- and second-order expansion are shown for different multiplicative functions in (a), (b), and (c) for the total number of elements $N = 8$ (blue curves) and $N = 64$ (red curves).

6.8.5 The PUNQ-S3 reservoir model

As documented by Barthelmann et al [9], Klimke [80], Petvipusit et al [125] and the results from Fig. 6.6, the Adaptive Sparse Grid Interpolation (ASGI) generally provides a better convergence rate in comparison to the non-dimensional adaptivity sparse grid approach. The comparison of the ASGI and the RSGI techniques for optimisation in a reservoir model was reported in Chapter 4 where the injection rates of CO_2 sequestration were optimised. We showed that the ASGI is an efficient technique for the PUNQ-S3 reservoir model. In this chapter, we use the same ASGI technique to compare with the WCut-HDMR approach. We first describe the objective function and well control configurations, followed by optimisation results of the ASGI and the WCut-HDMR using PUNQ-S3 model.

6.8.5.1 Objective function for CO_2 sequestration

We used the modified break-even tax credit (BT) function given in Eq. 5.26 for the CO_2 optimisation. The optimal solution is achieved through the minimisation of the BT function by allocating the CO_2 injection rates at each step for each well. The objective function is formulated under an optimisation framework in the following form:

$$\begin{aligned} &\text{Minimise:} && BT(\mathbf{x},\mathbf{y}) \\ &\text{Subject to:} && z_i(\mathbf{x},\mathbf{y}) = 0, \quad i = 1,2,\ldots,N_z \\ &&& g_i(\mathbf{x}) = 0, \quad\;\; i = 1,2,\ldots,N_g \\ &&& h_i(\mathbf{x}) \leq 0, \quad\;\; i = 1,2,\ldots,N_h \end{aligned} \tag{6.21}$$

where $z_i(\mathbf{x},\mathbf{y})$ denotes the flow equations of CO_2 in each cell i for the total of N_z cells as a function of injection rates $\mathbf{x}$ and petrophysical variable $\mathbf{y}$ defined permeability, porosity, and anisotropy ratio; $g_i(\mathbf{x})$ and $h_i(\mathbf{x})$ present linear (or non-linear) equality and inequality constraints of variable $\mathbf{x}$, respectively; the total numbers of equality and inequality constraints are denoted by N_g and N_h, respectively.

6.8.5.2 Model and well control descriptions

The PUNQ-S3 model described in Section 4.5.2.1 was chosen for this study to test the WCut-HDMR algorithm. The properties of the aquifer were given in Table 4.3. Two different geological models are used in this section, each of which is assumed to present a known geology. The petrophysical properties including permeability, porosity, and anisotropy ratio of both geological models were generated from the Gaussian covariance function [31] with mean and variance shown in Table. 6.2, of which the spatial distributions were displayed in Fig. 4.18. Since geological uncertainty will not be considered in this chapter, the optimisation result from each model is independent and interpreted as a separate model.

For well control configurations, two injectors were operated by injection rate controls with constraints on the bottom hole flowing pressure (BHP). The location of the wells was previously shown in Fig. 4.15. In this section, we modify the well configurations in Section 4.5.2.3 for ten injection periods, each of which is 3 years. The total injection period

Table 6.2: Geostatistical parameters used to generate Gaussian random fields for the PUNQ-S3 model

Model	k_h^{**} (md)			Porosity			$k_v \backslash k_h$
	Mean (μ)	Variance (σ^2)	Corr*	Mean (μ)	Variance (σ^2)	Corr*	
Model 1	100	0.16	12	0.15	0.0144	15	0.5
Model 2	500	0.16	12	0.25	0.0144	15	0.1

*Correlation length of the covariance function, ** Horizontal permeability

is 30 years, and the post-injection period is 50 years. The injection ratio $(x_1, \ldots, x_{10})$ of each injector is used as the input variables instead of injection rates. By using the ratios of the injection rates, dimension of the problem can be reduced from 20 to 10. This is due to the fact that the amount of injecting CO_2 is usually constrained by the total supplied from industrial sources (e.g. power plants).

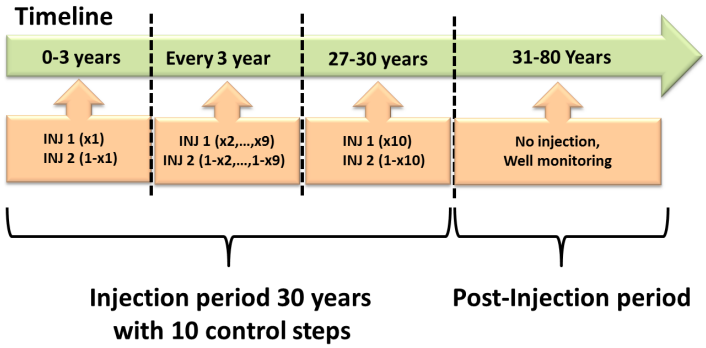


Figure 6.18: Well control configurations for two injectors; the injection ratios of the injector 1 at the first and the d^{th} injection periods are denoted by x_1 and x_d, respectively. The total injection period is 30 years with 10 control steps for each well.

The injection ratios $(x_1, \ldots, x_{10})$ of injector 1 at different injection periods are shown in Fig. 6.18. For example, the injection ratio (x_1) is obtained from the injection rate of a particular well divided by the total supplied rate. Thus, at the same injection period, the sum of the injection ratio of each well is unity at all time $(x_1 + (1 - x_1) = 1)$; x_2 to x_n are obtained in a similar way to x_1. The detailed well configuration parameters are listed in Table 6.3.

Table 6.3: Well configuration parameters for the PUNQ-S3 model

Description	Parameter values
Total supplied rate (Sm^3/day)	5.0×10^6
Minimum injection rate (Sm^3/day)	0
Maximum injection rate per well (Sm^3/day)	2.5×10^6
Maximum BHP (MPa)	60
Injection period	30 years
Post-injection period	50 years
Time period per control step	3 years

6.8.5.3 Optimisation results

In this section, we first present optimisation results for CO_2 sequestration using different numbers of control steps. Then we compare the WCut-HDMR approach with the ASGI technique where the break-even tax credit (BT) and the mass of trapped CO_2 are shown. Last, we discuss the potential benefits of using the WCut-HDMR for optimisation in comparison to other control modes such as the bottom hole flowing pressure (BHP) and constant injection rate controls using the PUNQ-S3 model. It should, however, be noted that the surrogate in this work is only used to guide the optimisation process where the optimal solution is obtained directly from the numerical flow simulation. This ensures that the comparison is not affected by the approximation errors arising from the surrogate. Particle swarm optimisation (PSO) [78] was used in this study to obtain the optimal solution but only run on the approximate model represented by the WCut-HDMR. The PSO parameters and the stopping criteria of the WCut-HDMR are provided in Tables 6.4 and 6.5, respectively.

Table 6.4: Initial parameters of the PSO algorithm

No.	Parameters	Numerical values
1.	generations (iterations)	300
2.	population size	50
3.	function tolerance value of BT	1e-6
4.	variable tolerance	1e-6
5.	social attraction	1.25
6.	cognitive attraction	0.5

Table 6.5: Stopping criteria for the WCut-HDMR approach

Criteria	Numerical values
Difference in the optimal value ($y^{i+1} - y^i$)	0.01
Euclidean distance of the input variables (l_2-norm)	0.01
The percentage ratio of the hyper-volume domain to the initial domain	1%
Maximum iterations	8
Maximum number of collocation points	3000

Degree of freedom of injection strategies. In this example, we first compare different degrees of freedom using 10 and 20 control steps where the ASGI technique is used to build a surrogate to assist optimisation process. Later, we will demonstrate that using 20 control steps could impact the optimal BT and the mass of trapped CO_2 in comparison to 10 control steps. It should be noted that the 10 and 20 controls are represented by 5 and 10 independent variables, respectively, in the optimisation algorithm due to the total injection constraints.

Optimal break-even tax credit. Figure 6.19(a) presents optimal BT values plotted versus the number of collocation points of the ASGI surrogate. The dashed and solid curves were obtained from the optimal strategy using 10 and 20 control steps, respectively. The number of collocation points was obtained by varying the interpolation level of the ASGI surrogate where the PSO was used to find the optimal BT. We see that the BT value from

the 10-control is lower than the one from the 20-control for the range of 1,000 collocation points. However, for the range of 8,000 points, the BT value from the 20-control is lower than the one from the 10-control.

The corresponding control steps of the optimal BT from the last iteration are displayed in Fig. 6.19(b) where only the injection ratios of the injector 1 are shown. The injection ratio of the injector 2 (not shown here) is simply obtained by subtracting the injection ratio of the injector 1 from 1 (see Fig. **??**). In Fig. 6.19(b), the injection ratios of the 10-control and the 20-control strategy are different and from the result in Fig. 6.19(a) we know that the 20-control strategy provides a better BT value than the 10-control strategy. This is because the global optimal solution obtained from problems with a greater number of independent input variables could not, in general, be worse than the solution obtained using the smaller one since using a higher degree of freedom enhances controllability and allows more control options in the system. However, having many input variables comes with the cost of function evaluations and also increases the complexity of the problem so that the global optimal solution sometimes is difficult to find. Though it is possible to combine different control scales for the injection rate optimisation as suggested in [94, 118], it is not within the scope of this study. In contrast, in this study, we demonstrate that the optimal results can be improved by considering more control variables for optimisation. In the following, we will discuss the impact of these controls on the various phases of CO_2 in the aquifer.

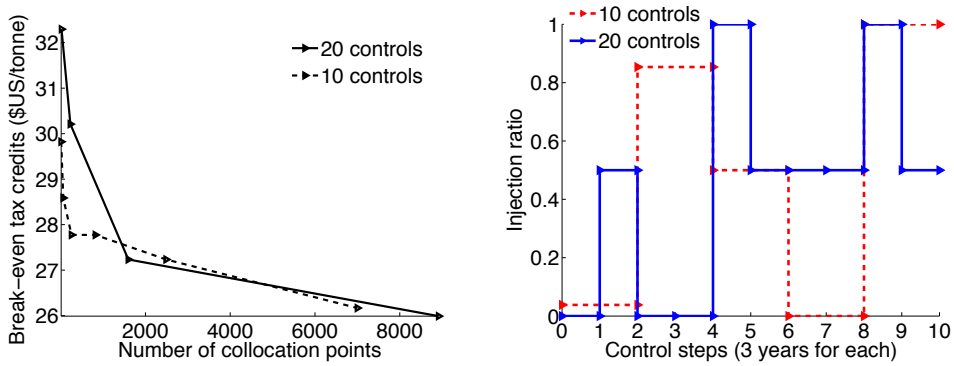
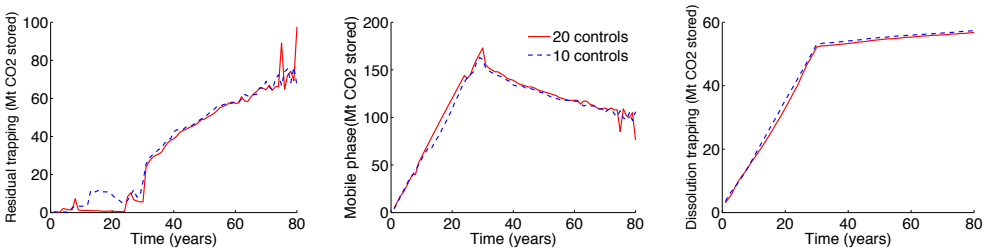


Figure 6.19: a) The break-even tax credit (BT) values plotted versus the number of collocation points from the ASGI approach using 10 and 20 control steps are shown with the dashed and the solid curves, respectively. b) The optimal injection strategies of the 10 and 20 controls are displayed, the break-even tax credits of which are presented in the last iteration of figure a).

The mass of CO_2 in various phases. Figure 6.20 presents the mass of CO_2 in various phases in the aquifer, i.e. residual (capillary) trapping, dissolution trapping, and mobile phase of CO_2, using 10 and 20 injection steps. The optimal strategy of the 20-control was selected from the minimum BT value shown in Fig. 6.19(a) at 8,000 points, whereas the 10-control was obtained from the average between two consecutive steps of the 20-control strategy. This average is to emulate the strategy of 20 control steps with fewer degrees

of freedom. As can be seen, using injection strategies with more degrees of freedom (the solid curves) provides much greater capillary trapping in Fig. 6.20(a) and less mobile phase of CO_2 (see Fig. 6.20(b)), which results in a lower BT value even though the 10-control strategy provides slightly more dissolution trapping, see Fig. 6.20(c). Based on the results in Figs. 6.19 and 6.20, it is apparent that using 20 control steps for optimisation in this example is a better injection strategy in terms of BT and the mass of trapped CO_2. Also, it can be inferred that the optimisation of high-dimensional problems could provide better results. To this end, in the next examples, we will focus on the surrogate-assisted optimisation technique of 10 input variables (20 control steps) using the WCut-HDMR approach.



(a) Residule trapping (b) Mobile phase of CO_2 (c) Dissolution trapping

Figure 6.20: The mass of CO_2 in various phases, (a) capillary trapping, (b) dissolution trapping, and (c) mobile phase of CO_2, obtained from the optimal injection strategy using 10 and 20 control steps shown with the dashed and solid curves, respectively.

WCut-HDMR for CO_2 optimisation. The optimisation results from the WCut-HDMR and the ASGI methods are compared in this section using two geological models from PUNQ-S3. Since we have not considered model uncertainty in this chapter, each model realisation is independent and the optimisation results are interpreted separately.

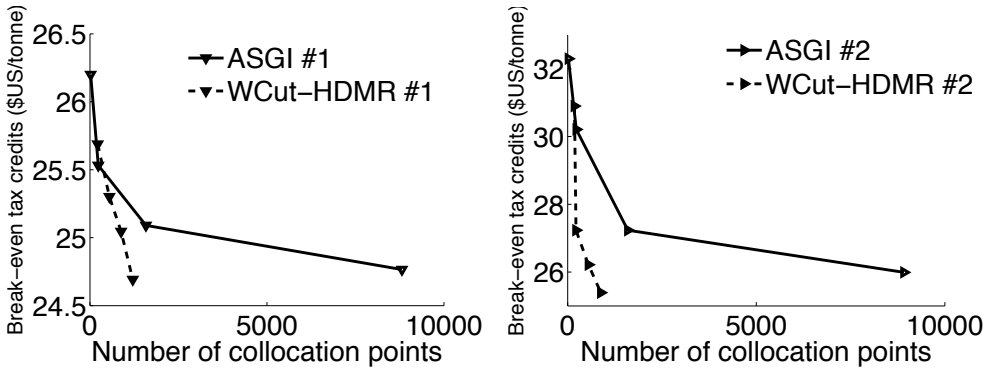
From Figs. 6.21(a) and 6.21(b), the optimal BT values obtained from the WCut-HDMR technique (the dashed curve) are less than the one from the ASGI approach for two models. This indicates that the proposed WCut-HDMR method is an efficient surrogate-assisted optimisation technique since it provides the lowest BT value at the minimum number of collocation points, in comparison to the ASGI technique. Only the first-order WCut-HDMR is displayed in this figure from which the alpha value (Eq. 6.11) is 0.4 in Fig. 6.21(a) and 0.6 in Fig. 6.21(b). The other numerical settings for both models 1 and 2 are shown in Table. 6.6.

The corresponding injection ratios of both WCut-HDMR and ASGI are shown for two model realisations in Figs. 6.22(a) and 6.22(b). We recall that the injection ratios shown here belong to injector 1. Hence, a ratio equal to zero in the figure means that CO_2 is injected only from injector 2, and vice versa for an injection ratio equal to 1. In Fig. 6.22(a), it can be seen that the control steps determined using the WCut-HDMR are equal to 0.5 for about the first 15 years, which means that it balances the injection rates between the two injectors equally in this period. For the later period (after 15 years), the WCut-HDMR method allocates more CO_2 to injector 2 as does the ASGI approach. However, the WCut-HDMR provides different CO_2 injection rates for both wells in this period to create a pulse-

Table 6.6: Numerical settings for the WCut-HDMR approach

No.	Parameters	Numerical values
1.	Maximum number of points for the low-order surrogate (during exploration phase)	200
2.	Absolute hierarchical surpluses (Eq. 4.37)	1e-6
3.	Relative hierarchical surpluses (Eq. 4.38)	1e-6
4.	Maximum number of points for building the component functions	1000
5.	Maximum interpolation level for the component functions	4
6.	Maximum order of HDMR	2

shaped injection, which then enhances trapping efficiency. This pulse-shaped injection during the late injection period obtained from both WCut-HDMR and ASGI are also observed in Fig. 6.22(b) where the WCut-HDMR tends to allocate more CO_2 at injector 2 compared to the ASGI method. This type of injection process was also reported in [73] as the pulse-shape allows the interruption of continuous flow to enhance the trapping process.

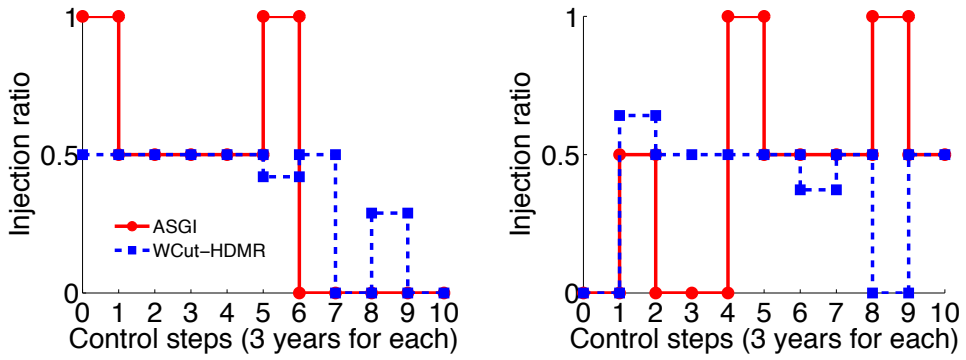


(a) Model 1 (b) Model 2

Figure 6.21: The relationship of optimal break-even tax credits and the number of collocation points are shown. The first-order WCut-HDMR is compared with the ASGI approach for two geological realisations (a) and (b) in this plot, where the alpha value of the WCut-HDMR method is set at 0.4 and 0.6 in (a) and (b), respectively.

Figure 6.23 presents the optimal BT obtained from both first- and second-order of the WCut-HDMR approach for two models. For model 1 in Fig. 6.23(a), the BT value at the last iteration of the second-order WCut-HDMR is less than the one obtained from the first order, but using more collocation points. However, using the second order method in model 2 shown in Fig. 6.23(b) does not improve the BT of the first-order case.

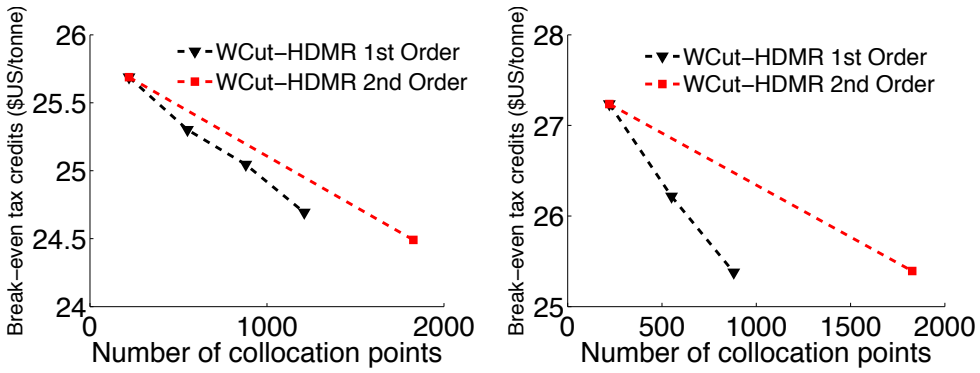
From the results in Fig. 6.23, it can be seen that there are two reasons why the second-order WCut-HDMR does not perform significantly better than the first order method. Firstly, the BT function may not be a purely multiplicative function and hence the second-order WCut-HDMR is not necessary to represent the pairwise interactions between the input variables, as discussed in Section 6.8.3. Secondly, even if the BT function has a multiplicative structure, only some bivariate component functions $f_{ij}(x_i, x_j)$ will contribute to the model output. Therefore, having all bivariate terms in the HDMR structure requires more colloca-



(a) Model 1 (b) Model 2

Figure 6.22: The optimal injection strategies obtained from the first-order WCut-HDMR and the ASGI approach for two geological realisations (a) and (b) are shown. Only the injection ratio for injector 1 is shown.

tion points than what is necessary. This issue will be addressed in the next chapter through adaptive estimation of the important bivariate terms.



