

Error aware analysis of multi-scale reactivity models for chemical surface reactions

**An Adaptive and a Multilevel Adaptive Sparse Grid approach to address global
uncertainty and sensitivity**

Vom Fachbereich Mathematik und Informatik der Freien Universität Berlin
zur Erlangung der Würde eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
genehmigte Dissertation

vorgelegt von

Sandra Döpking

Berlin 2021

Erstgutachter: Dr. S. Matera

Fritz Haber Institute, Max Planck Society, Berlin

Zweitgutachter: Prof. Dr. K. Reuter

Fritz Haber Institute, Max Planck Society, Berlin

Tag der Disputation: 28.01.2022

Teile dieser Dissertation wurden veröffentlicht:

„Addressing global uncertainty and sensitivity in first-principles based microkinetic models by an adaptive sparse grid approach “,

Sandra Döpking, Craig P. Plaisance, Daniel Strobusch, Karsten Reuter, Christoph Scheurer, and Sebastian Matera,

Journal of Chemical Physics, **148**, issue **3**, 034102

©2018 Author(s). Published by AIP Publishing (2018), <https://doi.org/10.1063/1.5004770>

„Multilevel Adaptive Sparse Grid Quadrature for Monte Carlo Models “,

Sandra Döpking and Sebastian Matera,

arXiv preprint arXiv:1810.00810, (2018) , <https://arxiv.org/abs/1810.00810>

In the field of heterogeneous catalysis, first-principle-based microkinetic modeling has been proven to be an essential tool to provide a deeper understanding of the microscopic interplay between reactions. It avoids the bias of being fitted to experimental data, which allows us to extract information about the materials' properties that cannot be drawn from experimental data. Unfortunately, the catalytic models draw information from electronic structure theory (e.g. Density Functional Theory) which contains a sizable error due to intrinsic approximations to make the computational costs feasible. Although the errors are commonly accepted and known, this work will analyse how significant the impact of these errors can be on the model outcome. We first explain how these errors are propagated into a model outcome, e.g., turnover-frequency (TOF), and how significant the outcome is impacted. Secondly, we quantify the propagation of single errors by a local and global sensitivity analysis, including a discussion of their dis-/advantages for a catalytic model.

The global approach requires the numerical quadrature of high dimensional integrals as the catalytic model often depends on multiple parameters. This, we tackle with a local and dimension-adaptive Sparse Grid (SG) approach. SGs have shown to be very useful for medium dimensional problems since their adaptivity feature allows for an accurate surrogate model with a modest number of points. Despite the models' high dimensionality, the outcome is mostly dominated by a fraction of the input parameter, which implies a high refinement in only a fraction of the dimensions (dimension-adaptive). Additionally, the kinetic data shows characteristics of sharp transitions between "non-active" and "active" areas, which need a higher order of refinement (local-adaptive). The efficiency of the adaptive SG is tested on different toy models and a realistic first principle model, including the Sensitivity Analysis. Results show that for catalytic models, a local derivative-based sensitivity analysis gives only limited information. However, the global approach can identify the important parameters and allows extracting information from more complex models in more detail.

The Sparse Grid approach is useful for reducing the total number of points, but what if evaluating the point itself is very expensive? The second part of this work concentrates on solving high dimensional integrals for models whose evaluations are costly due to, e.g. being only implicitly given by a Monte Carlo model. The evaluation contains an error due

to finite sampling. To lower the error, we would have to increase computational effort for a high number of samples. To tackle this problem, we extend the SG method with a multilevel approach to lower the cost. Unlike existing approaches, we will not use the telescoping sum but utilise the sparse grid's intrinsically given hierarchical structure. We assume that not all the SG points need the same accuracy but that we can double the points' variance and halve the drawn samples with every refinement step. We demonstrate the methodology on different toy models and a realistic kinetic Monte Carlo system for CO oxidation. Therefore, we compare the non- multilevel adaptive Sparse Grid (ASG) with the Multilevel Adaptive Sparse Grid (MLASG). Results show that with the multilevel extension we can save up to two orders of magnitude without challenging the accuracy of the surrogate model compared to a non-multilevel SG.

Contents

List of Symbols	1
List of Abbreviations	5
1 Introduction	9
2 Parameter Uncertainty and Sensitivity Analysis	15
2.1 Parameter Uncertainty and first-principles models	16
2.2 Uncertainty Quantification	22
2.2.1 Local Sensitivity Analysis (LSA)	25
2.2.2 Global Sensitivity Analysis (GSA)	27
3 Sparse Grid surrogate model	31
3.1 Problem formulation	31
3.2 Sparse Grids	34
3.2.1 One-dimensional hierarchical basis functions	34
3.2.2 High-dimensional Sparse Grid spaces	40
3.2.3 Sensitivity Indices	43
3.3 Adaptivity	45
3.3.1 Refinement Criteria	46
3.3.2 Dimension Adaptivity	47
3.3.3 Local Adaptivity	52
3.3.4 Combination: Dimension and Local Adaptivity	54
3.4 Testing refinement strategies	55
3.4.1 Spherical edge test function	58
3.4.2 Sigmoid function	61
3.4.3 Continuous function	64
3.4.4 Oscillatory function	67

3.4.5	Higher dimensional functions	68
4	<i>Oxygen Evolution Model</i>	75
4.1	Sparse Grid	83
4.2	Sensitivity Analysis	87
4.2.1	Total Sensitivity Index TSI	88
4.2.2	Higher Order Sensitivity Indices	92
4.3	Discussion	95
4.4	Conclusion	103
5	<i>Multilevel Refinement</i>	105
5.1	Parametric Monte Carlo Models	106
5.2	Multilevel idea	107
5.3	Sparse Grid	110
5.3.1	Adaptivity	110
5.4	Multilevel Adaptive Sparse Grid	111
5.5	An Illustrative model	115
5.5.1	2D-case	116
5.5.2	7D-case	120
5.5.3	7D-2D-case	122
5.6	Test examples	122
5.7	Parametric Monte Carlo Model: CO oxidation model	129
5.8	Global Sensitivity Analysis	135
5.9	Conclusion	138
6	<i>Conclusion and Outlook</i>	139
7	<i>Appendix</i>	173
8	<i>Deutsche Zusammenfassung (German Summary)</i>	187
9	<i>Eidesstattliche Erklärung (Declaration)</i>	189
10	<i>Acknowledgements</i>	191

List of Symbols

Greek

- φ basis function
- ϵ accuracy/threshold
- η overpotential
- γ local errors
- ν hierarchical surplus
- α one dimension for the Sensitivity Analysis
- Ω domain
- β position in the sites' chain (lattice model)
- ρ random number
- σ standard deviation

Latin

- \mathcal{A} ancestors, set of points
- V_L approximation space of level L
- c user defined constant for noise input
- \mathcal{C} children, set of points
- \mathcal{E} computational cost
- Covcovariance
- C^* upper bound for the variance of the functions

evaluations

D	dimension
$\Delta\Delta E$	energy error
ΔE	activation energy for process
\mathbf{e}	unit vector
E	expected value
f, g	functions
h	Planck constant
W	hierarchical difference spaces
i	count index
I	integral
\mathcal{I}	hierarchical index set
k	rate constant for process
k_B	Boltzman constant
K	equilibrium constant
l	level index
L	maximum level index
M	finite number of samples
\mathbb{N}	set of natural numbers
\mathbb{R}_0^+	set of non negative real numbers
\mathbb{R}	set of real numbers

- A matrix
- N sample size (spatial approximation)
- $n(l)$ number of the sparse grid point at level l
- s artificial noise
- \mathcal{O} Landau Symbol
- p probability density function
- P probability
- \mathcal{P} parents, set of points
- r spatial approximation level
- R highest spatial approximation level
- Y random variable
- S Sensitivity Index
- \mathcal{S} set of points
- t time
- T temperature
- TS transition state
- u approximation constructed by a sparse grid
- Varvariance
- $\mathbf{x}, \mathbf{i}, \mathbf{l}$ vectors
- w weight
- x input parameter

Z partition distribution

Symbols

$|\cdot|_\infty$ infinity Norm

$|\cdot|_1$ L_1 -Norm

$\overline{\cdot}$ statistical average

\oplus, \bigoplus tensor product

List of Abbreviations

TOF Turnover Frequency

TST Transition State Theory

hTST harmonic Transition State Theory

TS Transition State

PES Potential Energy Surface

DFT Density Functional Theory

kMC kinetic Monte Carlo

PDF probability density function

SA Sensitivity Analysis

LSA Local Sensitivity Analysis

GSA Global Sensitivity Analysis

DRC Degree of Rate Control

QoI Quantity of Interest

OAT one-factor-at-a-time

STD standard deviation

SI Sensitivity Index

TSI Total Sensitivity Index

MD Molecular Dynamics

UQ Uncertainty Quantification

- MC** Monte Carlo
- QMC** Quasi Monte Carlo
- ME** Master Equation
- SG** Sparse Grid
- ANOVA** Analysis of Variance
- FG** Full Grid
- BF** basis function
- GSG** Generalized Sparse Grid
- LA** local adaptivity
- DA** dimension adaptivity
- NoP** number of points
- OER** oxygen evolution reaction
- CFD** computational fluid dynamics
- ML** Multilevel
- MLMC** Multilevel Monte Carlo
- MLSG** Multilevel Sparse Grid
- SC** Stochastic Collocation
- MLQMC** Multilevel Quasi Monte Carlo
- PDE** partial differential equation
- ASG** adaptive Sparse Grid
- MLASG** Multilevel Adaptive Sparse Grid

MIMC Multi-Index Monte Carlo

MISC Multi-Index Stochastic Collocation

MLQ Multilevel Quadrature

FSG Full Sparse Grid

CPU Central Processing Unit

EC equilibrium constant

1 Introduction

Since the introduction of heterogeneous catalysis in 1835, the field has developed many catalytic processes for the industrial appliance. The Oswald process to oxidise ammonia or the Haber process to produce ammonia are a few of the famous examples. One application we can find in our everyday life is exhaust gas cleaning in cars. It purifies the gas by reducing the major toxins, e.g. carbon monoxide, nitrogen dioxide and hydrocarbons. Therefore, the three-way catalyst enables the toxins' conversion into a non-toxic product, e.g. water, carbon dioxide and nitrogen. Consisting of multiple reactive centres, the significant reactions always happen on the surface [1]. This is only one of the many applications, but it describes how a heterogeneous catalyst works. The system consists of a catalyst and reactants in two different phases, mostly a solid catalyst and a gaseous or liquid phase of reactants. Although the reaction is theoretically possible in a purely gaseous phase, the reaction has a very high energy barrier that needs to be overcome so that the reaction happens on a very long unfeasible timescale. To increase the conversion (number of reactions per time), the catalyst enables or supports the reaction by lowering the energy barrier through weakening bonds between atoms or increasing the local concentration of the educts. The catalyst itself does not take place in the reaction and therefore stays unchanged. This is a significant advantage over homogeneous catalysts. The separation of the catalyst and the products is easy in a heterogeneous setting, as well as the processing and cleaning. On the contrary to this, the surface area of a solid catalyst is limited, which gives a natural limit to the conversion. Many approaches have been developed to increase efficiency, like modifying the catalyst surface to provide more active centres. Others rely on a molecular understanding of the catalytic system or selectivity of all available reaction paths [2, 3].

Historically, these catalytic processes have been developed primarily by trial and error. Fundamental studies directed at a basic understanding of single processes or the chemistry were added later on, e.g. by Langmuir or Hinshelwood [4]. They added fundamental kinetics and thermodynamic concepts of surface reactions to provide a molecular-level explanation of how they function. This also includes detecting key reactants that are only active at a rarely formed construction of the catalyst or further understanding specific reactions, like dissociative adsorption. This evolved in the 20th century toward better characterisation of the catalyst by experimental or theoretical methods or the improved study of reaction pathways leading to more effective and more complex catalysts or catalyst systems [5, 6]. Nowadays, the requirements of catalysts grow more complex, the challenges increase, and effectiveness also includes the aim for minimal waste and an environmentally friendly solution. To tackle these

challenges, we need a more extensive knowledge of molecular details on an experimental and theoretical base [2].

Some of today's innovations in the field of heterogeneous catalysis aim to increase the efficiency of the chemical reaction network. The problem with this is that the conversion on the macroscopic scale depends not only on the outer parameters like pressure and temperature but also on the microscopic processes happening on the surface. Optimising the external parameters reaches a limit of improvement when lacking understanding on the microscopic level.

Therefore chemical kinetic models have been developed to provide insights into the experimental behaviour and address questions that cannot be answered solely by experimental data [7–9]. Kinetic models map the macroscopic physical and chemical state to rates of formation on a mesoscopic scale. In general, there are two types of kinetic models. First, the phenomenological model captures the essential features of observed rates and does not provide chemical details. Second, the models are based on elementary steps, like microkinetic models. They are aligned with the underlying elementary chemical steps. Elementary step kinetic models can be further distinguished between top-down and bottom-up approaches that either rely on experimental data to obtain kinetic parameters and molecular insights or rely on molecular-level information to construct kinetic parameters used for simulations to match their prediction to experimental observations. Nowadays, models are often constructed by a combined approach of bottom-up and top-down [9, 10]. The knowledge derives from an interplay of theory, prior chemical knowledge, experimental observation and modelling [11] In this work, we will concentrate on the bottom-up models, but all approaches have gained usage caused by the expanded capabilities, and predictive accuracy of them [9, 12–17]. Their ability to account for individual catalytic sites' structure and composition effects produce a qualitative or even semiquantitative agreement with experimental observations. Unfortunately a lot of these models also carry uncertainties. The model inaccuracies emerge from two types of errors, either a deficiency of the model itself or of the underlying physics or computations, which we will elaborate on in this work. These approaches use energetic information on adsorption energies and reaction barriers derived from predictive quality electronic structure theories to simulate elementary reaction mechanisms. Within this framework, these models avoid relying on resource-intensive experiments and give mechanistic insights without the bias of fitting to experimental data. [18–22]

Besides all the advantages of first-principle models, the used energetic information is inherently inaccurate. Electronic structure theories computationally rely on intrinsic approximations to make the computational cost feasible. Consequently, the energetics used by the model contain errors that affect the model outcome. For models of heterogeneous catalysis,

the energetics are mostly provided by Density Functional Theory (DFT) in the generalised gradient approximation since the data is most accurate at a feasible computational effort. While this error typically is around 10% for reaction barriers and adsorption energies, it can have a tremendous impact on the outcome of the kinetic model leading to uncertainties of orders of magnitude for the activity at typical operation temperatures [23, 24]. Nevertheless, the error of a single energy are often figured with $\approx 0.2\text{eV}$ [7]. While this size of the error is commonly agreed on, the discussion about the impact of such an error is limited [24–27]. We shortly display the impact of the error for a known model, the CO oxidation on RuO₂(110) [28–30]. Our example shows that an energetic error of $\approx 0.2\text{eV}$ causes the models' input parameter to be wrong by a factor of around 2 000. The uncertainty can be fully propagated to the model outcome, e.g. the Turnover Frequency (TOF), which leaves us with the question of how much information we can actually draw from these numbers.

Furthermore, we want to address how we could decrease the overall uncertainty in our model outcome. Therefore, we need to know which of our model's input parameter errors has the highest contribution to the outcome's total variance. Assuming that only a fraction of all input parameters is important, it would allow for a selective refinement with, e.g. employing a more accurate, higher-level electronic structure theory [31, 32] for these parameters without producing unfeasible costs. Information like this enables identifying critical factors for the materials' performance and further insights into microscopic catalytic performance.

Different methods of Sensitivity Analysis (SA) address the selective impact of single input parameters on the model outcome. The most common approach is the derivative-based Local Sensitivity Analysis (LSA). It analyses the model response derivatives with respect to the input parameters [33–35]. Although the LSA is a very intuitive and low-cost approach, it often identifies incorrect importances quantitatively and qualitatively. Its underlying assumption of linear behaviour does not apply to the kinetic data that we are working with. Kinetic data often exhibits rapid jumps between non-active and active areas, which are highly non-linear. A more promising approach is a Global Sensitivity Analysis (GSA), based on a Analysis of Variance (ANOVA) decomposition [36, 37]. Approaches like these, and also anchored ANOVA or multibody-expansion [38] try to represent the model's response as a sum of minimal low-dimensional functions. The method intrinsically implies that only the important dimensions are included in the representation. It has the advantage to take the uncertainty domains into account so that the results can be interpreted as an induced uncertainty. Such non-derivative based methods are also applicable to models for which parameter derivatives can only be estimated at very high computational costs, e.g. by a kinetic Monte Carlo (kMC) simulation [39].

For the GSA, we need to solve high dimensional integrals. Although many methods, like

Monte Carlo (MC) or Quasi Monte Carlo (QMC) have proven to be effective in high dimensions, the GSA includes multiple functional approximations. Each Sensitivity Index has to be approximated separately, which is only computationally feasible if the number of MC points are limited [36]. Such MC and QMC methods scale unfavourably with the number of input parameters so that they become quickly computationally unfeasible, especially if singular evaluations are already costly. Therefore, we use an adaptive SG approach that constructs a surrogate model whose evaluations are cheaper than the model evaluations. SGs show a high convergence rate for medium dimensional cases. Although kinetic models are high dimensional, the dependence on only a fraction of the parameter implies an intrinsic lower-dimensional problem. The classical Full Sparse Grid (FSG) approach for a high dimensional problem would not suffer from the *curse of dimensionality*. However, it would produce a lot of unnecessary points, which we can avoid with adaptive refinement. In our case, we know the typical behaviour of kinetic data. In the logarithmic settings, the kinetic data mostly behaves linearly with localised areas of rapid jumps between "non-active" and "active" regimes. Both regimes can be active by definition, but the jump indicates a high difference in conversion rates.

We use a combined technique of dimension adaptivity to avoid unimportant refinement of parameter domains and local adaptivity for the kinetic data's local non-linear behaviour. Based on the piecewise-linear basis function, we can approximate the areas of linear behaviour with a minimal amount of points and produce only a high density of points in important dimensions and areas of non-linearity. Jakeman and Roberts first introduced the combined method we use in this work [40]. We modify it to allow for larger numbers of parallel kinetic model evaluations.

Being based on an expansion of piecewise linear basis function, the SG constructs a surrogate model we can use for the GSA method to analyse a Sensitivity Index (SI), such as the Total Sensitivity Index (TSI) or the second-order sensitivity indices. While the TSI gives us an overall impact of one specific parameter, the second-order sensitivity indices give us an idea of which parameters are interacting toward the model outcome. With these measurements, we can achieve a very detailed picture of the microkinetic behaviour.

In this work, we will use both kinds of Sensitivity Analysis', LSA and GSA, for a first-principles based model, the Oxygen evolution on Co_3O_4 [32] to show their difference and limitations. The results underline that both methods can identify the most important parameters. Still, the LSA tends to underestimate the amount of non-trivial parameters and overestimate the impact of these parameters on the outcome. The GSA, on the other hand, can identify all important parameters for the whole uncertainty domain. We perform the SA at different regimes of the model to provide an accurate picture of the results at non-active areas and

active areas. In the end, we want to give a theoretical idea of the effect a prior [GSA](#) can have on the accuracy of the model outcome.

In the second part of this work, we will tackle the problem when the single reduction of points for the [SG](#) is still too little and constructing a surrogate model comes with an increased computational cost. Like in many applications of uncertainty quantification, finance or molecular simulations, the integrands for the integrals are only implicitly given and a significant source of computational costs. In our kinetic modelling case, the integrands are provided by a [kMC](#) model, whose function evaluation comes with an additional error caused by the finite sampling. To lower the evaluation error, we would have to increase the number of samples with an increased computational effort. Although the [SG](#) reduces the number of possible expensive function evaluations, it does not provide a measure of how accurate the function evaluations have to be. This leads to the question of how much effort should be spent on every data point and whether we should spend the same effort for every point.

A popular method to tackle this problem is a Multilevel ([ML](#)) approach which balances the simulation with the quadrature error [41]. The idea consists of utilising a hierarchy of numerical approximations of the underlying model and construct the finest approximation just by the sum of the lowest level approximation and the differences between consecutive levels. The general findings show that most simulations are performed on the lowest approximation. Only decreasing numbers of simulations are performed for increasing approximation accuracy, which lowers the overall costs. Two common methods are the Multilevel Monte Carlo ([MLMC](#)) and the Multilevel Quasi Monte Carlo ([MLQMC](#)). The latter is especially often used in the field of uncertainty quantification, e.g. for the treatment of partial differential equations with random input [42–44]. The multilevel idea was also adapted for stochastic collocation with [SG](#) [45, 46]. In recent years, the classical [ML](#) idea was extended with a Multi-Index idea that uses different spatial approximations for different dimensions, allowing for only using the most effective mixed differences. Examples for MIMC and MISC can be found in references [47, 48]. Most of this work wants to approximate expected values. In the case of uncertainty quantification, this means solving integrals. In our case, we want to apply [GSA](#) to our kinetic model that does not include expected values but not linear functionals. That is why we want to construct a surrogate model with a multilevel approach.

In 2001 Heinrich addressed a [ML](#) Monte Carlo method to construct an interpolation of a parametric Monte Carlo model [49]. He tackles the question, whether [MC](#) is still an option if we want to approximate whole functions. The function evaluations are approximated by a [MC](#) algorithm. He shows that one spatial approximation of the function domain gives a computational cost of $\mathcal{O}(nM)$, where n describes the density of the spatial grid and M the number of Monte Carlo samples. With extending the one spatial grid to multiple spatial grids

of different resolutions, he can adapt the MC samples for every grid and lower the number of the total MC samples M ; thus, the computational cost only scales with M , $\mathcal{O}(M)$. Unfortunately for high dimensions, like for kinetic data, this can become very costly, especially if not all dimensions require the same resolution of the grid. For the lower intrinsic dimensionality of the kinetic data, we need a method to lower the overall points for the spatial approximation and balance the computational effort for each grid point.

To tackle this problem, we introduce a Multilevel Adaptive Sparse Grid (MLASG) method that utilises the intrinsic multilevel structure of the SG to determine the sampling effort of the kMC model. Therefore we will shortly explain the classical ML approach combined with SGs and that with a stochastic sampling error, the classical method will only produce random refinement for the SG. Then we will continue with the explanation on how to utilize the intrinsic multilevel SG structure. Analog to the SG construction before, we use the piecewise linear basis function. Still, we only use local adaptivity to predefine the variance of single points and find that we can halve the sampling effort with every refinement step without challenging the surrogate model's accuracy. For the kinetic model, we can save up to a factor of 200 of the computational costs with the multilevel extension of the model. Although other approaches can be efficient for approximating the high-dimensional integrals, the surrogate model of the MLASG can be used for further analysis, such as GSA.

2 *Parameter Uncertainty and Sensitivity Analysis*

Parameter uncertainty has appeared in many different fields over time. One of the early ones, the papers of Heckman and Meyers, Meyers and Schenker, contains the field of risk management [50, 51] and both papers dealing with aggregated losses. While they artificially add uncertainty to parameters to get a better understanding of the model, nowadays, a lot of scientific computational models suffer from uncertain input parameters [52].

Problems in chemistry or physics are highly complex and can hardly be formalised without approximations, or the use of experimental data [23]. This simplifies the problem in such a way that a model performs with a reasonable computational effort. However, within this simplification lies the source of the parameter uncertainty. These uncertainties can have different reasons, whether the model uses data for parameters of varying models or the model's data is limited or includes an approximation error of the underlying high-fidelity model. Most approximations imply an error propagated to the model outcome; thus, most models in engineering and science are subject to parameter uncertainty. Unfortunately, most of the work on these models neglect such uncertainty, although its impact on the model results is not negligible [26, 27]. How much these uncertainties play a role depends mainly on the problem settings. For instance, the simulation of a laminar flow of water carries only a small error due to the uncertainties in the transport and thermochemical coefficients because these have experimentally been determined with high accuracy. In contrast, chemical kinetic models which have been parametrised with quantum chemical methods carry significant parameter uncertainties due to the inherent approximations in quantum-chemical methods. In this case, propagated input parameter uncertainties can cause variances of two or more orders of magnitude in the model output[23, 24].

We want to examine the problem of uncertainty propagation in first-principles quantum chemistry-based models (in the future just termed first principles-model) and show a method to identify the input parameters with the highest impact on the model outcome. First-principles models are often utilised for modelling surface reactivity by using energetic information from electronic structure calculations. These models provide detailed insights on catalytic behaviour without using experimental data. Unfortunately, the electronic structure theory calculations rely on intrinsic approximation to perform at feasible computational costs. Therefore, the energies contain errors that are propagated into the model outcome. In the long run, the approach of identifying the important parameters has the potential to reduce the overall

computational effort and to provide more insights into the catalytic mechanism. An initial identification enables to spend more computational effort on fewer parameters for higher accurate data, and problems like the rate-determining step can be tackled [23, 24].

This chapter first demonstrates the uncertainty propagation in the case of first-principles models and its impact on the model outcome to a point where the information of the model is minimal. Secondly, two of the most common ways to quantify the propagation of particular parameter uncertainties are introduced, the Local Sensitivity Analysis (LSA) and variance-based Global Sensitivity Analysis (GSA), with a discussion about their use for first-principles models.

2.1 Parameter Uncertainty and first-principles models

In the field of heterogeneous catalysis, first-principles based kinetic models have received increasing interest for modelling surface reactivity [7]. These approaches use energetic information on adsorption energies and reaction barriers derived from predictive-quality electronic structure theories to simulate elementary reaction mechanisms. Since these models do not rely on experimental data, we can predict material's properties, like the Turnover Frequency (TOF), a measure for the efficiency of a catalyst, or coverages of the surface and allows the addressing of questions that can not be answered just by experimental data.[18–22]. The predictive quality of these approaches and the provided further understanding of surface chemistry cause their high popularity.

Besides all the advantages of first-principles models, the used energetic information is inherently inaccurate. Electronic structure theories rely on intrinsic approximations to achieve feasible costs. The resulting energetics for Density Functional Theory (DFT) in the generalized gradient approximation typically contain a large approximation error of around 0.2 eV or more [24–27]. To underline the impact of the energy errors on the first-principles model outcome, we will briefly explain how a first-principles kinetic Monte Carlo (kMC) [53] model uses the DFT-binding-energies. Therefore we will briefly elaborate on defining rate constants based on Transition State Theory (TST).

To model the chemical kinetics on a mesoscopic scale, we have to coarse grain the continuous molecular motion and define chemical reactions that describe the transition of one metastable state of the system to another. Considering a heterogeneous catalytic model, we define fluid reactants, whose reactions include adsorption/desorption, diffusion of the catalytic surface and the reaction as shown in figure 2.1. Such a state \mathbf{i} of figure 2.2 describes a metastable state of the system, which is the currently occupied local minimum of the Potential Energy

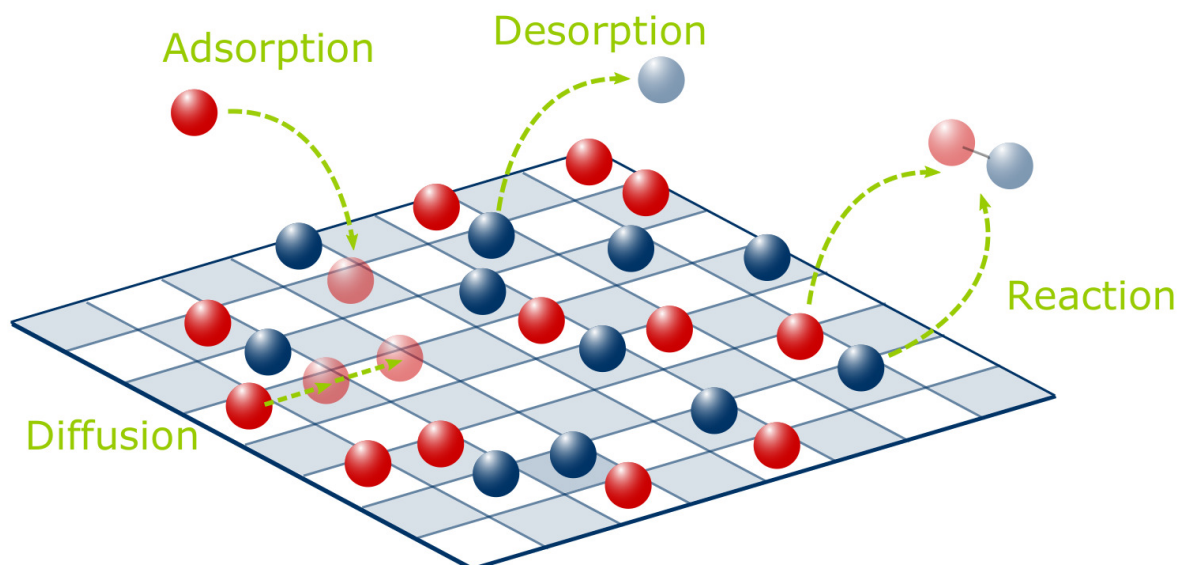


Figure 2.1: Description of the four considered kinds of elementary reactions : I) Ad-/Desorption of reactants (blue and red balls) on the surface II) Diffusion of a reactant over the surface and III) Reaction of two reactants, including a desorption of the surface.

Surface (PES). We want the sequence of chemical reactions for modelling, so the evolution of meso-states visited by the system. Next to the state-changing reactions, we also have vibrational motion in the basin around every minimum, which does not change the system's state. Every trajectory is signed to a certain meso-state and at time t and will leave the assigned area at a different time. These details are not available in a coarse-grained description, so that we can not define a sequence of meso-states where we know which meso-state followed the last one. Instead, we need to decide which information are needed to describe the evolution of the mesoscopic state with a limited number of simulations on the microscopic scale. In our case, the reactions steps are rare events compared to the underlying motions that do not change the system. The vibrational motion happens on a shorter timescale than the elementary reaction caused by the higher energy that needs to be overcome to leave the meso-state. This means there are many oscillations before the next rare event happens. The elementary reactions happen on a shorter timescale compared to the timescale of an unchanged meso-state. The time between consecutive events can be magnitudes longer [13]. Hence, the system forgets the former meso-state, and the subsequent transition is only based on the current state. Based on that, the evolution and the transition of one meso-state to another can be described as a Markov process [54]. Since most catalytic reaction models do not adsorb atoms at random positions, but at certain adsorption sites, we examine a lattice structure of these sites on the surface, as it is shown in figure 2.1. This gives us a vector of n sites, defined as a state

$\mathbf{i} = (i_1, i_2, \dots, i_n)$ with different configurations i_n , whether a species is adsorbed or the site is empty. The main idea of modeling surface reactions is to describe the evolution of the states visited by the system, hence, a sequence of elementary processes, that changes the configuration of the surface.