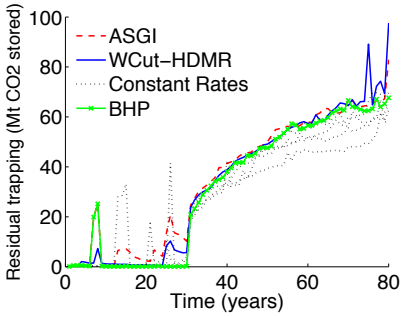
(a) Model 1 (b) Model 2

Figure 6.23: The BT values plotted versus the number of collocation points using both the first and second order WCut-HDMR approach are shown for two realisations (a) and (b). All other parameters except the order of WCut-HDMR are set the same as in Fig. 6.21.

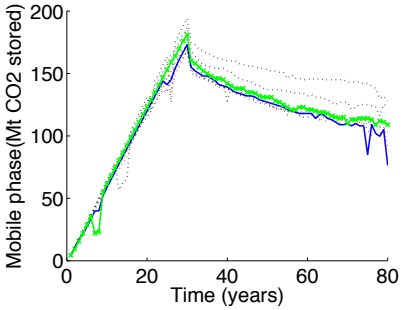
It is also important to compare the other control modes of CO_2 injection such as the BHP and the constant rate controls, with the WCut-HDMR approach. For this reason, we selected the optimal injection strategies obtained from the ASGI and the WCut-HDMR techniques using an equivalent number of collocation points, 229 and 221 respectively from model 2 in Fig. 6.22(b). This selection of the collocation points ensures that we compare different techniques at equivalent computational cost.

The result shown in Fig. 6.24 displays various phases of CO_2 for different injection modes constrained by the total mass of CO_2 injection. In comparison to the ASGI approach and the other types of control modes, WCut-HDMR provides the maximum mass of CO_2 capillary trapped, see Fig. 6.24(a) and minimum mobile phases of CO_2 as shown in Fig. 6.24(b). Moreover, the results of the WCut-HDMR and the ASGI techniques are

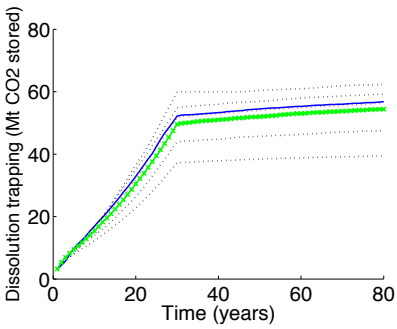
significantly better than the other control modes, which also confirms the value of optimisation over non-optimal strategies using the BHP and the fixed rate controls. In Fig. 6.24(c), we see that using the constant rate control could be sufficient to increase the mass of CO_2 dissolved in brine, but it is not a good strategy to enhance the capillary trapping capacity, see the BT value in Table 6.7.



(a) Residual trapping



(b) Mobile phase of CO_2



(c) Dissolution trapping

Figure 6.24: The mass of CO_2 in various phases, shown in (a) Residual (capillary) trapping, (b) dissolution trapping, and (c) mobile phase of CO_2, are obtained from different control modes. The BHP and the constant rate controls are represented by the square-marker and the dotted curves, respectively, while the solid and the dashed curves present the ASGI and the WCut-HDMR techniques, respectively.

The numerical values of different control modes are provided in Table 6.7 where the constant rate controls are represented by the ratio of injector 1, as described by x_1 to x_{10} in Fig. 6.18. For example, Rate 0.0 means that well 1 is shut in and only well 2 is operated at the maximum rate equal to the total injection, and vice versa for Rate 1.0. According to Table 6.7, we see that the WCut-HDMR approach improves the BT value by 20.92% and trapping capacities by 26.45% in comparison to BHP control. Moreover, the optimal BT value and trapping capacities are improved approximately by 10% from the results of the ASGI approach.

6.9 Summary

We have developed an efficient optimisation technique by using a hierarchical adaptive sparse grid using high-dimensional model representation (ASGI-HDMR) with a walking cut-point to construct a surrogate-assisted optimisation technique for CO_2 injection. The developed techniques have been evaluated computationally and the results of this computational experiments are summarized as follows:

1. Adaptive Sparse Grid Interpolation (ASGI) can suffer from the curse of dimensionality when it is used to approximate high-dimensional functions. In contrast, High-Dimensional Model Representation (HDMR) is suitable for multivariate functions due to the fact that the HDMR structure uses a combination of low-order correlation effects from different input variables to effectively represent the model output. Hence, using the HDMR for high-dimensional functions requires fewer function evaluations than using the ASGI approach.

2. We used the ASGI technique with the HDMR expansion to formulate a hierarchical adaptive sparse grid with high-dimensional model representation (ASGI-HDMR) for efficient surrogate construction of high-dimensional functions. This technique uses low-dimensional component functions to represent a higher-dimensional function. We observed that better convergence rates are obtained from the ASGI-HDMR approach in comparison to the ASGI approach for several analytical functions with an additive structure.

3. For analytical functions with a multiplicative structure, the order of HDMR expansion dominates the interpolation accuracy. Hence, for a highly multiplicative function structure, a better approximation could be achieved through the use of a high-order expansion, rather than increasing the interpolation level. However, we later demonstrated that the subdomain HDMR, which exploits the low-order expansion of the ASGI-HDMR with a smaller input domain, can be used efficiently to approximate highly multiplicative functions.

4. We proposed the use of subdomain HDMR with a moving cut-point (WCut-HDMR) for a surrogate-assisted optimisation technique. The surrogate using the WCut-HDMR approach is a surface built locally around the local optimal point obtained from the previous iteration. This locally optimal point is then used as the new cut-point in the Cut-HDMR for the next iteration. This allows the cut-point to move closer to the optimal solution while keeping the minimum number of collocation points (simulation runs). Additionally, the WCut-HDMR technique can be used to balance the exploration and exploitation phases of the optimisation process, making it an efficient approach for most practical problems with many input variables.

5. Numerical evaluations using the PUNQ-S3 model revealed the potential benefits of using the WCut-HDMR approach for CO_2 sequestration in comparison to the ASGI technique and other control modes, such as BHP and constant rate controls, for a large number of input variables.

Table 6.7: Comparison of different methods and control modes of CO_2 injection

Techniques	BT***		The mass of CO_2			
	$US /tonne	Diff** %	Res+ Dis (Mt)	Diff** %	Mobile phase (Mt)	Diff** %
WCut-HDMR	**27.24**	**-20.92**	**154.4**	**26.45**	**76.3**	**-30**
ASGI	30.21	-12.28	139.2	14	91.4	-16.15
BHP*	*34.44*	*0*	*122.1*	*0.0*	*109.0*	*0.0*
Rate 0.0	33.01	-4.16	127.4	4.34	103.0	-5.5
Rate 0.25	32.65	-5.20	128.8	5.49	101.0	-7.34
Rate 0.5	34.44	0	122.1	0.0	109.0	0.0
Rate 0.75	39.56	14.86	106.3	-12.94	124.0	13.76
Rate 1	41.19	19.59	102.1	-16.38	128.0	17.43

*The BHP control is the reference case.

**The percentage difference in comaparison to the BHP control mode

***Break-even tax credits

CHAPTER 7

ADAPTIVE HIGH-DIMENSIONAL MODEL REPRESENTATION

7.1 Introduction

As discussed in the previous chapter, the ASGI-HDMR is a promising surrogate technique that can be used to approximate high-dimensional functions from the sum of hierarchical lower-dimensional functions — the component functions. However, this chapter focuses on the component functions that have significant impacts on the model output so that we can construct the ASGI-HDMR surrogate with fewer component functions and collocation points. This idea leads to decompositional adaptivity of the ASGI-HDMR.

In this Chapter, we first introduce a principle behind decompositional adaptivity, followed by the use of sensitivity indices to identify the important component functions. Later, the algorithm used to approximate the sensitivity indices is presented. Last, we describe the use of sensitivity indices for the ASGI-HDMR method.

7.2 Brief background: adaptive schemes for high-dimensional model representation

Generally, the idea behind HDMR stems from the principle that multivariate continuous functions can be represented by relatively lower-order variate functions. This decomposition structure enables us to identify the important component functions to the model output so that only the influential terms are used, and the less significant terms are discarded, making it an efficient technique for surrogate construction. The selection process of the important terms could be achieved through prior knowledge for some physical systems. However, it is often not known in many practical problems. Thus, finding the influential input terms of the model output becomes a key element to improve the efficiency of the HDMR approach. We introduce here an adaptive HDMR (AHDMR), which automatically identifies the influential input terms and thus requires less computational cost (collocation points) than the non-adaptive HDMR approach.

In recent years, adaptivity schemes for HDMR have been extensively studied for the efficient representation of multivariate functions. For example, Ma and Zabaras [96] considered two approaches for HDMR adaptivity: effective dimensions (the significant component functions), and superposition dimensions (the truncation order of the HDMR). They determined the effective dimensions based on the relative weight of univariate component functions where only the weights greater than the threshold value were selected. Later, Hu

et al [69] applied a different approach to select the component functions based on a decomposition and reselection method. They found this method to be efficient because it uses a reselection procedure which ensures that the selected terms contribute to the model output and hence the less important terms are discarded at the next selection process.

The underlying assumption for both methods [69, 96] was restricted to the admissibility relation which implies that any interaction of important factors is also important. Even though this is a reasonable assumption, it may not be valid for some applications. For example, given $f(x_1, x_2, x_3) = g(x_1, x_3) + g(x_1) + g(x_2)$, using the univariate component functions to find the effective dimension would provide x_1 and x_2 as effective dimensions. Hence, the adaptive HDMR would give $f(x_1, x_2, x_3) \approx g(x_1) + g(x_2) + g(x_1, x_2)$; the important term $g(x_1, x_3)$ is not captured from this scheme whereas the term $g(x_1, x_2)$ in the adaptive HDMR is obviously not an important term. For this reason, an adaptive scheme was proposed by Labovsky and Gunzburger [88] to independently identify the interaction terms using a finite difference approach.

Selecting the influential component functions for adaptive HDMR can be broadly viewed as a sensitivity analysis approach which quantifies relative contributions to the output response. As mentioned previously, Labovsky and Gunzburger [88] used the gradient based approach to find important component functions. This method, similar to other local sensitivity analyses, is very fast to compute; however, it only captures linear relationships of the system and often depends on the chosen perturbation size at the local neighbourhood.

In contrast, we propose the use of global sensitivity analysis [146] to consider the full range of input variables and to capture the nonlinear mapping of the input-output relationship of the system. In this thesis, we will use the term AHDMR to generally represent the component-function adaptivity of the ASGI-HDMR surrogate.

7.3 Sobol indices for sensitivity analysis

An ANOVA expansion is generally a variance-based method which can be used for global sensitivity analysis [145]. The ANOVA expansion in this study is used to quantify the effective dimensions, which consequently leads to the selection of influential component functions of the Cut-HDMR expansion. Let $\mathcal{V}$ be the variance of function $f(\mathbf{x})$, the following variance term is given as:

$$\mathcal{V} = \int_{\Omega^n} [f(\mathbf{x}) - f_0]^2 \mathrm{d}\mu(\mathbf{x}) = \int_{\Omega^n} f^2(\mathbf{x}) \mathrm{d}\mu(\mathbf{x}) - f_0^2 \tag{7.1}$$

The variance $\mathcal{V}$ in Eq. 7.1 is the total variance arising from averaging the square of Eq. 6.2 and then subtracting by its mean square (f_0^2) subject to the orthogonality where the covariance between the component functions is zero. Then the variance decomposition is expressed in the following:

$$\mathcal{V} = \sum_{i=1}^{n} \mathcal{V}_i + \sum_{1 \le i < j \le n} \mathcal{V}_{ij} + \cdots + \mathcal{V}_{1,2\ldots,n} \tag{7.2}$$

where the partial variances of univariate and bivariate component functions with a uniform distribution of the range $[0,1]^n$ are represented by.

$$\mathcal{V}_i = \int_0^1 f_i^2(x_i) d(x_i) \tag{7.3}$$

$$\mathcal{V}_{ij} = \int_0^1 f_{ij}^2(x_i, x_j) d(x_i) d(x_j) \tag{7.4}$$

The first- and second-order sensitivity indices defined as Sobol indices S_i and S_{ij} are obtained by dividing both sides of Eq. 7.2 with the total variance $\mathcal{V}$, yields:

$$1 = \sum_{i=1}^{n} S_i(x_i) + \sum_{1 \le i < j \le n} S_{ij}(x_i, x_j) + \cdots + S_{1,2\ldots,n}(x_1, x_2, \ldots, x_n) \tag{7.5}$$

The Sobol indices are used as a measure to quantify how input uncertainty (the variance of input variables) can contribute to the output uncertainty (the variance of objective function) for the entire range of input domain [157]. The first-order Sobol index (S_i) determines the individual effect contributing to the model output without accounting for the interaction effect, whereas the second-order index (S_{ij}) reveals the interaction effect of two variables. We use the second-order Sobol index to identify the significant pairwise component functions for the AHDMR approach. With the use of Lebesgue measure during the optimisation framework, the component functions in Eq. 6.2 can be derived and expressed as a conditional expectation.

$$f_i(x_i) = \mathrm{E}(\mathbf{y}|x_i) - \mathrm{E}(\mathbf{y}) \tag{7.6}$$

$$f_{ij}(x_i, x_j) = \mathrm{E}(\mathbf{y}|x_i, x_j) - \mathrm{E}(\mathbf{y}|x_i) - \mathrm{E}(\mathbf{y}|x_j) - \mathrm{E}(\mathbf{y}) \tag{7.7}$$

Substituting $f_i(x_i)$ from Eq. 7.6 to Eq. 7.3 subject to orthogonal properties yields:

$$\mathcal{V}_i = \mathrm{Var}[\mathrm{E}(\mathbf{y}|x_i)] \tag{7.8}$$

Substituting $f_{ij}(x_i, x_j)$ from Eq. 7.7 to Eq. 7.4 yields:

$$\mathcal{V}_{ij} = \mathrm{Var}[\mathrm{E}(\mathbf{y}|x_i, x_j)] \tag{7.9}$$

Hence, the first- and the second-order Sobol indices with the conditional expectation are:

$$S_i = \frac{\mathrm{Var}[\mathrm{E}(\mathbf{y}|x_i)]}{\mathrm{Var}(\mathbf{y})} \tag{7.10}$$

$$S_{ij} = \frac{\mathrm{Var}[\mathrm{E}(\mathbf{y}|x_i, x_j)]}{\mathrm{Var}(\mathbf{y})} \tag{7.11}$$

Equations 7.10 and 7.11 provide sensitivity indices, whose values indicate how the input variables impact the model output. Large values indicate a large impact on the output. Even though the Sobol indices are useful for a global sensitivity analysis [145], evaluating them is computationally intensive since it requires high-dimensional quadrature to evaluate the variance term. For example, if a Monte Carlo approach is used to estimate $\mathrm{E}(\mathbf{y}|x_i)$ with N

evaluations, estimating S_i would need N^2 evaluations. In recent decades, much research has focused extensively on the estimator for the sensitivity index (e.g. [74, 144, 163]). For example, Saltelli [144] developed an algorithm based on Sobol quasi-random sequence [18] to estimate the Sobol indices; this method was also extended by Saltelli et al [146] for estimating the total effect. In this work, we make use of the ASGI surrogate together with the algorithm proposed in [144] to find the Sobol indices, then evaluating these indices becomes relatively cheap.

7.4 Algorithm to compute Sobol indices

The algorithm proposed by Saltelli [144] reduces the number of sampling points to $M(2+N)$ with M and N representing the number of samples and the number of input variables, respectively. The detailed method used to estimate first-order sensitivity indices can be found in [144, 145, 157]. In addition, we slightly modify this algorithm to estimate the second-order Sobol indices:

1. Given two $M \times N$ matrices A and B contain M samples of N input variables, which are obtained from quasi-random numbers [18]:

$$\mathbf{A} = \begin{pmatrix} a_1^1 & \dots & a_1^i & \dots & a_1^j & \dots & a_1^N \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_M^1 & \dots & a_M^i & \dots & a_M^j & \dots & a_M^N \end{pmatrix} \quad \mathbf{B} = \begin{pmatrix} b_1^1 & \dots & b_1^i & \dots & b_1^j & \dots & b_1^N \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ b_M^1 & \dots & b_M^i & \dots & b_M^j & \dots & b_M^N \end{pmatrix} \tag{7.12}$$

2. Define a^i and a^j as $M \times 1$ column vectors in $\mathbf{A}$ where $i \neq j$, whose index $i, j \in \{1, \dots, N\}$, then create $M \times N$ matrix $\mathbf{C}_{ij}$ to be identical to the matrix $\mathbf{B}$ except that column i and j are replaced by column a^i and a^j from the matrix $\mathbf{A}$.

$$\mathbf{C}_{ij} = \begin{bmatrix} b^1 & \cdots & a^i & \cdots & a^j & \cdots & b^N \end{bmatrix} \tag{7.13}$$

3. Compute the model response from the ASGI surrogate $\mathcal{Q}(\cdot)$, yields $M \times 1$ vectors as follows:

$$y_A = \mathcal{Q}(\mathbf{A}), \qquad y_B = \mathcal{Q}(\mathbf{B}), \qquad y_{C_{ij}} = \mathcal{Q}(\mathbf{C}_{ij}) \tag{7.14}$$

4. The second-order Sobol indices (Eq. 7.11) is estimated by:

$$S_{ij} = \frac{\text{Var}[\text{E}(\mathbf{y}|x_i, x_j)]}{\text{Var}(\mathbf{y})} = \frac{\frac{1}{M} y_A^T y_{C_{ij}} - f_0^2}{\frac{1}{M} y_A^T y_A - f_0^2} = \frac{\frac{1}{M} \sum_{r=1}^{M} y_A^r y_{C_{ij}}^r - f_0^2}{\frac{1}{M} \sum_{r=1}^{M} y_A^r y_A^r - f_0^2} \tag{7.15}$$

where $f_0^2 = \frac{1}{M} y_A^T y_B = \frac{1}{M} \sum_{r=1}^{M} y_A^r y_B^r$. The principle behind this algorithm is based on the intuition that the product $y_A^T y_{C_{ij}}$ of the influential pairwise term ij would be large (or small), which then provides the S_{ij} to be large for the significant pairwise term. In contrast, if the pairwise term is not influential, the product $y_A^T y_{C_{ij}}$ would be randomly distributed near the f_0 term, and consequently the term S_{ij} will be small. Thus, we will select only the pairwise terms that have relatively higher S_{ij} values to be used in the AHDMR expansion.

7.5 Adaptive decomposition ASGI-HDMR based on the Sobol indices.

In this section, we will define the criteria used to select the important terms based on the Sobol indices calculated from Eq. 7.15. Let $\mathbf{S}$ be a set of second-order Sobol indices with $s^k \in \mathbf{S}$ and $k \in \{1, \ldots, \#\mathbf{S}\}$. The cardinality of the set $\mathbf{S}$, presented by $\#\mathbf{S}$, is equal to $\frac{N!}{(N-2)!2!}$ when N denotes the total number of input variables. Then we define the normalised value of the Sobol indices in the following:

$$s^k_{norm} = \frac{s^k - s^{min}}{s^{max} - s^{min}} \tag{7.16}$$

where $s^k_{norm} \in \mathbf{S}_{norm}$ and $\mathbf{S}_{norm}$ is a set of normalised Sobol indices; the minimum and the maximum value of the element in the set $\mathbf{S}$ are represented by s^{min} and s^{max}, respectively. To select important terms based on the normalised Sobol indices, we define

$$\begin{aligned} &\mathbf{S}_{selected} := \{s_{select} | s_{select} \in \mathbf{S}_{norm},\ s_{select} \geq \varepsilon_s\} \\ &\text{where:} \quad \varepsilon_s \in [0, 1] \end{aligned} \tag{7.17}$$

where the parameter ε_s is the cut-off that decides the minimum value of the normalised Sobol indices in the set $\mathbf{S}_{selected}$. Intuitively, the greater the cut-off ε_s is, the fewer the component functions of the second-order terms. If the cut-off ε_s is equal 0, the HDMR expansion is fully represented by both first and second orders without decompositional adaptivity.

Given a set $\mathcal{T}_1 := \{\{1\}, \{2\}, \ldots, \{N\}\}$, the elements of which are 1-subsets of $\mathcal{D}$, and let $\mathcal{T}_2$ be a set of 2-subsets of $\mathcal{D}$ with only selected pairwise indices associated with each member of the set $\mathbf{S}_{selected}$, then we can define a set of important indices as $\mathbf{u^a} := \mathcal{T}_1 \cup \mathcal{T}_2$. For example, if the important pairwise indices are $\{1, 2\}, \{2, 3\}$ and $\{8, 9\}$, then $\mathbf{u^a} = \{\{1\}, \ldots, \{N\}, \{1, 2\}, \{2, 3\}, \{8, 9\}\}$. Thus, we can rewrite Eq. 6.9 for the adaptive high-dimensional model representation (AHDMR) in the following form:

$$f(\mathbf{x}_\Omega) = \sum_{\mathbf{p} \in \mathbf{u^a}} \mathcal{Q}_{q,n=\#\mathbf{p}}(f^{cut}_{\mathbf{p}}(\mathbf{x}_\mathbf{p})) \tag{7.18}$$

It should be noted that univariate and bivariate component functions are considered in the HDMR expansion in this thesis where only the important bivariate terms are identified for dimensional adaptivity. Hence, Eq. 7.18 is simply a truncated HDMR with first and second order component functions. In the next section, we will present an optimisation technique that makes use of the ASGI-HDMR and the AHDMR expansions.

7.6 An adaptive walking Cut-HDMR technique

To enhance the computational efficiency of the WCut-HDMR algorithm, we reduce the number of collocation points (simulation runs) by using AHDMR instead of ASGI-HDMR in step 6 of Table. 6.1. This leads to a modification in the preparation process from which the Sobol indices are calculated. This proposed method is basically the moving surrogate built from the hierarchical adaptive sparse grid using Cut-HDMR technique by adding Sobol

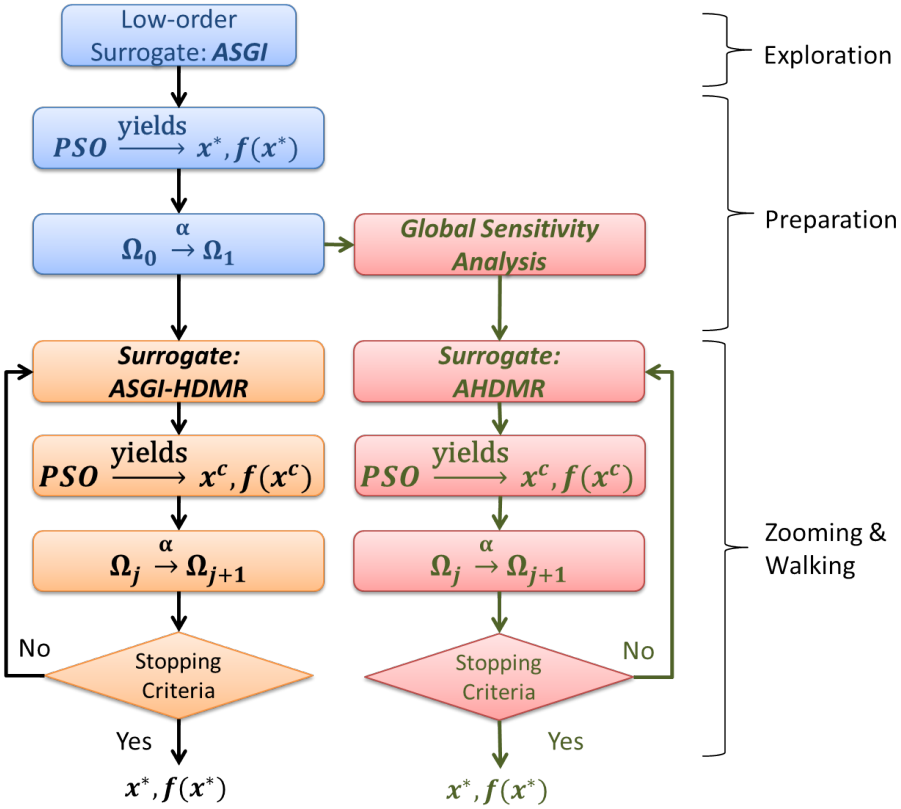


Figure 7.1: A flowchart of WCut-HDMR and WCut-AHDMR algorithms. The four steps in both algorithms are shown: exploration, preparation, zooming, and walking. The workflow of zooming and walking procedures on the left presents the WCut-HDMR (the orange boxes) whereas the one on the right shows the WCut-AHDMR (the red boxes).

indices for the selection of the component functions. We refer this technique to as the WCut-AHDMR.

According to the AHDMR approach, the WCut-AHDMR algorithm is given in the following.

$$\arg\min_{\mathbf{x}\in\Omega_j} \sum_{\mathbf{p}\in\mathbf{u}^{\mathbf{a}}} \mathcal{Q}_{q,n=\#\mathbf{p}}(f_{\mathbf{p}}^{cut}(\mathbf{x}_{\mathbf{p}})) \tag{7.19}$$

where $\mathbf{u}^{\mathbf{a}}$ given in Eq. 7.18 is a set of indices for the AHDMR structure; the other parameters are described in Eq. 6.12.

A workflow of the WCut-HDMR and the WCut-AHDMR with the procedure described in Table 6.1 is shown in Figure 7.1. The ASGI surrogate is built first in the exploration phase, and then the ASGI-HDMR and ASGI-AHDMR are constructed during the zooming and walking phases where the WCut-HDMR (the orange boxes) is presented on the left while the WCut-AHDMR (the red boxes) is shown on the right. Global sensitivity analysis is added in the preparation phase for the WCut-AHDMR algorithm.

7.7 Applications : adaptive high-dimensional model representation

This section presents the applications of the AHDMR approach in comparison to the non-adaptive HDMR technique, where we use simple mathematical functions to demonstrate several characteristics of the AHDMR expansion. Later, we will present efficient optimisation techniques using the WCut-AHDMR method with reservoir PUNQ-S3 models where different optimisation techniques are compared with our proposed WCut-AHDMR approach.

7.7.1 Numerical studies using mathematical functions

In this section, we present convergence rates of the AHDMR and the non-adaptive HDMR methods using three functions with different degrees of input variable interactions: bivariate, trivariate and multiplicative functions. For the AHDMR method, we perform a second-order sensitivity analysis on each function to find effective pairwise variables by using the Sobol estimator from Eq. 7.15. The effective terms are then included in the Cut-HDMR expansion for the AHDMR. The numerical parameters used in both AHDMR and HDMR methods are given below in Table 7.1.

Table 7.1: Numerical settings for the AHDMR and non-AHDMR surrogates

No.	Parameters	Numerical values
1.	Number of sampling point M in Eq. 7.15	4000
2.	Absolute hierarchical surplus ($\varepsilon_{w_q}^{abs}$)	1e-6
3.	Relative hierarchical surplus ($\varepsilon_{w_q}^{rel}$)	1e-6
4.	Maximum order of AHDMR	2
5.	Maximum order of Cut-HDMR	4
6.	Cut point (x_c) used in HDMR	the centroid of the input domain

7.7.1.1 Bivariate function

We consider the following bivariate function with multiple pairwise interaction terms over the range $[0, 2]^d$.

$$f(\mathbf{x}) = \sum_{i=1}^{d} x_i^2 - \sum_{i=1}^{d/2} x_{2i-1}^4 x_{2i}^4 \tag{7.20}$$

where d is the number of input variables which is equal to 10. In this example, we use the normalised error from Eq. 6.14 with 100 sampling points ($N_s = 100$). The normalised norm (Eq. 6.14) is:

$$\varepsilon_{L_2} = \frac{\sqrt{\sum_{j=1}^{N_s} (f_{exact}(\mathbf{x}^j) - f_{approx}(\mathbf{x}^j))^2}}{\sqrt{\sum_{j=1}^{N_s} f_{exact}(\mathbf{x}^j)^2}}$$

From the bivariate function in Eq. 7.20, there are 5 terms that account for the pairwise interactions (i.e. x_1x_2, x_3x_4, x_5x_6, x_7x_8, x_9x_{10}). It is, therefore, expected that the proposed

adaptive scheme using the second-order Sobol indices can detect these component functions and discard the others, hence reducing the number of collocation points in comparison to the HDMR method.

We illustrate this point in Fig. 7.2 by presenting the error decay and the number of collocation points used for building the surrogate; the square- and the circle-marker curves represent the non-adaptive HDMR and the AHDMR, respectively. The error convergence of the known pairwise interaction terms is used as a reference case (the circle-marker curve) in Fig. 7.2(a), whereas the AHDMR using the Sobol indices is shown in Fig. 7.2(b). As can be seen in Fig. 7.2(a), the reference case requires fewer collocation points than the non-adaptive HDMR at the same error measure. This is due to the fact that some of the collocation points in the non-adaptive HDMR expansion are used to construct component functions which do not contribute to the model output, and consequently it requires more sampling points.

Comparing Figs. 7.2(a) and 7.2(b) shows that the error decay of the AHDMR in Fig. 7.2(b) provides a better convergence rate than the non-adaptive HDMR, and is similar to the reference case in Fig. 7.2(a). This result infers that the Sobol indices can be used to detect the important component functions without prior knowledge of dominating interaction terms, which is the case in most applications. The normalised second-order Sobol indices used in Fig. 7.2(b) are presented in Fig. 7.2(c) where the circle- and square-marker curves represent the changes of the normalised S_{ij} in terms of first and second derivatives. As can be seen in Fig. 7.2(c), using the cut-off ε_s between 0.2 and 0.8 for the AHDMR would result in the same number of effective terms (5 terms) as the reference case. From this example, it should be noted that we only need the second-order HDMR to represent the bivariate function, but we present convergence rates up to fourth order for the reference case, and to second order for AHDMR case to show the convergence.

The bivariate function was also used in [69] for AHDMR studies but with different measure criteria in which the authors made an assumption that the selection of higher-order indices relies on the lower-order indices. For instance, if $\mathbf{u}$ denotes a set of coordinates for the lower-order component functions and a set $\mathbf{v}$ defines the coordinates for the higher-order component functions, then $\mathbf{v} \subseteq \mathbf{u}$. This means that the higher-order terms are identified dependent on the lower-order terms. In most practical application, this dependency is not in general the case. In contrast, using the Sobol indices for the AHDMR does not rely on the lower-order terms, making it more practical for most applications.

7.7.1.2 Trivariate function

We slightly modified a function from [69] to demonstrate trivariate interactions terms in order to present: 1) error decay rates of HDMR for different levels of interpolation, 2) convergence rates between the AHDMR and HDMR approaches. The trivariate function used over the range $[0, 2]^d$ for $d = 10$ dimensions is given in the following form:

$$f(\mathbf{x}) = \sum_{i=1}^{d} x_i^2 - \sum_{i=1}^{d-2} x_i^3 x_{i+1}^3 x_{i+2}^3 \tag{7.21}$$

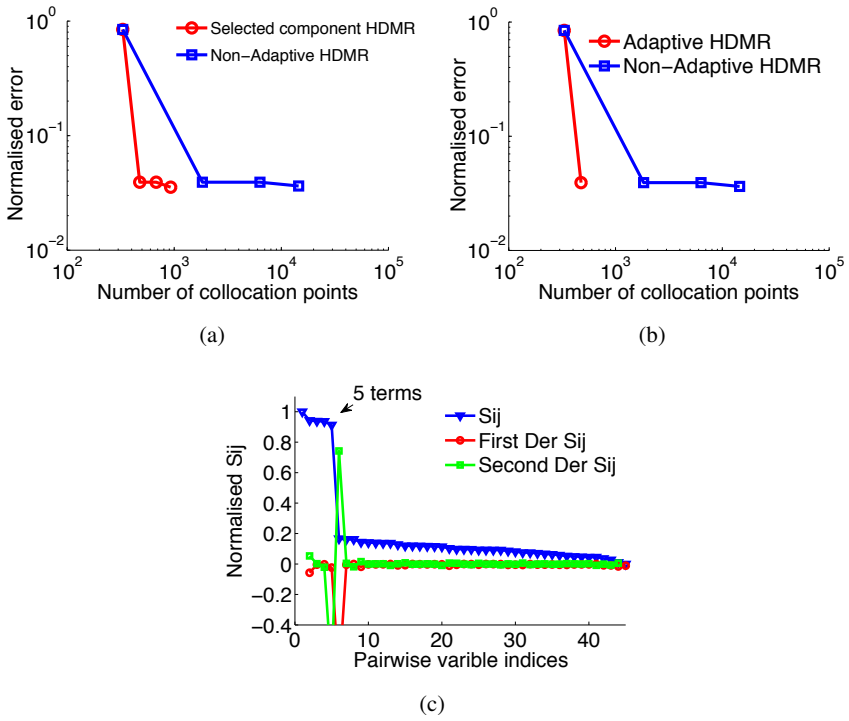


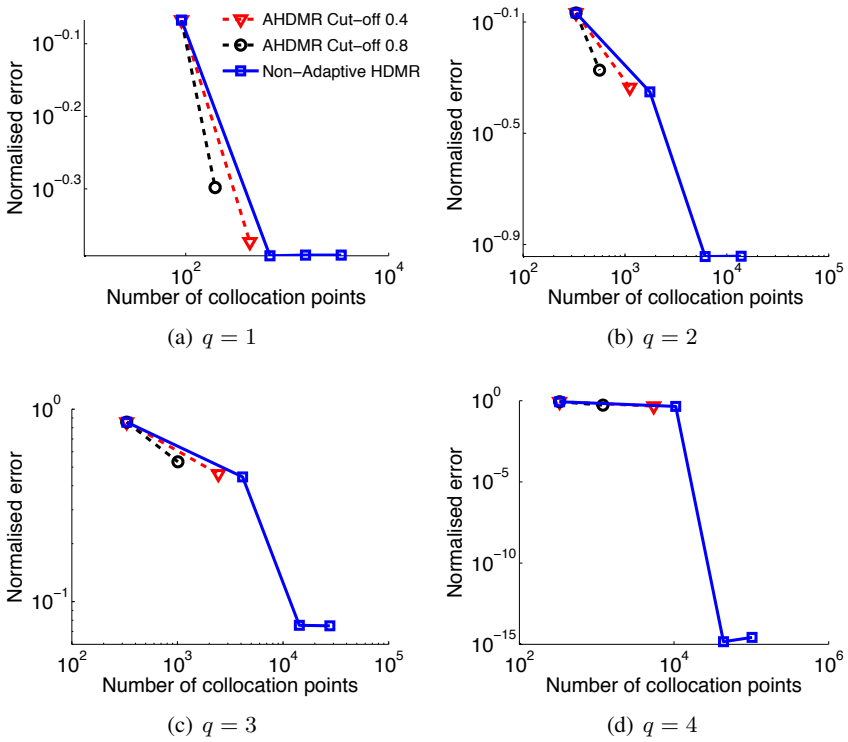
Figure 7.2: Error decays of the bivariate function plotted versus the number of collocation points: the non-adaptive HDMR is compared with selected component HDMR in (a) and with the Adaptive HDMR in (b). The pairwise variable indices and the normalised S_{ij} (Eq. 7.16) are displayed in (c), where the circle- and square-marker curve present first- and second-derivatives (Der) of the normalised S_{ij} with respect to the variable index.

where d presents the number of input variables. Figure 7.3 presents error convergences of the trivariate function for the HDMR and AHDMR approaches. Different levels of interpolation were obtained by varying the hierarchical surplus and the maximum number of collocation points of the component functions, while the number of collocation points of each curve was obtained by varying the order of the HDMR expansion. The full component HDMR (fixed component functions) is truncated after the forth order (shown by the square markers) while the AHDMR, adaptive component functions, is truncated after the second order.

Figures 7.3(a) – 7.3(d) present interpolation levels of the HDMR expansion from $q = 1$ to $q = 4$, where q is the Smolyak parameter (Section 4.4.2). The interpolation level is used for building the surrogate of the component functions in the HDMR. By increasing the interpolation level (from Figs. 7.3(a) – 7.3(d)), we see that the interpolation level improves the accuracy of the HDMR and can also affect the convergence rates. However, for the same interpolation level (within the same figure), the error decay is dependent on the order of the HDMR. As expected, the HDMR is converged at the third order for the trivariate function in Figs. 7.3(b) – 7.3(d).

In addition, two cut-offs (ϵ_s) in each figure represent the degree of adaptivity of the HDMR expansion. For instance, the AHDMR with $\epsilon_s = 0.4$ means that the normalised second-order Sobol indices greater than or equal to 0.4 are used in the HDMR expansion

and the others are discarded. A smaller cut-off value indicates that more pairwise interaction terms are included in the HDMR. This can be seen in Fig. 7.3 where the circle-marker curve ($\epsilon_s = 0.4$) requires fewer collocation points than the triangle-marker curve. If the HDMR is truncated after the second order, we see that using the AHDMR is more efficient than using the non-adaptive HDMR as it has a better convergence rate with a slightly poor accuracy. Even though the AHDMR is not better than the non-adaptive HDMR truncated after the second order, we showed in Section 6.5 that the subdomain HDMR can be used to efficiently approximate highly multiplicative functions.



(a) $q = 1$ (b) $q = 2$

(c) $q = 3$ (d) $q = 4$

Figure 7.3: Error decays of the trivariate function plotted versus the number of collocation points: the non-adaptive HDMR (the square-marker curve) is compared with the Adaptive HDMR. The level of interpolation is varied by using Smolyak parameters from $q = 1$ to $q = 4$ shown in (a)-(d). Two cut-offs ϵ_s are used at 0.4 (triangle-marker) and 0.8 (circle-marker), respectively.

7.7.1.3 Multivariate function

For multivariate function structure, Eq. 6.18 is used with 10 input variables over the range $[0, 2]^{10}$. The function is shown again below.

$$f(\mathbf{x}) = (\sum_{i=1}^{10} x_i)^{10} \tag{7.22}$$

In this example, the multiplicative function is chosen for two reasons: (1) to study the interpolation level and the error convergence of the non-adaptive HDMR expansion as shown in

Fig. 7.4, and (2) to compare the AHDMR and the non-adaptive HDMR with various cut-off values as shown in Fig. 7.5.