Since the detailed information about which process is happening next or at what point of time t it is happening is not available, only the probability of finding the surface in state \mathbf{i} at time t is given by $P(\mathbf{i}, t)$. Based on the fact that we are dealing with Markov processes, the probability obeys the Master Equation (ME) [30, 55].

The kMC algorithms provides a numerical solution to the chemical ME

$$\frac{\partial P(\mathbf{i}, t)}{\partial t} = \sum_{\mathbf{j}} k_{\mathbf{ij}} \cdot P(\mathbf{j}, t) - \sum_{\mathbf{j}} k_{\mathbf{ji}} \cdot P(\mathbf{i}, t) \quad (2.1)$$

that describes the dynamic system evolution by averaging over the shorttime dynamics and and only considering the rare events, which allows for times scales longer than seconds. $k_{\mathbf{ij}}$ are the rate constants for every possible transition out of state \mathbf{i} and $k_{\mathbf{ji}}$ for the reversed reaction.

The multiple numbers of states cause the ME to be too high-dimensional to solve it analytically. Therefore kMC-methods provide a numerical solution by generating an ensemble of trajectories. It is based on the fact that the ME describes the dynamics of jump processes, and the expected values are obtained by statistical averaging over these trajectories. Each trajectory propagates the system from state to state so that with averaging over the entire ensemble yields probability functions $P(\mathbf{i}, t)$ for all states \mathbf{i} . The sequence of transitions, precisely the selection of processes, and the transition times are randomly selected based on the transition probability, which is contained in the rate constants $k_{\mathbf{ji}}$.

In detail, there are different kMC methods to tackle different problems, e.g. off-lattice vs on-lattice problems or null event algorithms vs rejection-free algorithms. A good overview of the different kMC methods is given by Chatterjee and Vlachos [56]. Overall the kMC algorithm provides a sequence of time steps when the configuration is updated. In the case of a null event algorithm, it allows for events where the configuration stays unchanged, making the connection of the time-step to real-time complicated. On the other hand, we work with a rejection-free algorithm that adjusts the time step so that one reaction happens at each time step. In the following, we will shortly recap the algorithm.

It starts with a configuration of the catalytic surface \mathbf{i} . Then every possible elementary process to leave the state \mathbf{i} is considered and normalized by determining the total rate $k_{tot} = \sum_{\mathbf{j}=1} k_{\mathbf{ij}}$ of all possible processes. In practice the kMC uses a random number $\rho_1 \in [0, 1]$ to select the

process of all R processes with

$$\sum_{j=1}^{r-1} k_j \leq \rho_1 \cdot k_{tot} \leq \sum_{j=r}^R k_j \quad (2.2)$$

with a timestep $\Delta t = -\frac{\ln(\rho_2)}{k_{tot}}$.

So the crucial requirement is the definition of a set of possible processes and their rate constants. This means for every atom or molecule interacting with the surface the rate constants for every single reaction step have to be calculated.

Some reaction rate theory methods use only the underlying potential energy surface properties to derive the rate constant. It only requires static total energy calculations and avoids dynamical simulations.

As previous work shows, the TST with its harmonic approximation is sufficient to describe highly active catalytic surface processes [13, 57, 58]. The basic idea behind the TST is to determine the rate constants by examining the interactions of atoms, given by the PES. A reaction path, as it is described in figure 2.2, from an initial state \mathbf{i} , to a final state \mathbf{j} is characterised by a transition state, which is often a saddle point of the PES. It separates the two minima of the \mathbf{i} and the \mathbf{j} and defines a barrier that has to be overcome for the reaction path. Further details on the TST and the harmonic Transition State Theory (hTST) can be found in [59]. The important feature of the hTST is the reduction of information to the activation energy that defines the difference between the minimum of state \mathbf{i} and the transition state, which requires relatively few evaluations of the PES. In the hTST theory rate constants are generally defined as

$$k = \frac{k_B T}{h} \frac{Z_{TS}}{Z_i} \exp\left(\frac{\Delta E_{process}}{k_B T}\right) \quad (2.3)$$

where ΔE is the potential energy difference between state \mathbf{i} and the TS. Furthermore, Z_{TS} and Z_i are the harmonic vibrational partition function in the TS and the initial state \mathbf{i} , k_B the Boltzmann constant and T the surface temperature. Although in detail rate constants equations for different processes differ in their dependencies [57], all rate constants have the same general dependency on the activation barrier and follow the form of

$$k_{process} \propto \exp\left(\frac{\Delta E_{process}}{k_B T}\right), \quad (2.4)$$

which resembles the form of the Arrhenius equation definition $k = A \exp\left(\frac{\Delta E_{process}}{k_B T}\right)$ (with A as a preexponential factor), but uses first-principles instead of experimental data to determine

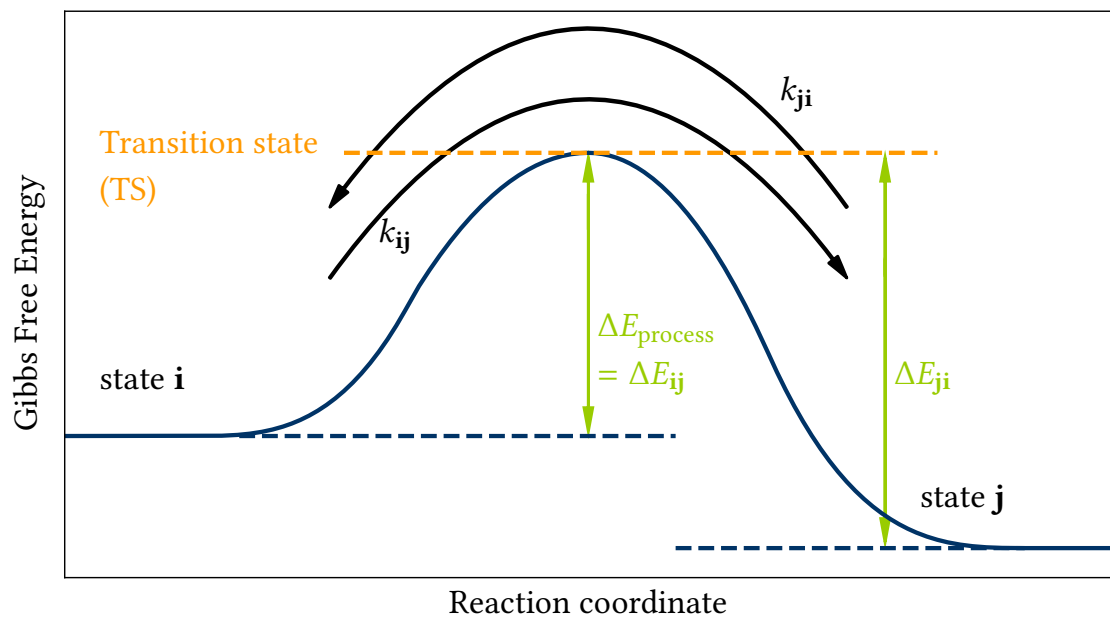


Figure 2.2: Energy diagram of a reaction from state **i** to state **j**, including the description of the TS as the saddlepoint of the PES. The system has to overcome the energy barrier ΔE_{ij} for the forward reaction and ΔE_{ji} for the backward reaction

T \ $\Delta\Delta E$	0.1 eV	0.2 eV	0.3 eV
293 K	52.5	2755.0	144605.3
400 K	18.2	331.0	6023.3
600 K	6.9	47.9	331.0
800 K	4.3	18.2	77.6

Table 2.1: Factor k_{error} , caused by the energy error $\Delta\Delta E$ $k \propto e^{\left(\frac{\Delta E_{\text{process}} \pm \Delta\Delta E}{k_B T}\right)}$ resulting a parameter variation of $k = \left[\frac{k_0}{k_{\text{error}}}, k_0 \cdot k_{\text{error}}\right]$ with $k_{\text{error}} = e^{\left(\frac{\Delta\Delta E}{k_B T}\right)}$. The effect of the error uncertainties lowers with increasing temperature, so high temperature models are less influenced by the effect. With most models performing between 500 K-600 K the effect on each rate constant is at least one order of magnitude difference.

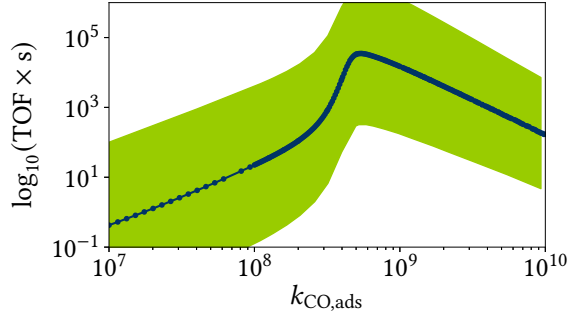


Figure 2.3: Model output of the CO oxidation on RuO₂ model for $T = 298.15$ K. The conversion, measured as $\log_{10}(\text{TOF} \times \text{s})$, is plotted over the CO adsorption rate constant $k_{\text{CO,ads}}$. For nominal values of all input parameters the conversion is displayed as the blue dots. The green shaded area describes the standard deviation of $\log_{10}(\text{TOF} \times \text{s})$ by assuming uniformly distributed DFT-errors in range of $[-0.2, 0.2]$ eV.

the rate constants. For the last step to calculate the rate constants the activation energies $\Delta E_{\text{process}}$ are required. As mentioned before, a common method in the electronic structure theory community is the DFT to calculate these energies. Former studies have shown that the underlying approximations cause an error of 0.2 eV-0.3 eV [6, 60] in the energies, with absolute values of 0.1 eV-5 eV, depending on the functional and the chemical substance [57, 61]. By propagating these errors to the rate constants, depending on the temperature, we can observe a factor of up to 6 orders of magnitude in variation in the rate constants, as displayed in table 2.1. The effect of the error propagation is decreased with rising temperature, but with most models performing at 500 K-600 K the factor is at least one order of magnitude. Nevertheless, these variations have an even higher effect on the model outcome. An increase in the complexity of the model often implies an increasing amount of erroneous input parameters. In that context, the model uncertainty can reach up to 5 orders of magnitude even for high temperatures, so we are faced with the question: How reliable are these results?

Figure 2.3 shows schematically the impact of an error = 0.2 eV on the result of a simplified kMC for the CO oxidation on RuO₂ surface. [29, 30, 55]. The model has 7 input parameters k , all containing an energy error $\Delta\Delta E = 0.2$ eV. The blue dots indicate the logarithm of the TOF at nominal values of k , plotted over one input parameter $k_{\text{CO,ads}}$ and the green shaded area the standard deviation of $\log_{10}(\text{TOF})$ caused by the DFT error calculated with 1e5 Monte Carlo points. It shows that the nominal value of the TOF varies by multiple orders of magnitude

making it difficult to extract information.

In the following chapters, we will discuss how we can efficiently identify which uncertainty input parameter has the most significant impact on the outcome and how we can extract the most information from the error-prone data regarding mechanical insights or rate-determining steps.

2.2 Uncertainty Quantification

The topic of Uncertainty Quantification (UQ) deals with the translation of the uncertainty of input parameters into the corresponding uncertainty in model outputs and to analyse the relative importance of sources of uncertainty. To tackle this problem, we use a Sensitivity Analysis (SA) to quantify the propagation of the uncertain input parameter into the quantity of interest (TOF, coverages). By doing so, we are able to extract qualitative information about the reaction network, like identifying important elementary reactions in a complex kinetic model that provides the potential for a reduced model with equivalent predictive power [29, 30, 62]. The objectives of an SA can be numerous. It provides a deeper understanding of the model with respect to the input parameter, which corresponds to the elementary steps of the catalytic system. The SA gives us the rate determine step/s that is responsible for the overall conversion. A step further are relations between elementary steps, given by higher-order SA, which provides a picture of a reaction network. On the other hand, it also provides a good characterisation of input parameters that can be neglected. This could mean an enormous dimension reduction for complex chemical systems, providing additional information and reducing the computational effort. The model can be reconstructed with only a fraction of the dimension. This becomes very important if single input parameters are very expensive like DFT-calculations. A SA enables the possibility to start with a model that has insufficient input data and analyse the dependencies. With the reduced amount of dimensions, we can increase the accuracy of the input data, which are important for the model. This will keep the computational effort as low as possible.

There are multiple methods for a SA, and the most common general topics are the LSA and GSA. We will introduce both methods in detail later on, and we will explain why the derivative-based LSA is no alternative for the variance-based GSA we use. Compared to the LSA, which only considers small input perturbations around the nominal value, the GSA considers the whole variations of the inputs [63]. In general, there are many different GSA approaches depending on the underlying problem. The application range from analysing

the input of genetic parameters for a crop model prediction [64, 65] to the most significant input for an aircraft infrared simulation model [65, 66]. So over time, multiple methods, like heuristic, graphic tool or Monte Carlo techniques, have been developed [67–71]. Most of them were developed for models with independent input parameters, but often they were extended with dependency between components [72–79].

One of the topics is screening methods which are qualitative methods for studying sensitivities. These are often applied to models with a high amount of input parameters because the discretisation of inputs in levels allows for fast exploration of code behaviour. The most substantial effect of these methods is the minimal amount of points used for non-influential inputs. This makes the method suitable for an early stage of the SA to exclude dimensions before applying a more costly method. In detail, there are many methods to achieve this exploration, but the most common one is the Morris method, where initial model evaluations are randomly selected over the whole domain [80]. From these nominal points, one grid jump at a time is performed, with a size according to the refinement level and the dimension randomly selected. The change of the output is then analysed to define a rank of input parameters. A second widespread approach is the importance measure. These can be divided into a linear models approach and a functional decomposition of variance models. The linear approach comes in handy when the size of samples is limited so that the evaluations are fitted to a linear model that can then be analysed. In this case, the sensitivity measures vary from a Standard Regression Coefficient to a Pearson correlation coefficient [65]. If the model is non-linear and non-monotonic, we can use the functional decomposition of variance, which is the idea we use in this work and will be explained in detail later on. This method is based on representing the correlation between the input parameter and the output as a sum of elementary functions [81]. Under the assumption that variables are mutually independent, the functional decomposition of the variance is available so that singular contributions of the specific input parameter to the variance can be determined.

These are only two examples, but in [65] Iooss and Vlachos give a good overview of the standard methods, including deep methods or methods for multiple outputs [82–85]. Furthermore, Rocquigny et al. provide a good intention when to use which method in [68].

The job of the SA is to characterise the relation between the uncertain input parameter and the model output and to analyse which input parameter affects the model outcome the most [86]. This information can be used to decrease the Quantity of Interest (QoI)-variance by decreasing the uncertainty of the single input parameter, e.g. in the case of DFT-calculations by using a higher-level electronic structure theory or turning to experimental data. Furthermore, the input parameters have an atomistic interpretation in the microscopic model, whose importance classification gives us a detailed insight into a reaction network and a possibility

to identify descriptors in material's screening [23].

Before we can apply any SA, we start with a priori generic input-output relation, that can either be an explicit function y or a relation only implicitly given by a computational model for more complex problems. In both cases the general relation is similar and we can write a model response for the deterministic model as $f : \mathbb{R}^D \rightarrow \mathbb{R}$

$$y = f(\mathbf{x}) = f(x_1, x_2, \dots, x_D), \quad (2.5)$$

where y represents the output and $\mathbf{x} = (x_1, \dots, x_D)$ the input with D parameters, each containing an individual uncertainty Δx_i .

LSA methods, mostly derivative based, are the straightforward approach of the sensitivity concepts. Their inexpensive computational costs and direct chemical interpretation of the impact of one parameter x_i makes it favorable for a lot of engineering models. Unfortunately such methods are only effective if the model is linear and the parameter uncertainties are rather small. That is why variance based approaches can be more effective.

In such cases of the SA, the methods require a model for the parameter variation to define whether initial relations between parameters are known or singular parameter are more likely than others, i.e. a joint probability distribution on the D - dimensional parameter domain [23]. This is usually given in terms of a probability density function (PDF) $p : \mathbb{R}^D \rightarrow \mathbb{R}_0^+$,

$$p(x_1, \dots, x_D). \quad (2.6)$$

The information provided by the PDF is critical since it can significantly influence the SA. Although in practice, uncertain parameters are often correlated, the common approach is to assume uncorrelated parameters. On the one hand, most of the time, the correlation is not previously known or not easy to formulate and on the other hand, a large set of SA methods do not work with correlated parameters, e.g. Sobol's method [87], FAST [88], or methods based on the correlation ratio [89].

In the case of experimental data, the PDF p might represent statistical errors in the parameters that are originated from the experiment. If the data is obtained from independent experiments, the PDF is often taken as a product of Gaussian distributions with means and standard deviations fitted to experimental findings [90]. The errors in electronic structure theory, e.g. DFT, behave systematically and not statistically. Thus, a repetition of the calculation implies no improvement on the parameter estimates [23].

For our case, the PDF has to reflect the information about the parameters and possible outcomes. Since only the individual errors $\Delta \mathbf{x}$ of the input parameters are known but no further information about initial relation or the probability distribution, we have to assume the

worst case scenario. Based on the limited knowledge we define the parameters as uniformly distributed over the hypercube $\Omega = \mathbf{x}_{nom} + [-\Delta\mathbf{x}, +\Delta\mathbf{x}]$, centered at the nominal parameters setting value \mathbf{x}_{nom} ,

$$p(\mathbf{x}) = \begin{cases} \frac{1}{(2\Delta\mathbf{x})} & \text{if } \mathbf{x} \in \Omega \\ 0 & \text{else.} \end{cases} \quad (2.7)$$

Among all PDFs which obey the bound constraints, this PDF maximises the entropy and minimises the information content. As mentioned above, also in the case of DFT-calculation, errors of different material classes or underlying correlation, e.g. Brønsted-Evans-Polanyi correlation, can change the shape of the PDF [6]. Unfortunately, without further information, we can not employ anything else than the general bounds.

Since we want to employ SA methods, that are based on the variance and variance decomposition, we need to define the expected value $E(y)$ and the parameter uncertainty induced variance $\text{Var}(y)$ of the QoI. With the PDF definition we can specify

$$E(y) = \int_{\Omega} f(\mathbf{x})p(\mathbf{x})dx^D \quad (2.8)$$

$$\text{Var}(y) = \int_{\Omega} (f(\mathbf{x}) - E(y))^2 p(\mathbf{x})dx^D, \quad (2.9)$$

over the D -dimensional hypercube Ω . Even though we want to concentrate on variance based methods for the SA, there are different ways to perform an SA and to explore the parameter domain, spanned by the individual uncertainties. Especially analyzing the dependence of the QoI on the input parameter is a critical characteristics for different methods. Therefore we will introduce two common methods to tackle the uncertainty propagation problem for first-principles kinetic models which are essentially different in computational complexity and information content: the LSA and the GSA.

2.2.1 Local Sensitivity Analysis (LSA)

The LSA is the most prevalent approach, especially in the applied engineering field, due to its straightforward and inexpensive calculations providing a quantity of information[33–35]. The concept of the LSA is to analyse how sensitive the model output y is to small perturbation in the input value vector $\mathbf{x} = (x_1, \dots, x_D)$. The general Sensitivity Index (SI) S consists of estimating the first order derivatives $\frac{\partial y}{\partial x_i}(x_1, \dots, x_D)$, which is mostly normalized with the means or the standard deviations σ of y and x_i [86].

Therefore, the LSA quantifies the effect of one single input parameter i , while the rest of the

input parameters are fixed to the nominal value \mathbf{x}_{nom} . Methods that only look at one factor isolated are called one-factor-at-a-time (OAT) methods and require linear relation of the data to be effective. Furthermore, identification of interactions between parameters is not possible with an LSA. Nonetheless, local approaches have a high variety of applications, e.g. solving inverse problems, parametric sensitivity or efficiently exclude non-important parameters to reduce models [86]. Depending on the structure of the model, the derivative based methods can give rather limited information. For example, for an additive model $\sum_i^D a_i \cdot x_i$, with $a_i = const.$, the derivative based LSA would show an equal dependence on every parameter [91], without taken the uncertainty range into account. So these methods work effectively only for models with a small uncertainty range.

In our case, we have further knowledge about the parameter, i.e. a known range of uncertainty which can enter the LSA. The idea is similar to the so-called sigma normalised derivatives, where the derivatives are weighted with the parameter uncertainty factor to produce a ranking within the sensitivities. So will concentrate on the variance and its decomposition into contributions by individual input uncertainties. Such a decomposition gives an explicit relation on which parameter influences the QoI the most. Since every parameter pertains to an elementary reaction, the decomposition gives us an insight on which microscopic aspects control the catalytic reactivity under the assumption that the parameters carry uncertainties [23].

If we expect that the parameters are statistically independent we could assume a linear dependence of the response $f(y)$ on the input \mathbf{x} and the PDF resolves in the product form $p(x_1, \dots, x_D) = (p_1 \times p_2 \times \dots \times p_D)$. With this, the output variance $\text{Var}(y)$ decomposes into contribution of different input parameter

$$\text{Var}(y) = \sum_i^D S_{i,\text{LSA}} \quad \text{with} \quad S_{i,\text{LSA}} = \left| \frac{\partial y}{\partial x_i}(\mathbf{x}_{nom}) \right|^2 \text{Var}(x_i), \quad (2.10)$$

with $\text{Var}(x_i)$ as the variance of x_i . This local sensitivity definition has a adequate, statistical interpretation. The $S_{i,\text{LSA}}$ measures the variance contribution and therefore the importance of a single input parameter uncertainty. So if the derivative and the uncertainty with respect to x_i is small or large, the $S_{i,\text{LSA}}$ will be equally small or large, symbolizing the importance of that input parameter. In any other case, the balance between the local dependence of $f(\mathbf{x})$ on x_i and the error induced uncertainty will define the $S_{i,\text{LSA}}$. Despite the overall success of LSA, microkinetic models are usually highly non-linear and the parameter uncertainties are far from being small. Thus, derivative based approaches, which assume linear behaviour, are likely to assign incorrect importances to single parameters. Furthermore, the linear sensitivity can change from zero to a sizable value within reasonable variations of the input parameter,

as recent studies have shown [26, 35].

According to this, a global sensitivity analysis would be more suitable for kinetic models.

2.2.2 Global Sensitivity Analysis (GSA)

While the LSA provides information close to the nominal value, the GSA takes the whole parameter variation into account [37]. Numerous derivative-based approaches have been extended towards larger parameter spaces, like the Morris method of 1991 [80], which partitioned the domain in equally big intervals and performs path simulations for the SI, whose mean value can be considered as a global sensitivity measure. However, this does not easily overcome the problem of non-linearity of the underlying model or the difficulty if the derivatives are not easily accessible, e.g. for kMC models. Furthermore, the SI do not display the induced uncertainty directly.

Therefore, we need a model-free approach and focus on variance-based methods that are suitable for complex non-linear and non-additive models [91]. One drawback can be the higher computational cost, but it allows for parameter interaction and interpretation of sensitivity measures in terms of an induced uncertainty with a readily comprehensible relation between the LSA and the GSA. Furthermore, it can provide not only dependencies of the single input parameter but also of combination with one another, resulting in a dependency network.

In the non-linear settings of the model, the equivalent of the decomposition Eq. (2.10) is the so-called Analysis of Variance (ANOVA) decomposition [36]. The idea is based on the functional decomposition scheme, where the original function is decomposed in multiple lower order functions, which can reconstruct the original function. The decomposition is a feature to achieve further understanding on identity of the constituent components, reflecting physical processes. In the context of SA and based on the product PDF $\mathbf{p} = (p_1, \dots, p_D) = (p_1 \times p_2 \times \dots \times p_D)$, we can decompose our model response $f(\mathbf{x})$ into summands of increasing dimensions

$$f(x_1, \dots, x_D) = f_0 + \sum_{i=1}^D f_{1,i}(x_i) + \sum_{1 \leq i < j \leq D} f_{2,ij}(x_i, x_j) + \dots + f_D(x_1, \dots, x_D). \quad (2.11)$$

In total, the number of terms increases exponentially with the dimension D , 2^D , and can get computationally expensive for large D . Eq. (2.11) is also known as the Hoeffding decomposition, with the requirement of D statistically independent parameters $\mathbf{x} = (x_1, \dots, x_D)$ and a pairwise orthogonality of the terms $f_0, f_{1,i}, f_{2,ij}$, etc. with respect to the scalar product

$(f, g) = \int f g p dx^D$. Therefore the unicity condition for Eq. (2.11) is given by

$$\int_0^1 f_{1,2,\dots,s}(x_1, x_2, \dots, x_s) dx_i = 0, \quad \text{with } 1 \leq s \leq D. \quad (2.12)$$

f_0 is the mean and all other $f_{1,i}, f_{2,ij}$, etc. have a mean value of zero. The functions $f_{1,i}, f_{2,ij}, \dots, f_D$ are obtained from

$$\begin{aligned} f_0 &= \mathbf{E}(y) \\ f_{1,i} &= \mathbf{E}_{x_{\sim i}}(y|x_i) - f_0 \\ f_{2,ij} &= \mathbf{E}_{x_{\sim i, \sim j}}(y|x_i, x_j) - f_{1,i} - f_{1,j} - f_0 \end{aligned} \quad (2.13)$$

and so on for higher orders, with the convention that $x_{\sim i, \sim j}$ indicates the expected value over the whole parameter domain except for the parameter in the subset (i, j) .

The orthogonality of the functions also implies the ANOVA decomposition, which was shown by Sobol in 1990 [92]; thus, the total variance $\text{Var}(y)$ can be partitioned similar to the function decomposition Eq. (2.11)

$$\text{Var}(y) = \sum_i^D \text{Var}_i + \sum_{1 \leq j < i \leq D} \text{Var}_{i,j} + \dots + \text{Var}_D, \quad (2.14)$$

where the contributions $\text{Var}_{1,i}, \text{Var}_{2,ij}$, etc. are the partial variances to the corresponding function terms $f_{1,i}, f_{2,ij}$, etc.

$$\begin{aligned} \text{Var}_i &= \text{Var}(f_i(x_i)) = \text{Var}_{x_i}(\mathbf{E}_{x_{\sim i}}(y|x_i)) \\ \text{Var}_{ij} &= \text{Var}(f_{2,ij}(x_i, x_j)) \\ &= \text{Var}_{x_i, x_j}(\mathbf{E}_{x_{\sim i, \sim j}}(y|x_i, x_j)) - \text{Var}_{x_i}(\mathbf{E}_{x_{\sim i}}(y|x_i)) - \text{Var}_{x_j}(\mathbf{E}_{x_{\sim j}}(y|x_j)). \end{aligned} \quad (2.15)$$

Terms for the higher order and the ANOVA decomposition itself (Eq. (2.14)) hold for independent parameters. In our case this is implied by the product PDF, but extensions of the ANOVA decomposition exist towards dependent parameters. [73, 75, 78, 93–95]. Nevertheless, the function (Eq. (2.11)) and the variance decomposition (Eq. (2.14)) allow us to explicitly address the impact of individual parameters on the model output. Only the terms $f_{1,i}, f_{2,ij}$, etc., where i appears are influenced by the parameter x_i , thus only the corresponding variances $\text{Var}_{1,i}, \text{Var}_{2,ij}$, etc. are affected by the parameter uncertainty. The contributions $\text{Var}_{1,i}$ are called the first order index or the main effect with respect to x_i and take the isolated impact

of one input parameter into account [96]

$$S_i = \text{Var}_i = \text{Var}_{x_i}(\mathbb{E}_{x_{\sim i}}(y|x_i)). \quad (2.16)$$

In other words, S_i defines the expected reduction of the overall variance $\text{Var}(y)$, if x_i can be fixed and does not contain an uncertainty [96]. Higher-order terms consider interactions between different input parameters, whereby the second-order index $\text{Var}_{2,ij}$ displays a measure for the interactions of parameters i and j [96]. S_i describes the main effect by one parameter, and the second-order indices describe the impact that cannot be explained by an isolated singular parameter but by a coupling effect of parameters. They do influence not only the overall output but also the importance of other input parameters.

If all contributions higher than the first-order index are zero, the function $f(\mathbf{x})$ can be represented by set of one-dimensional functions similar to the decomposition of LSA. If the underlying function $f(\mathbf{x})$ is linear, the Sensitivity indices for the local and the global approach are equal.

The number of terms for the ANOVA increases exponentially (in analogy to the growth of function terms) with the dimensionality, producing an enormous quantity of information given by the myriad of SI-numbers. This makes it difficult to produce a clear interpretation of all the parameters. That is why we want to limit the amount of SI-numbers. Overall, we are interested in the effective impact of individual input parameters x_i . However, using the isolated first order SI does not provide this information, because a significant portion of the output variation can be assigned to interactions of higher order. In fact, the first order sensitivity can be zero, although the parameter has a big impact. If we now consider all the contributions in which x_i appears, we will account for all possible interactions. Summing over these terms leads to the Total Sensitivity Index (TSI)

$$S_{T,i} = \text{Var}_{1,i} + \sum_{j \neq i} \text{Var}_{2,ij} + \dots + \text{Var}_D, \quad (2.17)$$

which captures the expected value of the total variance if all input parameters are fixed except for i . One way to interpret this relation is to consider it as the difference of the total variance and all the first order effects of all input parameter except for x_i ,

$$S_{T,i} = \text{Var}(y) - \text{Var}_{x_{\sim i}}(\mathbb{E}_{x_i}(y|x_{\sim i})) = \mathbb{E}_{x_{\sim i}}(\text{Var}_{x_i}(y|x_{\sim i})). \quad (2.18)$$

Due to the known identity $\text{Var}(y) = \text{Var}_{x_{\sim i}}(\mathbb{E}_{x_i}(y|x_{\sim i})) + \mathbb{E}_{x_{\sim i}}(\text{Var}_{x_i}(y|x_{\sim i}))$ the $S_{T,i}$ can be reformulated as the expected variance, that would be left, if all parameters except for x_i are fixed [97].