For the first part, the error convergence of the non-adaptive HDMR is plotted versus the number of collocation points in Fig. 7.4 where each curve represents different levels of interpolation used in the HDMR expansion. The number of collocation points for each curve was varied by increasing the order of the HDMR up to the fourth order. In Fig 7.4, it can be seen that an increase in the interpolation level from $q = 1$ to $q = 4$ can improve the accuracy of the approximation as q increases. In particular, using a higher interpolation level improves the approximation of the high-order component functions (order ≥ 3). As can be seen, the normalised errors of the third and fourth order are smaller when the level q increases. The result in this example indicates that the interpolation level affects the approximation of the high-order component functions. However, the interpolation level in our proposed algorithm, the ASGI-HDMR, is automatically adjusted through the hierarchical surpluses which select the level q that satisfies the error threshold.

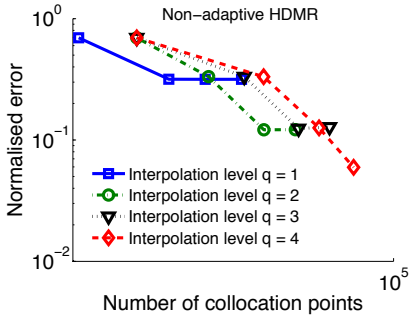


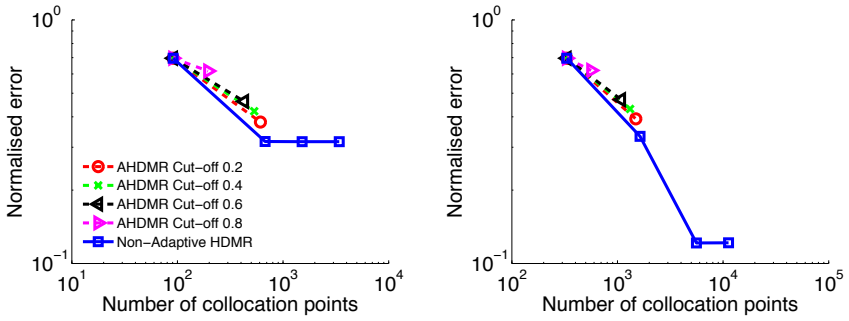
Figure 7.4: Error decays of the multiplicative function and the number of collocation points. Each curve shows the error of the non-adaptive HDMR approach at different interpolation levels $q = 1$ to $q = 4$.

For the second part, we present error convergences and the number of collocation points in Fig. 7.5 where the non-adaptive HDMR and the AHDMR are compared for different cut-off values. As described previously, the cut-off ϵ_s is used to determine the significant pairwise interaction terms. In this example, we use a second-order expansion for the AHDMR and a forth-order expansion for the non-adaptive HDMR to be the same as in Fig. 7.4. For the purpose of comparing both methods, we will discuss the convergence rates up to the second order.

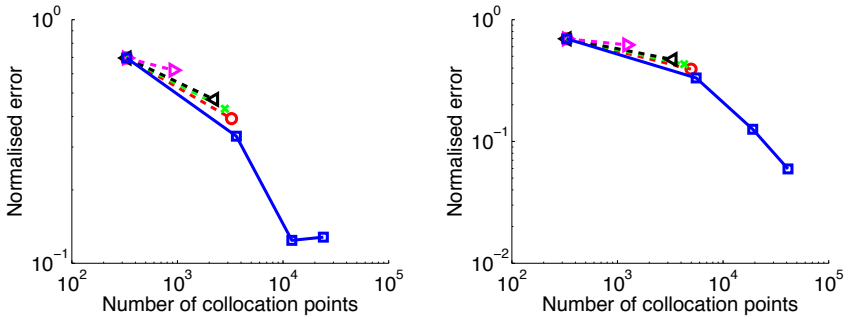
As shown in Fig. 7.5(a), the error curves of the AHDMR with various cut-offs ϵ_s are worse than the non-adaptive HDMR as the normalised errors from the second-order expansion of the AHDMR are greater than the one from the non-adaptive HDMR. This is because the multiplicative function consists of all interaction terms and all the terms contribute to the model output. Hence, the AHDMR is not a good candidate to approximate this function unless all interaction terms are included in the expansion. For such cases, the cut-offs ε_s in this example can be used as a trade-off between the number of collocation points and the accuracy of approximation, as can be seen from the circle-, cross-, triangle-, and square-marker curves in Fig. 7.4. Though it has shown that the interaction effects of many input variables from physical applications e.g. molecular physics [148] and physical chemistry [90] are insignificant, the truncated HDMR (and also the AHDMR in this example) is better used to approximate problems with weak interaction effects from high-order cooperation of input variables.

In Figs. 7.5(a) – 7.5(d), we see that increasing the interpolation level does not significantly improve the accuracy of the AHDMR, similar to the results of the non-adaptive

HDMR in Fig. 7.4. As described previously, using a high interpolation level could improve the surrogate quality for high-dimensional functions except for the AHDMR as the interpolation level is only used for low-order component functions (up to a second-order expansion). Therefore, we suggest using the subdomain HDMR to approximate highly multiplicative functions, rather than increasing interpolation level or increasing the error thresholds. The idea of the subdomain HDMR used in the optimisation framework will be applied to the reservoir PUNQ-S3 in the following section.



(a) $q = 1$ (b) $q = 2$



(c) $q = 3$ (d) $q = 4$

Figure 7.5: Error decays of the multiplicative function and the number of collocation points. The non-adaptive HDMR (the square-marker curve) is compared with the AHDMR using various cut-offs ϵ_s. The level of interpolation is varied by using Smolyak parameters from $q = 1$ to $q = 4$ shown in (a)-(d). The AHDMR with cut-offs 0.2, 0.4, 0.6 and 0.8 are shown by the dashed curves using circle-, cross-, triangle-, and square-markers, respectively.

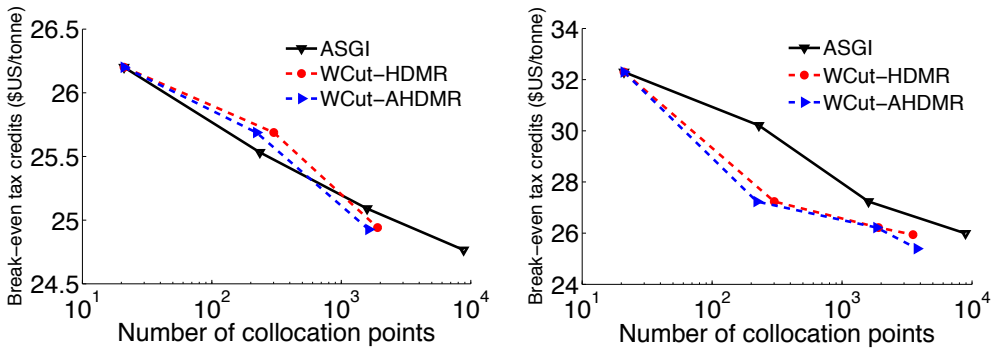
7.7.2 The PUNQ-S3 reservoir model

In this section, we present the use of the WCut-AHDMR method, described in Section 7.6, in comparison to the WCut-HDMR using the reservoir PUNQ-S3 models. The detailed descriptions, objective function, and initial model setup were described in Section 6.8.5, from which two models (named as model 1 and model 2) of the PUNQ-S3 are used.

Optimal results using the break-even tax credit function are presented first, followed by the study of alpha (α) and sensitivity indices used in the algorithm. Then, we present the effect of the low-order surrogate on the number of effective pairwise variables. Last, a comparison of the WCut-AHDMR and other control modes are shown.

7.7.2.1 Economic performance criteria: break-even tax credit.

Optimal BT values obtained from different surrogate building techniques (the ASGI, the WCut-HDMR, and the WCut-AHDMR) are displayed in Fig. 7.6. The number of collocation points shown on the x-axis is the number of points used to build the surrogate — this shows the computational cost of building the surrogate. From the figure, we expect the best surrogate approach to provide minimum values for both BT and the number of collocation points. As can be seen in Fig. 7.6(a), both WCut-HDMR and WCut-AHDMR approaches provide better BT values (lower BT values) than the ASGI technique at about 1,600 simulation runs, while the WCut-AHDMR is slightly better than the WCut-HDMR in terms of the number of collocation points. However, at about 200 points, the ASGI performs slightly better than the other two approaches. Results from model 2 are shown in Fig. 7.6(b) where the WCut-AHDMR performs much better than the ASGI and also shows an improvement in the BT value in comparison to the WCut-HDMR approach. This result generally indicates that the Sobol indices procedure, used to detect the significant component function terms, can improve the performance of the WCut-AHDMR algorithm while requiring fewer simulations.



(a) Model 1 (b) Model 2

Figure 7.6: The break-even tax credit (BT) values plotted versus the number of collocation points are displayed with three surrogate building techniques: the ASGI, the WCut-HDMR, and the WCut-AHDMR for two geological models a) and b), where the alpha value (Eq. 6.11) of both WCut-HDMR and WCut-AHDMR methods are set at 0.6 and 0.4 in (a) and (b), respectively. The cut-off value ϵ_s (Eq. 7.17) is set at 0.3 and 0 for WCut-AHDMR and WCut-HDMR respectively.

It is also important to compare the performance of the ASGI, WCut-HDMR, and WCut-HDMR methods with other optimisation algorithms. PSO [78] was chosen in this study because it has been successfully applied in many realistic applications, see [130]. Also, PSO is one of the metaheuristic approaches that does not require assumptions about the characteristics of the problem being solved, such as differentiability and/or smoothness of the function. In this study, we used the PSO algorithm directly with the reservoir flow simulation software without using a surrogate. The PSO parameters in this study are described in Table. 7.2.

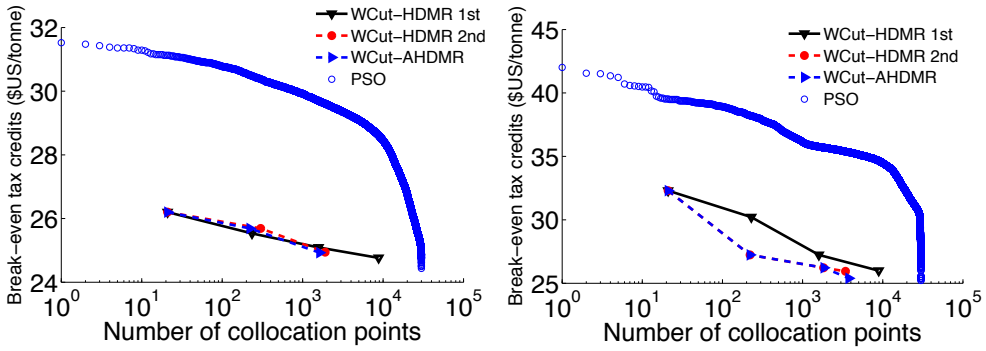
Figure 7.7 presents the optimal BT values plotted versus the number of collocation points for models 1 and 2. The results from Fig. 7.6 are replotted here to compare the BT

values from PSO as shown by the circle makers. Instead of showing each iteration of PSO, we sorted the objective function values obtained from each function evaluation of PSO in a descending order. It can be seen from the figure that spectral interpolation methods i.e. the ASGI, the WCut-HDMR, and the WCut-AHDMR show better convergence rates than PSO. The PSO was stopped at 30,000 function evaluations — the maximum number of function evaluations for the stopping criteria. However, the optimal solution obtained from PSO is slightly better than the one obtained from our approach, but with significantly more function evaluations. The optimal BT values of each method from models 1 and 2 are shown in Tables 7.3 and 7.4 respectively.

From Tables 7.3 and 7.4, we can claim that the global optimum of the BT from models 1 and 2 using 30,000 simulations are 24.43 and 25.23 $US/tonne of CO_2, respectively. Also, the optimum BT values from the WCut-AHDMR approach are very close to the global optimum solution for both models 1 and 2 using significantly fewer simulation runs. In the next section, we will present the effect of domain shrinking criteria of the WCut-AHDMR to the optimal solution.

Table 7.2: Initial parameter of the PSO algorithm

No.	Parameters	Numerical values
1.	generations (iterations)	600
2.	population size	50
3.	function tolerance	1e-6
4.	variable tolerance	1e-6
5.	social attraction	1.25
6.	cognitive attraction	0.5



(a) Model 1 (b) Model 2

Figure 7.7: The break-even tax credit (BT) values plotted versus the number of collocation points are displayed with four optimisation techniques: the ASGI, the WCut-HDMR, the WCut-AHDMR and PSO for two geological models a) and b). The ASGI, the WCut-HDMR, and the WCut-AHDMR obtained from Fig. 7.6 are replotted here for the purposes of comparison.

Table 7.3: Optimal BT values obtained from the last iteration for model 1 in Fig.7.7(a).

Techniques	BT values $US/tonne	Number of function evaluations (collocation points)
ASGI	24.77	8813
WCut-HDMR	24.94	1907
WCut-AHDMR	24.93	1628
PSO	24.43	30,000

Table 7.4: Optimal BT values obtained from the last iteration for model 2 in Fig.7.7(b).

Techniques	BT values $US/tonne	Number of function evaluations (collocation points)
ASGI	26	8917
WCut-HDMR	25.94	3513
WCut-AHDMR	25.39	3779
PSO	25.23	30,000

7.7.2.2 The effect of α in the WCut-AHDMR algorithm

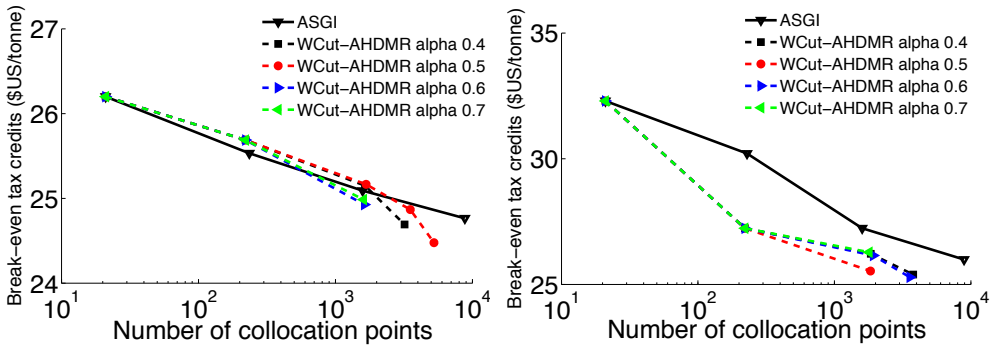
The alpha (α) of the WCut-AHDMR algorithm (Eq. 6.11) is used to determine the domain size of the AHDMR surrogate, which is a key factor to balance the exploration and exploitation of the optimisation process. In this section, we consider the impact of different alpha values on the optimal solution.

Figure 7.8 presents variation of the BT values from the WCut-AHDMR approach using different alpha values for models 1 and 2. Intuitively, a greater alpha value means that the WCut-AHDMR allows the optimisation algorithm to search in a larger domain, creating chances for another diverse but promising solution. This however may slow the rate of finding the optimal solution and may, in some cases, sacrifice the accuracy of the surrogate. In contrast, a smaller alpha value enables the algorithm to probe a restricted domain for a better solution, which may increase the rate of convergence but limit diversification of the solution.

We see from both Figs. 7.8(a) and 7.8(b) that the optimal BT values are better using an alpha value of 0.5. The results here show that alpha has an influential effect on the optimisation results, indicating that the optimal result is sensitive to the searching criteria of the WCut-AHDMR algorithm. It is also shown that an alpha value of around 0.5 provides the best convergence rate for both models. We will discuss an approach for selecting alpha in Section 8.2.3.

7.7.2.3 Sobol indices and the WCut-AHDMR algorithm

Since the proposed WCut-AHDMR approach provides the low-order surrogate model during the exploration phase (see the flowchart in Fig. 7.1), we can use this surrogate to find Sobol indices. We used the Sobol estimator in Eq. 7.15 to obtain the first-order Sobol indices for significant injection steps with respect to the BT function, and the second-order Sobol indices for significant interactions of the injection controls. In this section, we present both first- and second-order Sobol Indices for a sensitivity analysis of the BT.



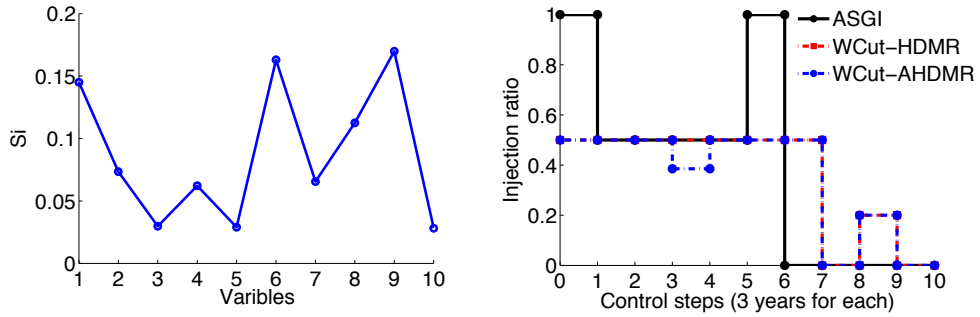
(a) Model 1 (b) Model 2

Figure 7.8: The break-even tax (BT) values plotted versus the number of collocation points are displayed using the ASGI and the WCut-AHDMR for various alpha values for models 1 and 2 in a) and b).

First-order Sobol sensitivity indices. We recall from Fig. 7.6 that the BT values are the optimal solutions obtained from different surrogate building approaches, from which the best BT was obtained from the WCut-AHDMR approach and the worst from the ASGI method. Two models are used here: Figs. 7.9 and 7.10 present the results from models 1 and 2, respectively. In both figures, the first-order Sobol indices (the main effect) and the associated injection steps are shown; only the injection ratios of injector 1 are displayed since the injection ratios of injector 2 can be simply obtained by subtracting the ratios of injector 1 from 1 (see Fig. **??**).

For model 1, we see from Fig. 7.9(a) that the variables 1, 6, and 9 have a higher Sobol index than the others. This means that the BT function is sensitive to changes in these control variables. Therefore, adjusting these control variables might result in a change of the model output — the BT. In Fig. 7.9(b), we see that both the WCut-AHDMR and WCut-HDMR approaches have control steps 1, 6, 7, and 9 that are different from the ASGI, and most of these control steps correspond to the important variables shown in Fig. 7.9(a). Therefore, we can infer that the control steps 1, 6, and 9 from the WCut-AHDMR and the WCut-HDMR contribute to the better BT value, see Fig. 7.6). Additionally, step 4 in Fig. 7.9(a) is not an important control variable since it has a relatively smaller number of the main effects. Thus, altering step 4 (in Fig. 7.9(b)) of the WCut-AHDMR does not improve the BT value.

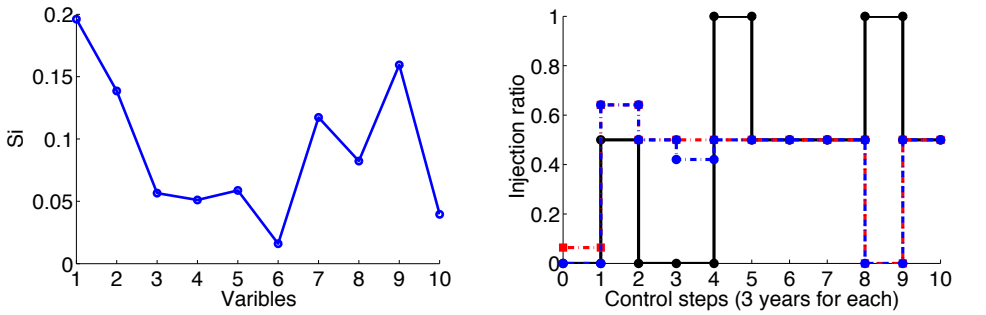
For model 2, Fig. 7.10(a) shows that the variables 1, 9, and 2 (in order of Sobol index ranking) are the important control steps; therefore, changing these steps can impact the BT value. As can be seen in Fig. 7.10(b), the injection steps 9 and 2 from both the WCut-AHDMR and WCut-HDMR approaches are different from the ASGI. Thus, it may imply that both methods target the adjustment on these steps in order to change (improve) the BT value, and consequently using these control steps leads to the lower BT value (Fig. 7.6(b)). In addition, the difference in the BT value from both the WCut-HDMR and WCut-AHDMR methods results principally from the difference in control step 1 since this injection step is one of the most significant control variables. The results from Figs 7.9 and 7.10 imply that the WCut-AHDMR method targets the injection rates most likely at the influential control



(a) First-order Sobol indices (b) Injection control steps

Figure 7.9: a) The first-order Sobol indices (S_i in Eq. 7.10) evaluated from the low-order surrogate during the exploration phase. b) Injection control steps for three different approaches: the ASGI, the WCut-HDMR, and the WCut-AHDMR. This result was obtained from model 1.

steps and hence it helps to guide the search to a better optimal solution while using fewer simulation runs than PSO and the ASGI method.



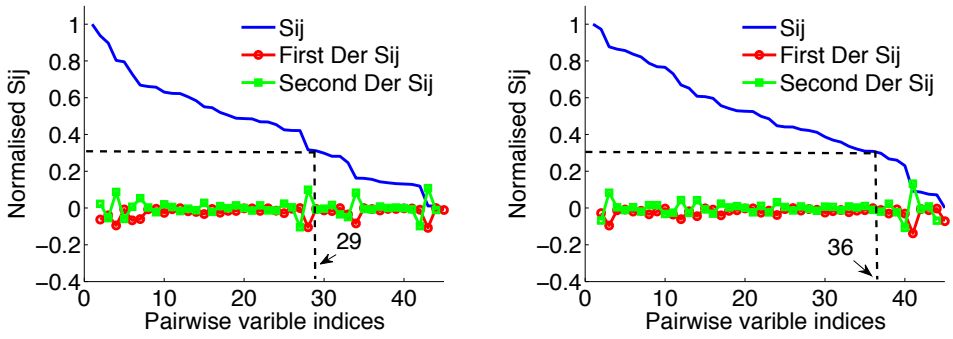
(a) First-order Sobol indices (b) Injection control steps

Figure 7.10: a) The first-order Sobol indices (S_i in Eq. 7.10) evaluated from the low-order surrogate during the exploration phase. b) Injection control steps for three different approaches: the ASGI, the WCut-HDMR, and the WCut-AHDMR. This result was obtained from model 2.

Second-order Sobol sensitivity indices. In Section 7.3, we showed that a larger second-order Sobol index implies the significant impact of the pairwise variables. Based on the second-order Sobol indices, we can decide which important terms are included in the Cut-HDMR expansion for efficient surrogate construction.

In this work, we present the use of normalised Sobol indices (Eq. 7.16) with the cut-off parameter (ϵ_s in Eq. 7.17) to determine the influential terms. Figure 7.11 presents how the normalised S_{ij} were selected in the AHDMR algorithm for both models 1 and 2. The solid curve denotes the normalised Sobol indices while the circle- and square-marker curves display the first and second derivative of the normalised S_{ij}. The derivatives illustrate the change of the normalised S_{ij} value with respect to the pairwise variable index. The cut-off value is selected based on the largest (either first or second) derivatives of the normalised

S_{ij}. If the cut-off ϵ_s is set to zero, all pairwise component functions $f_{ij}(x_i, x_j)$ will be included in the Cut-HDMR structure and hence there is no decompositional adaptivity. On the other hand, if the cut-off ϵ_s is set to 1, only the most significant term will be included and all other terms will be discarded. These cut-off values at 0 and 1 demonstrate extreme cases of the decompositional adaptivity. Thus, they represent the adaptivity degree of the Cut-HDMR expansion. In this study, we use the cut-off at 0.3 (dashed lines) for both model realisations. There are 10 injection steps in this study; therefore, the total number of the pairwise terms is 45 (${}^{10}C_2 = \binom{10}{2}$). Based on the cut-off value at 0.3, it can be seen in Fig. 7.11 that we can obtain 29 and 36 influential pairwise component functions for models 1 and 2, respectively. These influential pairwise terms were used in the AHDMR expansion for the WCut-AHDMR algorithm, of which the optimal BT were previously shown in Fig. 7.6.



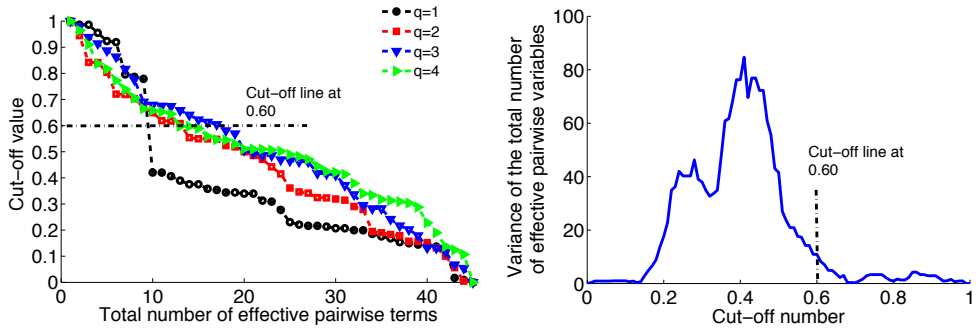
(a) Model 1 (b) Model 2

Figure 7.11: Normalised second-order Sobol indices plotted versus the index of pairwise variables for two realisations a) and b). The circle- and square-marker represents the changes of the normalised S_{ij} in terms of first and second derivatives (Der) with respect to the pairwise variable index.

7.7.2.4 The effect of low-order surrogates in the WCut-AHDMR algorithm.

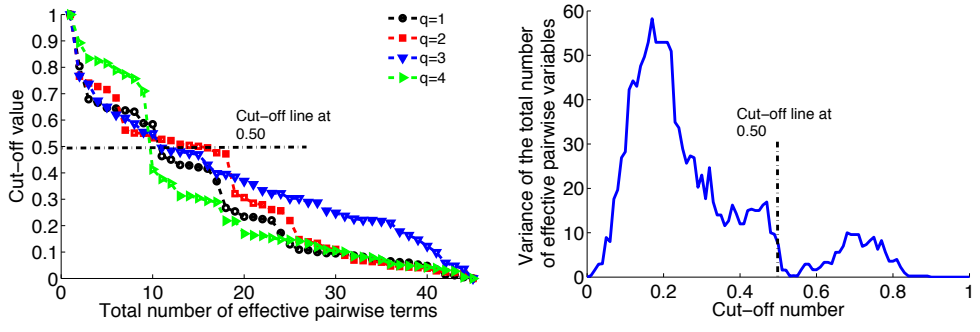
In the previous section, we demonstrated how the low-order surrogate in the exploration phase could be used to assist the global sensitivity to the input variables. However, in this section, we study how the low-order surrogates using different interpolation levels could affect the total number of effective pairwise variables. In order to avoid confusion, we refer to the low-order surrogate as the exploration surrogate in this section. To this end, different exploration surrogates were obtained by varying the interpolation level from $q = 1$ to $q = 4$. Figures 7.12 and 7.13 present the relationship of the cut-off values ϵ_s and the total number of effective pairwise variables for models 1 and 2, respectively.

For model 1, the curves in Fig. 7.12(a) represent the exploration surrogates using various interpolation levels. We see that the total number of effective pairwise variables depends on the interpolation level of the surrogate — there is a large variation in the total number of effective pairwise interactions when the cut-off value is smaller than 0.6. In contrast, the variation of the total number of effective pairwise variables is relatively small when the cut-



(a) (b)

Figure 7.12: The relationship of the cut-off values and the total number of effective pairwise terms with four levels of interpolation is shown in (a) where the associated variance of the total number of effective components is presented in (b). Model 1 is used in this figure.



(a) (b)

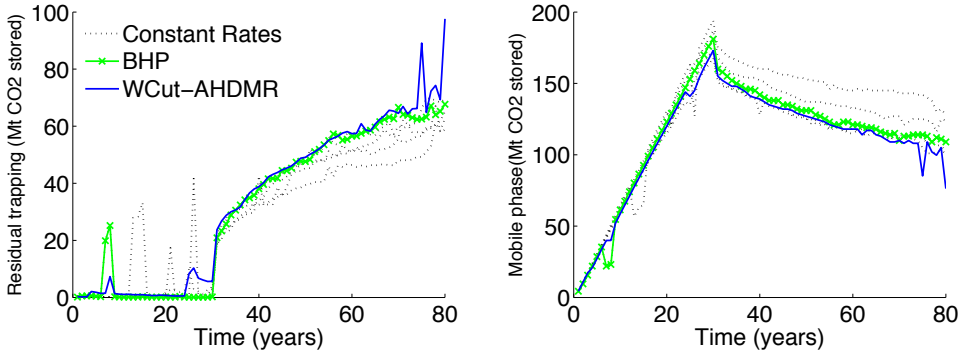
Figure 7.13: The relationship of the cut-off values and the total number of effective pairwise terms with four levels of interpolation is shown in (a) where the associated variance of the total number of effective components is presented in (b). Model 2 is used in this figure.

off value is greater than 0.6. The variance and the cut-off value are shown in Fig. 7.12(b) where the variance of the total number of effective pairwise variables highlights the sensitivity of the total number of effective terms to the interpolation level. In this figure, we used the cut-off line at 0.6 to separate the region in which the interpolation level of the surrogate affects the total number of the effective variables.

For model 2, Figure 7.13 shows similar results to model 1 with the cut-off line at 0.5. We can see that the total number of effective pairwise variables becomes more sensitive to the interpolation level when the cut-off value is smaller than 0.5. The results from both models show that the choice of cut-off thresholds might impact the number of effective components and subsequently the quality of the surrogate need to compensate for missing terms in the HDMR decomposition.

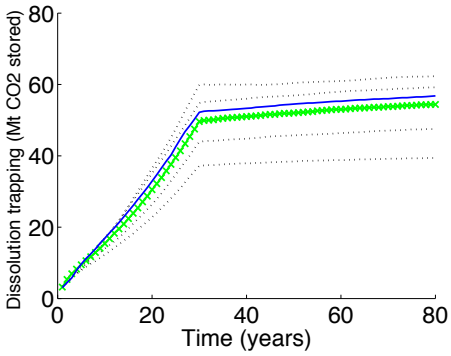
7.7.2.5 Comparison of the WCut-AHDMR method and other control modes

Figure 7.14 displays the various phases of CO_2 obtained from the optimal injection strategy, BHP controls, and constant rate controls at different injection ratios of 0, 0.25, 0.75, and 1; all control rates were constrained by the total mass of CO_2 injection in this study. The optimal strategy is obtained from the WCut-AHDMR approach from model 2 whose BT value is shown in Fig. 7.6(b). In comparison to BHP and constant rate controls, the WCut-AHDMR approach provides the maximum mass of CO_2 capillary trapped, see Fig. 7.14(a), and the minimum amount of mobile CO_2, see Fig. 7.14(b). Moreover, the results obtained from the WCut-AHDMR are significantly better than the other control modes, which also confirms the value of optimisation over non-optimal strategies using BHP and fixed rate controls. In Fig. 7.14(c), we observe that using a constant rate could be sufficient to increase the mass of CO_2 dissolved in brine, but obviously it is not a good injection strategy to enhance the amount of immobile CO_2, both capillary and dissolution trappings.



(a) Residual trapping

(b) Mobile phase of CO_2



(c) Dissolution trapping

Figure 7.14: The mass of CO_2 in various phases, shown in (a) residual (capillary) trapping, (b) dissolution trapping, and (c) mobile phase of CO_2, are obtained from different control modes. The BHP and the constant rate controls are represented by the cross-marker and the dotted curves, respectively. The solid curve presents the proposed WCut-AHDMR technique.

7.8 Summary

We have developed an efficient surrogate-assisted optimisation technique using Adaptive High-Dimensional Model Representation (AHDMR) with walking cut points to find the optimal CO_2 injection strategy. Our results are summarised as follows:

1. The adaptive ASGI-HDMR approach uses effective lower-dimensional functions to represent a high-dimensional function from which the Sobol index is used to identify the effective pairwise terms in the HDMR structure.

2. On bivariate test function, better convergence rates were obtained from the AHDMR approach in comparison to the ASGI-HDMR approach. Therefore, building the surrogate with the AHDMR technique in this example results in fewer points (simulation runs) than the ASGI-HDMR for equivalent errors.

3. On trivariate test function, the AHDMR was more computationally efficient than the non-adaptive HDMR with a slightly poor accuracy when its expansion was truncated after the second order. Even though using the HDMR up to third-order expansion could well represent the trivariate function, this is not necessary if subdomain HDMR is employed.

4. The results from the multiplicative function showed that the AHDMR method was worse than the non-adaptive HDMR. This is because all component functions contributes to the model output and hence using the AHDMR will not improve the rate of convergence. However, we reported that the cut-off number in the AHDMR expansion could be used as a trade-off between the accuracy of the surrogate and the cost of computation.

5. For a surrogate-assisted optimisation method, we developed the WCut-AHDMR technique which was modified from the WCut-HDMR approach by including a global sensitivity analysis to identify the effective pairwise component functions. The surrogate using the WCut-AHDMR is a surface built locally with a reduced domain. The centroid of the surface is iteratively updated based on the optimal point obtained from the previous iteration, and the new surface with the updated centroid is built by shrinking the search domain. This process is repeated iteratively while balancing the trade-off between exploration and exploitation in the optimisation algorithm. By using sensitivity indices in the algorithm, the results based on the BT function showed that the WCut-AHDMR not only provides a better optimal solution, but also requires fewer simulation runs than the WCut-HDMR.

6. The WCut-AHDMR method tested on the PUNQ-S3 model revealed the potential benefits of using this optimisation technique for CO_2 sequestration in deep saline aquifers in comparison to the WCut-HDMR, the ASGI, and other control modes, such as BHP and constant rates.

CHAPTER 8

SUMMARY AND RECOMMENDATIONS

This thesis addressed two key areas: (i) how to assist decision making under uncertainty in Chapters 4 and 5, and (ii) the development of optimisation algorithms in Chapters 6 and 7. The techniques presented in the thesis are based on the surrogate-assisted optimisation approach with application to CO_2 sequestration. The main results are summarised in this chapter with a suggestion for future work at the end.

8.1 Summary

Successful CO_2 sequestration relies on operation strategies considering economic performance criteria, sufficient capacity and long-term security of storage formations. Unfortunately, these criteria are generally conflicting objectives, and often difficult to guarantee when sufficient geological data from storage sites is not available. In this thesis, we addressed these issues and showed how to make efficient and effective decisions for CO_2 injection strategies in deep saline formations under geological uncertainty. Injection rates of CO_2 were allocated among injection wells for optimisation.

The thesis began with Chapter 2 with an introduction to climate change and carbon capture and storage (CCS). We briefly reviewed the evidence that climate change is closely related to greenhouse gas emissions (GHGs) which can be mitigated by sequestration of CO_2 in geological formations such as deep saline aquifers. We discussed trapping mechanisms which render the CO_2 immobile, and hence safely stored, in aquifers.

Uncertainty quantification (UQ) was addressed in Chapter 3. We presented the main ingredients of UQ, types and sources of uncertainty, and described the type of uncertainty present in the geological description of aquifers. To mitigate the impact of geological uncertainty, we introduced the use of robust optimisation that considers both the objective function value and its variation. A literature review related to sensitivity analysis of CO_2 trapped in aquifers to various reservoir parameters was presented, and recent optimisation techniques that enhanced the trapping mechanism were also reviewed.

The main contribution of this thesis was presented in Chapter 4 onward presenting decision making and optimisation techniques for CO_2 injection strategies under uncertainty. The research was presented in three parts in four chapters:

1. Surrogate-assisted optimisation in Chapter 4.
2. Multi-criterion optimisation in Chapters 5.
3. Surrogate technique for high-dimensional functions in Chapters 6 and 7.

Surrogate-assisted optimisation

Computational optimisation applied to CO_2 sequestration requires multiple function evaluations of the computationally intensive flow equations. Moreover, this computational cost increases rapidly when dealing with the impact of geological uncertainty.

We introduced the surrogate building technique that assists the optimisation process so that finding a robust injection strategy is feasible. Theoretical and approximate approaches using the sparse grid sampling method were presented in this chapter where the basic principles of polynomial interpolation, and the types of chosen grid nodes were described. Smolyak and sparse grid interpolation approaches were used to approximate the objective function for optimisation in this work. For efficient construction of the surrogate, dimensional adaptivity was applied to automatically fill the sampling points along the dimensions where greater resolution is needed. The adaptive sparse grid interpolation approach was implemented in Chapter 4 to find a robust CO_2 injection strategy.

Chapter 4 presented CO_2 applications using the adaptive sparse grid interpolation method. A break-even tax credit (BT) function was formulated to take into account the economic criteria of a CO_2 sequestration project. Later, we proposed the use of the utility function by using the mean and variance of the BT function to cope with geological uncertainty. Both adaptive and non-adaptive sparse grid interpolation were used to approximate the utility function and the optimisation procedure using particle swarm optimisation (PSO) was run on the utility function to find a robust solution.

From simple mathematical test functions and the PUNQ-S3 model, we found that adaptive sparse grid interpolation (ASGI) is more efficient than the regular (non-adaptive) sparse grid. Therefore, ASGI was used within two surrogate building approaches: the one-to-one and the mean-std techniques. The later technique is the better one for building the utility function for optimisation. Later, we reported that the robust solution was necessary for decision making under uncertainty as it was insensitive to the impact of geological uncertainty while providing the minimum value of the BT function. Results were shown for various utility parameters which can be adjusted based on the risk attitude of decision makers.

Multi-criterion optimisation

We presented the use of multi-objective optimisation (MOO) techniques to find robust injection strategies with a trade-off between conflicting criteria. We discussed the basic principles and general ideas of multi-optimisation. The concepts of domination, a Pareto-optimal set, and dependency of chosen criteria were introduced and discussed, from which we emphasised that the use of MOO would be more efficient than single-objective optimisation if there exist conflicts among objectives such that the Pareto-optimal set is necessary for decision making. A short survey of MOO approaches was given with detailed descriptions of a non-dominated sorting algorithm (NSGA-II). Also, we presented the MOO approach under uncertainty and discussed the use of ASGI for a surrogate-assisted optimisation approach.

Applications to CO_2 sequestration were presented with two main sections. Firstly, the utility function of the mean and the variance of the BT function in Chapter 4 was replaced by two multiple objective functions and the MOO approach was used to solve for the robust

solution. Since the MOO approach relies on parametric estimation of the mean and the variance, we presented three types of statistical measures and showed the impact on decision making using different parametric probability distributions. Secondly, in addition to economic criteria, we considered the sweep efficiency of CO_2 flooding and the risk of CO_2 leakage from storage aquifers. The use of the MOO approach under geological uncertainty showed that the mass of trapped CO_2 and the risk of CO_2 leakage are trade-off solutions which can be controlled by using an appropriate injection strategy. Later in this chapter, we proposed the preference-based bias, and minimax approaches to assist decision making with multiple trade-off solutions. To this end, we developed tools for robust injection strategies when the economic considerations, CO_2 sweep efficiency, and the risk of CO_2 leakage were taken into account simultaneously.

Optimisation techniques for high-dimensional functions

Our previous work was limited to problems with a small number of control variables. For problems with many input variables, the difficulty of constructing accurate high-dimensional surrogate models is attributed to the curse of dimensionality — the computational time increases exponentially with the number of input variables. In Chapters 6 and 7, we proposed surrogate building techniques for the optimisation of CO_2 injection strategies with many input parameters.