These definitions differ from the common SIs and their first definition by Homma and Saltelli [98] since we omit the normalisation with the total variance $\text{Var}(y)$. In our context, we want to analyse and compare the LSA and GSA sensitivity indices under the aspect of an induced uncertainty. The definition of TSI as the expected variance of the uncertainty of x_i allows us the targeted interpretation of the induced uncertainty. $\text{Var}_{x_i}(y|x_{\sim i})$ describes the variance caused by varying x_i , while all other parameters are fixed, which is the non-linear analogon to the LSA-SI. So the TSI $S_{T,i}$ is the variance induced by x_i - uncertainty, averaged over all other parameters. The TSI is not a decomposition of the total variance in a general non-linear setting, as their sum can be larger. Only under the circumstances that the target function f is a sum of one-dimensional functions, the TSIs would sum up to the total variance. Then the first order S_i and the TSI will agree and in that case the same will be true for the $S_{i,LSA}$ and $S_{T,i}$ if the underlying function is linear[23]. In this context, it should be mentioned that although the LSA and GSA agree if the function $f(x)$ is linear, the opposite argument is not valid.

For our purpose, we want to apply the SA to highly complex systems, which often contain numerous uncertain parameters. Unlike the linear case, the variance and the TSI need to be obtained by averaging over the parameter distribution. For complex reaction networks with numerous elementary reactions, this requires the approximation of functionals, including the solution of multiple high dimensional integrals.

3 *Sparse Grid surrogate model*

The need for an effective operating on functions over high dimensional domains arises in different fields not only in the field of Uncertainty Quantification (UQ) [99], but also in other scientific, engineering, and socio-economic fields. Besides UQ, it is a big topic in finance [100] or molecular simulations [101]. Often today's problems can consist of multiple dimensions, so that the approximation of integrals can be very costly. The quadratures have to be performed numerically as analytical solutions are not available, or simulations only implicitly give the response. In such cases, the evaluation of the integrand causes significant computational costs so that the high accuracy of the approximation needs to be reached with a limited number of evaluations. The high dimensionality arises, e.g., from small-time steps in time discretizations or - as in the field of stochastic reactivity models - from numerous state variables. Each of them carries an uncertainty, which causes the need for a parametrical integration over the model output for the UQ [23, 52]. In finance, the integrals often occur in the context of option pricing, like Asia options, zero-coupon bonds, and collateralized mortgage obligations, to approximate the price of the functional derivatives [100]. For problems of high dimensional integrals, Monte Carlo (MC) or Quasi Monte Carlo (QMC) can be a solution without a high computational cost.

For the Global Sensitivity Analysis (GSA), the use of the classical approaches is more difficult when the single evaluation is at a high computational cost. MC and QMC can estimate the sensitivity indices for all dimensions, but with a high number of function evaluations, which makes it less suitable for a model with only a fraction of important dimensions. This worsens if we also want to estimate higher-order sensitivity indices. We need a surrogate model whose characteristics correspond to the kinetic model and whose evaluation is cheap to tackle this problem. With this, we have a function approximation and can perform the GSA without high computational costs.

3.1 *Problem formulation*

As a starting point, we consider a function $f : \Omega \rightarrow \mathbb{R}$ with $\Omega \in \mathbb{R}^D$ whose domain Ω is defined as a hypercube $\Omega = [0, 1]^D$. To perform a GSA we need an effective surrogate model

of

$$f(x_1, x_2, \dots, x_D) \quad (3.1)$$

in a medium to a high dimensional domain, i.e., D is larger than 5. In the context of just approximating the high dimensional integrals,

$$I = \int_{\Omega} f(x) dx^D \quad (3.2)$$

classical numerical integration methods, often based on the product of one-dimensional rules [102], are suffering from the *curse of dimensionality* [103, 104]. Hence, the computational effort, strongly correlated to the points N , for a predefined accuracy ϵ increases exponentially with the dimension D , $\epsilon(N) = \mathcal{O}(N^{-r/D})$, for functions with bounded total derivatives of order r [104, 105]. This quickly becomes unfeasible for higher dimensions unless the function shows a certain smoothness, which is usually not fulfilled. One of the famous approaches that does not fall under the curse are the MC methods, where the sample points are chosen randomly from a uniform distribution on Ω [106]. Despite the advantage of being independent on the dimensionality of the problem, the convergence rate of $\epsilon(N) = \mathcal{O}(N^{-1/2})$ with an increasing number of points N , i.e., the number of the function evaluation, is often considered to be infeasibly slow [107, 108]. For instance, to decrease the error by a factor of 10, 100 times as many sampling points N are needed. This caused the necessity of methods that show a higher convergence rate than MC like QMC or Sparse Grid (SG) methods [104, 109]. Both of them are deterministic schemes and depend less strongly on the dimensionality than comparable methods [107, 110] which makes them most effective for medium dimensional problems. The QMC scheme achieves a convergence rate of $\mathcal{O}((\log N)^D/N)$ or even better for a sufficiently smooth f and the Sparse Grid scheme has a convergence rate of $\mathcal{O}(N^{-r}(\log N)^{r(D+1)})$ for bounded mixed derivatives of order r , with r depending on the integrand and the basis function [40]. While the QMC method works exactly like the MC approach, except that the QMC uses pre-described sampling points rather than randomized ones, the SG approach constructs a surrogate model based on a basis set expansions.

The Sparse Grid approach recently reached popularity in the field of uncertainty propagation and sensitivity analysis, mainly in the context of partial differential equations [111–113]. Still, its foundation was already built in the nineties [114–116] and similar ideas date even further back to the sixties [117]. By taking the smoothness of a function and the lower intrinsic dimensionality into account, the SGs avoid the curse of dimensionality and achieve a faster convergence rate. This makes the approach very suitable for the field of uncertainty quantification, where not every dimension shows the same importance [118].

In this work, we concentrate on the Sparse Grid (SG) approach to solve the high-dimensional

integrals and benefit from the key features of the **SG**. Under the purpose of using it for sensitivity analysis and estimating an Analysis of Variance (**ANOVA**) decomposition

$$f(x_1, \dots, x_D) = f_0 + \sum_{i=1}^D f_{1,i}(x_i) + \sum_{1 \leq i < j \leq D} f_{2,ij}(x_i, x_j) + \dots + f_D(x_1, \dots, x_D) \quad (3.3)$$

$$\text{Var}(y) = \sum_i^D \text{Var}_i + \sum_{1 \leq j < j \leq D} \text{Var}_{i,j} + \dots \text{Var}_D, \quad (3.4)$$

previous research showed that the **MC** and **QMC** approach scale unfavourably with the dimension [36, 119]. Each of the variance terms in Eq. 3.4 requires multiple **MC** simulations and with increased dimensionality this results in unfeasible for ,e.g. kinetic Monte Carlo (**kMC**)-models, where a single evaluation is already expensive [120].

If we extend the idea of Sensitivity Analysis (**SA**) to multiscale modelling where we couple the catalytic results to computational fluid dynamics (**CFD**) simulation of a flow stream in a reactor, the importance of an accurate but cheaper function evaluation rises. Here we would couple a low-fidelity-**kMC** model with high-fidelity-**CFD** simulations. Therefore, we need an accurate function approximation, which is not possible with the classical **MC** and **QMC** approach, but with an **SG** approach [104]

$$u(x) \approx f(x). \quad (3.5)$$

The Sparse Grid approach constructs such a surrogate model based on a product of one-dimensional basis function by omitting the *higher order cross terms* and the corresponding grid points, which do not significantly increase the accuracy of the model.

As we want to perform a **GSA** we expect for models introduced in section 2.1, a hierarchy of importance regarding the dimensions since not every parameter has the same impact on the model outcome. A fraction of the elementary steps are rate-determining and, therefore, will influence the overall conversion of the kinetic model so that only a fraction of the dimension need a grid with a high resolution. With a dimension-adaptive approach, we can omit unnecessary points in dimensions with no impact and reduce overall cost. Furthermore, the kinetic data shows only locally features that need a higher resolution. The local features are phase transitions of a non-active surface to an active surface and vice versa. With a local adaptation of the **SG**, we can approximate the overall domain with a few grid points and increase the resolution in areas of this phase transitions. This means that the **SG** surrogate model needs a varying spatial resolution, which is achieved by adaptive refinement and ensures a reduction of expensive model evaluations.

3.2 Sparse Grids

A full grid approach constructs a D -dimensional integrand on a regular grid with equidistant points as a product of one-dimensional basis functions. With increasing dimensions, this quickly becomes unfeasible without the right choice of basis function. In contrast, the sparse grid employs hierarchical basis functions to construct a surrogate model and omits basis functions with a negligible contribution to the overall interpolation. In the following chapter, we start with the one-dimensional quadrature rule and the basis functions. Based on this, we construct the D -dimensional SG with some necessary notation for the different refinement strategies, which follow in section 3.3 and show some examples.

3.2.1 One-dimensional hierarchical basis functions

First, we consider a function $f : \Omega \rightarrow \mathbb{R}$, defined on the domain $\Omega = [-0.5, 0.5]$. We might not know the closed form of f but we can evaluate f at certain spatial points. To construct an interpoland of the function f , we have to discretize the spatial domain Ω . For a full grid approach, we would choose a regular grid with equidistant points of a mesh width of $h_l := 2^{-l}$, at which we can evaluate function f [121]. With increasing l , the mesh grid becomes finer which results in a higher accuracy of the interpoland and higher number of points such that l can be defined as the refinement level. Based on a choice of basis function (BF) φ and the one-dimensional interpolation rule, we can construct the interpolation $u(x)$ with

$$f(x) \approx u(x) = \sum_i v_i \varphi_i(x) \quad (3.6)$$

$$u(x_i) = f(x_i) \quad (3.7)$$

where the expansion coefficients v are determined by the interpolation requirement Eq. (3.7) for the grid points x_i . With adjusting the refinement level, we can control the accuracy of the interpolation and thus the amount of points N . If we extend the full grid to the D -dimensional case we have to deal with not N , but N^D points, which easily suffers from the curse of dimensionality [104, 122]. Although a full grid can be constructed with the right choice of basis functions, displaying the function characteristics with the minimal amount of points, with increasing dimensions the number of points will become unfeasible. In contrast, the sparse grid idea omits the basis function, which contributes little to the accuracy of the surrogate model, primarily the higher-order cross terms.

The sparse grid approach employs a hierarchical decomposition of the underlying spatial approximation and uses hierarchical basis functions for equation (3.6) to ensure an adaptive interpolation. The idea behind hierarchical basis functions is to start with a coarse resolution of the interpolation, constructed with a minimal amount of points. Then only more BFs are added if we find the function not well represented. Then we keep the old BFs and add new points, whose BF interpolate the difference between the old coarse and the newly refined interpolation. This procedure is repeated until the interpolation difference is reasonably small so that the new added BFs have only a small contribution to the overall interpolation. With this, the method ensures that only necessary BFs are included.

Since the SG approach wants to use a limited amount of BFs (and the corresponding grid points), the choice of BFs is very important. Depending on the characteristics of the underlying function f , different kinds of BFs can be more or less efficient. The most common choice are piecewise linear functions or spline basis function [112, 123, 124], but the SG approach is not restricted to them and employs other basis functions such as polynomial [123], wavelets [125] or prewavelets [104].

Our purpose is to construct a surrogate model for kinetic data, but it can be applied to problems with similar characteristics. To explain the characteristics of the underlying data, we illustrate the dependence of an observable, in this case, the Turnover Frequency (TOF), on the uncertainty of an input parameter, in figure 3.1. This is a schematic example of the reduced model for the CO oxidation on Rutheniumdioxide(110) from Gelß et al., where we displayed the TOF dependence over the adsorption rate of CO over a range of $[10^7, 10^{10}]$ s^{-1} . As the figure shows, the TOF does not have an overall linear dependence on the reaction coordinate. It can instead be described as an interplay between almost linear behaviour and rapid local changes. This local non-linear behaviour corresponds to the phase transitions between domains with varying catalytic surface activity.

Using piecewise linear, locally supported basis functions seems to be the most promising approach based on the kinetic characteristics. If we considered higher-order splines as BFs for these characteristics, we would produce overshoots at the local non-linearities, which requires further refinement points to correct these. We are testing both sets of BFs for a rather extreme example of a target function g with two discontinuities at $x = \pm 0.2$ and a constant function elsewhere.

$$g(x) = \begin{cases} 0 & \text{for } x \leq -0.2 \text{ or } x \geq 0.2 \\ 1 & \text{for } -0.2 < x < 0.2. \end{cases} \quad (3.8)$$

Figure 3.2 shows the two different interpolation schemes for the spline BFs of order $p = 2$ and the piecewise linear basis function. Even though the discontinuity is a drastic case of the rapid local changes, the spline interpolation shows the problem of approximating sharp

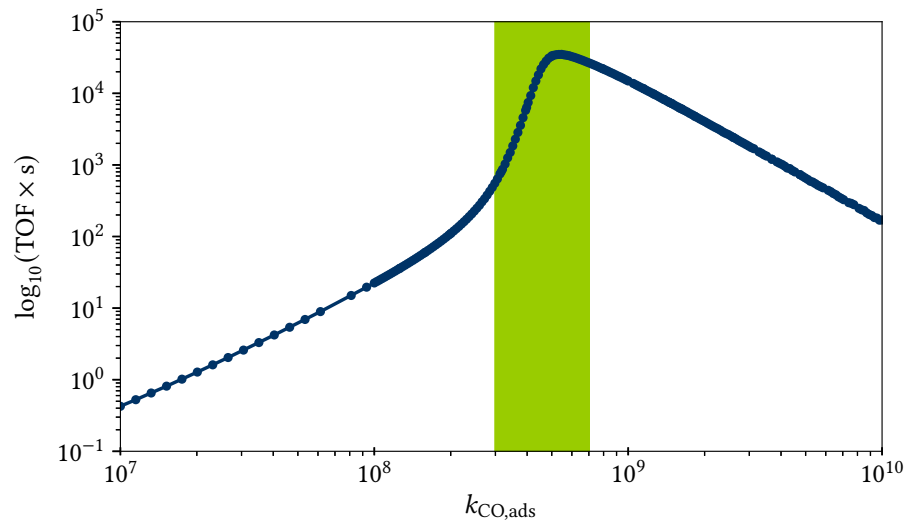


Figure 3.1: This is the CO oxidation on Rutheniumdioxide(110) dependent on the adsorption rate of CO. The logarithm of the Turnover frequency (TOF) behavior plotted over the reaction coordinate, which spans the uncertainty of one system parameter. Along the reaction coordinate the TOF values highly variate and show areas of linear behaviour and also of non linear behaviour █.

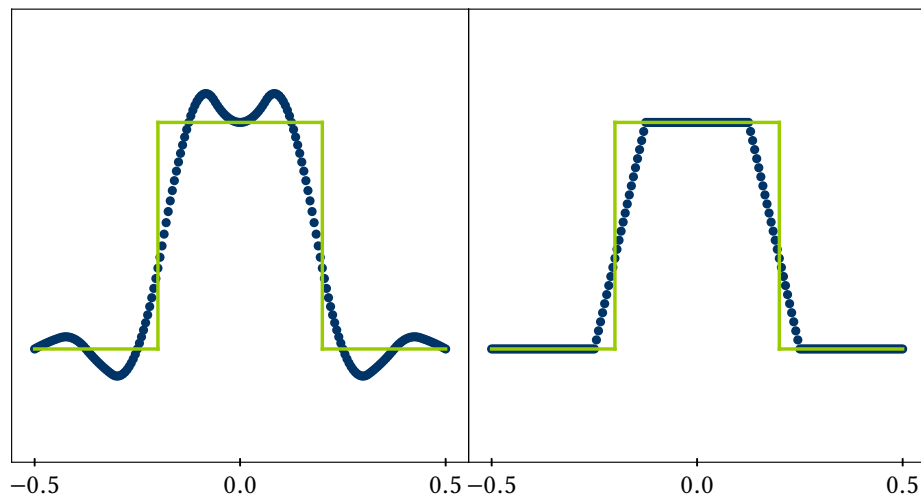


Figure 3.2: Sparse Grid Interpolation for locally discontinuous function $g(x)$ █ with splines of different orders for level $l = 2$. *Left panel* : Interpolation of $p = 2$ order splines, with significant overshoots towards the discontinuity. *Right panel* Interpolation of $p = 1$ order splines, with a closer adaptation of the characteristics of the underlying function.

changes with higher-order splines. While the linear basis functions in the right panel adapt the constant characteristics accurately and only need additional basis functions in the area of the discontinuities, the overshoot of the spline basis functions in the left panel produces an error in the area of the constant function. This would need further refinement to compensate for the deviation. Because of this, we choose piecewise linear basis functions, which are depicted in figure 3.3 and are defined by

$$\varphi_{l,i} = \begin{cases} 1 & \text{for } l = 0 \\ \max(1 - 2^l|x - x_{l,i}|, 0) & \text{for } l > 0, \end{cases} \quad (3.9)$$

with a *level index* l and a *count index* i [126, 127]. The basis functions have a local support and are centered at grid points $x_{0,0} = 0$, $x_{1,0/1} = \mp 0.5$, $x_{2,0/1} = \mp 0.25$, and $x_{l,i} = x_{l-1, \lfloor i/2 \rfloor} - (-1)^i/2^l$, which are the evaluation points for the interpolation of f . The *level index* describes the hierarchical order of the basis function belonging to the hierarchical decomposed space, whereas the *count index* i lets us uniquely identify a BF in a hierarchical difference spaces. By defining a hierarchical index set

$$\mathcal{I}_l = \begin{cases} \{i \in \mathbb{N} : 0 \leq i \leq 2^l - 1\} & \text{for } l = 0, 1 \\ \{i \in \mathbb{N} : 0 \leq i \leq 2^{l-1} - 1\} & \text{for } l > 1 \end{cases} \quad (3.10)$$

we obtain hierarchical difference spaces

$$W_l = \text{span}\{\varphi_{l,i} | i \in \mathcal{I}_l\}, \quad (3.11)$$

which are used to construct the approximation space V_L of the function f as a direct sum of W_l

$$V_L = W_0 \oplus W_1 \oplus \dots \oplus W_L = \bigoplus_{l \leq L} W_l, \quad (3.12)$$

based on a full grid [118]. L defines the maximum refinement level, thus the accuracy of the surrogate model, whereby the surrogate model (the interpolant) of $L \rightarrow \infty$ converges against f .

The piecewise linear basis functions and the connection to the hierarchical subspaces W_l are displayed in the left panel of figure 3.3 for the refinement levels $l = 0$ to $l = 3$. This underlines, why the *level index* l is called refinement level. For $l = 0$ the subspace W_0 is spanned by a single constant basis function, whose local support spans over the whole domain $\Omega = [-0.5, 0.5]$. With increasing level index, the amount of basis function increases while the local support decreases, resulting in a higher resolution. Starting with the coarsest resolution

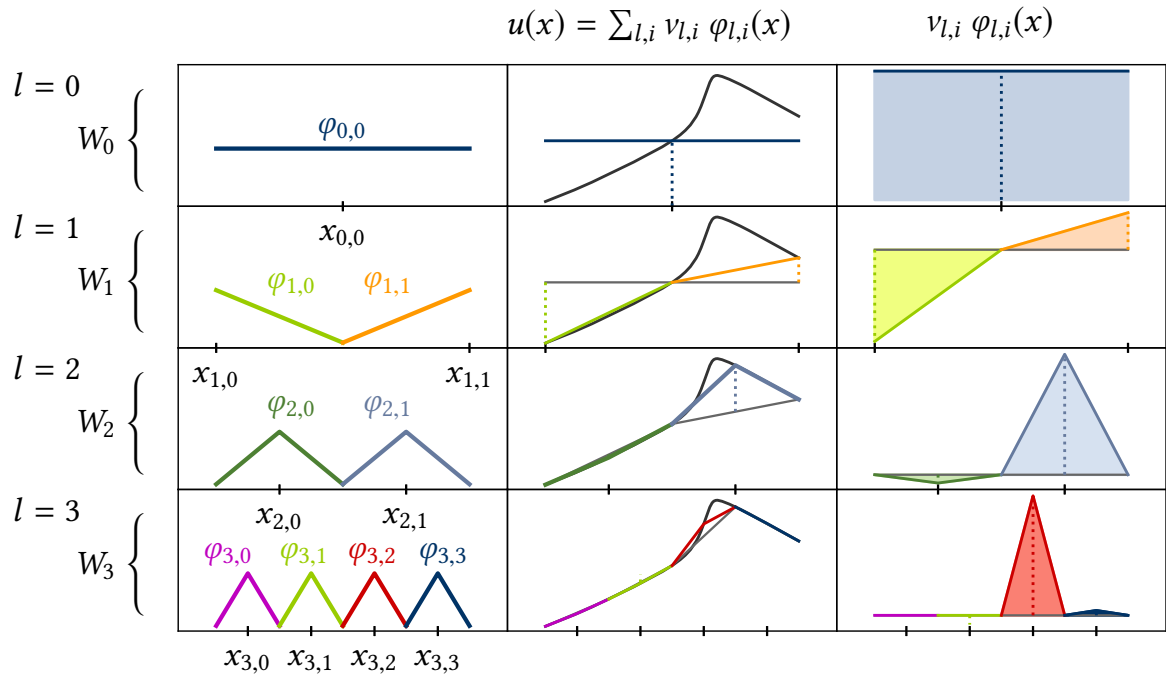


Figure 3.3: Illustration of the hierarchical interpolation in one dimension with increasing level index. *Left Panel:* Basis functions added at the corresponding levels l . *Middle Panel:* The original function $f(x)$ and the hierarchical interpolation $u(x)$. *Right Panel:* Illustration of the contribution of the basis functions added at each level $v_{l,i} \varphi_{l,i}$. The contributions decrease with increasing level, and, at the finest level, only those basis functions close to the sharp non-linearity still have significant contributions to the interpolation.

W_0 , the space W_{l+1} spanned by the BFs added in the $(l + 1)$ th refinement step is considered as the refinement of the space W_l spanned by the BFs of the previous refinement step. Note that the BFs from the same subspace have disjoint supports of the same size, which union equals Ω all BF [122].

In general, the subdomain for piecewise linear basis functions covered by one basis function can be defined as

$$\Omega_{l,i} = \begin{cases} [x_{l,i} - 0.5, x_{l,i} + 0.5] & \text{for } l = 0 \\ [x_{l,i}, x_{l,i} \pm 0.5] & \text{for } l = 1 \\ [x_{l,i} - 2^{-l}, x_{l,i} + 2^{-l}] & \text{for } l > 1. \end{cases} \quad (3.13)$$

which displays the decreasing support with increasing l . In the context of refinement and consecutive levels, the relationship of *parents* and *children* plays an important role. By definition children refine the interpolation area spanned by their parents. The children of a BF $\varphi_{l,i}$ are BFs of $l + 1$ with an overlapping support. For the one dimensional case these two BFs are $\varphi_{l+1,2i}$ and $\varphi_{l+1,2i+1}$. BFs from $l - 1$ with an overlapping support $\Omega_{l,i}$ are then called parents of $\varphi_{l,i}$. Eq. (3.14) and Eq. (3.15) describe the sets of points, defined as *children* or *parents* [40, 118].

$$\mathcal{C}_{l,i} = \{(m, j) \mid \Omega_{m,j} \subset \Omega_{l,i} \text{ with } m = l + 1\} = \{(m, j) \mid m = l + 1 \wedge j = 2i(+1)\} \quad (3.14)$$

$$\mathcal{P}_{l,i} = \{(m, j) \mid \Omega_{l,i} \subset \Omega_{m,j}\} = \{(m, j) \mid l = m + 1 \wedge i = 2j(+1)\} \quad (3.15)$$

We call all BFs of a smaller refinement level with an overlapping support the *ancestors* $\mathcal{A}_{l,i}$ (Eq. 3.16) and of higher refinement level *descendants* $\mathcal{D}_{l,i}$. Their indices are recursively defined with Eq. (3.14) and Eq. (3.15).

$$\mathcal{A}_{l,i} = \{(\tilde{l}, \tilde{i}) \mid \Omega_{l,i} \subset \Omega_{\tilde{l},\tilde{i}}, \text{ for all } \tilde{l} < l\}. \quad (3.16)$$

Figure 3.3 depicts the construction of the sparse grid surrogate model according to the underlying function. The common Sparse Grid method assumes values of zero on the boundary $\delta\Omega$; hence it does not consider points on the boundary [122]. Therefore, classical hat functions are used [122]. In our case, we can not assure zero values on the boundary, which would cause numerous refinements with the classical hat functions. The zero value of the basis functions (BF) induces a high error towards the boundary, including the necessity of refinement and an increased density of points. We chose the basis function for the first level with centre points directly on the boundary to avoid this. With this, we assure the correct values on the boundary and avoid unnecessary refinement points. [124]

As depicted in the left panel of figure 3.3, the refinement of the grid starts with a single grid point $x_{0,0}$ in the centre of the integration interval. Level $l = 1$ adds those sparse grid points located on the boundary and their corresponding BFs. If we want to refine the approximation further, we add BFs of higher levels by adding additional nodes between the existing nodes [122, 127].

Based on the BFs of the subspaces, the interpolation $u(x) \in V_L$ can be written as a one dimensional interpolation

$$u(x) = \sum_{l \leq L, i \in \mathcal{I}_l} v_{l,i} \varphi_{l,i} \quad (3.17)$$

where the hierarchical surplus $v_{l,i}$ is uniquely indexed with the same index set (l, i) as the corresponding basis function. The middle panel of figure 3.3 shows the interpolation for the target function using all the BFs up to the refinement level l . With increasing l and the finer resolution of the basis functions, the characteristics of the underlying function are effectively adapted to the point at $l = 3$, where only little differences are visible between the interpolation and f . As expected, not all basis function in one hierarchical difference space have the same impact on the approximation. This can be seen in the right panel, where the contribution of a single BF, denoted as $v_{l,i} \cdot \varphi_{l,i}$ is shown. As the hierarchical structure implies, the contribution of a single BF decreases with the refinement level, such that the support of the $l = 3$ BF are only a fourth of the $l = 0$ BF. Furthermore, the BFs in one difference space may have unequal contribution to the approximation. During the refinement from $l = 1$ to $l = 2$, the linear characteristics are already accurately approximated while the non-linear behavior needs further refinement. Therefore, $\varphi_{2,0}$ does not improve the approximation and can be neglected. This effect is even stronger in the refinement step from $l = 2$ to $l = 3$, where only the BFs of $\varphi_{3,2}$ has an impact on the interpolation. Taking only BFs with a contribution above a predefined threshold into account is one of the main feature used for local refinement.

3.2.2 High-dimensional Sparse Grid spaces

We can now extend the previously introduced interpolation method to a function f that is defined on a D -dimensional hypercube $\Omega = [-0.5, 0.5]^D$. The BFs are extended to the D -dimensional BFs via a tensor product approach,

$$\varphi_{\mathbf{l}, \mathbf{i}}(\mathbf{x}) = \prod_{\alpha=1}^D \varphi_{l_\alpha, i_\alpha}(x_\alpha) \text{ with } \varphi_{l,i} \in \Omega \quad (3.18)$$

with \mathbf{l} and \mathbf{i} as the D -dimensional *level index* and *count index* indicating level and index for each dimension. The definition $\varphi_{l_\alpha, i_\alpha}(x_\alpha)$ indicates the α entry of $\varphi_{\mathbf{l}, \mathbf{i}}(\mathbf{x})$. In analogy to eq.(3.18), the other notation can be transferred to the higher dimensional case. Based on the subspaces $W_{\mathbf{l}}$

$$W_{\mathbf{l}} = \text{span}\{\varphi_{\mathbf{l}, \mathbf{i}} \mid \mathbf{i} \in \mathcal{I}_{\mathbf{l}}\} \quad (3.19)$$

$$\text{with } \mathcal{I}_{\mathbf{l}} = \begin{cases} \{\mathbf{i} : 0 \leq i_j \leq 2^l - 1, \text{ for } l = 0, 1; 1 \leq j \leq d\} \\ \{\mathbf{i} : 0 \leq i_j \leq 2^{l-1} - 1, \text{ for } l > 1; 1 \leq j \leq d\} \end{cases} \quad (3.20)$$

the sparse grid approximation space V_L of (total) level L of the univariate function can be expressed as

$$V_L = \bigoplus_{|\mathbf{l}_1| < L, \mathbf{i} \in \mathcal{I}_{\mathbf{l}_1}} W_{\mathbf{l}_1}. \quad (3.21)$$

A Sparse Grid approximation is then obtained by

$$f(\mathbf{x}) \approx u_L(\mathbf{x}) = \sum_{|\mathbf{l}_1|=1, \mathbf{i} \in \mathcal{I}_{\mathbf{l}_1}} v_{\mathbf{l}_1, \mathbf{i}} \varphi_{\mathbf{l}_1, \mathbf{i}} + \sum_{|\mathbf{l}_1|=2, \mathbf{i} \in \mathcal{I}_{\mathbf{l}_1}} v_{\mathbf{l}_1, \mathbf{i}} \varphi_{\mathbf{l}_1, \mathbf{i}} + \dots \quad (3.22)$$

$$= \sum_{|\mathbf{l}_1| \leq L, \mathbf{i} \in \mathcal{I}_{\mathbf{l}_1}} v_{\mathbf{l}_1, \mathbf{i}} \varphi_{\mathbf{l}_1, \mathbf{i}} \quad (3.23)$$

employing only the BFs whose L_1 -norm of the level index \mathbf{l} is not exceeding the predefined maximum refinement level L [23]. The hierarchical surplus $v_{\mathbf{l}_1, \mathbf{i}}$ for the high dimensional case are based on the same interpolation structure as Eq. (3.17) for the 1 dimensional case. So $v_{\mathbf{l}_1, \mathbf{i}}$ for one BF is defined by the difference between the function evaluation and the coarser interpolation at $\mathbf{x}_{\mathbf{l}_1, \mathbf{i}}$

$$f(\mathbf{x}_{\mathbf{l}_1, \mathbf{i}}) = u_L(\mathbf{x}_{\mathbf{l}_1, \mathbf{i}}) \quad (3.24)$$

$$= u_{L-1}(\mathbf{x}_{\mathbf{l}_1, \mathbf{i}}) + v_{\mathbf{l}_1, \mathbf{i}} \varphi_{\mathbf{l}_1, \mathbf{i}}(\mathbf{x}_{\mathbf{l}_1, \mathbf{i}}) \quad \text{with } \varphi_{\mathbf{l}_1, \mathbf{i}}(\mathbf{x}_{\mathbf{l}_1, \mathbf{i}}) = 1$$

$$v_{\mathbf{l}_1, \mathbf{i}} = f(\mathbf{x}_{\mathbf{l}_1, \mathbf{i}}) - u_{L-1}(\mathbf{x}_{\mathbf{l}_1, \mathbf{i}}). \quad (3.25)$$

This relation is based on our choice of BF, where we do not have an overlapping support of BF in the same $W_{\mathbf{l}}$. Therefore the function is locally only interpolated by one BF per refinement step. The difference between a SG approximation and a full grid approximation can be explained in terms of eq 3.23. For a full grid, we would have to consider all subspaces $W_{\mathbf{l}}$ up to an L_∞ -norm of L , not the L_1 -norm. This would mean to include all combinations of the one-dimensional basis functions, also the expensive *high order cross term*. By setting the maximum refinement level to an upper bound for all dimensions together, the SG approach

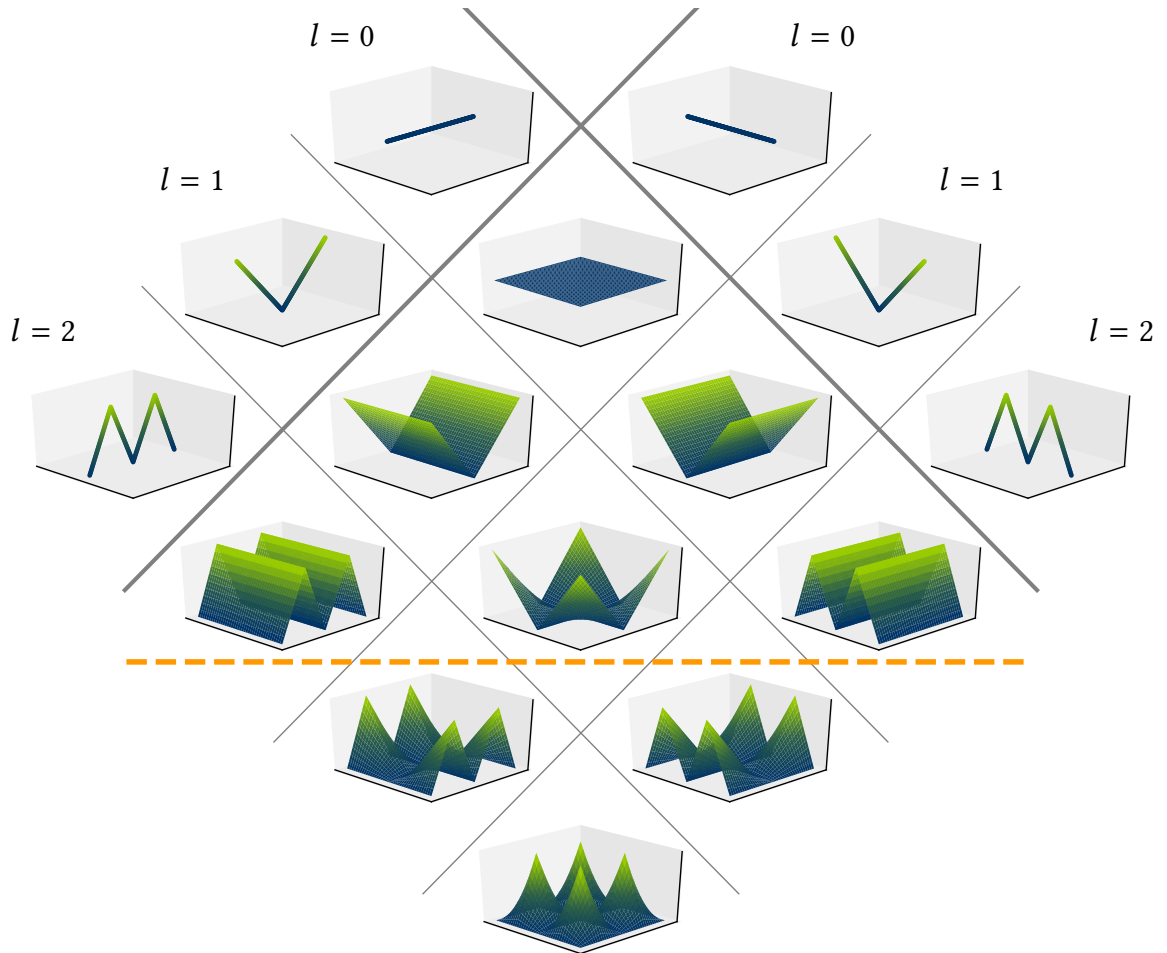


Figure 3.4: Hierarchical difference spaces for the 2D Sparse Grid Interpolation according to the refinement level. The top left and the top right row show the one dimensional basis function for each dimension and level. In the inner diamond grid, the 2D tensor products of the one dimensional functions are displayed. An interpolation constructed by the shown basis function has the maximum level of $L = 2$. For the SG, all hierarchical difference spaces above the orange dotted line --- are included, thus it consists of spaces with an L_1 -norm $|I|_1 \leq 2$. In comparison to that, a full grid includes all spaces with a maximum norm $|I|_\infty \leq 2$, which includes also the spaces below the orange dotted line --- .

overcomes the curse of dimensionality. Note here that the SG in a one-dimensional case equals an Full Grid (FG) approach. Figure 3.4 displays the construction of the higher dimensional basis functions based on the one dimensional case. The two outer rows outline the 1D-BF for $l = [0, 1, 2]$ and the tensor product of two BFs each is shown in the inner diamond mesh. Each horizontal row describes a higher dimensional subspace according to the L_1 -norm of the refinement level. Assuming a maximum level $L = 2$, the sparse grid would be spanned by all BFs above the orange line, thus $|l|_1 \leq 2$. In contrast to that, a FG approach would also include the functions below the orange line as the maximum norm has to be equal to 2.

The key concept for constructing the SG in the high dimensional case is the hierarchical ordering of the sparse grid BFs in terms of *parents* and *children*. If the interpolation of the parents is not sufficient enough, the SG will locally add points in the subdomain spanned by the parents. The additional points, called children, will interpolate the difference between the function evaluation and the results of the coarser interpolation. In the one-dimensional case, we call those BFs from Eq. (3.14) and Eq. (3.15) are the base for the high dimensional case. The indices of the children $\mathcal{C}_{\mathbf{m},\mathbf{j}}$ and the parents $\mathcal{P}_{\mathbf{l},\mathbf{i}}$ of a BF $\varphi_{\mathbf{l},\mathbf{i}}$ are given by the sets

$$\mathcal{C}_{\mathbf{l},\mathbf{i}} = \{(\mathbf{m}, \mathbf{j}) \mid \exists d, b \in [1, D] : \mathbf{m} = \mathbf{l} + \mathbf{e}_d \wedge j_d = 2 i_d(+1) \wedge j_b = i_b \text{ for } b \neq d\} \quad (3.26)$$

$$\mathcal{P}_{\mathbf{l},\mathbf{i}} = \{(\mathbf{m}, \mathbf{j}) \mid \exists d, b \in [1, D] : \mathbf{l} = \mathbf{m} + \mathbf{e}_d \wedge i_d = 2 j_d(+1) \wedge j_b = i_b \text{ for } b \neq d\}. \quad (3.27)$$

$$(3.28)$$

These sets display the refinement relation between two consecutive levels which build the base for any adaptive strategy. For further relations, the 1-dimensional case gives the idea of *ancestors* of one BF in Eq. (3.16). Based on this and the Eq. (3.26,3.27) the *ancestors* in the high dimensional case are recursively defined as the union of $\mathcal{P}_{\mathbf{l},\mathbf{i}}$ and *ascendents* of the BF contained therein.

3.2.3 Sensitivity Indices

Based on equation (3.23) we can define the quadrature and approximate the integral of function f by using the SG interpolant

$$\int_{\Omega} u_L(\mathbf{x}) dx^D = \sum_{|l|_1 \leq L, \mathbf{i} \in \mathcal{I}_1} v_{\mathbf{l},\mathbf{i}} w_{\mathbf{l},\mathbf{i}} \quad (3.29)$$

$$\text{with } w_{\mathbf{l},\mathbf{i}} = \int_{\Omega} \varphi_{\mathbf{l},\mathbf{i}}(\mathbf{x}) dx^D \quad (3.30)$$

with $w_{\mathbf{l},i}$ as the integration weight attributed to the BF $\varphi_{\mathbf{l},i}$. With this, we can efficiently approximate the expected values and variances for the Total Sensitivity Index (TSI) over the higher-dimensional space. The expected value equals the quadrature equation (3.29). For the variance, we have to approximate the second moment

$$E(y) \approx \int_{\Omega} u(\mathbf{x}) dx^D = \sum_{|\mathbf{l}|_1 \leq L, \mathbf{i} \in \mathcal{I}_1} v_{\mathbf{l},i} w_{\mathbf{l},i} \quad (3.31)$$

$$E(y^2) \approx \int_{\Omega} u(\mathbf{x})^2 dx^D = \int_{\Omega} \left(\sum_{\mathbf{l},i} v_{\mathbf{l},i} \varphi_{\mathbf{l},i}(\mathbf{x}) \right)^2 dx^D \quad (3.32)$$

$$= \sum_{\mathbf{l},i,\mathbf{l}',i'} v_{\mathbf{l},i} v_{\mathbf{l}',i'} w_{\mathbf{l},i,\mathbf{l}',i'} \quad (3.33)$$

$$\text{with } w_{\mathbf{l},i,\mathbf{l}',i'} = \int_{\Omega} \varphi_{\mathbf{l},i}(\mathbf{x}) \varphi_{\mathbf{l}',i'}(\mathbf{x}) dx^D. \quad (3.34)$$

Since we use uniform distributed parameters, we can estimate the sensitivity index TSI from Eq. (2.18) with respect to a parameter x_{α} with

$$S_{T,\alpha} = E_{x_{\sim\alpha}}(\text{Var}_{x_{\alpha}}(y|x_{\sim\alpha})), \quad (3.35)$$

$$\begin{aligned} S_{T,\alpha} &\approx \int_{\Omega_{\sim\alpha}} \left(\int_{\Omega_{\alpha}} u(\mathbf{x})^2 dx_{\alpha} \right) dx_{\sim\alpha}^{D-1} - \int_{\Omega_{\sim\alpha}} \left(\int_{\Omega_{\alpha}} u(\mathbf{x}) dx_{\alpha} \right)^2 dx_{\sim\alpha}^{D-1} \\ &= \int_{\Omega} \left(\sum_{\mathbf{l},i} v_{\mathbf{l},i} \varphi_{\mathbf{l},i}(\mathbf{x}) \right)^2 dx^D - \int_{\Omega_{\sim\alpha}} \left(\int_{\Omega_{\alpha}} \sum_{\mathbf{l},i} v_{\mathbf{l},i} \varphi_{\mathbf{l},i}(\mathbf{x}) dx_{\alpha} \right)^2 dx_{\sim\alpha}^{D-1} \end{aligned} \quad (3.36)$$

where $\int_{\Omega_{\sim\alpha}} \cdots dx_{\sim\alpha}^{D-1}$ denotes the integral over the whole domain, except for the dimension of α and $\int_{\Omega_{\alpha}} \cdots dx_{\alpha}$ the integral over the dimension α . Utilizing the product shape of the BFs (3.18) and the second moment (3.34), we can rewrite the TSI as

$$S_{T,\alpha} \approx \sum_{\mathbf{l},i,\mathbf{l}',i'} v_{\mathbf{l},i} v_{\mathbf{l}',i'} \prod_{\beta \neq \alpha} w_{\mathbf{l},i,\mathbf{l}',i'}^{\beta} \left(w_{\mathbf{l},i,\mathbf{l}',i'}^{\alpha} - w_{\mathbf{l},i}^{\alpha} w_{\mathbf{l}',i'}^{\alpha} \right) \quad (3.37)$$

with respect to the parameter x_{α} and where

$$w_{\mathbf{l},i}^{\alpha} = \int_{-0.5}^{0.5} \varphi_{l_{\alpha},i_{\alpha}}(x) dx \quad (3.38)$$

and

$$w_{\mathbf{l},i,\mathbf{l}',i'}^{\beta} = \int_{-0.5}^{0.5} \varphi_{l_{\beta},i_{\beta}}(x) \varphi_{l'_{\beta},i'_{\beta}}(x) dx. \quad (3.39)$$

3.3 Adaptivity

Although we are now able to construct the Sensitivity Index based on our SG model, we need many function evaluations to construct such a model in a high dimensional space. The next step is to reach the goal of adaptive refinement in the high dimensional case for which the relation of *parents* and *children* is essential. As indicated in the right panel of figure 3.3, not every basis function, even in the Sparse grid space, has the same impact on the accuracy of the surrogate model. While regions of linear behaviour are accurately displayed with a minimal amount of BFs, the non-linear behaviour requires more refinement. To tackle this problem, the SG approach allows for adaptive refinement since the multidimensional approach inherits the hierarchical structure of the one-dimensional BFs (3.6). The most common refinement strategies are *local-adaptive refinement* and *dimension-adaptive refinement*. The dimension-adaptive refinement was first introduced by Griebel in 1998 [127] and later on further developed with Gerstner [102] to a Generalized Sparse Grid (GSG) approach. The method generalizes the traditional isotropic SG approach that allows for computational effort in the important dimensions, which mostly means higher variations [100, 128]. For our example of parameter uncertainty in a kinetic model, determining the Sensitivity Index for every parameter implies that not every parameter uncertainty has the same impact on the outcome. It is rather the case that a kinetic model consists of numerous parameters, but the intrinsic dimensionality is a lot smaller such that only a handful of parameters are important. For the computational approach, this means that some dimensions are easier to approximate than others.

The local-adaptive refinement involves a similar approach as the dimension-adaptive refinement but focusing on spatial resolution. Functions can have a spatially varying complexity - as it can be seen in figure 3.3 where non-linearity only appears locally. The goal for the local refinement is to add only points in spatial areas of rough approximations to avoid unnecessary refinement in the subdomains of an accurate approximation. The relationship of parents and children is essential for that so that adding a child to the final SG is dependent on the error indicator of the parent. The first local-adaptive refinement was also introduced by Griebel [127], but since then, numerous approaches of local refinement have been developed. Moreover, combining the two refinement strategies gained much popularity in the last decade [23, 40, 118, 129].

Since our model contains a hierarchy of important dimensions and varying spatial complexity, we concentrate on both refinement strategies. First, we will introduce a proper refinement criterion and the local and the dimension-adaptive refinement strategies separately. Second, we explain how both can be combined for an effective refinement of our problem [23, 40, 118,

130].

3.3.1 Refinement Criteria

Since we have no prior knowledge of important dimensions or subdomains of high variability, we must adapt to these characteristics during the computational procedure. For an effective refinement, a proper refinement criterion is crucial. Depending on the application and the function f , some refinement criteria are more or less effective. In principle, we want to increase the interpolation accuracy, which means we want to refine the BFs that show a high interpolation error. Practically, the exact interpolation error is unknown and cannot be estimated without multiple expensive function evaluations. Thus, it can only be approximated based on an indicator of the local interpolation error due to omission of the descendants of a BFs of W_1 . Independent of the measurement for the local error, the hierarchical structure of the sparse grid implies that, if the grid point (\mathbf{I}, \mathbf{i}) has a small contribution, we expect this to also be the case for all its *descendants* [42]. In consequence, if the local error criteria exceed a predefined accuracy, no further points in the domain $\Omega_{\mathbf{I}, \mathbf{i}}$ are included. There are many different measures for the local error but the most common are the hierarchical surplus $v_{\mathbf{I}, \mathbf{i}}$ or the contribution $v_{\mathbf{I}, \mathbf{i}} \varphi_{\mathbf{I}, \mathbf{i}}$ of a BF. The hierarchical surplus $v_{\mathbf{I}, \mathbf{i}}$ describes the difference between the function evaluation and the sparse grid interpolation at the next coarser level at a grid point. Although the hierarchical surplus is synonymous with the definition of a local interpolation error, adding BFs according to the hierarchical surplus can be inefficient depending on the functions' characteristics. When dealing with a highly smooth function, the hierarchical surplus can be sufficient for determining suitable refinement BFs whereas the surplus decreases very slowly with respect to the impact of the BF in areas of rapid jumps. Figure 3.5 and the table 3.1 depict the problem with the hierarchical surplus during the refinement of non linear behavior. Despite the constant improvement of the interpolation along with the refinement, the hierarchical surplus documents a very slow decrease for the accuracy of the interpolation [40]. If we measure the L_1 -norm of the interpolation error and this value exceeds the predefined accuracy ϵ , the method will produce unnecessary points towards the non-linear behaviour such that the hierarchical surplus becomes disproportional to the interpolation error. Hence, we employ the L_1 -norm of the surplus and the weight of the BF as an error measure

$$\gamma_{\mathbf{I}, \mathbf{i}} := |v_{\mathbf{I}, \mathbf{i}} w_{\mathbf{I}, \mathbf{i}}| \quad (3.40)$$

for a single BF with $w_{\mathbf{I}, \mathbf{i}}$ as the integration weight and the L_1 -norm of $\varphi_{\mathbf{I}, \mathbf{i}}$. Considering that we are interested to achieve a high accuracy also for the quadrature with a minimal amount of

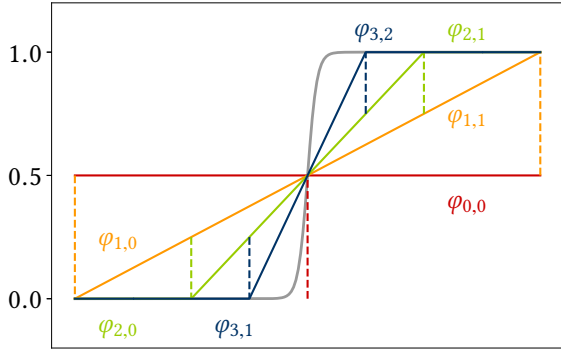


Figure 3.5: Refinement procedure and interpolation of a target function with a jump characteristic with the basis functions for different refinement levels $l = [0, 1, 2, 3]$.

Basis function	$ v_{l,i} $	$ v_{l,i} w_{l,i} $
$\varphi_{0,0}$	0.5	0.5
$\varphi_{1,0}$	0.5	0.125
$\varphi_{1,1}$	0.5	0.125
$\varphi_{2,0}$	0.25	0.0625
$\varphi_{2,1}$	0.25	0.0625
$\varphi_{3,0}$	0	0
$\varphi_{3,1}$	0.25	0.03125
$\varphi_{3,2}$	0.25	0.03125
$\varphi_{3,3}$	0	0

Table 3.1: Different refinement criteria for the refinement of the target function in figure 3.5. The middle column documents the hierarchical surplus, i.e., the maximum norm of the local error, which decreases very slowly with the increasing level. Compared to that, the right column shows the L_1 -norm of the local error with a constant decrease with the refinement.

points, the $|v_{l,i} w_{l,i}|$ gives the impact of this BF on the quadrature, since $\|f\| = \|\sum v_{li} \varphi_{li}\| \leq \sum \|v_{li} \varphi_{li}\| = \sum |v_{li}| * \|\varphi_{li}\|$. So even when the hierarchical surplus decreases slowly, the increasing level refinement induces a decreasing weight. This is also underlined by the example of figure 3.5 and table 3.1, where the classical surplus fails but the weighted surplus shows a decreasing trend along with the refinement. We can avoid costly and unnecessary refinement points towards non-linear behaviour with the weighted surplus compared to the single surplus.

3.3.2 Dimension Adaptivity

The classical Sparse Grid approach includes all hierarchical subspaces W_1 with $|\mathbf{l}|_1 \leq L$. Such an approach will not suffer from the *curse of dimensionality* but is still unfeasible for high dimensional problems. The construction scheme showed in section 3.2.2 treats all dimensions equally, although some dimensions contribute just little to the interpolation. These dimensions are not known a priori but have to be determined during the computational procedure. For that, we based our method on the GSG approach of Gerstner and Griebel [102] that identifies the difference spaces W_1 that contribute significantly to the variability of the function f .

A difference space is indicated by the level index \mathbf{l} and has D $\{\mathbf{l} + \mathbf{e}_d : 1 \leq d \leq D\}$ refinement levels, with the unit vector \mathbf{e}_d for the dimension d . Similar to the definition of the BFs, the refinement levels are called the *children* of level \mathbf{l} and thus every level index has a maximum of D *parent* levels. In the original GSG definition these level dependencies are called *forward* and *backward neighbor*. Furthermore the definition of *admissibility* is a key point of the GSG approach. A level index \mathbf{l} is only admissible if

$$\mathbf{l} - \mathbf{e}_d \in \mathcal{S} \text{ for } 1 \leq d \leq D, l_d > 0 \quad (3.41)$$

with \mathcal{S} as a set of all level indices of the sparse grid. The basic idea is to only include the level \mathbf{l} if the parent levels $\{\mathbf{l} - \mathbf{e}_d : 1 \leq d \leq D\}$ are admissible and have a big contribution to the SG interpolant $u(\mathbf{x})$. If that is not the case, we omit the difference space and the corresponding BFs and grid points. In the case of a high contribution of one parent space $W_{\mathbf{l} - \mathbf{e}_i}$ and an insignificant contribution of a second parent space $W_{\mathbf{l} - \mathbf{e}_j}$, it is highly probable that the children will also have an insignificant contribution and can be excluded. This means only if both hierarchical difference spaces needs further refinement, the children space will have a high impact on the accuracy of the SG. The contribution definition of the GSG approach by Griebel and Gerstner sums over the local errors of the singular basis functions to define the contribution of $W_{\mathbf{l}}$ [40, 102, 127]

$$\gamma_{\mathbf{l}} := \sum_{\mathbf{l}_i \in \mathcal{I}_{\mathbf{l}}} \gamma_{\mathbf{l}_i}. \quad (3.42)$$

The original approach chooses the \mathbf{l} with the largest $\gamma_{\mathbf{l}}$ and includes all $\{\mathbf{l} + \mathbf{e}_k : 1 \leq k \leq D\}$ children, if the error of the children are big enough. This procedure is repeated until the predefined threshold is met. For our study, we modified the approach to refine multiple level indices simultaneously and employ the advantage of parallel computing. We start at a SG whose highest refinement level have a L_1 -norm of L and instead of only including points for the refinement space with the highest error indicator we look at all refinement spaces for the currently highest levels. With every refinement step we consider all new refinement spaces of $|\mathbf{l}|_1 = L + 1$ and check whether all the parents level of each level have a high contribution. Therefore we do not use the L_1 -norm of the local error vector, we choose the maximum norm

$$\gamma_{\mathbf{l}} := \max_{\mathbf{l}_i \in \mathcal{I}_{\mathbf{l}}} \gamma_{\mathbf{l}_i} \quad (3.43)$$

and include all of the levels \mathbf{l} whose parents level have minimum one point with a $\gamma_{\mathbf{l}}$ above a predefined threshold ϵ . That ensures that only points are included whose parents have a high contribution and excludes levels whose parents are not in the SG and therefor have a $\gamma_{\mathbf{l}} = 0$. Since we base the level error indicator on singular point error indicator, we are able

to combine refinement strategies based on refinement spaces and point descriptions. The resulting refinement set for the dimension adaptivity \mathcal{S}_{dim}

$$\mathcal{S}_{\text{dim}} := \left\{ (\mathbf{l}, \mathbf{i}) \mid \min_{m \in \mathbf{l} - \mathbf{e}_d} (\max_{j \in \mathcal{I}_m} (|v_{\mathbf{m},j} w_{\mathbf{m},j}|)) > \epsilon \right\} \quad (3.44)$$

includes all index tuples (\mathbf{l}, \mathbf{i}) whose parents' level exist and where all of the parent levels have at least one point with an error above the threshold. The outline of the method is shown in the figure 3.6 for a 2D case. Similar to the already presented functions, we used a sigmoid function, shown in the lower left panel of figure 3.6, as the underlying function with the non linearity at $x = 0.25$

$$f(x_1, x_2) := \frac{1}{1 + \exp((-x_1 + 0.25) * 100)}. \quad (3.45)$$

Since the function depends only on one parameter dimension, we expect that the dimension adaptivity removes all grid points in the x_2 -direction. The initial set consists only of the singular set of $\mathcal{S} = \{\mathbf{0}, \mathbf{0}\}$ - with $\mathbf{0} = \{0, 0\}$ for $D = 2$ - which needs to be refined. This set equals the refinement space $W_{\mathbf{0}}$ for which we consider all possible children levels $\{\mathbf{l} + \mathbf{e}_k : 1 \leq k \leq 2\}$ and check for each one if they are admissible, i.e. every parent level of them exists. For the first step, the only parent set existing is the target $W_{\mathbf{0}}$ space so all sets of children are produced. For the second refinement step, the new sets $W_1 = \{W_{(1,0)}, W_{(0,1)}\}$ are checked for refinement. Since at this point the sets equal a classical sparse grid approach, all parents are admissible so all children sets $W_{(2,0)}, W_{(0,2)}, W_{(1,1)}$ are considered. Next, we have to check whether the refinement criteria of at least one grid point in every parent set exceeds the threshold ϵ . The children set $W_{(2,0)}$ has only $W_{(1,0)}$ as a parent whose points are located along the non linearity and show a high refinement criteria. In contrast to that, the $W_{(0,2)}, W_{(1,1)}$ sets have $W_{(0,1)}$ as a parents set, whose points lay perpendicular to the target function. Thus, the points have a local error indicator of 0. This leads to the expected behavior of neglecting the refinement step in the second dimension. Further refinement follows the same scheme, except that refinement spaces in the second dimension are neglected because of the non admissibility of the parent sets. With this, the method avoids to include higher order refinement spaces without the coarser refinement spaces.

The resulting SG can be seen in the lower right panel of figure 3.6 with a high density of points along the first dimension and only a coarse resolution for the second dimension. Although the method adapts the intrinsic dimensionality effectively, the subspaces of local behaviour are unnecessarily accurate. Since the jump lies only in the right part of the space, the high density of points along the axis (in the left part) produces unnecessary computational effort. To avoid these is the subject for local refinement.

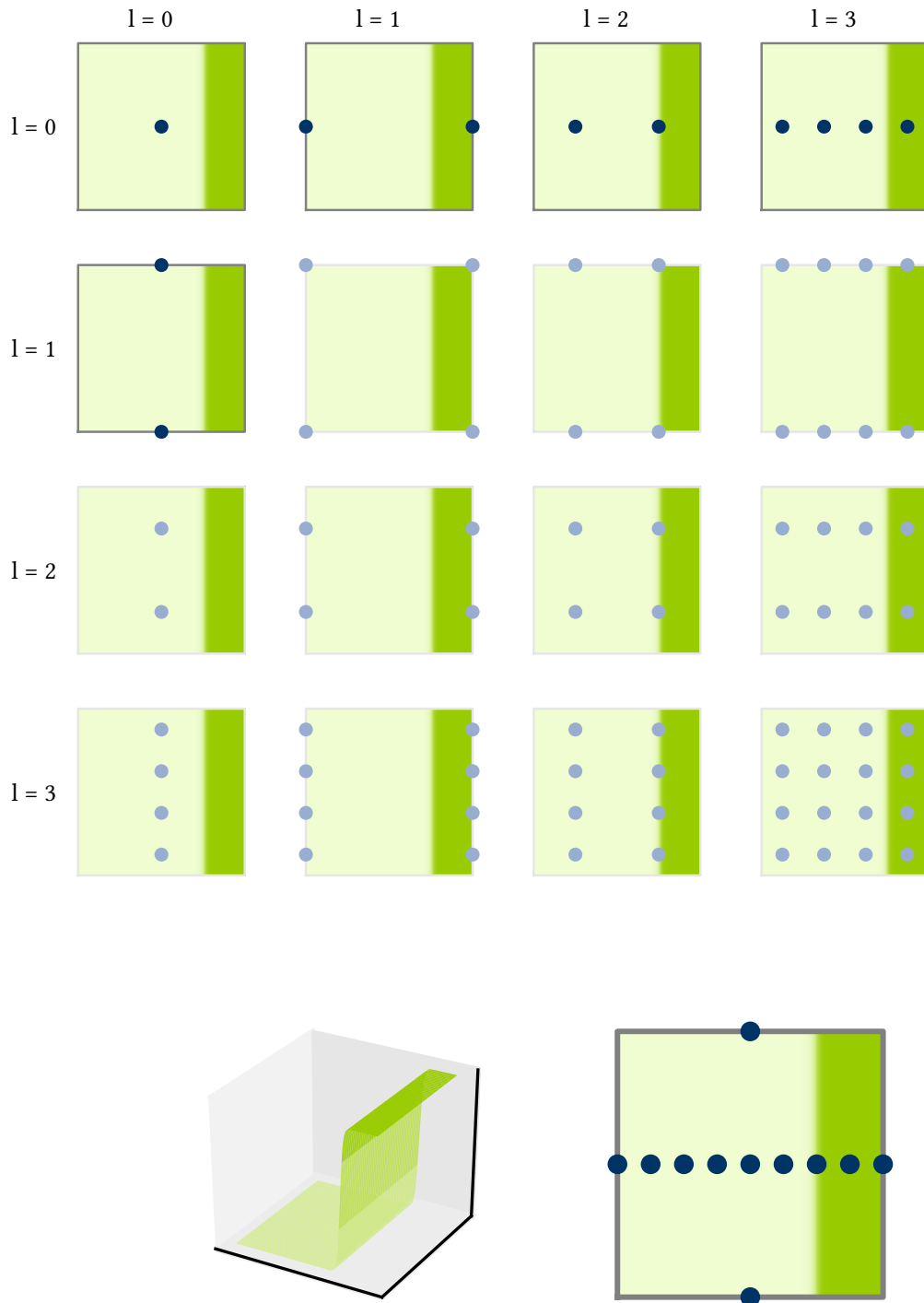


Figure 3.6: Scheme for the dimension adaptivity. *Upper panel:* All possible hierarchical difference spaces up to $|l|_{\infty} \leq 3$ in 2D and the corresponding grid points. Solid color symbols the refinement spaces, considered by the dimension adaptivity while the opaque color symbols the spaces that are neglected by the method. Thus all refinement spaces according to dimension 1 are included while the second dimension is only refined once. *Lower panel, left:* Target function of a sigmoid character that needs to be approximated by the Sparse Grid. *Lower panel, right:* Final Sparse Grid with dimension adaptivity.