Chapter 6 presented the use of high-dimensional model representation (HDMR) to efficiently map the input-output of functions in a hierarchical structure of component functions with different input variables. This technique was used to build the surrogate of the objective function having many input variables. Several HDMR techniques were discussed in the chapter and the Cut-HDMR coupled with the ASGI method was used to assist CO_2 optimisation. In the application section, we demonstrated the use of HDMR with simple mathematical functions and the PUNQ-S3 model. Based on the curse of dimensionality study, it was shown that the hierarchical adaptive sparse grid using HDMR was suitable for building a surrogate having many input variables. Even though the degree of input variable interactions of the function is important for the selection of the HDMR order, we found that the low-order HDMR is a promising approach when the domain of input variables becomes small. Later, we showed that the subdomain HDMR with a moving cut-point for optimisation of the PUNQ-S3 model provides better optimal results in comparison to the ASGI approach. In the end, we presented that the proposed WCut-HDMR approach resulted in the maximum mass of CO_2 trapped in various phases, in comparison to the other control modes.

In order to increase computational efficiency of the algorithm, we developed an adaptive scheme for the WCut-HDMR method using second-order Sobol indices in Chapter 7. We demonstrated the use of Sobol indices to adaptively select the effective component functions in the HDMR expansion. Mathematical test functions and the PUNQ-S3 models were presented. For simple mathematical test functions, we found that the Sobol indices efficiently reveal the effective component functions in the HDMR expansion, leading to the adaptive HDMR (AHDMR). For the PUNQ-S3 model, the AHDMR method combined with a mov-

ing cut-point for optimisation was better than the WCut-HDMR, the ASGI, and the PSO methods, providing a promising tool for optimisation with many input variables.

Final remarks

In the thesis, we have addressed three key factors: robust optimisation, multi-criterion optimisation, and optimisation with many input variables, to assist the decision making process for CO_2 sequestration under uncertainty. The proposed techniques provide alternative and efficient tools for the design of CO_2 injection in deep saline aquifers.

8.2 Research directions for future work

8.2.1 Decision making and robust optimisation

Mean-variance method

Mean and variance were used in the thesis for robust optimisation either in the form of a utility function or multiple objective functions. We showed that the mean and the variance of the objective function are conflicting criteria for which the Pareto-optimal front exists and the trade-off solution can be obtained. However, for effective decision making using the mean-variance approach, we found that some criteria need to be satisfied.

Figure 8.1 presents the idea for the mean and standard deviation of the objective function when both are minimised. The left figure presents two extreme solutions of the Pareto-optimal front on the mean-std criterion space; these two solutions can be expressed in three cases for the PDF of the objective function, as shown in the blue and red distributions in A, B, and C; the dashed lines and the arrows show the mean and the Value-at-Risk (VaR) of the distribution, respectively. In case A, the VaR of the blue distribution is greater than the mean of the red distribution; this is a general characteristic of the mean-std approach from which a trade-off between mean and standard deviation exists. However, in case C, we see that the blue solution is always better than the red as the worst case (VaR) of the blue is still better than the best case of the red distribution. Hence, the mean-std approach in case C does not provide useful information as a rational decision maker would prefer the blue solution to the red. Future work could investigate the necessary criteria of the mean-std approach so that it is useful for decision making. For this reason, we suggest that the VaR equal to the mean of the red distribution (Case B) can be used as a necessary condition.

Parametric Pareto-optimal front

The Pareto front was discussed previously where we showed that it is a key ingredient for multi-criterion decision making. For future work, we suggest that the Pareto front can be modelled as a parametric function that contains the input parameters of interest. By doing this, a sensitivity analysis of the Pareto front to the input parameters can be obtained. For example, we can include types of geophysical properties in the model so that we can investigate which parameters have the most influential effect on the Pareto front.

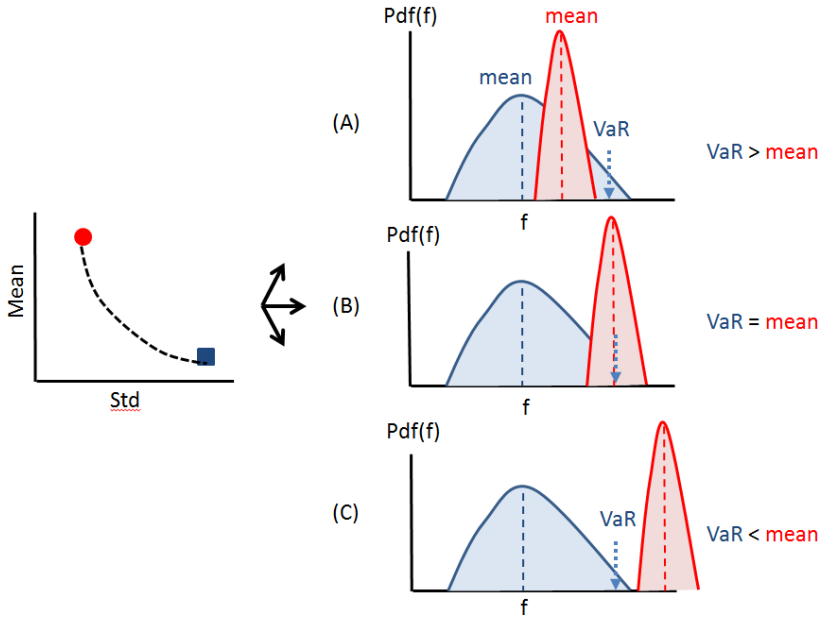


Figure 8.1: A suggested necessary condition of the mean-std approach for decision making: the left figure presents the Pareto-optimal front with two extreme solutions while the right figure represents three scenarios of the PDFs of the objective function. The Value-at-Risk and the mean of the distribution are denoted by the VaR and the dashed line, respectively.

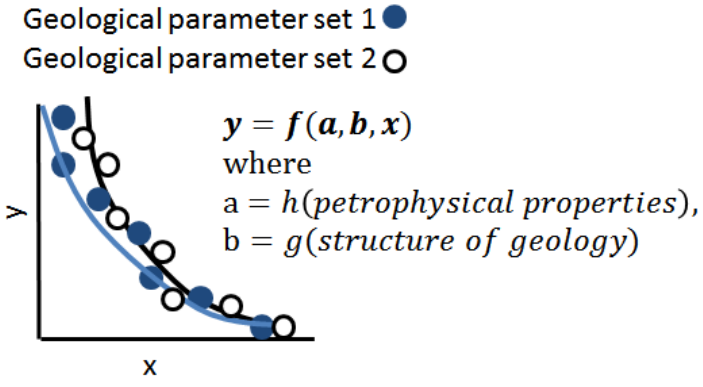


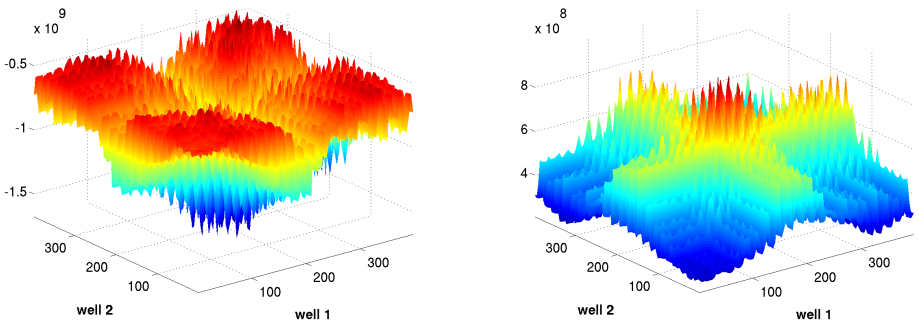
Figure 8.2: Demonstration of sensitivity analysis of Pareto-optimal front to different sets of geological properties of reservoir models; x and y present two conflicting objectives and the solid curves represent the Pareto-optimal front obtained from the regression of the solutions.

Figure 8.2 presents two conflicting criteria x and y with two sets of Pareto-optimal solutions (the blue and the hollow circles). Each set of solutions, represented by different colours, is obtained from different geological models. For example, the blue circles represent solutions from geological parameter set 1, while the hollow circles are from geological parameter set 2. The Pareto-optimal front, represented by the solid curves, is a regression model which is formulated by a function $y = f(a, b, x)$, where a and b are functions of the geological parameters, such as petrophysical properties and structural geology of aquifers, respectively. Therefore, based on a, b, and a given definition of variation of f, we can analytically quantify the significant input variables (e.g. petrophysical properties, the structure of geology) to multi-criterion decision making.

8.2.2 Surrogate building techniques: adaptive sparse grid

Well placement optimisation: integer variable problems

A smooth function (mixed derivatives are bounded) is required for the error decay of the polynomial interpolation to converge, meaning that the interpolating polynomial is more accurate as more sampling points are used. However, for some applications such as well placement problems for instance, the optimising variables (well location) are positive integers and hence the objective function might not be smooth. For this reason, a transformation of the input variables might be required. Figure 8.3 presents the response surface of the objective functions in well placement optimisation where wells 1 and 2 represent well locations defined by the grid box index. From this figure, we found that the mean and standard deviation of the NPV obtained from different geological realisations of the PUNQ-S3 model are not smooth as the input variables are all integers. Therefore, for this preliminary work detailed in [27], we transformed the integer variables (well locations) to be approximately close to continuous variables such that the objective function is smoother. Then, we used adaptive sparse grid interpolation to approximate the objective function on the presumably continuous domain. For future work, we will further investigate transformation techniques for integer variables for polynomial interpolation.



(a) Mean NPV

(b) NPV standard deviation

Figure 8.3: Three-dimensional response surface of the mean and the standard deviation of the NPV for well placement problems, from [27].

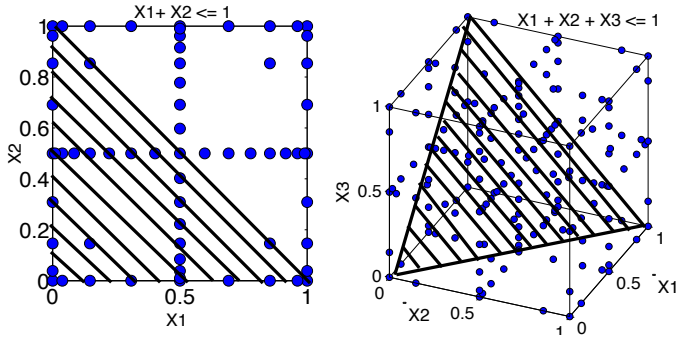


Figure 8.4: The layout of Chebyshev-based grid points with constrained input variables. Left figure: the shaded area shows the condition $x_1 + x_2 \leq 1$. Right figure: the volume behind the cross lines represents the condition $x_1 + x_2 + x_3 \leq 1$.

Constrained surrogates for optimisation problems

The layout of Chebyshev-based grid points is shown in Fig. 4.10 where we see that the domain range of input variables is bounded by 0 and 1. In this thesis, constraints on input variables have not been considered. However, for the problems with many injection wells for CO_2 sequestration, these constraints need to be included in future work so that the surrogate can be built with respect to the input constraints. We present this idea in Fig. 8.4. The shaded area on the left figure satisfies the condition $x_1 + x_2 \leq 1$ while the volume behind the cross lines on the right figure satisfies the condition $x_1 + x_2 + x_3 \leq 1$. These conditions represent the constraints on the total CO_2 injection rates as only a certain amount of CO_2 is supplied from industrial sources.

Multi-criteria optimisation

For the surrogate-assisted multi-objective optimisation technique, we built a separate surrogate for each objective function, as described in Section 5.5. Building a separate surrogate demands considerable computational resources in multi-objective optimisation. However, it is possible to use the information from both objective functions (the hierarchical surpluses of both objective functions) to determine the location of the sampling points. In other words, the adaptive sparse grid will take error measurements from both objective functions and hence the surrogate could be built based on the function with a greater degree of non-linearity. To this end, future work should include the use of dimensional adaptivity of the surrogate based on multiple output for efficient surrogate construction.

8.2.3 WCut-AHDMR algorithm

We list several parameters below that can be improved to increase the efficiency of the WCut-AHDMR algorithm.

Damping strategy for alpha (Eq. 6.11)

We see from Section 7.7.2.2 that the alpha value in the WCut-AHDMR algorithm could impact the optimal solution as it is used to balance the exploration and exploitation of the searching process of optimisation. The selection of alpha can be improved further in future work by using damping parameters (λ) similar to the Levenberg-Marquardt algorithm [99]. The damping strategy, in general, is used to update the alpha value based on the iteration step in the optimisation, which could improve the overall performance of the WCut-AHDMR algorithm.

Anisotropic alpha

In this thesis, we used a constant alpha value in all dimensions — an isotropic alpha. Future work could investigation the use of an anisotropic alpha — alpha values in all directions may not be the same — and its impact on the performance of the algorithm. Since a sensitivity analysis of the input variables is available from the WCut-AHDMR algorithm, one may use this sensitivity information to independently update the alpha in each dimension.

A comparison of WCut-AHDMR and ensemble-based optimisation

Ensemble-Based Optimisation (EnOpt) [29] is an approximate-gradient method where the search direction is obtained from Monte-Carlo sampling of the Gaussian distribution with a given mean and standard deviation. The standard deviation in EnOpt and alpha used in the WCut-AHDMR are similar. A comparison between the WCut-AHDMR and the EnOpt algorithms should be included in future work as EnOpt has been shown to be a promising approach in recent years [29, 124]. Also, both methods have advantages and disadvantages that are worth considering. For example, using EnOpt is more likely to get trapped at local optima than the WCut-AHDMR method as its search space is limited to the local neighbourhood whereas the WCut-AHDMR uses the information from the exploration surrogate to search for the optimum solution. However, the computational cost of the WCut-AHDMR algorithm depends on the number of input variables. In contrast, EnOpt can efficiently handle optimisation problems with many input variables.

8.2.4 Global sensitivity analysis

We used global sensitivity analysis (GSA) in Section 7.3 to investigate the model parameters to which the performance criterion of CO_2 sequestration is sensitive. We showed that the advantages of using GSA are as follows: Firstly, the sensitivity study accounts for the entire domain of the parameters of interest, in contrast to a local sensitivity analysis. Secondly, it is a model-independent approach where the linearity of the model inputs and outputs

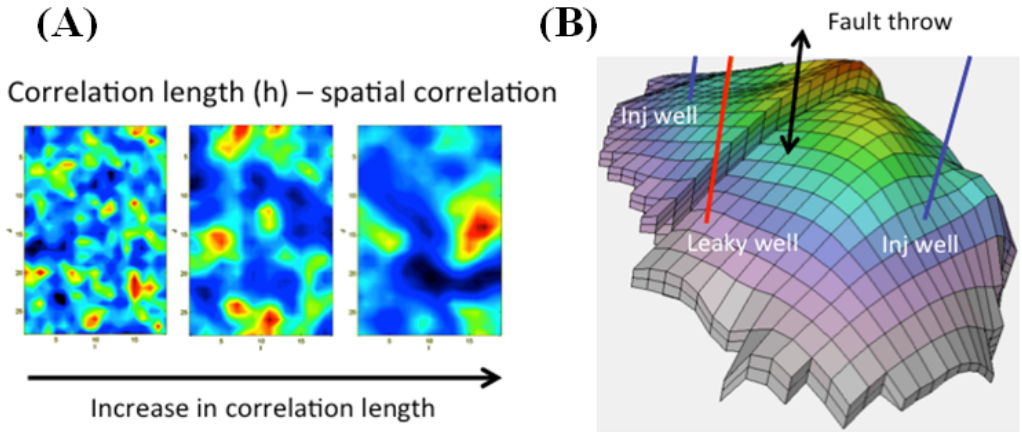


Figure 8.5: The parameters used in the global sensitivity analysis; these parameters are permeability, porosity, and fault throw. a) The correlation length represents the connectivity of the petrophysical properties; the figure shows different correlation lengths using the same mean and standard deviation of permeability. b) Fault throw demonstration: the throw of the fault in the PUNQ-S3 represents the uncertainty of the aquifer structure.

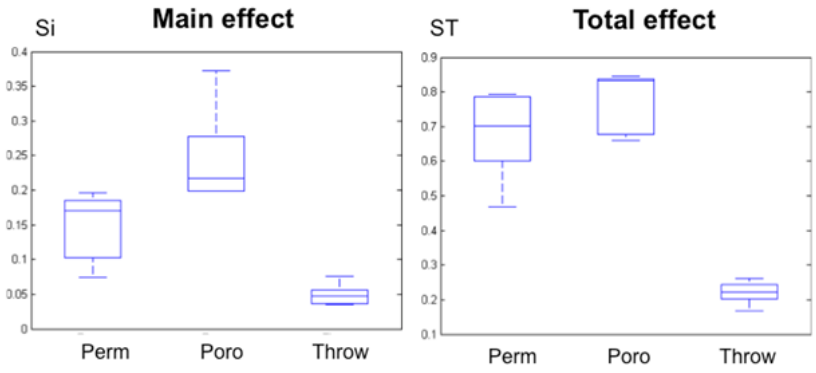


Figure 8.6: Box plots of the main and total effects. The main effect represents the uncertainty of individual parameters that contributes to the output uncertainty. In contrast, the total effect represents the contribution of output uncertainty from the interaction between the individual input parameters.

is not necessarily known a priori. Lastly, the GSA reveals the degree of input parameter interactions.

In our preliminary results, we used GSA to identify several aquifer properties of the PUNQ-S3 model in relation to the BT function. The parameters of interest are permeability, porosity, and structural features of the aquifer (fault throw). Figure 8.5 presents the parameters used in the GSA study: correlation length and fault throw.

Figure 8.6 displays the box plots of permeability, porosity, and fault throw for the main and total effects. The main effect represents the independent effect of the parameter alone, while the total effect presents the interaction of the particular parameters. The range of the boxes presents the variation from different injection strategies. From the figure, we see that the fault throw appears to be less sensitive in comparison to permeability and porosity for both the main and total effects. Additionally, the output contribution from the fault throw

appears to be insensitive to injection strategy as it has the smallest box range. However, this preliminary work needs to further study of the effect of different structural features and geophysical properties. For future work, we suggest the following:

- A sensitivity study of break-even tax credits to correlation lengths, sandbody connectivity, petrophysical properties, and structural features of saline aquifers using the GSA.

- A sensitivity study of the risk of CO_2 leakage to correlation lengths, sandbody connectivity, petrophysical properties, and structural features of saline aquifers using the GSA.

- A sensitivity of geological realisations (a set of correlation lengths, sandbody connectivity, petrophysical properties, and structural features of saline aquifers) and various injection strategies to objective function will be studied. Then, the effective realisation (a greater value of the total sensitivity indices) will be ranked and used in the optimisation process. For example, given that x represents a vector for the injection strategy, y represents a vector of geological realisations where $y := (y_1, y_2, \ldots, y_n)$ for n realisations, and f is an objective function. Then f can be explicitly written as:

$$f(x, y) = g(x) + h(y) + r(x, y) \tag{8.1}$$

$$\text{where } r(x, y) = r_1(x, y_1) + r_2(x, y_2) + \ldots + r_n(x, y_n) \tag{8.2}$$

  from which GSA makes use of the function $r(x, y)$ to detect the geological realisation that has the most interaction with injection strategy. This technique is based on the importance of group variables, proposed in [162].

CHAPTER A

ERROR ANALYSIS OF LAGRANGE POLYNOMIAL INTERPOLATION

In this section, we first define terms that will be used throughout this study, followed by a discussion about the choice of grid points that can impact the interpolation quality. For simplicity of the analysis, we consider a univariate function with a bounded interval $[-1, 1]$.