Algorithm 1: Dimension Adaptivity

```
 $\mathcal{S} = \{(\mathbf{0}, \mathbf{0})\}$   
 $\mathcal{S}_{dim} = \{(\mathbf{0}, \mathbf{0})\}$   
 $L = 1$   
while  $\mathcal{S}_{dim} \neq \emptyset$  do  
   $\mathcal{S}_{dim} = \emptyset$   
  for  $(\mathbf{i}, \mathbf{l})$  in  $\mathcal{S}$  with  $|\mathbf{l}|_1 = L$  do  
    for  $k := 1, \dots, D$  do  
      refine = False  
       $\mathbf{l}_c = \mathbf{l} + \mathbf{e}_k$   
      for  $n := 1, \dots, D$  do  
         $\mathbf{m} = \mathbf{l}_c - \mathbf{e}_n$   
        if  $\mathbf{m} \in \mathcal{S}$  then  
           $\gamma_{\mathbf{m}} := \max(\{\gamma_{\mathbf{m},j} : j \in \mathcal{I}_{\mathbf{m}}\})$   
          if  $\gamma_{\mathbf{m}} \leq \epsilon$  then  
            refine = True  
            break  
        if refine == True then  
           $\mathcal{C} := \{(\mathbf{i}, \mathbf{l}_c), \text{ for } \mathbf{i} \in \mathcal{I}_{\mathbf{l}_c}\}$   
           $\mathcal{S}_{dim} = \mathcal{S}_{dim} \cup \mathcal{C}$   
     $L = L+1$   
   $\mathcal{S} = \mathcal{S} \cup \mathcal{S}_{dim}$ 
```

3.3.3 Local Adaptivity

Similar to the dimension-adaptive refinement, the local adaptivity also tries to minimize the points for an accurate interpolation. Since the Full Sparse Grid treats every dimension and interaction the same, it can not adapt to local phenomena, like strong non-linearity. The sparse grid method should only refine in subdomains of high variability and the relationship of singular points belonging to consecutive levels imply a selective refinement. In the dimension refinement, we concentrated on the refinement spaces-wise. For the local refinement, we are using BFs explicitly and the definition of *parents* and *children*. We assume that if the contribution of a parent is small, the contribution of the children are at least equally small. For some functions we expect for higher levels an exponential decrease of the surplusses, so that the children will have a lower impact as their parent. As a measure for the contribution we use the indicator $\gamma_{\mathbf{l},\mathbf{i}}$ from Eq. (3.40) for every single BF. We calculate $\gamma_{\mathbf{l},\mathbf{i}}$ of all BFs $|\mathbf{l}|_1 = L$, where L is the maximum level of the sparse grid at the current refinement step, and for those BFs, whose indicator exceeds the predefined threshold ϵ we include the children [40, 112]. This means, with every refinement step we include BFs that correspond to the refinement set

$$\mathcal{S}_{\text{loc}} = \left\{ (\mathbf{l}, \mathbf{i}) \mid |\mathbf{l}|_1 = L + 1 \wedge \max_{(\mathbf{j}, \mathbf{m}) \in \mathcal{P}(\mathbf{l}, \mathbf{i})} (|v_{\mathbf{j}, \mathbf{m}} \cdot w_{\mathbf{j}, \mathbf{m}}|) > \epsilon \right\} \quad (3.46)$$

where $\mathcal{P}(\mathbf{l}, \mathbf{i})$ denotes the set of parents of the BF $\varphi_{\mathbf{l}, \mathbf{i}}$ with the definition of Eq. (3.15).

The local refinement of the sparse grid is done by refining the difference space, and we consider all spaces with the current highest refinement level $|\mathbf{l}|_1 = L$ and check the local error for every BF within these spaces. As we indicated in section 3.3.2, for all the refinement spaces in this step, we do not include all BFs of $|\mathbf{l}|_1 = L + 1$ but only those whose parents have a high contribution. Figure 3.7 shows these characteristics of the refinement strategy for the underlying function. These characteristics are the same as the target function Eq. (3.45) of the dimension adaptivity, as the lower-left panel shows. All possible refinement spaces up to a level of $|\mathbf{l}|_\infty = 3$ are displayed which would construct a full grid if we considered all of them. As mentioned before, the refinement also operates on the hierarchical difference spaces $W_{\mathbf{l}}$ with $|\mathbf{l}|_1 = L$, similar to the dimension adaptivity. The initial set consists only of the singular set of $\mathcal{S} = \{\mathbf{0}, \mathbf{0}\}$ with $\mathbf{0} = \{0, 0\}$. In order to locally refine this set, we consider all possible refinement spaces with $|\mathbf{l}|_1 = 1$, $W_{\mathbf{l}} = \{W_{(1,0)}, W_{(0,1)}\}$. Next the parents of every individual point $\{(\mathbf{i}, \mathbf{l}), |\mathbf{l}|_1 = 1\}$ are calculated with Eq. (3.15) and checked whether the refinement criteria exceeds the threshold ϵ . The first row of the figure 3.7 denotes these with red crosses (✖). Since for the first step, the parent is the point to be refined itself, all children are produced. This procedure is repeated until the accuracy meet the expectations.

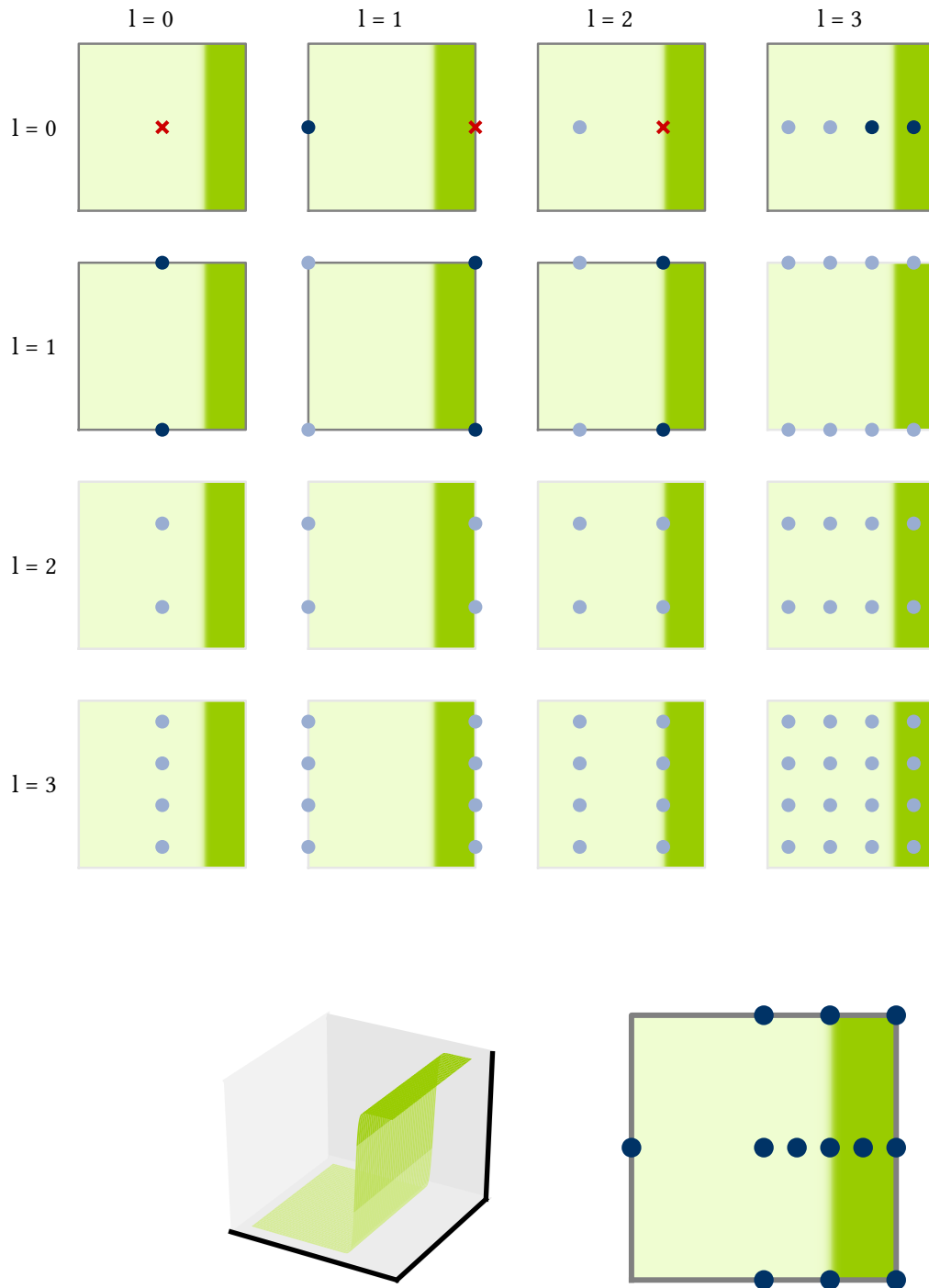






Figure 3.7: Scheme for the local adaptivity. *Upper panel:* All possible hierarchical difference spaces up to $|\mathbf{l}|_\infty \leq 3$ in $2D$ and the corresponding grid points. The solid darkblue points \bullet and the red crosses \times map out the points included by the local adaptivity. The opaque light blue \bullet color symbols the points that are neglected by the method. The red crosses denote the points, which carry a high local error and need to be refined. Thus the method concentrates the points closest to non linear behavior. *Lower panel, left:* Target function of a sigmoid character, that needs to be approximated by the Sparse Grid. *Lower panel, right:* Final Sparse Grid with local adaptivity.

Hence, for the second step, all refinement spaces with $|\mathbf{l}|_1 = 2$ are checked for refinement. The space of $W_{(0,2)}$ is excluded, since no parent in $W_{(0,1)}$ has a high refinement criteria. For the spaces of $\{\mathbf{l} = (2, 0), (1, 1)\}$, the points in $\mathbf{l} = (1, 0)$ are significant. Points in the area of constant behavior (denoted by the light green area ) have a small refinement criteria and thus all their children are neglected (denoted by the lightblue points ). In contrast to that, the points along the non linear behavior have a high refinement criteria (red crosses ) and therefore all the children are included in the sparse grid (denoted with the darkblue points ). Continuing with this strategy we arrive at the final SG points for function f , shown in the lower right panel of figure 3.7. Similar to the dimension-adaptive approach some unnecessary points in the second dimension that do not contribute to the accuracy of the surrogate model can also be found here. Hence, we used a combination of the dimension-adaptive and the local-adaptive strategy to minimize the computational effort.

Algorithm 2: Local Adaptivity

```

 $\mathcal{S} = \{(\mathbf{0}, \mathbf{0})\}$ 
 $\mathcal{S}_{loc} = \{(\mathbf{0}, \mathbf{0})\}$ 
 $L = 1$ 
while  $\mathcal{S}_{loc} \neq \emptyset$  do
     $\mathcal{S}_{loc} = \emptyset$ 
    for  $(\mathbf{i}, \mathbf{l})$  in  $\mathcal{S}$  with  $|\mathbf{l}|_1 = L$  do
        if  $\gamma_{\mathbf{i}, \mathbf{l}} \leq \epsilon$  then
             $\mathcal{C} := \text{findChildren}((\mathbf{i}, \mathbf{l}))$ 
             $\mathcal{S}_{loc} = \mathcal{S}_{loc} \cup \mathcal{C}$ 
     $L = L + 1$ 
     $\mathcal{S} = \mathcal{S} \cup \mathcal{S}_{loc}$ 

```

3.3.4 Combination: Dimension and Local Adaptivity

Although both refinement strategies reduce the number of sparse grid points compared to the original sparse grid approach, the target sigmoid function is not displayed efficiently with either of the methods. As the figure 3.6 and 3.7 show, both of the refinement strategy produce unnecessary points. For the 2D- case, this does not have a big impact on the sparse

grid's performance, but for functions in a higher dimensional case, this will be a bottleneck in the performance. Since the problems we are dealing with show the characteristic of low intrinsic dimensionality and locally rapid changes, we propose a combination of both refinement strategies, similar to the *h*-GSG method [40]. Both methods declare different points as unnecessary regarding the refinement goal. Since we want the minimum amount of points, we only want to include points declared as important by both refinement strategies, combined with the same error indicator. Therefore we chose all different spaces that need to be refined at $|I|_1 = L$ and check if their refinement spaces have parents spaces with a high error indicator, forming a new set of refinement spaces. For every possible new grid point in this refinement spaces, we then check whether the parents $\mathcal{P}(\mathbf{i}, \mathbf{I})$ have a local error that is above the threshold ϵ . Only if the contribution of one parent is high enough, we include the point. This is how we can include more refinement spaces than one and employ parallel computing for the different points. In general, we take the intersection of the two refinement sets \mathcal{S}_{dim} (Eq. 3.44) and \mathcal{S}_{loc} (Eq. 3.46).

$$\mathcal{S}_{\text{dim,loc}} = \mathcal{S}_{\text{dim}} \cap \mathcal{S}_{\text{loc}}. \quad (3.47)$$

The algorithm 3 shows the detailed method of the dim-loc adaptivity. For our example of the sigmoid function, this means a further reduction of points. Figure 3.8 displays the approach for the target function. Therefore we chose an initial grid (●) of $|I|_1 = 1$, since the first subspace is always refined. The red crosses ✖ show all potential sparse grid points of the refinement step. If we produce the refinement sets according to the dimension adaptivity (upper right panel) and the local adaptivity (lower left panel), the considered points reduce to 2 and 3 points, respectively. However, the intersection of both sets shows how many points are actually needed to increase the interpolation accuracy, which is only a single point along with the non-linearity.

3.4 Testing refinement strategies

Before applying the refinement strategies to a first-principles kinetic model, as introduced in section 2.1, we test the performance of the different strategies in four test cases, illustrated in the figure 3.9 for 2D. Two of the cases will have similar characteristics as kinetic data for the TOF and the coverage of one species, described in the upper row of the figure 3.9. The other two are classical cases for integration problems to show how good the methods work for already known problems. Therefore we use the shifted continuous function and the

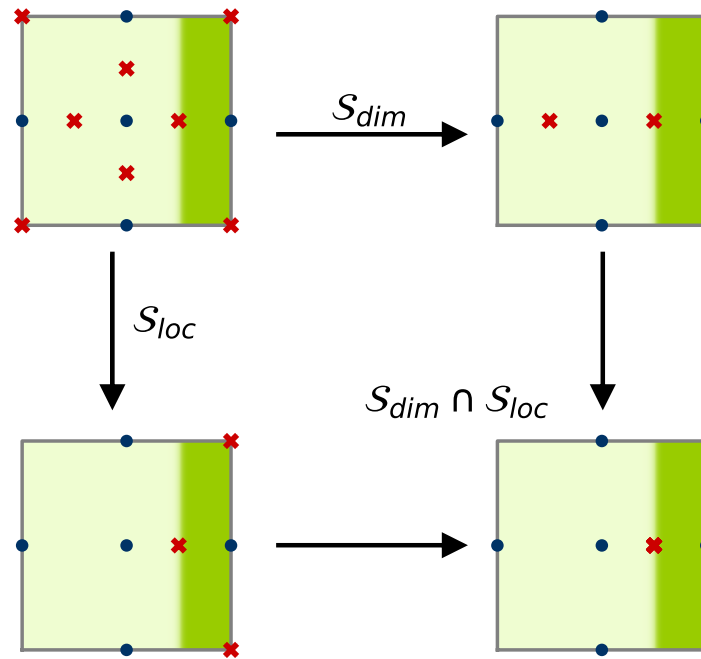


Figure 3.8: Combination technique of the dimension and local adaptivity. *Upper left panel:* Initial grid of $|l|_1 = 1$ \bullet and all possible refinement points of $|l|_1 = 2$ \times . *Upper right panel:* Reduction of refinement points according to the dimension adaptivity; thus only points in the first dimension along the variation of the function are considered. *Lower left panel:* Reduction of refinement points according to the local adaptivity; thus only point in the area of the variation are considered. *Lower right panel:* The intersection of both refinement sets is added to the Sparse Grid construction.

Algorithm 3: Dimension and Local Adaptivity

```

 $\mathcal{S} = \{\mathbf{0}, \mathbf{0}\}$ 
 $\mathcal{S}_{dim,loc} = \{\mathbf{0}, \mathbf{0}\}$ 
 $L = 0$ 
while  $\mathcal{S}_{dim,loc} \neq \emptyset$  do
     $\mathcal{S}_{dim,loc} = \emptyset$ 
    for  $(\mathbf{i}, \mathbf{l})$  in  $\mathcal{S}$  with  $|\mathbf{l}|_1 = L$  do
        if  $\gamma_{\mathbf{l}} > \epsilon$  then
            for  $k := 1, \dots, D$  do
                 $\mathbf{l}_c = \mathbf{l} + \mathbf{e}_k$ 
                if all of  $\mathbf{m} := \mathbf{l}_c - \mathbf{e}_n \forall n := 1, \dots, D$  have one point with  $|\gamma_{\mathbf{m},j}| > \epsilon$ , with
                     $j \in \mathcal{I}_{\mathbf{m}}$  then
                        for  $i$  in  $I_{\mathbf{l}}$  do
                            if  $|\gamma_{\mathbf{l},i}| > \epsilon$  then
                                 $\mathcal{S}_{loc} = \text{findAxialChildren}(\mathbf{i}, k)$ 
                                 $\mathcal{S}_{dim,loc} = \mathcal{S}_{dim,loc} \cup \mathcal{S}_{loc}$ 
             $\mathcal{S} = \mathcal{S} \cup \mathcal{S}_{dim,loc}$ 
             $L = L + 1$ 

```

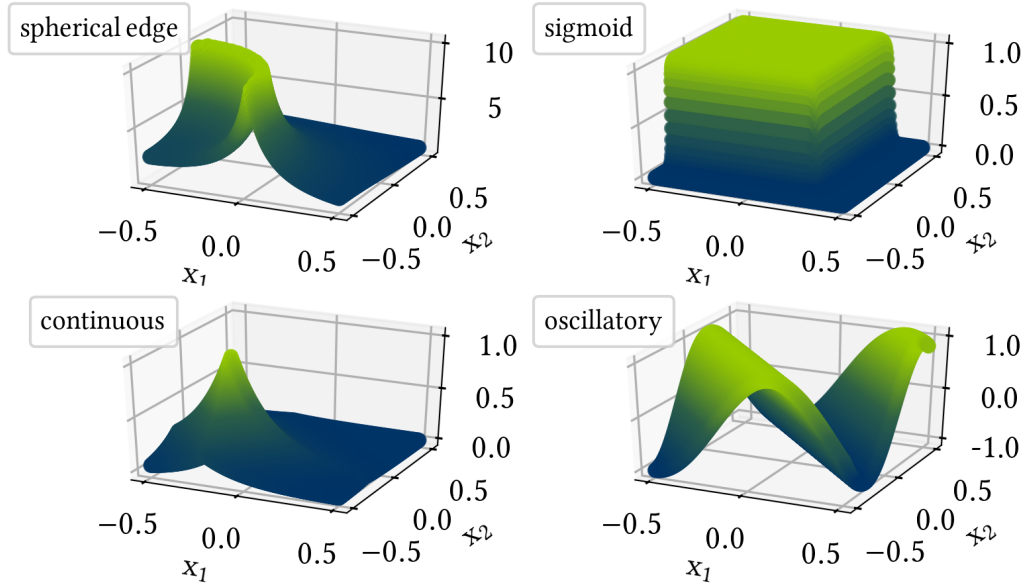


Figure 3.9: Four test functions for the refinement strategies. *Upper panel:* Significant functions for kinetic data that show the characteristics of the TOF of kinetic data (spherical edge, *left figure*) and the coverage for one species (sigmoid, *right figure*). *Lower panel:* Classical examples for integration problems, introduced by Genz. *Left figure* shows the shifted peak of the continuous function and the *right figure* shows the oscillatory function.

oscillatory function, introduced by Genz in 1984 [131], which are displayed in the lower row of figure 3.9. To illustrate the dimension-adaptive methods, we test the functions in different dimensions, $4D$ and $8D$, including 2 and 6 dummy dimensions, that do not impact the function value. To outline the characteristics for all three refinement methods, we will display the final sparse grid of each method for $D = 2$ corresponding to a threshold of $\epsilon = 0.01$. For examining the accuracy of the interpolation, we analyse the measures of the L_1 -norm of the interpolation error and the quadrature error.

3.4.1 Spherical edge test function

The overall goal is to find a method that is most effective for kinetic data. Therefore we choose a function whose characteristics are similar to that kind of data, i.e., with local rapid transition between regimes of a very smooth (almost linear) behaviour. Like the following function in $2D$

$$f_{SE}(x_1, x_2) = \frac{1}{|0.3 - (x_1 + 0.5)^2 - (x_2 + 0.5)^2| + 0.1}. \quad (3.48)$$

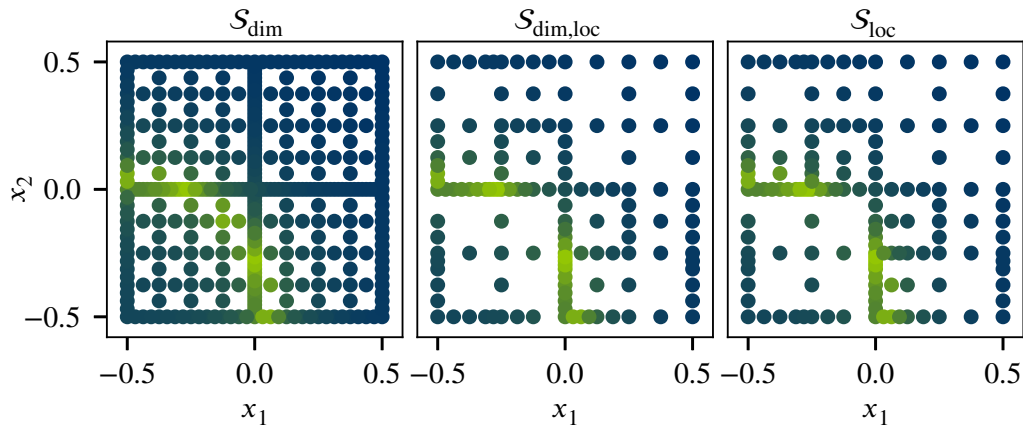


Figure 3.10: Final Sparse Grid for the approximation of f_{SE} (spherical edge) with a threshold of $\epsilon = 0.01$. The lighter green dots indicate the kink of the function. *Left panel* Grid points for the DA refinement strategy which equals a SG without adaptivity. *Right panel* Grid points for the LA refinement strategy, which resolves in a higher resolution in the area of the non-linear kink. *Middle panel* Grid points for the combined refinement technique whose refinement is mostly dominated by the LA and results in a similar grid as LA.

As shown in the upper left panel of figure 3.9, f_{SE} has a kink located on a radius of 0.5 around $[-0.5, -0.5]$, which previously has been used by Jakeman and Roberts [40]. Since the function depends on all dimensions in the 2D case, we expect a limited impact of the dimension-adaptive refinement for the SG construction. The final SG of the three different refinement strategies for the 2D case are displayed in figure 3.10 with the dimension adaptivity (DA) on the left side and the local adaptivity (LA) on the right side. The combined refinement strategy is located in the middle. All refinement strategies are tested with a threshold of $\epsilon = 0.01$.

The dimension-adaptive refinement refines uniformly since the function is dependent on both dimensions; thus, the intrinsic dimensionality is 2. However, the local adaptivity adapts the kink with a finer grid and resolves the smooth behaviour with fewer points. The combined technique grid is similar to the final grid of the LA. As we expected, the reduction of points is dominated by the LA and differs from the final LA grid only by 6% fewer points. A reduction of points is one important criterion of adaptivity. Another one is the accuracy of the surrogate model. Detailed data can be found in the table 7.1 in the Appendix.

As a final measure for the refinement accuracy, we are primarily interested in the quadrature error, but considering using the surrogate model for a sensitivity analysis, we also include the L_1 -norm of the interpolation error. As mentioned before, the advantage of the dimension adaptivity is very limited for the 2D case, so that we include two higher-dimensional cases, 4D and 8D, but with an intrinsic dimensionality of 2. Thus, we introduce 2 and 6 dummy

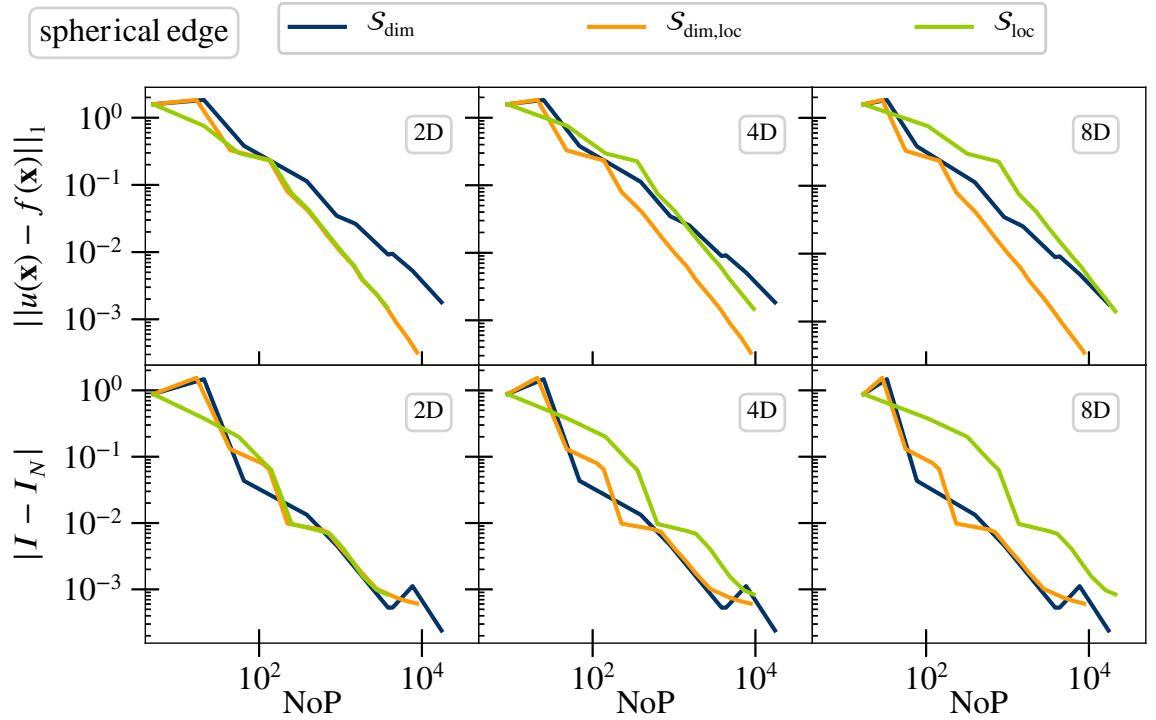


Figure 3.11: Performance of the three refinement methods for the L_1 -norm of the interpolation error (*upper panel*) and the quadrature error (*lower panel*) in case of the spherical edge function. Displayed are three different dimensional cases. All of them have an intrinsic dimensionality of 2, and in the case of $D = 4$ two and the case of $D = 8$, six dummy dimensions are added. In all 6 cases, the combined technique $\mathcal{S}_{\text{dim,loc}}$ performs the best. The increasing dimensionality worsens the performance of the \mathcal{S}_{loc} method due to the increasing refinement points in all dimensions. Contrary to that, the DA method can not efficiently adapt the function's local features but gives a two-dimensional full sparse grid.

parameters that have no impact on $f_{KM}(\mathbf{x})$ and can be excluded by the DA. These cases are closely related to the real kinetic data since the outcome does not show a sensitivity to most parameters in a kinetic model. Figure 3.11 shows the results of all three cases, 2D, 4D and 8D, plotted over the number of points (NoP) for decreasing tolerances. The *upper panel* displays the results for the L_1 -norm of the interpolation error and the *lower panel* the integration error. The final grid results (figure 3.10) already indicated the results for the L_1 -norm of the interpolation. For higher tolerances, i.e., smaller NoP, all three refinement strategies differ marginally from each other. However, for decreasing tolerance, the dimension adaptivity shows the worst of all three performances. The DA reaches a high accuracy, but the NoP are higher compared to the LA and the combined technique ($\mathcal{S}_{\text{dim,loc}}$). The LA and the combined technique exhibit almost an identical performance, which is analogue to the grid results. With

increasing outer dimensionality, the impact of the **DA** becomes stronger so that with a $D = 8$, the **LA** and **DA** end up at the same accuracy of the interpolation error with the same **NoP**. However, we have to point out that the **LA** method shows a steeper slope, which leads to the conclusion that it will perform better with decreasing tolerance than the **DA**. Not only does the combined technique performs best for all three cases of different dimensionality, but the results are almost identical, which shows the efficiency of the adaptation. The data in table 7.1 in the Appendix display less than a 10% deviation in points and less than a 0.1% deviation in the L_1 -norm of the interpolation and the quadrature error.