Let us first define a function $f : [-1, 1] \to \mathbb{R}$ on a closed interval where $f \in C^{\alpha}([-1, 1])$, and C^{α} represents a linear space of a continuous function with derivatives up to order α. We use $\alpha = 0$ for a continuous function in this analysis, unless otherwise stated. Given that an input space X contains $\{x_i\}_{i=0}^{n}$ where $-1 \leq x_0 \leq x_1 \leq \ldots \leq x_n \leq 1$, then the maximum or uniform norm is defined as:

$$\|f\|_{\infty} = \max_{x \in [-1,1]} |f(x)| \tag{A.1}$$

Given $\mathbb{P}_n$ and p_n denote a set of polynomials of degree at most n on a closed interval $[-1, 1]$ and an interpolating polynomial of degree n, respectively. Then, we can define the interpolation error as

$$E_{n,\infty} = \|f - p_n(f)\|_{\infty}, \; n = 0, 1, \ldots \tag{A.2}$$

If the best approximation polynomial is denoted by $p_n^* \in \mathbb{P}_n$, then the error of the best approximation polynomial is:

$$E_{n,\infty}^* = \|f - p_n^*(f)\|_{\infty} = \min\{\|f - p\|_{\infty}, p \in \mathbb{P}_n\} \tag{A.3}$$

For given polynomials of degree n, p_n^* is the best approximation polynomial because it has the minimum of a maximum distance between the function and the polynomial. The interpolation error $E_{n,\infty}$ and the best approximation error is related to a Lebesgue constant [138] which is used as an upper-bound for the interpolation error. The Lebesgue constant is expressed in the following, the details of which can be found in [134, 138, 158].

$$E_{n,\infty} \leq (1 + \Lambda_n(X)) E_{n,\infty}^* \tag{A.4}$$

where $\Lambda_n(X)$ is the Lebesgue constant given a fixed polynomial degree n; the Lebesgue constant [134, 138] can be expressed as:

$$\Lambda_n(X) = \left\| \sum_{i=0}^{n} |\ell_i^{(n)}(X)| \right\|_{\infty} \tag{A.5}$$

where $\ell_i^{(n)}$ is the characteristic polynomial with $n + 1$ points or the Lagrange basis function defined in Eq. 4.6. According to Eqs. A.4 and A.5, we see that the interpolation error

of the polynomial p_n, $E_{n,\infty}$, has the upper bound by the factor of $(1 + \Lambda_n(X))$ of the best approximation error $E^*_{n,\infty}$. Additionally, the Lebesgue constant depends only on a selection of grid points and does not depend on $f(x_i)$. It can also be seen that there may exist an optimal grid pattern that minimises $\Lambda_n(X)$. However, it is a difficult task to minimise the Lebesgue constant as the explicit form is generally not known. For that reason, Erdős [44] generalised $\Lambda_n(X)$, and later it was shown in [21] that the optimal Lebesgue constant can be estimated for any grid pattern X in a bounded interval, such that

$$\frac{1}{2} + \frac{2}{\pi}\log(n+1) < \Lambda_n(X^*) < \frac{3}{4} + \frac{2}{\pi}\log(n+1)\,, \quad n \geq 1 \tag{A.6}$$

For any continuous function $f \in C^0([a,b])$ in a bounded interval $[a,b]$, Eq. A.6 implies that there exist some grid patterns that the polynomial interpolation does not have a uniform convergence as $\Lambda_n(X) \to \infty$ when $n \to \infty$. The detailed survey of different grid patterns can be found in [21, 158]. However, we present only two types of grid distribution in this study: an equidistant and a non-equidistant grid pattern.

Given that X^{eq} are equally spaced points of the Lagrange interpolation in a bounded interval $[-1, 1]$, defined as

$$X^{eq} \;=\; \begin{cases} x_i \;=\; -1 + \frac{2(i-1)}{n-1} & ;i = 1, 2, \ldots, n \text{ for } n > 1 \\ x_1 \;=\; 0.5 & \text{for } n = 1 \end{cases} \tag{A.7}$$

Gutknecht and Gautschi [62] showed that the Lebesgue constant of the equally spaced points (X^e) has the asymptotic series

$$\Lambda_n(X^{eq}) \backsim \frac{2^{n+1}}{e\, n \log n} \quad \text{as} \quad n \to \infty \tag{A.8}$$

where e is the Euler's number (≈ 2.7183) and the $\backsim$-symbol expresses asymptotic equivalence, such that $g(N) \backsim f(N)$ if and only if $g(N)/f(N) \to 1$ as $N \to \infty$. In contrast to equally spaced points, the Lagrange polynomial using non-equidistant points, such as the Chebyshev Gauss-Lobatto (CGL) grid pattern, provides a much better Lebesgue constant. For the CGL grid pattern, we define T for a set of one-dimensional grids in the following:

$$T = \{x_i = -\cos(i\pi/n) : i = 0, 1, \ldots, \pi\ ;\ n = 1, 2, \ldots\} \tag{A.9}$$

The corresponding Lebesgue constant $\Lambda_n(T)$ of the CGL grid is shown with the upper bound in the following [138]:

$$\Lambda_n(T) \leq 1 + \frac{2}{\pi}\log(n+1) \quad \text{for } n \geq 1 \tag{A.10}$$

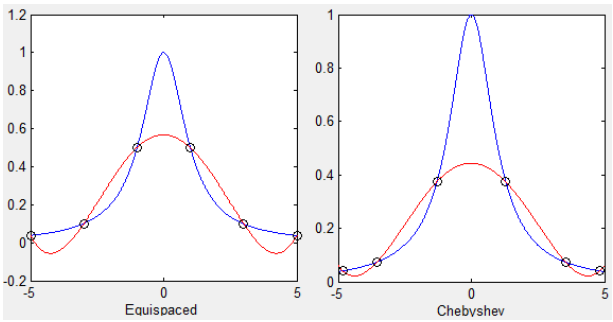
We see that the Lebesgue constant of the CGL grid is in the range of the optimal Lebesgue constant $\Lambda_n(X^*)$ shown in Eq. A.6; this confirms that the CGL has much better interpolation properties than the equally spaced grid pattern (Eq. A.8). The Lebesgue constant values of the CGL grids at different numbers of points can be found in [20, 80, 127]. The CGL is one of the grid patterns that provide Lebesgue constants very close to the optimal one. Thus, we will use the CGL grid pattern for the Barycentric interpolation in this study. It, however, should be noted that the other grid patterns such as the extended Chebyshev, and

augmented Chebyshev (see [68, 158]), which has Lebesgue constants close to the optimal value, can also be used.

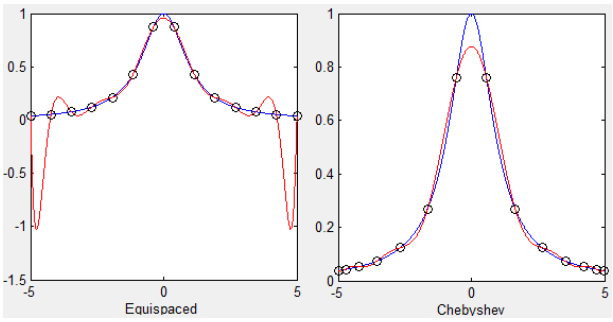
To demonstrate the impact of grid patterns on interpolation, we use the equispaced and the CGL pattern to interpolate the following Runge function [142] over the range $[-5, 5]$.

$$f(x) = \frac{1}{1 + \alpha^2 x^2} \quad (\alpha = 1 \text{ in this example}) \tag{A.11}$$

Figure A.1 illustrates the Runge function in blue and the Lagrange polynomial (in red) using different types of grid patterns for 5 and 14 points (the circles). As can be seen, increasing the number of equally spaced points (left column of the Figure) does not improve the quality of interpolation. This also implies that increasing the degree of the polynomial of the equispaced grids cannot always improve the quality of the polynomial interpolation. Additionally, the oscillation from the polynomial can introduce more errors at the edge of the interval. In contrast, the Chebyshev-based grid, the non-equidistant grid pattern, provides better interpolation as it reduces the oscillation of the interpolating polynomial at the edges of the interval (right column of Fig. A.1), known as the Runge phenomenon [142]. For a smooth function, we also see that increasing the number of CGL points improves the interpolation quality of the polynomial.



(a) n = 5 points



(b) n = 14 points

Figure A.1: Lagrange polynomial of the Runge function using equispaced and the CGL points for 5 points (a) and 14 points (b). The circle markers represent the data points while the blue and the red curves present the Runge function and the Lagrange polynomial, respectively.

BIBLIOGRAPHY

[1] Alexander GJ, Baptista AM (2002) Economic implications of using a mean-VaR model for portfolio selection: A comparison with mean-variance analysis. Journal of Economic Dynamics and Control 26(7-8):1159–1193

[2] Alış ÖF, Rabitz H (2001) Efficient implementation of high dimensional model representations. Journal of Mathematical Chemistry 29(2):127–142

[3] Allen J, Choi H, Rolander N (2005) What is Robust Design ? http://www.srl.gatech.edu/

[4] Ambrose WA, Lakshminarasimhan S, Holtz MH, Núñez López V, Hovorka SD, Duncan I (2007) Geologic factors controlling CO_2 storage capacity and permanence: case studies based on experience with heterogeneity in oil and gas reservoirs applied to CO_2 storage. Environmental Geology 54(8):1619–1633

[5] Anchliya A, Ehlig-Economides CA, Jafarpour B (2012) Aquifer Management To Accelerate CO_2 Dissolution and Trapping. SPE Journal 17(3):805–816

[6] Andrew M, Bijeljic B, Blunt MJ (2013) Pore-scale imaging of geological carbon dioxide storage under in situ conditions. Geophysical Research Letters 40(15):3915–3918

[7] Bachu S, Adams J (2003) Sequestration of CO_2 in geological media in response to climate change: capacity of deep saline aquifers to sequester CO_2 in solution. Energy Conversion and Management 44(20):3151–3175

[8] Balu A, Rao B (2012) Multicut-high dimensional model representation for structural reliability bounds estimation under mixed uncertainties. Computer-Aided Civil and Infrastructure Engineering 27(6):419–438

[9] Barthelmann V, Novak E, Ritter K (2000) High dimensional polynomial interpolation on sparse grids. Advances in Computational Mathematics 12:273–288

[10] Bellman R (1961) Adaptive control processes: a guided tour, vol 4. Princeton University Press

[11] Bennaceur K, Gielen D, Kerr T, Tam C (2008) CO_2 capture and storage: a key carbon abatement option. OECD

[12] Bentham M, Kirby M (2005) CO_2 storage in saline aquifers. Oil & Gas Science and Technology 60(3):559–567

[13] Berrut JP, Trefethen LN (2004) Barycentric Lagrange interpolation. SIAM Review 46(3):501–517

[14] Beyer HG, Sendhoff B (2007) Robust optimization–a comprehensive survey. Computer Methods in Applied Mechanics and Engineering 196(33):3190–3218

[15] Birkholzer J, Zhou Q, Tsang C (2009) Large-scale impact of CO_2 storage in deep saline aquifers: A sensitivity study on pressure response in stratified systems. International Journal of Greenhouse Gas Control 3(2):181–194

[16] Bohachevsky IO, Johnson ME, Stein ML (1986) Generalized simulated annealing for function optimization. Technometrics 28(3):209–217

[17] Bowman AW, Azzalini A (1997) Applied Smoothing Techniques for Data Analysis. Oxford University Press, Oxford

[18] Bratley P, Fox BL (1988) Algorithm 659: Implementing Sobol's quasirandom sequence generator. ACM Trans Math Softw 14(1):88–100, URL http://doi.acm.org/10.1145/42288.214372

[19] Brouwer D, Jansen J (2004) Dynamic Optimization of Waterflooding With Smart Wells Using Optimal Control Theory. SPE Journal 9(4):1–14

[20] Brutman L (1978) On the Lebesgue function for polynomial interpolation. SIAM Journal on Numerical Analysis 15(4):694–704

[21] Brutman L (1997) Lebesgue functions for polynomial interpolation–a survey. Annals of Numerical Mathematics 4:111–127

[22] Bryant S, Lakshminarasimhan S, Pope G (2008) Buoyancy-Dominated Multiphase Flow and Its Effect on Geological Sequestration of CO_2. SPE Journal 13(4):22–26

[23] Buzzard GT (2012) Global sensitivity analysis using sparse grid interpolation and polynomial chaos. Reliability Engineering & System Safety 107:82–89

[24] Cameron Da, Durlofsky LJ (2012) Optimization of well placement, CO_2 injection rates, and brine cycling for geological carbon sequestration. International Journal of Greenhouse Gas Control 10:100–112

[25] Cao Y, Chen Z, Gunzburger M (2009) Anova expansions and efficient sampling methods for parameter dependent nonlinear pdes. International Journal of Numerical Analysis & Modeling 6(2)

[26] CCS (Accessed: Oct 2014) Carbon capture and sequestration (CCS), CO_2 solution. URL http://www.co2solutions.com/

[27] Chang Y, Petvipusit KR, Devegowda D (2015) Multi-objective optimization coupled with dimension-wise polynomial based method in smart well placement under model uncertainty (SPE 17329). Proceedings of Reservoir Simulation Symposium Conference, 23-25 February, Houston, Texas, USA

[28] Chen W (1999) Quality Utility – A Compromise Programming Approach To Robust Design. Journal of Mechanical Design 121:179–187

[29] Chen Y, Oliver DS, Zhang D (2009) Efficient ensemble-based closed-loop production optimization. SPE Journal 14(4):634–645

[30] Chernick MR (2007) Bootstrap Methods A guide for Practitioners and Researchers. Wiley; 2nd edition

[31] Chiles JP, Delfiner P (2012) Geostatistics: Modeling Spatial Uncertainty. John Wiley & Sons Inc; 2nd edition, DOI 10.1002/9781118136188.ch2, URL http://dx.doi.org/10.1002/9781118136188.ch2

[32] CO2CRC (Accessed: Jan 2012) URL http://www.co2crc.com.au/aboutccs/stor_trapping.html

[33] DAKOTA (Access: Oct 2014) The DAKOTA Project, Large-Scale Engineering OPtimisation and Uncertainty Analysis, Sandia National Laboratories. URL http://dakota.sandia.gov/packages.html

[34] Deb K (2001) Multi-Objective Optimization Using Evolutionary Algorithms. Wiley; 1st Edition

[35] Deb K, Gupta H (2006) Introducing robustness in multi-objective optimization. Evolutionary Computation 14(4):463–494

[36] Deb K, Pratap A, Agarwal S, Meyarivan T (2002) A fast and elitist multiobjective genetic algorithm: NSGA-II. Evolutionary Computation 6(2):182–197

[37] Dixon L, Szegö G (1978) The Application of Bayesian Methods for Seeking the Extremum. Towards global optimisation 2/ eds L. C. W. Dixon and G. P. Szegö, North-Holland Pub. Co.

[38] Drop-wave (Accessed: Oct 2014) Dropwave test objective function: Global optimization test functions index. URL http://infinity77.net/global_optimization/test_functions.html#test-functions-index

[39] Ebigbo A, Class H, Helmig R (2007) CO_2 leakage through an abandoned well: problem-oriented benchmarks. Computational Geosciences 11(2):103–115

[40] Efron B (1979) Bootstrap methods: Another look at the jackknife. The Annals of Statistics 7(1):1–26, URL http://dx.doi.org/10.1214/aos/1176344552

[41] Ennis-King J, Paterson L (2002) Engineering aspects of geological sequestration of carbon dioxide (SPE 77809). Proceedings of the SPE Asia Pacific Oil and Gas Conference and Exhibition, 8-10 October, Melbourne, Australia

[42] Ennis-King JP, Paterson L (2005) Role of convective mixing in the long-term storage of carbon dioxide in deep saline formations. SPE Journal 10(03):349–356

[43] EPA (2008) United States Environmental Protection Agency. EPA's report on the Environment URL http://www.epa.gov/ncea/roe/docs/roe_final/EPAROE_FINAL_2008.PDF

[44] Erdős P (1961) Problems and results on the theory of interpolation. ii. Acta Mathematica Hungarica 12(1):235–244

[45] Floris FJT, Bush MD, Cuypers M, Roggero F, Syversveen A (2001) Methods for quantifying the uncertainty of production forecasts: a comparative study. Petroleum Geoscience 7:87–96, DOI 10.1144/petgeo.7.S.S87

[46] Forrester A, Sobester A, Keane A (2008) Engineering design via surrogate modelling: a practical guide. John Wiley & Sons

[47] Forrester AI, Keane AJ (2009) Recent advances in surrogate-based optimization. Progress in Aerospace Sciences 45(1-3):50–79, DOI 10.1016/j.paerosci.2008.11.001

[48] Fowkes J (2011) Bayesian numerical analysis: Global optimization and other applications. PhD thesis, University of Oxford

[49] Fowlkes WY, Creveling CM (1995) Engineering methods for robust product design. Addison-Wesley

[50] Gao Z, Hesthaven JS (2010) On anova expansions and strategies for choosing the anchor point. Applied Mathematics and Computation 217(7):3274–3285

[51] Garcke J (2013) Sparse grids in a nutshell. In: Sparse grids and applications, Springer, pp 57–80

[52] Gasda SE, Bachu S, Celia MA (2004) Spatial characterization of the location of potentially leaky wells penetrating a deep saline aquifer in a mature sedimentary basin. Environmental Geology 46(6-7):707–720

[53] Geoffrion AM, Dyer JS, Feinberg A (1972) An interactive approach for multi-criterion optimization, with an application to the operation of an academic department. Management Science 19(4-part-1):357–368

[54] Gerstner T (1998) Numerical integration using sparse grids. Numerical Algorithms 18:209–232

[55] Gerstner T, Griebel M (2003) Dimension-Adaptive Tensor-Product Quadrature. Computing 71(1):65–87, DOI 10.1007/s00607-003-0015-5

[56] Goater AL (2011) Multiphase Flow simulation with application for CO_2 storage. Phd thesis, Imperial College London, URL http://www.mendeley.com/research/no-title-avail/

[57] Goda T, Sato K (2013) Global Optimization of Injection Well Placement Toward Higher Safety of CO_2 Geological Storage. Energy Procedia 37:4583 – 4590

[58] Griebel M (1998) Adaptive sparse grid multilevel methods for elliptic pdes based on finite differences. Computing 61(2):151–179

[59] Griebel M, Holtz M (2010) Dimension-wise integration of high-dimensional functions with applications to finance. Journal of Complexity 26(5):455–489

[60] Griebel M, Schneider M, Zenger C (1992) A combination technique for the solution of sparse grid problems. In: Bequwens R, Groen Pd (eds) Iterative Methods in Linear Algebra, Elsevier, North-Holland, pp 263–281

[61] Gronchi S (1986) On investment criteria based on the internal rate of return. Oxford Economic Papers 38(1):pp. 174–180, URL http://www.jstor.org/stable/2663324

[62] Gutknecht M, Gautschi W (1990) Lectures on numerical mathematics. URL http://link.springer.com/content/pdf/bfm%3A978-1-4612-3468-5%2F1.pdf

[63] Heath JE, Kobos PH, Roach JD, Dewers T, McKenna SA (2012) Geologic heterogeneity and economic uncertainty of subsurface carbon dioxide storage. SPE Economics & Management 4(01):32–41

[64] Hegland M (2003) Adaptive sparse grids. Australian Mathematical Society 44:335–353

[65] Heris SMK, Khaloozadeh H (2011) Open-and closed-loop multiobjective optimal strategies for HIV therapy using NSGA-II. Biomedical Engineering, IEEE Transactions on 58(6):1678–1685

[66] Hestenes MR, Stiefel E (1952) Methods of conjugate gradients for solving linear systems1. Journal cf Research of the National Bureau of Standards Vol 49(6):409–436

[67] Higham NJ (2004) The numerical stability of barycentric Lagrange interpolation. IMA Journal of Numerical Analysis 24(4):547–556

[68] Hoang N (2013) On node distributions for interpolation and spectral methods. ArXiv e-prints arXiv:13056104 URL http://arxiv.org/abs/1305.6104

[69] Hu C, Youn BD, Yoon H (2013) An adaptive dimension decomposition and reselection method for reliability analysis. Structural and Multidisciplinary Optimization 47(3):423–440

[70] Iaccarino G (2009) Quantification of uncertainty in flow simulations using probabilistic methods. Tech. rep., DTIC Document, URL http://web.stanford.edu/group/uq/docs/iaccarino_vki_notes.pdf

[71] Ide ST, Jessen K, Jr FMO (2007) Storage of CO_2 in saline aquifers: Effects of gravity, viscous, and capillary forces on amount and timing of trapping. International Journal of Greenhouse Gas Control 1(4):481–491

[72] Jahangiri HR, Zhang D (2011) Optimization of the Net Present Value of Carbon Dioxide Sequestration and Enhanced Oil Recovery (OTC 21985). Proceedings of the Offshore Technology Conference, 2-5 May, Houston, Texas, USA

[73] Jahangiri HR, Zhang D (2012) Ensemble based co-optimization of carbon dioxide sequestration and enhanced oil recovery. International Journal of Greenhouse Gas Control 8:22–33

[74] Jansen MJ (1999) Analysis of variance designs for model output. Computer Physics Communications 117(1):35–43

[75] Jones DR, Schonlau M, Welch WJ (1998) Efficient global optimization of expensive black-box functions. Journal of Global optimization 13(4):455–492

[76] Juanes R, Spiteri EJ, Orr FM, Blunt MJ (2006) Impact of relative permeability hysteresis on geological CO_2 storage. Water Resources Research 42(W12418):1–13

[77] Kahraman C (2012) Computational Intelligence Systems in Industrial Engineering: With Recent Theory and Applications, vol 6. Springer

[78] Kennedy J, Eberhart R (1995) Particle swarm optimization. Proceedings of ICNN'95 - International Conference on Neural Networks 4:1942–1948

[79] Killough J (1976) Reservoir Simulation With History-Dependent Saturation Functions. SPE Journal 16(1):37–48

[80] Klimke WA (2006) Uncertainty modeling using fuzzy arithmetic and sparse grids. PhD thesis, Universität Stuttgart

[81] Klimke WA, Wohlmuth B (2005) Algorithm 847. ACM Transactions on Mathematical Software 31(4):561–579

[82] Knowles J, Corne D (1999) The pareto archived evolution strategy: a new baseline algorithm for pareto multiobjective optimisation. In: Evolutionary Computation, 1999. CEC 99. Proceedings of the 1999 Congress on, vol 1, pp 98–105, DOI 10.1109/CEC.1999.781913

[83] Kolmogorov AN (1963) On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition. Amer Math Soc Transl 28:55–59

[84] Koziel S, Ciaurri D, Leifsson L (2011) Computational Optimization, Methods and Algorithms: Surrogate-Based Methods, Studies in Computational Intelligence, vol 356, Springer Berlin / Heidelberg, pp 33–59

[85] Krokhmal P, Palmquist J, Uryasev S (2002) Portfolio optimization with conditional value-at-risk objective and constraints. Journal of Risk 4:43–68

[86] Kumar A, Noh M, Pope G (2004) Reservoir simulation of CO_2 storage in deep saline aquifers (SPE 89343). Proceedings of SPE/DOE Symposium Improve Oil Recovery, 17-21 April, 2004, Tulsa, Oklahoma, USA

[87] Kumar N, Bryant S (2008) Optimizing injection intervals in vertical and horizontal wells for CO_2 sequestration (SPE 116661). Proceedings of SPE Annual Technical Conference and Exhibition, 21-24 September, Denver, Colorado, USA

[88] Labovsky A, Gunzburger M (2014) An efficient and accurate method for the identification of the most influential random parameters appearing in the input data for pdes. SIAM/ASA Journal on Uncertainty Quantification 2(1):82–105, DOI 10.1137/120883785, URL http://dx.doi.org/10.1137/120883785

[89] Lengler U, De Lucia M, Kühn M (2010) The impact of heterogeneity on the distribution of CO_2: Numerical simulation of CO_2 storage at ketzin. International Journal of Greenhouse Gas Control 4(6):1016–1025

[90] Li G, Rosenthal C, Rabitz H (2001) High dimensional model representations. The Journal of Physical Chemistry A 105(33):7765–7777, DOI 10.1021/jp010450t, URL http://pubs.acs.org/doi/abs/10.1021/jp010450t, http://pubs.acs.org/doi/pdf/10.1021/jp010450t

[91] Li G, Wang SW, Rabitz H (2002) Practical approaches to construct RS-HDMR component functions. The Journal of Physical Chemistry A 106(37):8721–8733

[92] Li G, Wang SW, Rabitz H, Wang S, Jaffé P (2002) Global uncertainty assessments by high dimensional model representations (HDMR). Chemical Engineering Science 57(21):4445–4460

[93] Li G, Schoendorf J, Ho TS, Rabitz H (2004) Multicut-HDMR with an application to an ionospheric model. Journal of Computational Chemistry 25(9):1149–1156

[94] Lien M, Brouwer DR, Manseth T, Jansen JD (2008) Multiscale Regularization of Flooding Optimization for Smart Field Management. SPE Journal 13(2):195–204

[95] Lilliefors HW (1967) On the Kolmogorov-Smirnov test for normality with mean and variance unknown. Journal of the American Statistical Association 62(318):pp. 399–402, URL http://www.jstor.org/stable/2283970

[96] Ma X, Zabaras N (2010) An adaptive high-dimensional stochastic model representation technique for the solution of stochastic partial differential equations. Journal of Computational Physics 229(10):3884 – 3915, DOI http://dx.doi.org/10.1016/j.jcp.2010.01.033, URL http://www.sciencedirect.com/science/article/pii/S0021999110000562

[97] MacQueen J (1967) Some methods for classification and analysis of multivariate observations. In: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, California, USA, vol 1, pp 281–297

[98] Marler R, Arora J (2004) Survey of multi-objective optimization methods for engineering. Structural and Multidisciplinary Optimization 26(6):369–395

[99] Marquardt DW (1963) An algorithm for least-squares estimation of nonlinear parameters. Journal of the Society for Industrial & Applied Mathematics 11(2):431–441

[100] MATLAB (2013) version 8.2.0.701 (R2013b). The MathWorks Inc., Natick, Massachusetts

[101] Matott LS, Rabideau AJ, Craig JR (2006) Pump-and-treat optimization using analytic element method flow models. Advances in Water Resources 29(5):760–775

[102] Mattson CA, Messac A (2005) Pareto frontier based concept selection under uncertainty, with visualization. Optimization and Engineering 6(1):85–115

[103] McNeil AJ, Frey R, Embrechts P (2010) Quantitative risk management: concepts, techniques, and tools. Princeton University Press

[104] Metz B, Davidson O, de Coninck H, Loos M, Meyer L (2005) IPCC special report on carbon dioxide capture and storage. Tech. rep., Intergovernmental Panel on Climate Change, Geneva (Switzerland). Working Group III

[105] Metz B, Davidson O, De Coninck H, Loos M, Meyer L (2005) IPCC Special report on carbon dioxide capture and storage, vol 2. Cambridge University Press

[106] Meyer CD (2000) Matrix analysis and applied linear algebra. SIAM

[107] Mo S, Akervoll I (2005) Modeling long-term CO_2 storage in aquifer with a black-oil reservoir simulator (SPE 93951). Proceedings of SPE/EPA/DOE Exploration and Production Environmental Conference, 7-9 March, Texas, USA

[108] Mulvey J, Vanderbei R (1995) Robust optimization of large-scale systems. Operations Research 43(2):264–281

[109] Nghiem L, Yang C, Shrivastava V, Kohse B, Hassam M, Card C (2009) Risk mitigation through the optimization of residual gas and solubility trapping for CO_2 storage in saline aquifers. Energy Procedia 1(1):3015–3022

[110] Nghiem L, Shrivastava V, Kohse B (2010) Simulation and Optimization of Trapping Processes for CO_2 Storage in Saline Aquifers. Journal of Canadian Petroleum Technology (8):15–22

[111] Nordbotten JM, Celia MA (2011) Geological Storage of CO_2: Modeling Approaches for Large-Scale Simulation. Wiley

[112] Nordbotten JM, Celia MA, Bachu S (2004) Analytical solutions for leakage rates through abandoned wells. Water Resources Research 40(W04204)

[113] Nordbotten JM, Celia MA, Bachu S (2005) Injection and storage of CO_2 in deep saline aquifers: Analytical solution for CO_2 plume evolution during injection. Transport in Porous Media 58(3):339–360

[114] Nordbotten JM, Celia MA, Bachu S, Dahle HK (2005) Semianalytical solution for CO_2 leakage through an abandoned well. Environmental Science & Technology 39(2):602–611

[115] Novak E, Ritter K (1996) High dimensional integration of smooth functions over cubes. Numerische Mathematik 75(1):79–97

[116] Obi EOI, Blunt MJ (2006) Streamline-based simulation of carbon dioxide storage in a North Sea aquifer. Water Resources Research 42(W03414)

[117] Oldenburg CM, Bryant SL, Nicot JP (2009) Certification framework based on effective trapping for geologic carbon sequestration. International Journal of Greenhouse Gas Control 3(4):444–457

[118] Oliveira DF, Reynolds AC (2014) An adaptive hierarchical multiscale algorithm for estimation of optimal well controls. SPE Journal 19(5):909–930

[119] Oliver DS, Reynolds AC, Liu N (2008) Inverse theory for petroleum reservoir characterization and history matching, vol 1. Cambridge University Press

[120] Onwunalu JE, Durlofsky L, et al (2011) A new well-pattern-optimization procedure for large-scale field development. SPE Journal 16(03):594–607

[121] Özay EK, Demiralp M (2012) Combined small scale high dimensional model representation. Journal of Mathematical Chemistry 50(7):2023–2042

[122] Panda S, Padhy NP (2008) Comparison of particle swarm optimization and genetic algorithm for facts-based controller design. Applied soft computing 8(4):1418–1427

[123] Parry ML (2007) Climate Change 2007: impacts, adaptation and vulnerability: contribution of Working Group II to the fourth assessment report of the Intergovernmental Panel on Climate Change, vol 4. Cambridge University Press

[124] Petvipusit KR (2011) Dynamic Well Scheduling and Well Type Optimization Using Ensemble-based Method (EnOpt). Msc Thesis at University of Oklahoma

[125] Petvipusit KR, Elsheikh AH, Laforce TC, King PR, Blunt MJ (2014) Robust optimisation of CO_2 sequestration strategies under geological uncertainty using adaptive sparse grid surrogates. Computational Geosciences 18(5):1–16, URL http://dx.doi.org/10.1007/s10596-014-9425-z

[126] Pflüger D (2010) Spatially adaptive sparse grids for high-dimensional problems. PhD thesis, Technische Universität Müunchen

[127] Phillips GM (2003) Interpolation and approximation by polynomials, vol 14. Springer

[128] Pohlheim H (Accessed: Oct 2014) Geatbx examples: Examples of objective functions (2005). URL http://www.geatbx.com/download/GEATbx_ObjFunExpl_v37.pdf

[129] Poli R (2008) Analysis of the publications on the applications of particle swarm optimisation. Journal of Artificial Evolution and Applications 2008, URL http://dx.doi.org/10.1155/2008/685175

[130] Poli R, Kennedy J, Blackwell T (2007) Particle swarm optimization. Swarm intelligence 1(1):33–57

[131] Psuade (Accessed: Jan 2012) Lawrence Livermore National Laboratry. URL https://computation.llnl.gov/casc/uncertainty_quantification/

[132] PUNQ-S3 (Accessed: Oct 2012) Department of earth science and engineering, Imperial College London. URL http://www3.imperial.ac.uk/earthscienceandengineering/research/perm/punq-s3model

[133] Qi R, Laforce TC, Blunt MJ (2009) Design of carbon dioxide storage in aquifers. International Journal of Greenhouse Gas Control 3(2):195–205

[134] Quarteroni A, Sacco R, Saleri F (2007) Numerical mathematics, vol 37. Springer

[135] Rabitz H, Alis O (1999) General foundations of high dimensional model representations. Journal of Mathematical Chemistry 25(2–3):197–233, DOI 10.1023/A:1019188517934, URL http://dx.doi.org/10.1023/A%3A1019188517934

[136] Rabitz H, Alis OF, Shorter J, Shim K (1999) Efficient input-output model representations. Computer Physics Communications 117(1–2):11–20, DOI http://dx.doi.org/10.1016/S0010-4655(98)00152-0, URL http://www.sciencedirect.com/science/article/pii/S0010465598001520

[137] RCP (Accessed: Nov 2014) RCP Database (version 2.0). URL http://tntcat.iiasa.ac.at:8787/RcpDb/dsd?Action=htmlpage&page=welcome#intro

[138] Rivlin T (1974) The Lebesgue constants for polynomial interpolation. In: Functional Analysis and its Applications, Springer, pp 422–437

[139] Rosenbauer RJ, Thomas B (2010) Developments and innovation in carbon dioxide (CO_2) capture and storage technology(Volume 2):Carbon dioxide sequestration in deep saline aquifers and formations. Woodhead Publishing Series in Energy No. 16

[140] Rosenthal RE (1985) Concepts, Theory, and Techniques Principles of Multiobjective Optimization. Decision Sciences 16(2):133–152

[141] Roy CJ, Oberkampf WL (2010) A complete framework for verification, validation, and uncertainty quantification in scientific computing. In: 48th AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition, 4-7 January, Orlando, USA

[142] Runge C (1901) Uber empirische Funktionen und die Interpolation zwischen äquidistanten Ordinaten. Zeitschrift für Mathematik und Physik (46):224–243, URL http://archive.org/details/zeitschriftfrma12runggoog

[143] Saddle-Surface (Accessed: Oct 2014) Saddle surface with normal planes in directions of principal curvatures. URL http://goo.gl/QZ1O7X

[144] Saltelli A (2002) Making best use of model evaluations to compute sensitivity indices. Computer Physics Communications 145(2):280–297

[145] Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, Saisana M, Tarantola S (2008) Global sensitivity analysis: the primer. John Wiley & Sons

[146] Saltelli A, Annoni P, Azzini I, Campolongo F, Ratto M, Tarantola S (2010) Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index. Computer Physics Communications 181(2):259–270

[147] Salzer HE (1972) Lagrangian interpolation at the Chebyshev points xn, $\nu \equiv \cos(\nu\pi/n)$, $\nu = 0\ (1)\ n$; some unnoted advantages. Computer Journal 15(2):156–159

[148] Schatz GC (1989) The analytical representation of electronic potential-energy surfaces. Reviews of Modern Physics 61(3):669–688

[149] Scheffe H (1999) The analysis of variance, vol 72. John Wiley & Sons

[150] Schlumberger (2010) Eclipse Technical Description Manual, CO_2 Storage

[151] Senel O, Chugunov N (2012) CO_2 Injection in a Saline Formation: How Do Additional Data Change Uncertainties in Our Reservoir Simulation Predictions? (CMTC 151454). Proceedings of Carbon Management Technology Conference, 7-9 February, Orlando, Florida, USA

[152] Shamshiri H, Jafarpour B (2012) Controlled CO_2 injection into heterogeneous geologic formations for improved solubility and residual trapping. Water Resources Research 48(W02530):1–15

[153] Shan S, Wang GG (2009) Survey of modeling and optimization strategies to solve high-dimensional design problems with computationally-expensive black-box functions. Structural and Multidisciplinary Optimization 41(2):219–241

[154] Shi L, Olafsson S (2008) Nested partitions method, theory and applications, vol 109. Springer

[155] Sifuentes W, Blunt MJ, Giddins M (2010) Modeling CO_2 Storage in Aquifers: Assessing the Key Contributors to Uncertainty. Journal of Petroleum Technology 62(7):71–72

[156] Smirnov NV (1944) Approximate laws of distribution of random variables from empirical data. Uspekhi Matematicheskikh Nauk pp 179–206, URL http://www.ams.org/mathscinet-getitem?mr=12387

[157] Smith RC (2013) Uncertainty Quantification: Theory, Implementation, and Applications, vol 12. SIAM

[158] Smith SJ (2006) Lebesgue constants in polynomial interpolation. In: Annales Mathematicae et Informaticae, Eszterházy Károly College, Institute of Mathematics and Computer Science, vol 33, pp 1787–5021

[159] Smolyak S (1963) Quadrature and interpolation formulas for tensor products of certain classes of functions. Soviet Math Dokl.4:240—-243

[160] Sobol' IM (1990) On sensitivity estimation for nonlinear mathematical models. Matematicheskoe Modelirovanie 2(1):112–118

[161] Sobol' IM (2003) Theorems and examples on high dimensional model representation. Reliability Engineering & System Safety 79(2):187 –193, DOI http://dx.doi.org/10.1016/S0951-8320(02)00229-6, URL http://www.sciencedirect.com/science/article/pii/S0951832002002296

[162] Sobol' IM, Kucherenko S (2010) A new derivative based importance criterion for groups of variables and its link with the global sensitivity indices. Computer Physics Communications 181(7):1212–1217

[163] Sobol' IM, Tarantola S, Gatelli D, Kucherenko S, Mauntz W (2007) Estimating the approximation error when fixing unessential factors in global sensitivity analysis. Reliability Engineering & System Safety 92(7):957–960

[164] Srinivas N, Deb K (1994) Muiltiobjective optimization using nondominated sorting in genetic algorithms. Evolutionary Computation 2(3):221–248

[165] Stocker TF, Qin D, Plattner GK, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (2013) Climate change 2013: The physical science basis. Intergovernmental Panel on Climate Change, Working Group I Contribution to the IPCC Fifth Assessment Report (AR5)

[166] Subbotin Y (1967) Piecewise-polynomial (spline) interpolation. Mathematical notes of the Academy of Sciences of the USSR 1(1):41–45, DOI 10.1007/BF01221723, URL http://dx.doi.org/10.1007/BF01221723

[167] Taleb NN (2010) The Black Swan: The Impact of the Highly Improbable Fragility. Random House LLC

[168] Tapia MGC, Coello CAC (2007) Applications of multi-objective evolutionary algorithms in economics and finance: A survey. IEEE Congress on Evolutionary Computation 7:532–539

[169] Tasmanian (Access: Oct 2014) The tasmanian project, Oak Ridge National Laboratories. URL http://tasmanian.ornl.gov/index.html

[170] Sparse grid Matlab Toolkit (Accessed: March 2012) URL http://www.ians.uni-stuttgart.de/spinterp

[171] Vapnik VN, Vapnik V (1998) Statistical learning theory, vol 2. Wiley New York

[172] Wang C, Li G, Reynolds AC (2009) Production Optimization in Closed-Loop Reservoir Management. SPE Journal 14(3):506–523

[173] Wayne G (2013) The beginner's guide to representative concentration pathways. URL http://www.skepticalscience.com/rcp.php

[174] Zandvliet M, Handels M, van Essen G, Brouwer R, Jansen JD (2008) Adjoint-Based Well-Placement Optimization Under Production Constraints. SPE Journal 13(4):26–28

[175] Zang C, Friswell M, Mottershead J (2005) A review of robust optimal design and its application in dynamics. Computers & Structures 83(4-5):315–326

[176] Zavala GR, Nebro AJ, Luna F, Coello Coello CA (2014) A survey of multi-objective metaheuristics applied to structural optimization. Structural and Multidisciplinary Optimization 49(4):537–558, DOI 10.1007/s00158-013-0996-4

[177] Zhang Z, Choi M, Karniadakis GE (2011) Anchor points matter in anova decomposition. In: Spectral and High Order Methods for Partial Differential Equations, Springer, pp 347–355

[178] Zhao D, Xue D (2011) A multi-surrogate approximation method for metamodeling. Engineering with Computers 27(2):139–153

[179] Zitzler E (1999) Evolutionary algorithms for multiobjective optimization: Methods and applications. Ph.D Thesis ETH 13398, Swiss Federal Institute of Technology Zurich