The refinement strategies show a different result for the accuracy of the expected value displayed in the lower panel. All refinement methods perform similarly for the outer dimension of 2, and no strong deviation can be seen. Although the **DA** shows a higher local error for the L_1 -norm, it approximates the expected value as accurately as the other methods. The full two-dimensional **SG** gives an accurate approximation of the integral by cancelling out the local errors. The impact of the dimension-adaptivity increases with higher dimensionality. For the $D = 8$ case with an intrinsic dimensionality of 2, the full **SG** \mathcal{S}_{dim} approximates the expected value as well as the combined technique. In contrast to that, the **LAs** performance declines with the dimensionality as it produces points in 6 dummy directions, which do not improve the accuracy.

Although the combined technique $\mathcal{S}_{\text{dim,loc}}$ show its strong advantage in the case of the interpolation error, it cannot outperform the **DA** in case of the expected value. This example already shows the limits for both classical approaches, **DA** and **LA**. While the \mathcal{S}_{loc} has problems with unimportant dimensions, the \mathcal{S}_{dim} fails to adapt local features, which makes it less favourable for the $2D$ case. Overall, the combined technique presents a solid performance in all six cases.

3.4.2 Sigmoid function

Most of the times, the **TOF** is the primary measure for the efficiency of the catalyst or the reaction, but for further understanding, the coverage of the surface can provide additional information. The surface coverage is highly sensitive to the input parameter and shows different regimes of dominant species according to various areas of our domain. These regimes can indicate a bottleneck of the reactions or give general information about active or non-active areas. In most cases, the surface is occupied with only one species. Thus the value of the coverages of one species is either constantly high or low. For the second model-related function, we use the sigmoid function, shown in the left upper panel of figure 3.9, whose

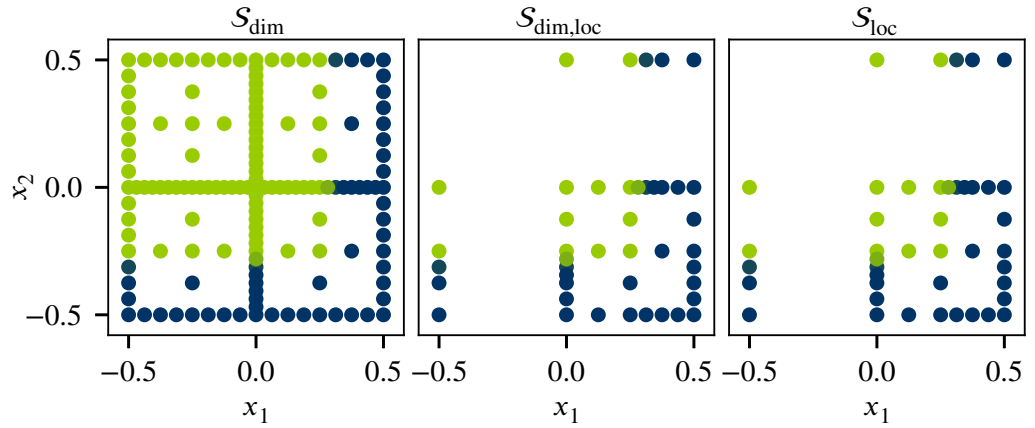


Figure 3.12: Final Sparse Grid for the approximation of the sigmoid function f_S with a threshold of $\epsilon = 0.01$. *Left panel* Grid points for the DA refinement strategy which equals a SG without adaptivity. *Right panel* Grid points for the LA refinement strategy, which resolves in a higher resolution in the area of the sharp changes and excludes the points in the constant areas. *Middle panel* Grid points for the combined refinement technique whose refinement is mostly dominated by the LA and results in a similar grid as LA.

characteristics often describe regimes' behaviour with one dominant species on the surface and sharp transitions in between. We choose constant plateaus to display a high occupation with rapid changes to a low occupation. The final function

$$f_S(x_1, x_2) = \frac{1}{1 + e^{(x_1 - 0.3) \cdot 100}} \cdot \frac{1}{1 + e^{(-x_2 - 0.3) \cdot 100}}, \quad (3.49)$$

has local changes not aligned with the axes and only varies between 0 and 1. Here, we also include the cases with 2 and 6 dummy dimensions. Figure 3.12 shows the drastic difference between the dimension-adaptive and the local-adaptive refinement strategies. For the 2D example, the dimension-adaptive method cannot produce fewer points than a full sparse grid since both dimensions are important. Detailed data are given in the table 7.2 in the Appendix. Most subdomains of the function have a constant value, which does not need a high amount of points. More effectively refined are the grids of the combined techniques and the local refinement strategy. These adapt the rapid changes at $[0.3, -0.3]$ and exclude the points in the constant areas. Only in the lower right quarter, where the corner of the sharp changes is located, do their grids need more refinement. Based on these results, the combined and the local method's performance should give similar results for the interpolation error and the expected value, which are shown in figure 3.13. The upper row shows the results for the L_1 -norm of the interpolation error in all three cases, 2D, 4D and 8D, and the lower row shows

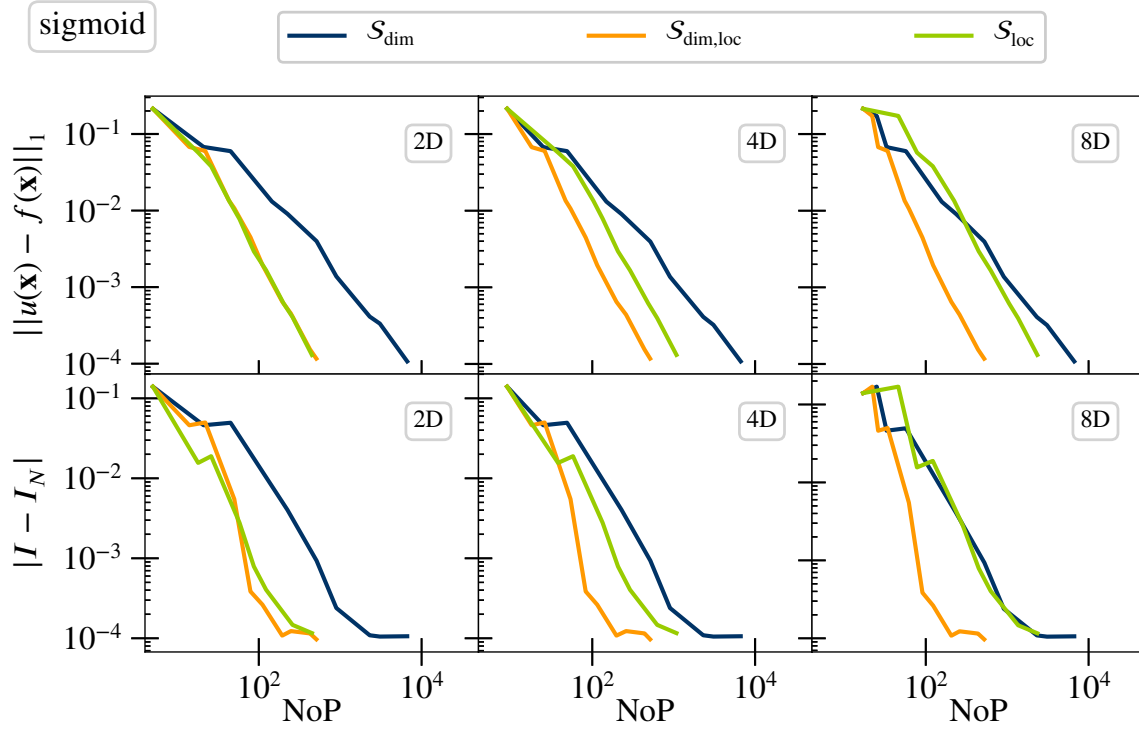


Figure 3.13: Performance of the three refinement methods for the L_1 -norm of the interpolation error (*upper panel*) and the quadrature error (*lower panel*) in case of the sigmoid function. Displayed are three different dimensional cases in which all of them have an intrinsic dimensionality of 2. In the case of $D = 4$ (*middel panel*) 2 and in the case of $D = 8$ (*right panel*) 6 dummy dimension added. In all 6 cases, the combined technique $\mathcal{S}_{\text{dim,loc}}$ performs the best. The increasing dimensionality worsens the performance of the \mathcal{S}_{loc} method due to the increasing refinement points in all dimensions. In contrary to that, the DA method can not efficiently adapt the local features of the function, but the intrinsic dimensionality and therefore gives a two-dimensional full sparse grid for all the dimensional cases. This leads to DA and LA showing the same refinement behaviour along the accuracy for $D = 8$ for the expected value.

the expected value error. As expected from the results of the final grid, the local-adaptive refinement strategy works better than the dimension-adaptive refinement in the 2D case. The DA uses an order of magnitude more points for an accuracy of 10^{-4} .

Especially noticeable is the almost identical performance for the interpolation error of the LA compared to the combined technique, showing that it is mainly influenced by the LA. The results are different if we move to a higher dimensionality of the function. These insignificant dimensions are adapted by the DA method, resulting in an identical L_1 -norm with increasing dimensionality in all cases. Similar results are visible for the combined technique, which effectively excludes the unimportant dimensions and shows the best performance throughout all cases. The performance of the LA, on the other hand, worsens with more dummy dimensions. Although the LA can implicitly exclude dimensions, the method produces children in all dimensions for every refined point, causing it to be less effective than the other methods. Regarding the expected value, the results show a significant difference to the kinetic model example results. Where in the previous example, the DA was able to approximate the expected value efficiently, it performs worse than the LA and the combined technique, in this case. In this example, the local features are so strong and sharp that the local adaptivity advantage leads to minimal points and, therefore, outperforms the dimension adaptivity. Analog to the cases of the interpolation error, the impact of the DA increases with dimension so that for $D = 8$ DA and LA show almost identical results. In summary, of all six cases, the combined technique performs the best and reduces the NoP up to one order of magnitude compared to the other refinement methods.

3.4.3 Continuous function

Recent studies [40, 102] have already shown the potential of sparse grids in the manner of the L_1 -norm and the quadrature for certain problems. Now we are going to choose two examples of general integration test problems [131, 132]. The first example is the continuous function, with a strong peak and four edges towards the corners of the domain

$$f_{\text{CF}}(x_1, x_2) = \exp\left(-\sum_{i=1}^2 5 \cdot |x_i + 0.2|\right), \quad (3.50)$$

where we shifted the peak to $[-0.2, -0.2]$, as it can be seen in the lower left panel of figure 3.9. The shift increases the complexity of the function and the difficulty for the characteristics adaptation. For the next two functions we limit the results to the 2D and the 8D cases to

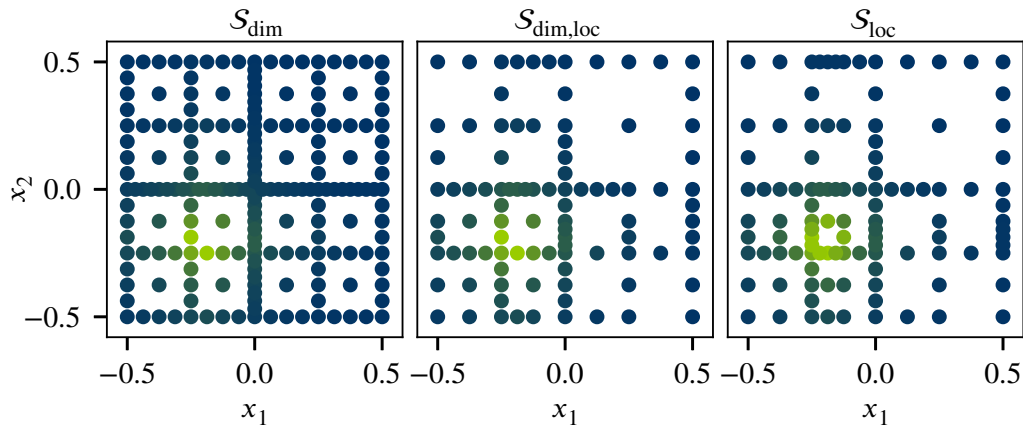


Figure 3.14: Final Sparse Grid for the approximation of f_{CF} (continuous function) with a threshold of $\epsilon = 0.01$. *Left panel* Grid points for the **DA** refinement strategy which equals a full SG for $D = 2$. *Right panel* Grid points for the **LA** refinement strategy which resolves in a higher resolution towards the non-linear peak at $[-0.2, -0.2]$. *Middle panel* Grid points for the combined refinement technique whose refinement is mostly dominated by the **LA** and results in a similar grid as **LA**, except for a smaller number of points. The combined technique uses 37% fewer points than the **LA** method.

capture the trends of the different methods. Figure 3.14 displays the final grids for the different refinement methods. Like the results before, the dimension refinement equals a full sparse grid since both dimensions are important. The similarity between the local refinement grid and the combined technique grid indicates that local adaptivity dominates the combined technique refinement. Consequently, the final grid shows a higher density of points in the area of the peak.

The detailed data, documented in table 7.3 in the Appendix, indicates an stronger impact of the **DA**, as the final points $\mathcal{S}_{loc} \setminus \mathcal{S}_{dim,loc}$ differ almost 40%. Therefore the combined technique effectively excludes points through the dimension-adaptive feature. Figure 3.15 outlines that the dimension-adaptive method provides a good approximation of the function with more **NoP** than the other methods. The combined method is half an order of magnitude better for the L_1 -norm of the interpolation error and less than half an order of magnitude better for the expected value than the dimension-adaptive method. Since the intrinsic dimensionality does not change, these results are valid for both dimensions. Although the peak is a very local feature, the edges are spanned over the major part of the domain, causing good results for the **DA**, whose points cover the whole domain.

For the 2D case, the **LA** method shows almost identical results as the combined technique for the L_1 -norm of the interpolation error. For the quadrature error, on the other hand, there is a deviation from the combined technique. It still performs slightly better than the **DA** method

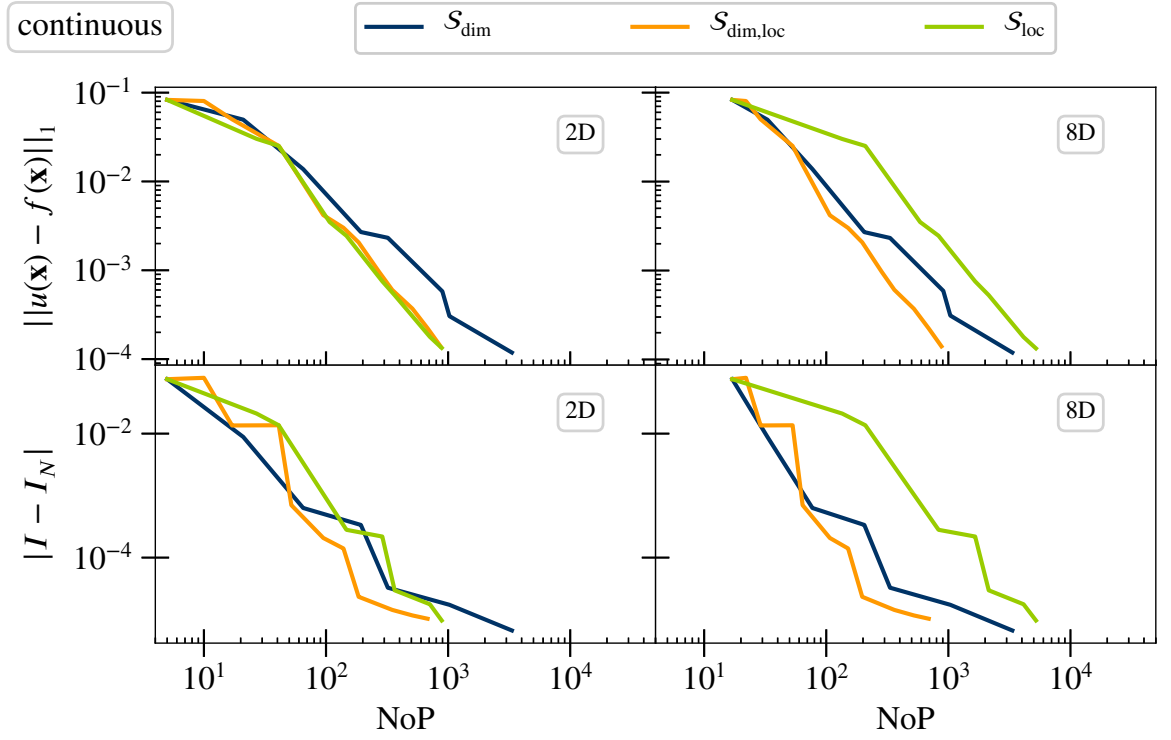


Figure 3.15: Performance of the three refinement methods for the L_1 -norm of the interpolation error (*upper panel*) and the quadrature error (*lower panel*) in case of f_{CF} (continuous function). Displayed are two different dimensional cases in which all of them have an intrinsic dimensionality of 2. In the case of $D = 8$ (*right panel*) 6 added dummy dimension. In all 4 cases, the combined technique $\mathcal{S}_{\text{dim,loc}}$ performs the best. The increasing dimensionality worsens the performance of the \mathcal{S}_{loc} method due to the increasing refinement points in all dimensions. Contrary to that, the DA method can not efficiently adapt the local features, which causes a factor 5 of the NoP to achieve the same accuracy as the combined technique for both measures.

for small tolerance, but the clear advantage, as in the L_1 -norm, cannot be determined. For the more extreme case of the six dummy parameters, the results become worse for the local refinement. It adapts the peak but refines in every dimension, and as we mentioned before, the features of the function are spanned over the whole domain, causing a significant shift in the NoP. For the expected value, the LA performs even worse. While for the 2D case, there was only a marginal difference, for the 8D case, it requires about one order of magnitude more points.

3.4.4 Oscillatory function

As the last example we choose a function whose characteristics are not favorable for local refinement strategies. The oscillatory function

$$f_{\text{OF}}(x_1, x_2) = \cos\left(\pi + \sum_{i=1}^2 5 \cdot (x_i + 0.5)\right), \quad (3.51)$$

shows no local features, but a dependence on the whole parameter space with a high degree of smoothness displayed in the lower right panel of figure 3.9. Figure 3.16 presents that all refinement strategies need almost a full sparse grid. The local and the combined refinement strategies have a higher density in the curved area of the function and the final grid of the DA method. However, it shows no adaptation, uses fewer points than the LA method. All detailed data are given in the table 7.4 in the Appendix.

Hence the combined technique to adapt the curves at the highest and lowest points of the oscillatory function with less of a dense grid than the LA method. For such a smooth function higher-order polynomial basis function would be a better choice. That is why the piecewise linear basis functions need extra refinement in the curved area. From these results, none of the adaptive refinement strategies has a clear advantage over the others. Figure 3.17 underlines this impression. All refinement results are almost identical for the 2D case and the interpolation error (upper left panel). The results show some fluctuations for the expected value (*lower left panel*), but no strategy strongly outperforms the others. For the 2D case, none of the methods shows a big advantage or disadvantage for a function with no local characteristics but depends on the whole parameter space.

Analog to the previous results for the increasing dimensionality, we can only see an effect in the LA results. While the performance of the DA and the combined technique are identical to the 2D case, the LA shows a shift to a higher number of points. Compared to the previous

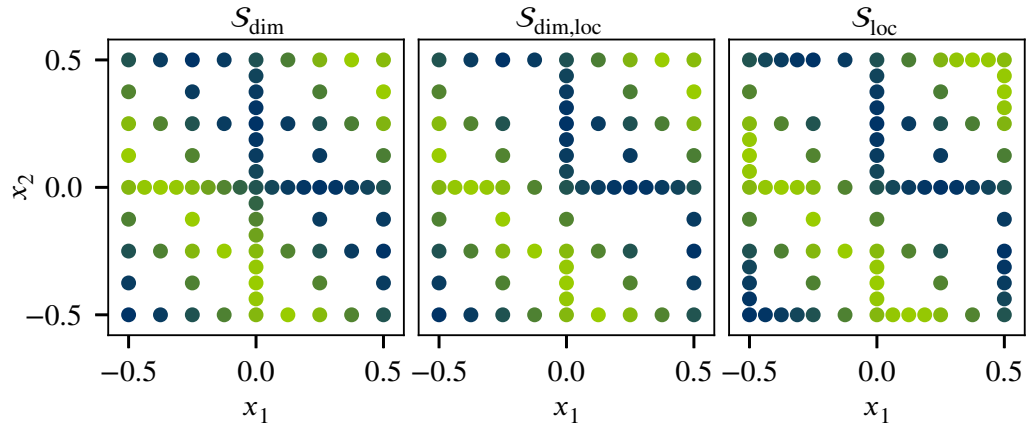


Figure 3.16: Final Sparse Grid for the approximation of f_{OF} (oscillatory function) with a threshold of $\epsilon = 0.01$. *Left panel* Grid points for the DA refinement strategy which equals a full SG but results in less points than the LA grid. *Right panel* Grid points for the LA refinement strategy, which resolves in a slightly higher resolution in the curved area. *Middle panel* Grid points for the combined refinement technique, which shows a slightly higher density in the area of the dense curves like the LA but with fewer points in total. This is caused by the excluding feature of the dimension adaptivity feature of the combined technique.

examples, this shift is even more significant because of the non-local features. Before, the increased number of points could be limited by efficiently adapting the local features, but, here, the S_{loc} show an almost full sparse grid. Hence, for increasing dimensionality, the local refinement will produce all children in eight dimensions for the 2D grid, shown in the right panel of figure 3.16.

3.4.5 Higher dimensional functions

We have analysed the performance of the different refinement strategies for kinetic-specific examples and classical functions. However, we also want to include examples with a higher intrinsic dimensionality than 2 and characteristics for an ANOVA decomposition. Therefore

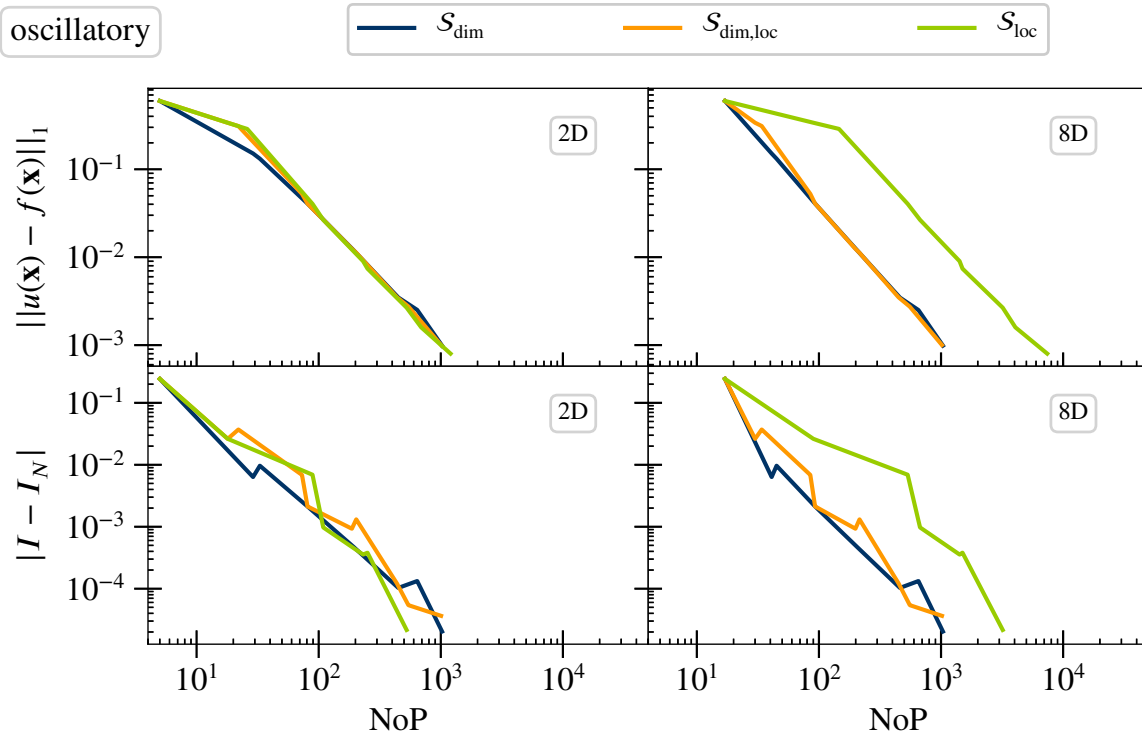


Figure 3.17: Performance of the three refinement methods for the L_1 -norm of the interpolation error (*upper panel*) and the quadrature error (*lower panel*) in case of the oscillatory function. Displayed are two different dimensional cases in which both of them have an intrinsic dimensionality of 2 and in the case of $D = 8$ (*right panel*) 6 added dummy dimension. For the 2D case, all of the methods perform equally well. For the expected value, it appears that the methods have some fluctuations, but none of the methods can outperform the others. Only for the increasing dimensionality, the performance of the LA method worsens due to the increasing refinement points in all dimensions.

we choose

$$f_{2D \text{ sum}} = \sum_{i=1}^D g(x_i, x_{i+1}), \quad D = 6, \quad \text{with } g(x_i, x_{i+1}) = f_{SE}(x_i, x_{i+1}) \quad (3.52)$$

$$f_{\text{prod func}} = \sum_{l=1}^D \prod_{i=1}^l g(x_i), \quad D = 4, \quad \text{with } g(x_i) = \frac{1}{1 + e^{(x_i - 0.3) \cdot 100}} \quad (3.53)$$

as two higher dimensional examples. The first function $f_{2D \text{ sum}}$ describes a sum of lower dimensional functions, whose cross references span a function dependency over the entire parameter domain. The function evaluations of the underlying function g are equal to the spherical edge test function (Eq. (3.48)) from subsection 3.4.1. The shape of the singular summands include a local kink dependent on two parameters.

The second function $f_{\text{prod func}}$ has a similar shape to the ANOVA decomposition. The sum over the products of increasing dimensionality can be compared to contributions of different parameters or parameter interactions. In this case, the underlying function g displays a sigmoid function similar to the example in subsection 3.4.2. Again, the refinement strategy needs to adapt to the local rapid jumps and an increasing dimensionality of the summands. Although both methods have a higher intrinsic dimensionality $D = 6, 4$ than the examples before, they still can be decomposed into smaller dimensional functions. This is comparable to kinetic models, where the data often shows different subdomains where different subsets of input parameters or different interactions between parameters are important.

Figure 3.18 shows the results for both functions for the L_1 - norm of the interpolation error $\|u(x) - f(x)\|_1$ (*upper row*) and the quadrature error $|I - I_N|$ (*lower row*). In the case of $f_{2D \text{ sum}}$ the DA refinement method shows an increasing interpolation error for the first refinement steps with a following decrease to $\|u(x) - f(x)\|_1 = 0.15$. The DA works best if some dimensions can be excluded in the first few refinement steps. However, the crossreferences in the sums avoid a general discard of dimensions and the local features of the underlying function f_{SE} need a few refinement steps to reach an improvement for the function. This and a good initial approximation by chance explain the initially increasing error.

During the refinement, the DA is able to exclude unecessary dimensions and decrease the interpolation error by almost two orders of magnitude. Nevertheless, the DA includes all the points of a refinement space and cannot selectively increase the point density in the area of the local kinks of the summands. Therefore, it needs more points than the combined method $\mathcal{S}_{\text{dim,loc}}$ that needs only a third of the NoP. Although, the DA performs worse than the combined method, it is a better method than the LA, for this function. The LA can adapt the local features but has to include all the points in every dimension, which causes more points than the DA to reach the same accuracy. Considering the single summands to be

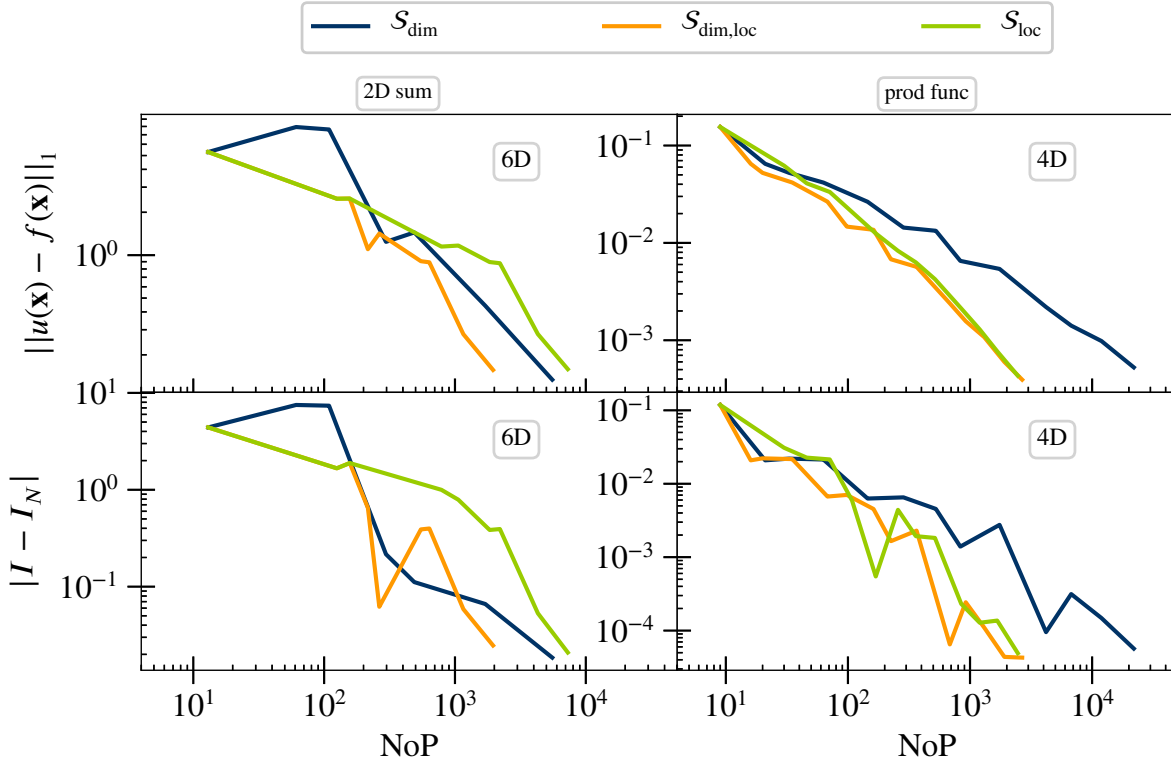


Figure 3.18: Results of the performance of the three refinement methods for the L_1 -norm of the interpolation error (*upper panel*) and the quadrature error (*lower panel*) in case of the $f_{2D \text{ sum}}$ function in the *left column* and the product function $f_{\text{prod func}}$ in the *right column*. The $f_{2D \text{ sum}}$ has a dimensionality of 6 and the product function a dimensionality of 4. Both functions have the characteristic of a parameter dependency that can be decomposed into smaller dimensional contributions. The combined method outperforms the other methods in all 4 cases.

two dimensional, it produces an excess of 4 points with every refined point. Therefore, the method can still decrease the interpolation error, just with a slower conversion rate. This also explains the better results for smaller NoP, where the LA can efficiently refine toward the local features, without the big impact of the point excess, yet. These two features of DA and LA explain why the combined method outperforms the other methods. The disadvantages of DA are balanced out by the additional local refinement and vice versa.

For the quadrature error in the left bottom panel of figure 3.18 we can see similar results, with some fluctuations for the combined method. All methods are able to decrease the quadrature error by more than two orders of magnitude to $|I - I_N| = 0.02$, with the combined method needing fewer points than the DA or the LA.

The right column of figure 3.18 shows the results for the four-dimensional product function Eq. (3.53), where the results differ from the previous example. For the L_1 -norm of the interpolation error, all methods can decrease the error by almost three orders of magnitude with the combined $\mathcal{S}_{\text{dim,loc}}$ and the LA method using around an order of magnitude fewer points than the DA. The increasing dimensionality of the summands prevents a big impact of the dimension adaptivity. The pure DA cannot approximate the local features of the underlying sigmoid function efficiently and reaches the same accuracy only with a larger number of points. The same argument explains why the LA achieves the same accuracy with fewer NoP. Since the summands have an increasing dimensionality up to $D = 4$, all dimensions are involved and the excess of points, as seen in other examples, is minimal. The LA can efficiently approximate the local features of the underlying sigmoid functions with the same amount of points as the combined technique. This means the combined method is dominated by the local adaptivity with only small deviations.

For the quadrature error (*right bottom panel*), all methods can decrease the error by more than three orders of magnitude and with less NoP for the combined method and the LA than the DA. Noticeable during the refinement are the fluctuations for all three methods due to the sharp jump in the sigmoid function. To adapt to these sharp jumps in higher dimensions, the SG needs a fine grid in the subdomain of the jump, but this can cause some over- or undershooting during the refinement for the quadrature. Moreover, it is possible that during the refinement, the LA has a smaller error compared to the combined method, but altogether they show the same conversion rate.

These higher dimensional results have shown that the dimension- and local-adaptive refinement strategy can outperform the singular DA and LA for higher and more complex functions. Furthermore, it shows the efficiency for the GSA, where we want to analyse the contribution of singular parameters or interactions between the parameter of our kinetic model. Especially the second example underlines the advantages of a SG surrogate model. Although the benefit

of the combined method over the pure [LA](#) is not so dominant, we have to consider that we used a function where every parameter has an impact. For our kinetic model, we assume that Eq. (3.53) would only describe the behaviour in a subdomain of our parameter domain. Therefore, the [DA](#) part would have a stronger impact than in this example.

In summary, these results have shown the robustness and the advantages of the dimension- and local adaptive refinement for a [SG](#) construction for approximating kinetic data. We have checked the performance for multiple characteristics of the data, e.g. the local kinks, the subset of important dimensions, the decomposition into smaller contribution, and found a superior performance of the combined technique. This makes us confident to apply the [SG](#)-refinement method to a real kinetic example in the next chapter.

4 Oxygen Evolution Model

After testing the refinement strategies, we want to apply the most promising refinement strategy to a real chemical model for the purpose of performing a Sensitivity Analysis (SA). For this, we investigate a first principles-based electrochemical model for the Oxygen evolution reaction on Cobalt(II, III) oxide, proposed by Plaisance et al. in Refs. [133] and [134]. The Cobalt material has shown to be a potential low-cost and earth-abundant material to perform the oxygen evolution reaction (OER). The OER can also be described as a photo- or electrocatalytic water splitting reaction that requires an overpotential [135, 136]. Core results have already been published in Ref. [23] and are explained here in more detail.

The purpose of the original OER paper is to analyze the structure sensitivity of the OER reaction, so four different structures are examined. In this work, we want to concentrate on the additional information content given by a SA and only consider the (110)-A surface site as a model to describe the OER mechanism and perform a SA on.

We examine the (110)-A surface structure of the Co_3O_4 that consists of two redox-active Cobalt atoms sharing a bridging Oxygen atom, whose activity for the OER mechanism is highly sensitive to the applied potential. We choose this model and this surface due to its complex structure but easily comprehensible analytical form, which makes the SA results suitable for chemical interpretation. Before we go into detail about the sensitivity of the model on the energetics parameters, we shortly introduce the model. For a more comprehensive explanation, we refer to the original articles by C. Pleasant, A. Santen, and K. Reuter [133], and [134]. Figure 4.1 shows the reaction circle of the Oxygen evolution schematically. The reaction pathway consists of six elementary steps, which can later be summarized into two irreversible steps. As a starting point (Structure 1), we have two bridging hydroxyl groups that will take part in the reaction, which are deprotonated in the first two steps (Structure 2/3) and producing a Cobalt atom with two bound oxos.

We create a hydroxyl group for the oxo (Structure 4) through water addition, which is turned into a superoxo by another deprotonation (Structure 5). Out of the superoxo ligand, we can form a desorb elementary oxygen molecule (Structure 6), whose vacancies are filled with another water addition (Structure 1).

As mentioned before the OER cycle on the (110)-A surface can be summarized in two irreversible non- electrochemical steps. Starting with an active center site of the Cobalt atom with two bound oxos, the first step 4.1 forms an O-O bond out of a bound oxo-species by nucleophilic addition. The second step 4.2 is based on an superoxo ligand, which forms an elementary Oxygen molecule O_2 by a water adsorption on the vacancies [134].

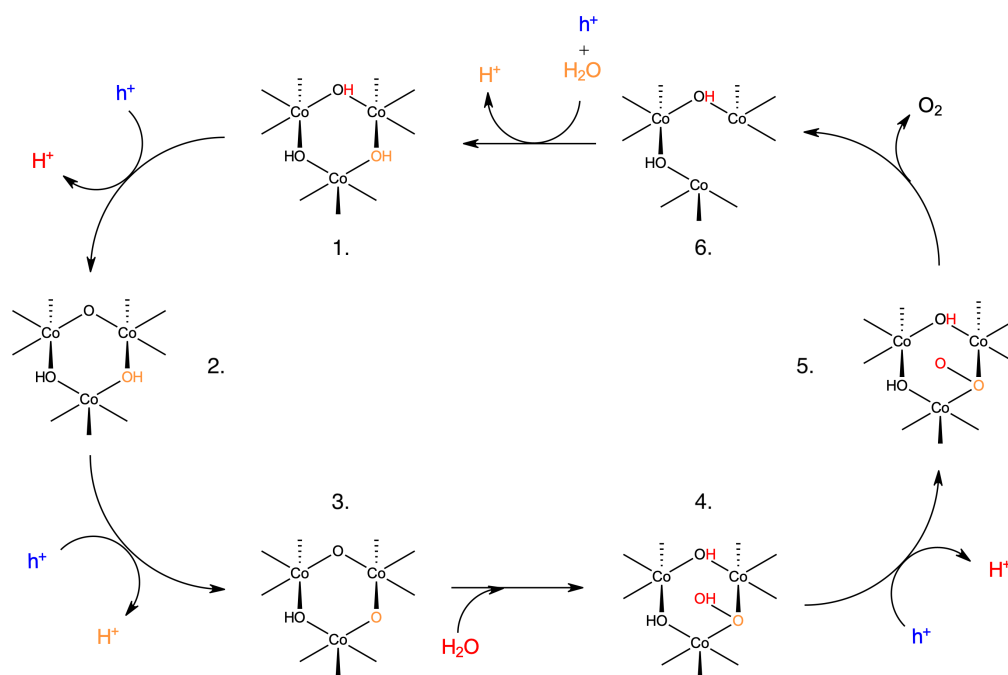
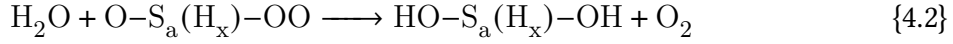
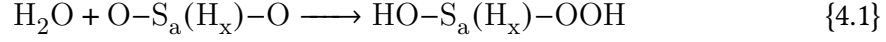


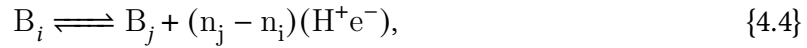
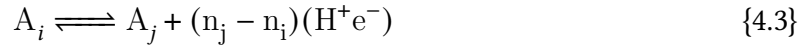
Figure 4.1: The cyclic Oxygen evolution reaction on the 110A site of Co_3O_4 . The structure consists of two redox-active Cobalt atoms bridged by an Oxygen atom. Structure 1) has two bridging hydroxyl groups that will take part in the reaction, which are deprotonated in the first two steps (Structure 2/3) and results in a Cobalt atom with two bound oxos. We create a hydroxyl group for the oxo (Structure 4) through water addition, which is turned into a superoxo by another deprotonation (Structure 5). The superoxo ligand desorbs as an elementary Oxygen molecule (Structure 6), whose vacancies are filled by another water addition (Structure 1).

Oxygen Evolution Mechanism



The notation $\text{Y}_1-\text{S}_a(\text{H}_x)-\text{Y}_2$ defines the arbitrary state of the active Cobalt center site S_a , to which the ligand species Y_1 and Y_2 are bound including x hydrogen atoms located as hydroxyl species on the surface ($0 \leq x \leq 12$) [23]. Both reactions are connected by quasi equilibrated ensembles of states involving the coupled transfer of electrons to the bulk electrode and protons to the bulk electrolyte.

Reaction 4.1 is therefore connected to reaction 4.2 by the state $A_i = \text{H}_y\text{O}-\text{S}_a(\text{H}_x)-\text{OOH}_z$ with $y, z = 0, 1$ and vice versa by $B_i = \text{H}_y\text{O}-\text{S}_a(\text{H}_x)-\text{OH}_z$ with $y, z = 0, 1$. The coupled transfer of electrons and protons,



is defined over the *degree of oxidation* $n_i = 14 - x - y - z$ of the states A_i and B_i .

The degree of oxidation describes the number of electrons and protons pairs that have to be removed to achieve this state in reference to state A_0, B_0 . The degree can vary between 0 and a maximum number of 14 due to a limited population for higher degrees of oxidation for this structure. Twelve of the electron/proton pairs can be removed from the centre atoms Co(III), oxidating them to Co(IV), in the surface layer and the remaining two come from Oxygen atoms. Also, these reactions are quasi equilibrated due to the low activation barriers of electron and proton transfer [137]. Of the two first steps (4.1, 4.2) only the first one is considered to be kinetically rate limiting. This means also for reaction 4.3 and 4.4 that we only have to consider the reaction 4.3, since the Oxygen release is an irreversible reaction.

As in most catalytic kinetic models the Quantity of Interest (QoI) of the model is the Turnover Frequency (TOF), which describes the number of chemical conversions per active site per second. Since we only have to consider one elementary reaction the TOF equals the total rate of this reaction. However, the reaction can occur through several Transition States, originating in the corresponding intermediate state A_i , as it can be seen in figure 4.2. These intermediate states are defined by the degree of oxidation, which means that the system reaches multiple oxidation states from which the reaction can happen. As a result, the TOF is

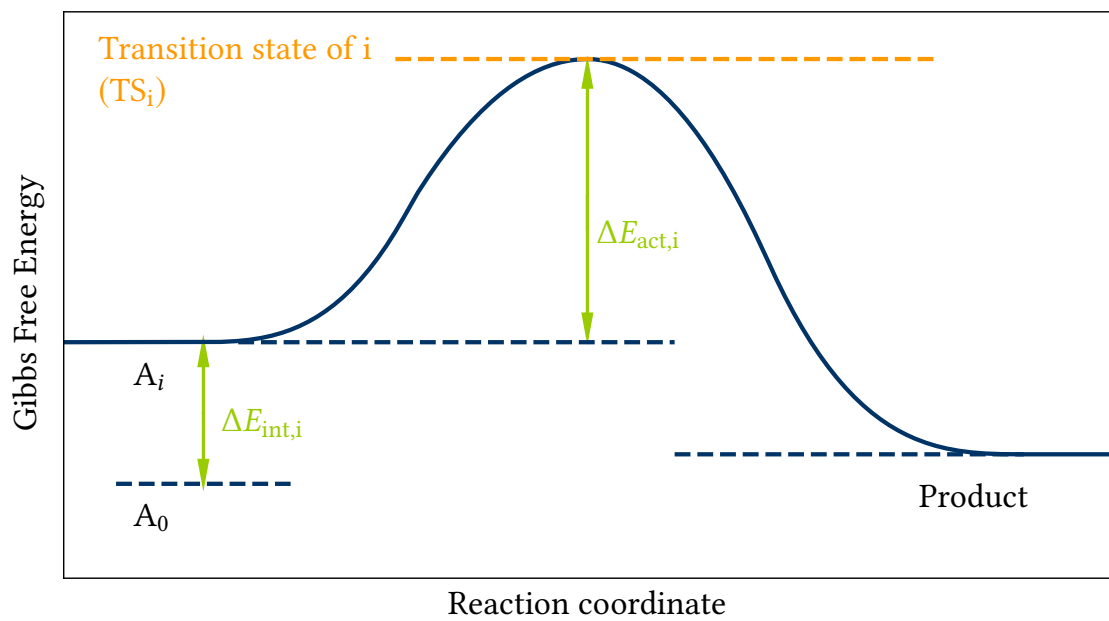


Figure 4.2: Adapted energy diagram for the OER reaction on Co_3O_4 . $\Delta E_{\text{int},i}$ indicates the free energy of the intermediate state A_i for the oxidation state i with respect to the reference state A_0 . The reaction can only happen from intermediate states A_i of even numbered states with an energy barrier $\Delta E_{\text{act},i}$ to arrive at the product state.

the sum over all single rates through the Transition State (TS) TS_i ,

$$\text{TOF} = \sum_i r_i = \sum_i k_i \theta_i, \quad (4.1)$$

where k_i describes the rate constant of the reaction 4.1 from the intermediate state A_i and θ_i is the probability to find the system in the state A_i before the rate limiting step. As it is described in chapter 2, the rate constant k is, according to the transition state theory, given by

$$k_i = \frac{k_B T}{h} \exp\left(\frac{\Delta E_{\text{act},i}}{k_B T}\right) \quad (4.2)$$

with T as the temperature, k_B as the Boltzmann constant. Since the step of constructing two bound oxos is slow we can assume the electrochemical exchange of h^+ and e^- with their surrounding to be in quasi-equilibrium. The exchange is responsible for the state A_i of the system, so how likely a high or low oxidation state is, follows a grand canonical ensemble. Therefore the propability to find the state A_i can be defined as a grand canonical distribution or Boltzmann distribution,

$$\theta_i = \frac{1}{Z} \exp\left(-\frac{\Delta E_{\text{int},i} - n_i e \eta}{k_B T}\right), \quad (4.3)$$

where the overpotential η can be defined as the sum of the chemical potentials of h^+ and e^- . Z is the grand partition function of all possible states of the system, which also serves as a normalisation factor

$$Z = \sum_i \exp\left(-\frac{\Delta E_{\text{int},i} - n_i e \eta}{k_B T}\right). \quad (4.4)$$

The activation free $\Delta E_{\text{act},i}$ energy, used in the rate constant expression Eq. (4.2), defines the energie difference between the transition state and the intermediate state A_i . The intermediate free energy $\Delta E_{\text{int},i}$ is the required energy to transform the state into A_i in reference to the state A_0 . To calculate the required energy, the free energy of a proton/electron is taken with respect to the equilibrium potential of the OER at standard conditions (1.23 V vs SHE) with $\eta = 0$. As the electrochemical models are highly dependent on the applied overpotential, the change in the free energy caused by the applied overpotential must be taken into account. The term $-n_i e \eta$ describes the deviation in the free energy of the intermediate states caused by the overpotential η compared to the OER equilibrium potential - with e as the unit of an electric charge[23]. Furthermore, the influence of the applied overpotential increases with the degree of oxidation n_i .

To explain how the parameter uncertainty influences the model output, we have to outline the model parameters in detail. For each degree of oxidation ($i = 0 \dots 14$), the model considers

intermediate energies		activation energies	
$\Delta E_{\text{int},0}$	0.00 eV	-	-
$\Delta E_{\text{int},1}$	0.44 eV	-	-
$\Delta E_{\text{int},2}$	0.88 eV	-	-
$\Delta E_{\text{int},3}$	1.64 eV	-	-
$\Delta E_{\text{int},4}$	2.33 eV	$\Delta E_{\text{act},4}$	0.60 eV
$\Delta E_{\text{int},5}$	3.09 eV	-	-
$\Delta E_{\text{int},6}$	3.79 eV	$\Delta E_{\text{act},6}$	0.59 eV
$\Delta E_{\text{int},7}$	4.59 eV	-	-
$\Delta E_{\text{int},8}$	5.38 eV	$\Delta E_{\text{act},8}$	0.62 eV
$\Delta E_{\text{int},9}$	6.22 eV	-	-
$\Delta E_{\text{int},10}$	7.05 eV	$\Delta E_{\text{act},10}$	0.60 eV
$\Delta E_{\text{int},11}$	7.88 eV	-	-
$\Delta E_{\text{int},12}$	8.71 eV	$\Delta E_{\text{act},12}$	0.56 eV
$\Delta E_{\text{int},13}$	9.77 eV	-	-
$\Delta E_{\text{int},14}$	10.83 eV	-	-

Table 4.1: Intermediate energies and activation barriers for the model for Oxygen evolution on the Co_3O_4 (110)-A surface, taken from Refs. [133] and [134]. Reproduced from [23], with the permission of AIP Publishing.

a single intermediate state, which is the one with the lowest free energy. As implied in the original paper [134], the pathways according to reaction 4.1, are only reasonably possible from even number intermediate states of $i = 4, 6, 8, 10, 12$. So altogether, the model consists of 20 elementary energies, 15 initial energies to reach the intermediate state from the reference state and five activation energies. Hence, the model considers 20 input parameters, displayed in the table 4.1. The reaction from odd number states are also possible, but, as argued by the authors of the model, for nominal parameters settings the odd-numbered intermediate states have such a low population over the range of condition that we examine, that these pathways are neglected. The low population is caused by a very low barrier to the next higher occupation state, due to the more stable configuration of even-numbered states.

All energies have been taken determined by our collaborators using the GGA+ U method (based on the Perdew-Burke-Ernzerhof Density Functional Theory (DFT) functional) [133,

138, 139]. The $+U$ correction was necessary to capture the effect of correlated 3d electrons and is the reason for the choice of the relatively high uncertainty of $\Delta\Delta E = 0.3\text{ eV}$ which we will employ throughout the following chapter. As mentioned before, reasonable errors for such chemical models are between 0.2 eV - 0.3 eV [60] from which we have chosen the upper bound. All DFT-calculation are performed with an $a+U$ correction term applied to the Cobalt 3d electrons [140–142] to avoid the over-delocalization of the 3d electrons. This over-delocalization is caused by a spurious self-interaction error in the standard semi-local exchange-correlation functionals [143]. The DFT+ U method requires specification of effective on-site Coulomb interaction, for which no agreed-upon technique exists, causing an additional ambiguity in the method [133]. As a consequence, we have to consider an increased uncertainty in the parameters, besides the general DFT error. Thus we assume the upper bound for the error of $[-0.3\text{ eV}, +0.3\text{ eV}]$ for each parameter. Ref. [133] provides a complete discussion of the DFT-functionals. Detailed information about the calculation and the choice of DFT+ U can be found in the original paper Ref. [133]

Since electrochemical models are not only dependent on material input parameters, the energetic parameters, but also external parameters, like the overpotential, we have to take it into account. We include the overpotential, not as an additional parameter in the SA, but conduct several SAs calculations at three different constant overpotential values. Figure 4.3 displays the dependency of the logarithm of the TOF on the overpotential η . The dark blue line indicates the TOF for the nominal values of the energies, documented in table 4.1, at room temperature (298.15 K). As the results shows, up to an overpotential of $\eta = 0.7\text{ V}$ the TOF increases almost exponentially and remains nearly constant for values between $\eta = 0.7\text{ V}$ and $\eta = 1.1\text{ V}$. With reaching the maximum value between $\eta = 1.0\text{ V}$ and $\eta = 1.1\text{ V}$, the TOF curve rapidly decreases for higher overpotential values and which have not be considered in the model. This may not be a physical effect but happens due to the absence of reaction pathways for high intermediate states A_{14} which is more dominantly present at higher overpotential values.

As described before the material parameters consist of uncertain data with an error of $\pm 0.3\text{ eV}$ due to the approximation in the DFT functionals of which we have no further knowledge and assume a uniformly distribution of the error over the hypercube $\Omega = \mathbf{x}_{nom} + [-0.3\text{ eV}, +0.3\text{ eV}]^D$, centered at the nominal values. This creates a probability density function (PDF),

$$p(\mathbf{x}) = \begin{cases} \frac{1}{(0.6\text{ eV})^D} & \text{if } \mathbf{x} \in \Omega \\ 0 & \text{else} \end{cases} \quad (4.5)$$

that minimizes the information content. To emphasize the impact of these errors on the TOF the figure 4.3 includes the green shaded area, describing the standard deviation (STD)

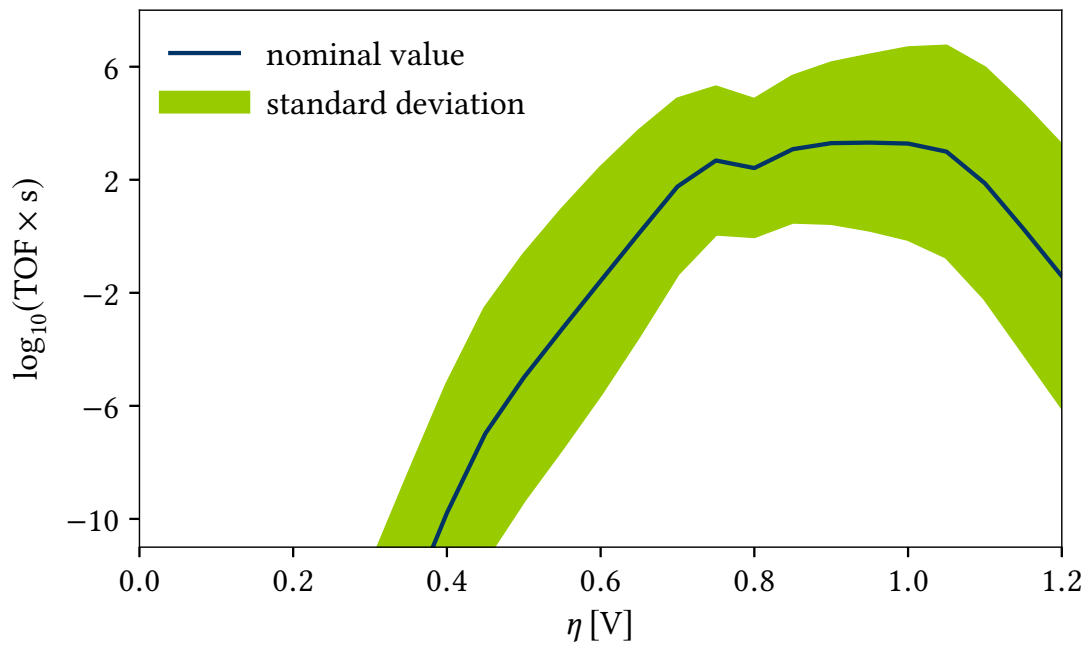


Figure 4.3: TOF as a function of the applied overpotential η for $T = 298.15$ K for the nominal energy values (blue line —). The green area depicts the standard deviation of the model output by assuming uniformly distributed DFT-errors in range of $[-0.3, 0.3]$ eV. Reproduced from [23], with the permission of AIP Publishing.

$\sigma = \sqrt{\text{Var}(y)}$ of the logarithm of the TOF caused by the DFT errors. For that, we performed 10^6 Monte Carlo points in the domain, spanned by Ω . The STD, displayed as a green shaded area, underlines that with our choice of PDF the results can deviate in more than two orders of magnitude from the nominal value of the TOF, making it difficult to extract information. The difference of the mean, $E(y)$, and the nominal value of the TOF can be as large as two orders of magnitude. Such a range can be the difference between an efficient and a non-efficient catalyst. As the 'true' curve cannot be obtained from statistical averaging since the uncertainty of y cannot be regarded as fluctuations, we have refrained from displaying the mean. Based on these results, it is unsure whether we can extract information from this model. One opportunity would be to decrease the DFT errors to increase the accuracy of the model outcome, but this is costly and only possible to a certain degree[60]. We choose to analyze the error-prone data regarding qualitative information, like rate-determining steps, and which parameters uncertainties have the biggest impact on the model uncertainty. Both can be tackled with a Sensitivity Analysis (SA). The results of the SA allow increasing the accuracy only of the important parameters by higher-order quantum-chemical methods or dedicated experiments. This will increase the accuracy of the model outcome without unnecessary costs. For the Global Sensitivity Analysis (GSA) we will employ our local- and dimensional adaptive sparse grid approach.

4.1 Sparse Grid

For the GSA of subsection 2.2.2, we need to construct a surrogate model to solve the integrals Eq. (3.36), which we will address using the Sparse Grid (SG) approach discussed in the previous chapter 3. As mentioned, the primary Quantity of Interest (QoI) is the turnover frequency. In chemical kinetics, it is common to investigate the logarithm of the TOF; see the definition of the Degree of Rate Control (DRC)[34]. This is caused by the exponential dependency on the energetics like barriers or (electro-) chemical potentials, defined in Eq. (4.2)) We, therefore, investigate the sensitivity of the logarithm of the TOF with respect to all intermediate and activation energies, also to have the GSA equivalent to the DRC. So we deal with a 20-dimensional system, consisting of 15 intermediate energies and 5 activation energies, listed in tab. 4.1. As an example, figure 4.4 displays a two dimensional subdomain of the 20 dimensional system, where the \log_{10} of the TOF is displayed as a function of the errors of two input parameters, $\Delta E_{\text{int},4}$ and $\Delta E_{\text{act},4}$. These two energies describe the reaction path through the intermediate state A_4 with an oxidation state of 4. The calculations are

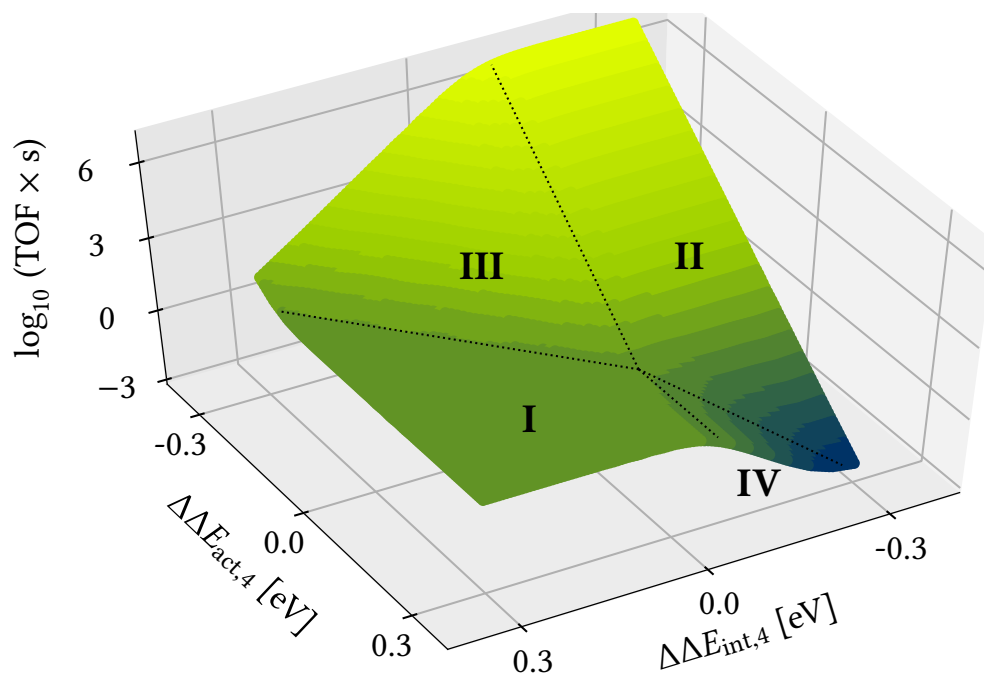


Figure 4.4: Two dimensional dependence of the TOF on the energy errors $\Delta\Delta E_{\text{int},4}$ and $\Delta\Delta E_{\text{act},4}$ at an overpotential of $\eta = 0.7$ V. (I) no dependence on $\Delta\Delta E_{\text{int},4}$ and $\Delta\Delta E_{\text{act},4}$, (II) only depending on $\Delta\Delta E_{\text{act},4}$, (III) depending on both, and (IV) only depending on $\Delta\Delta E_{\text{int},4}$. Reproduced from [23], with the permission of AIP Publishing.

performed at an overpotential of $\eta = 0.7$ and $T = 298.15$ K with all other input parameters fixed to the nominal values.

The interpolation does not show a singular linear dependency over the whole domain, but rather four domains with different dependencies on the input parameters. While the domains itself behave mostly linear, the edges between the domains are exactly the narrow subdomains where the function is non-linear. In detail, for the first regime (I) with medium to high $\Delta E_{\text{int},4}$ and $\Delta E_{\text{act},4}$, the response is independent of both parameters. In the second regime (II) with low $\Delta E_{\text{int},4}$, the TOF is solely dependent on $\Delta E_{\text{act},4}$. Compared to regime (II), regime (IV) shows the opposite dependency. For high $\Delta E_{\text{act},4}$ values the model shows a singular dependency on the $\Delta E_{\text{int},4}$. Only in the third regime (III) with low to medium $\Delta E_{\text{act},4}$ and medium to high $\Delta E_{\text{int},4}$ the response is dependent on both input parameters. The linear dependencies in the domains implies a use of piecewise linear basis functions for the SG (Sec. 3.2.1).

Next to the characteristic behaviour, we assume that in this 20-dimensional system, there will be a hierarchy of importance of the energies towards the turnover frequency. As the results of section 3.4 have shown, the combined technique of dimension and local adaptivity (LA) is the most suitable refinement strategy for such kinds of models.

Nevertheless, with a real chemical problem, we cannot expect perfect systematic behaviour

as we showed before in the test case (Sec. 3.4). We detected that this model has local features that go undetected by the refinement strategy based on some simulations, starting with a $|l|_1 = 1$ as an initial grid with $2D + 1$ points. In the case of the dimensional and local refinement, the strategy is very harsh with avoiding the unnecessary points without pushing the interpolation to a minimal threshold. This has the advantage that the refinement quickly adapts the important dimensions with minimal points, with the downside that if it is not adapted within the first step, it will be neglected. The SG refinement is only as good as the initial grid, so that we need a sufficiently sampled initial grid, which resolves all features well enough, such that the refinement strategy does not suffer from an early termination. To avoid this problem, we run a SG simulation with the dimension and local-adaptive refinement for a minimal threshold $\epsilon = 10^{-5}$, but with a limited refinement level of a maximum norm $|l|_\infty = 1$. With this, we include all the interactions between the dimensions in the first step and resolve important features on a coarse level [23]. For example, features like a rapid local increase in the corner of the Sparse Grid would not be detected by the usual initial grid with an L_1 -norm of $|l|_1 = 1$, unless the generated points indicate the feature as well. Using the results as a sufficiently accurate initial grid provides a more effective SG algorithm performed with higher thresholds.

In order to perform the sparse grid approach, we have to limit the values of overpotentials that we want to analyze. We will look at three interesting cases of $\eta = 0.4\text{ V}$, 0.7 V and 1.0 V and the corresponding SA. These three cases are signature examples for different kinetically active regimes of the model. $\eta = 0.4\text{ V}$ characterizes a regime with lower total activity but with an almost exponential increase. The medium overpotential $\eta = 0.7\text{ V}$ is the starting point of the highest active area with a relatively constant TOF and $\eta = 1.0\text{ V}$ marks the end of that kinetic regime. Since the TOF decreases for higher overpotentials and the results may not be physical, $\eta > 1.0\text{ V}$ are of less interest.

Due to the rising overpotential activating different reaction pathways, the complexity of the model differs according to the applied overpotential. However, the table 4.2 shows the initial grids for the $\eta = 0.4\text{ V}$, 0.7 V and 1.0 V .

As mentioned before and implied in figure 4.3, the model shows a lower reactivity for $\eta = 0.4\text{ V}$ than for $\eta = 0.7\text{ V}$. This could be due to the limited pathways at the lower η , but we would also expect less complexity for areas without phase transitions or fewer rate determining steps [34, 35]. Thereby the initial grid consist of only 149 points, compared to the 5162 points for $\eta = 0.7\text{ V}$. The same effect occurs for $\eta = 1.0\text{ V}$. Due to the absence of higher oxidation state pathways, the model reaches a point of limited possible pathways. So even though the TOF shows a high reactivity we assume a decreasing interaction of parameters, resulting in an initial grid of 256 points for $\eta = 1.0\text{ V}$. Although the SG is only a tool to obtain the expected

0.4 V	NoP	$E(\log_{10}(\text{TOF}))$	$\text{Var}(\log_{10}(\text{TOF}))$
initial grid	149		
0.1	157	-11.8005	20.41047734
0.01	230	-11.7179	20.80824933
0.001	467	-11.6742	20.92827296
0.0001	1238	-11.6692	20.95995558
0.7 V	NoP	$E(\log_{10}(\text{TOF}))$	$\text{Var}(\log_{10}(\text{TOF}))$
initial grid	5162		
0.1	5185	1.80462	8.18310458
0.01	6014	1.34399	9.327653459
0.001	13712	1.68762	9.498208983
0.0001	52357	1.61414	9.27517
1.0 V	NoP	$E(\log_{10}(\text{TOF}))$	$\text{Var}(\log_{10}(\text{TOF}))$
initial grid	256		
0.1	383	1.92823	9.85318734
0.01	550	2.01832	10.62025072
0.001	1416	1.89977	11.21293531
0.0001	4395	1.89208	11.19771616

Table 4.2: The SG results for the initial grid and the final grid for decreasing threshold $\epsilon = [0.1, 0.01, 0.001, 0.0001]$ and the estimates for the expected $E(\log_{10}(\text{TOF}))$ and the variance $\text{Var}(\log_{10}(\text{TOF}))$ for the SG results. Partially reproduced from supplementary material [23], with the permission of AIP Publishing.

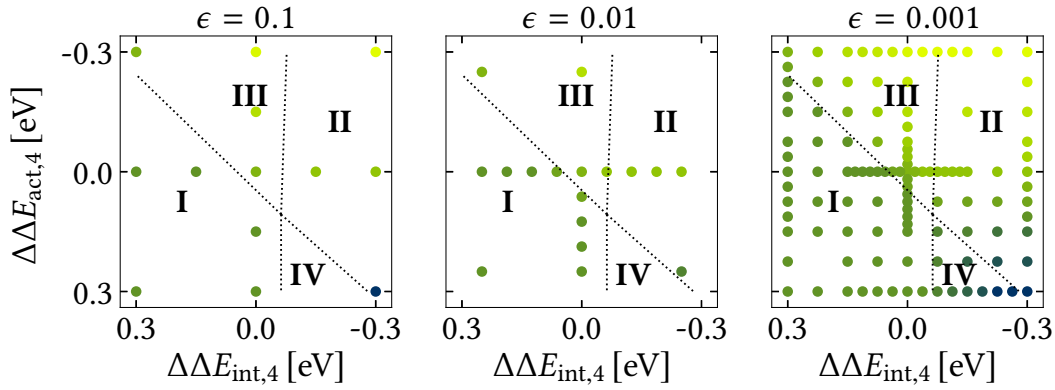


Figure 4.5: Adaptive grid for three applied tolerances $\epsilon = 0.1$ (left) $\epsilon = 0.01$ (middle) and $\epsilon = 0.001$ (right). Dotted line - - - indicates borders between four different regimes with different dependencies on the parameters $\Delta\Delta E_{\text{int},4}$ and $\Delta\Delta E_{\text{act},4}$. Refinement primarily happens close to the origin, where the four regimes meet. Partially reproduced from [23], with the permission of AIP Publishing.

value and the variance, we briefly want to discuss the SG results based on the 2D subdomain of figure 4.4 for $\eta = 0.7$ V. Based on the initial grid we show the final SG-grids for different thresholds $\epsilon = [0.1, 0.01, 0.001]$ in the figure 4.5. As we expected, the most refinement points are produced towards the non-linear edges between the different regimes and a significant amount at the crosspoint of all four regimes close to the origin. For decreasing threshold, this effect becomes significantly stronger.

Overall, these aspects underline only the LA of the refinement strategy. Since the dimension adaptivity (DA)'s effect is the strongest by excluding dimensions, it is difficult to show that in a 2D subplot. This is visible in the reduced final points for all three overpotential cases for different thresholds that are displayed in table 4.2.

4.2 Sensitivity Analysis

For the three overpotential examples we find the expected values and the variance documented in the table 4.2. The highest variance of the $\log_{10}(\text{TOF})$ of 20.9 can be found for the lowest overpotential ($\eta = 0.4$ V) with an expected value of $E(\log_{10}(\text{TOF})) = -11.7$. For $\eta = 0.7$ V the variance is 9.5 with $E(\log_{10}(\text{TOF})) = 1.7$ and for $\eta = 1.0$ V it is 11.3 with $E(\log_{10}(\text{TOF})) = 1.9$. This range of variances corresponds to TOF uncertainties of three to five order of magnitude. These results are taken from the SG surrogate model of $\epsilon = 0.001$, thus of a very low threshold and a very fine sparse grid. Nevertheless, the SG results for

higher threshold are already giving accurate estimation of the variance and the expected value ($\epsilon = 0.1$) and the reduction leads to minor changes of the numbers as shown in table 4.2. Solely the case of $\eta = 0.7 \text{ V}$ shows some bigger variations which is caused by the more complex structure of the model. Although the variations might have an impact on the SA the following results show that the SA is robust against this order of variations. To underline this, the figure 7.1 in the appendix show the convergence behaviour. This means that we do not need a very fine grid to estimate the necessary quantities, but even if we would consider lower threshold results, the number of points would still be moderate ($\eta = 0.4 \text{ V} : \sim 500$, $\eta = 0.7 \text{ V} : \sim 14000$, $\eta = 1.0 \text{ V} : \sim 1500$).

4.2.1 Total Sensitivity Index TSI

Based on the SG interpolation, we now investigate which of the parameters errors are responsible for the reported uncertainties in the model outcome, taking into account that we deal with three different kinetic behaviours of the model. For that, we include not only the GSA but also the Local Sensitivity Analysis (LSA), so that we can compare the results to a, mostly in the engineering science used, approach [33]. Figure 4.6 displays the total Sensitivity Index (SI) for all input parameters with the corresponding overpotential of $\eta = 0.4 \text{ V}$, 0.7 V and 1.0 V . The 20 Total Sensitivity Indices are separated into two classes corresponding to the intermediate energies $\Delta E_{\text{int},i}$ and the activation free energies $\Delta E_{\text{act},i}$. The grey areas of the activation energy panel indicate the intermediate states from which no reaction pathway is considered in the model; thus, only states of even numbers bigger than 2 are included for the SA.

In order to interpret the Total Sensitivity Index (TSI) results, we have to connect the uncertainty in the TOF results to the single variances caused by the input parameters. As we already mentioned, the TSI captures the expected variance caused by a single input parameter x_i . This means, a $S_{T,i} = 1$ equals a STD of 1 for the logarithm of the TOF and transferred for the non-logarithmic TOF a STD of 10. So an increased $S_{T,i}$ of 4, causes a STD of 2 for the logarithmic and a deviation of 100 for the nominal TOF values [23].

As figure 4.6 shows, the TSI indicate different dependencies for the different overpotentials. While for low overpotentials, which is displayed in the upper panel of figure 4.6, the TOF mostly depends on parameters whose state is of a low oxidation degree ($i < 5$), the importance shifts to states of higher degrees for increasing overpotential. From a physical point of view, this phenomenon makes sense since the system will exist predominantly in states of the

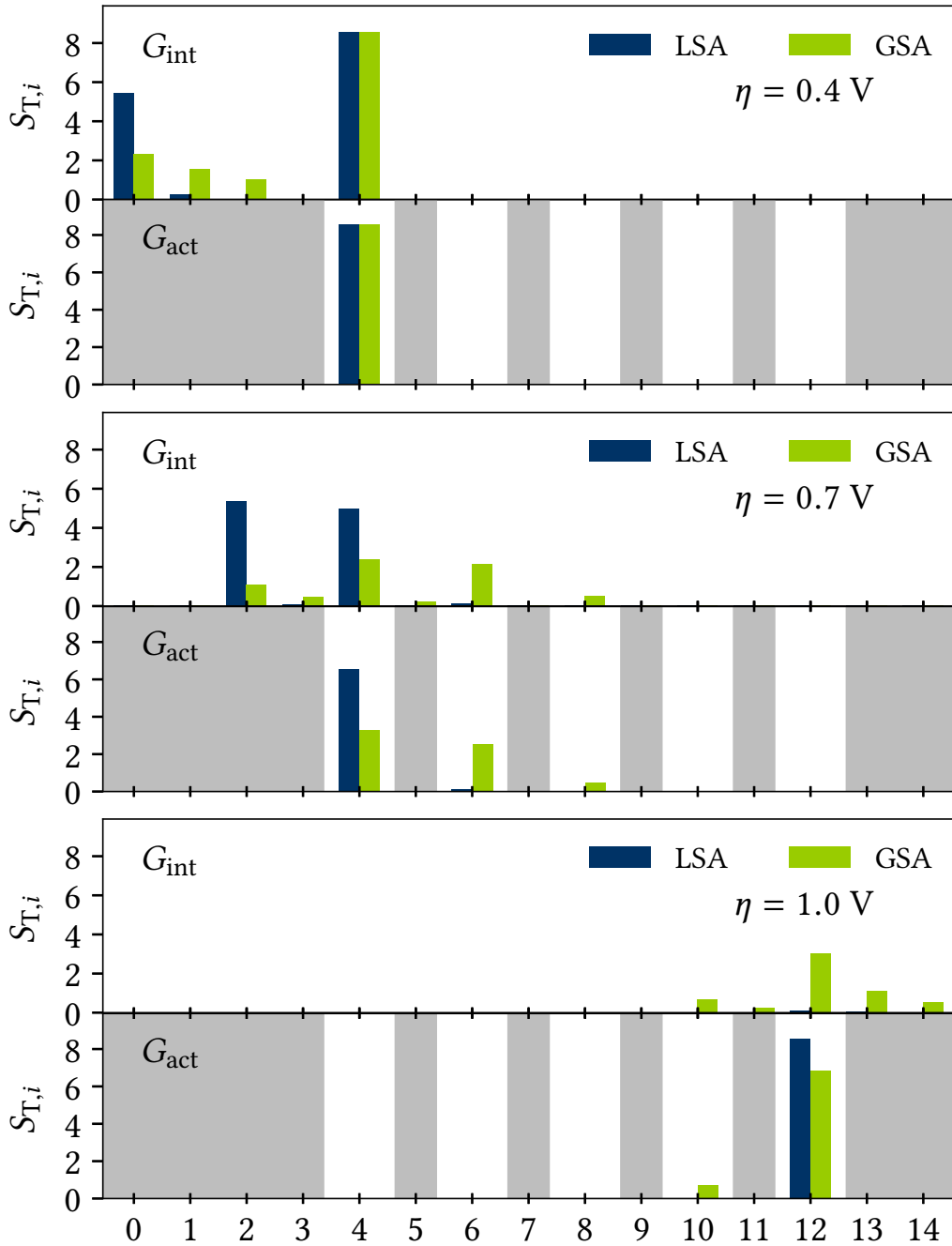


Figure 4.6: Local und global sensitivity indices $S_{T,i}$ for the 20 input parameters for the (110)-A structure of the Oxygen evolution reaction at three different overpotentials $\eta \in \{0.4, 0.7, 1.0\}$ V. The gray-shaded fields describe intermediate states that are not considered for a reaction in the model. Shown are the sensitivities of $\log_{10}(\text{TOF} \times s)$ with respect to the errors in $\Delta E_{int,i}$ and $\Delta E_{act,i}$. Reproduced from [23], with the permission of AIP Publishing.

lowest oxidation degree A_0 for a low applied overpotential, as the states of higher oxidation degrees are thermodynamically unfavourable. This means for the reaction network of water to Oxygen that the reaction 4.1 can only happen on the lowest possible degree of oxidation, A_4 . Even when taking the parameters uncertainties into account, the free energies of states with a higher degree of oxidation will not be low enough to contribute significantly to the probability, which would allow the surface to exist in that state. This means that over the parameters domain, the kinetic is only influenced by energies of the states of a lower degree of oxidation.

As the overpotential increases, as shown in the middle and lower panel of figure 4.6, the important states are shifted towards higher oxidation degrees, which dominate the probability distribution Eq. (4.3). By this means, the probability of the surface state is more widely distributed and does not have a significant probability of occurring in lower degrees of oxidation. In comparison, the LSA and the GSA show roughly a similar trend regarding the TSIs, yet a more detailed analysis reveals major discrepancies. The LSA often not only underestimates the number of important parameters but also overestimates the importance of a single parameter. For example in the case of $\eta = 0.4$ V (upper panel 4.6), the LSA indicates three parameters as very important and a small dependency on a fourth parameter, while the GSA shows, that the TOF depends on five parameters. For $\eta = 1.0$ V (lower panel 4.6), the LSA shows no dependencies on the intermediate energies, only on one activation free energy. In contrast to that, the GSA describes a broader distribution of important energies, in total seven. In the case of $\eta = 0.7$ V (middle panel 4.6), the LSA not only indicates three instead of nine parameters, but the importance itself is more than twice the value of the GSA, so that the TOF would strongly depend on only three parameters. Although the LSA may be the cheaper method, the convergence of the GSA, displayed in figure 7.1 in the appendix, underlines that the GSA does not need a highly converged grid to identify the important parameters. It identifies the parameters already with a very coarse sparse grid. Depending on the purpose of the GSA, the results can be obtained with a very small amount of grid points.

The discrepancies between the LSA and GSA are caused by non-linearities in the equation of the TOF

$$\text{TOF} = \frac{k_B T}{h} \frac{\sum_i \exp\left(-\frac{\Delta E_{\text{act},i} + \Delta E_{\text{int},i} - n_i e \eta}{k_B T}\right)}{\sum_i \exp\left(-\frac{\Delta E_{\text{int},j} - n_j e \eta}{k_B T}\right)} \quad (4.6)$$

based on eqs. (4.1) to (4.3). For a fixed value for the overpotential, the numerator and the denominator will likely be dominated by one exponential term, corresponding to an individual oxidation state i and j , respectively. This occurs due to the major energie uncertainties compared to the term $k_B T$. Taking this into account the logarithm of the TOF can be resembled

as piecewise linear functions

$$\log_{10}(\text{TOF}) = \log_{10}\left(\frac{k_B T}{h}\right) - \frac{\log_{10} e_{Euler}}{k_B T} (\Delta E_{\text{act},i} + \Delta E_{\text{int},i} - \Delta E_{\text{int},j}) + \frac{\log_{10} e_{Euler}}{k_B T} (n_i - n_j) e \eta \quad (4.7)$$

where e_{Euler} is the Euler's number. In this case the i corresponds to the transition state with the lowest free energy ($\Delta E_{\text{act},i} + \Delta E_{\text{int},i} - n_i e \eta$) and j to the intermediate state with the lowest free energy ($\Delta E_{\text{int},j} - n_j e \eta$). Hence the TOF should depend predominantly on the energies of the states i and j . Unfortunately this is only true if the uncertainty range captures only one kinetic regime, so that the dominant i and j stay the same in the domain of the uncertainty range. Then, the LSA and GSA results would be identical and only show non-vanishing sensitivities on $\Delta E_{\text{int},i}$, $\Delta E_{\text{int},j}$, $\Delta E_{\text{act},i}$. At $\eta = 0.4$ V these correspond to $i = 4$ and $j = 0$, at $\eta = 0.7$ V to $i = 4$ and $j = 2$ and at $\eta = 1.0$ V to $i = j = 12$. For the latter the LSA shows only a strong dependence on one activation energy. The vanishing SI towards all intermediate states energies are due to $i = j$ which causes $\Delta E_{\text{int},i} - \Delta E_{\text{int},j} = 0$ in Eq. (4.7).

In the case, that the uncertainty range includes multiple kinetic regimes, the GSA would identify the dominant energies $\Delta E_{\text{int},i}$, $\Delta E_{\text{int},j}$, $\Delta E_{\text{act},i}$ for each of these regimes. In contrast to that, the LSA would only identify the dominant energies for the regime at the nominal values unless we hit one of the boundaries between different regimes by chance. Furthermore, the LSA would overestimate the importance of the single energies, assuming that the current kinetic regime is true for the whole uncertainty range. The GSA captures the transition to other kinetic regimes at large enough deviations from the nominal values. This results in a broader dependence of different energies of different states [23, 37, 98].

Taking this consideration into account, we can assume that for $\eta = 0.4$ V the uncertainty range predominantly consist of one kinetic regime, because of the similar results of LSA and GSA. Both strategies identify $\Delta E_{\text{int},0}$, $\Delta E_{\text{int},4}$, $\Delta E_{\text{act},4}$ and the important reaction happening primarily from $i = 4$. The results for the more complex case of $\eta = 0.7$ V are very different. As mentioned before, the LSA overestimates the importance of the three identified energies and we see that the multiple other kinetic regimes include higher oxidation states. For the highest overpotential case of $\eta = 1.0$ V we see no big deviations in the results of the two strategies. This is due to the fact that the reactions mostly happen via the transition state of $i = 12$ over the whole uncertainty range.

4.2.2 Higher Order Sensitivity Indices

As the deviations between the LSA and the GSA show, the one-dimensional functions of the Analysis of Variance (ANOVA) decomposition have to be either non-linear or there are additional interactions between the different parameters as described in Subsection 2.2.2 [23, 98]. The first possibility would imply, that the variance can be decomposed into one-dimensional scans along the axes, and the deviation from linear behavior must be visible in these one-dimensional scans. More probable is the second possibility, which can not be proven or disproven by any local scans. The function can appear locally linear along the axes but behaves non-linear globally, which provides a lot of information. So far, we have just looked at the global contribution of single input parameter, but these can have interactions among each other, which are not directly discussed with the TSI. For a detailed discussion higher order SIs have to be included. As they correspond to the interaction of an increasing amount of parameters, we expect the highest contribution to be of pairwise interaction and concentrate mainly on the second order indices. Therefore, we first have to calculate first order sensitivity indices (Eq. (2.16))

$$\text{Var}_i = \text{Var}_{x_i}(\mathbb{E}_{x_{\sim i}}(y|x_i)) \quad (4.8)$$

in order to calculate the second order indices

$$\text{Var}_{ij} = \text{Var}_{x_i}(\mathbb{E}_{x_{\sim ij}}(y|x_i, x_j)) - \text{Var}_i - \text{Var}_j. \quad (4.9)$$

The ANOVA decomposition implies that the variance of the model output consists of single distributions of a different order. Performing the respective sums, we can estimate the residual contribution of higher orders than the second term. Table 4.3 shows these variance contributions for the first, the second-order and the remaining residual. The three cases have three very different conclusions. In the case of $\eta = 0.4$ V, with the highest variance of all three, the biggest contribution results from the first-order sensitivity indices, thus from singular parameters. Only 4.2% are due to interaction of higher-order, 3.7% from second-order indices. In contrast to that are the results for $\eta = 0.7$ V, where the first-order effects account for only 68.2% and second-order effects are responsible for 26.6% of the variance. This points to a function with an intrinsic multidimensional character [23, 37]. The results of case $\eta = 1.0$ V are in between the previous results. The dominant contributions (86.2%) are from first-order effects, but the second-order effects are still responsible for 11.4% and the higher-order effects for 2.4% of the variance. As a result, we only concentrate on the second-order indices since higher-order interactions do not significantly contribute. However, for other models, it can

η [V]	$\text{Var}(y)$	$\sum_i \text{Var}_i$	$\sum_i \sum_{j,j>i} \text{Var}_{ij}$
0.4	20.93	20.05	0.76
0.7	9.50	6.48	2.52
1.0	11.21	9.66	1.28

η [V]	$\frac{\sum_i \text{Var}_i}{\text{Var}(y)}$ [%]	$\frac{\sum_i \sum_{j,j>i} \text{Var}_{ij}}{\text{Var}(y)}$ [%]	$\frac{R}{\text{Var}(y)}$ [%]
0.4	95.8	3.7	0.5
0.7	68.2	26.6	5.2
1.0	86.2	11.4	2.4

Table 4.3: The total variance and the contribution of the first and second order decomposition, documented for three cases of overpotential η and displayed in the upper row of the table. The importance of either the first, second or the remaining residual R on the total variance are underlined by the percentage contribution are displayed in the lower row.

be interesting to analyze the higher-order indices [144]. Figure 4.7 shows the second-order indices for the three case, restricted to those parameters with a non vanishing TSI, as the Var_{ij} can not be bigger than the minimum of the individual TSIs $\min(S_{T,i}, S_{T,j})$ (Eq. (2.17)). So for $\eta = 0.4$ V, where the second-order interactions are responsible for less than 4% of the variance, we find the strongest coupling between the intermediate states of $i = 0$ and $j = 1$, which makes sense as the system is predominantly found in the lowest intermediate state. Furthermore, the indices show no interactions between intermediate and activation energies, only within the lower intermediate states energies. Thus the interaction of the intermediate state and the reaction on the fourth state are not important to the model outcome. The results for $\eta = 0.7$ V are more complex, as it shows the highest coupling where one of $\Delta E_{\text{act},4}$ or $\Delta E_{\text{act},6}$ is involved. Each activation energy couples the strongest with the corresponding intermediate state energy and vice versa. Only the intermediate energies of states from which no reaction is possible ($i = 2, 3$ and 5) interact exclusively with intermediate state energies. This effect can be reasoned with the definition of the logarithm of the TOF

$$\log_{10}(\text{TOF}) = -\log_{10} \left(\sum_{i=0}^{14} \exp\left(-\frac{\Delta E_{\text{int},i} - n_i e \eta}{k_B T}\right) \right) \quad (4.10)$$

$$+ \log_{10} \left(\sum_{i=2}^6 \exp\left(-\frac{\Delta E_{\text{int},2i} + \Delta E_{\text{act},2i} - n_i e \eta}{k_B T}\right) \right) \quad (4.11)$$

$$+ \log_{10} \left(\frac{k_B T}{h} \right) \quad (4.12)$$

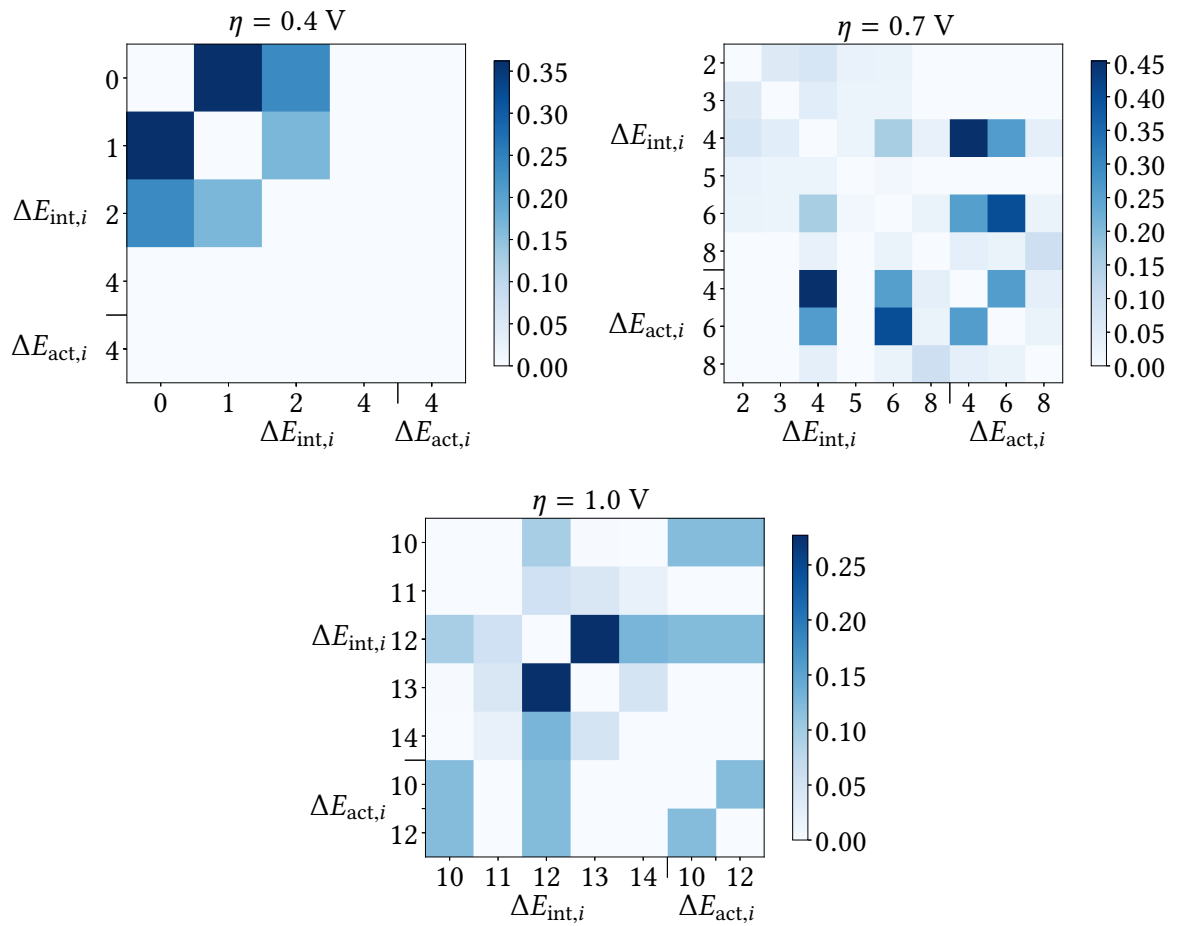


Figure 4.7: Second order indices with respect to the input parameters of the OER model on the (110)-A surface for an overpotential $\eta \in \{0.4, 0.7, 1.0\}$ V. Darker areas correspond to as stronger interaction between the two parameters. Partially reproduced from [23], with the permission of AIP Publishing.

where the first term includes the states 2,3 and 5 and does not depend on any activation energies. The second term, that depends on the activation energies, has no consideration of the states in question. For the last case of $\eta = 1.0\text{ V}$, we found the strongest interactions involving the intermediate state energy $\Delta E_{\text{int},12}$. The energy strongly couples with $\Delta E_{\text{int},13}$, but also with the activation energies. Furthermore, the results show that the intermediate energies of states that allow reactions have the highest second order indices. Similar to the case of $\eta = 0.7\text{ V}$, the energies of the intermediate states, of whose state no reaction is possible, interact only with other intermediate states, but not with the activation energies.

4.3 Discussion

Overall the uncertainty analysis draws attention to the impact of approximated DFT-energies and thus their errors in microkinetic modelling. Based on the initial variance and standard deviation of the model outcome, it is difficult to conclude whether the surface is highly active or not. Unfortunately, this model is only one of many examples whose quantity of interest, the TOF, carries uncertainties of several orders of magnitude. In our previous studies about first-principles kinetic Monte Carlo models, we show this exact difficulty [24].

However, using GSA it is still able to draw conclusions from the uncertain input data. Before discussing the results of the GSA, we are going to discuss the basis of the GSA, the performance of the SG approach. Overall, the locally and dimensionally adaptive SG-approach works well for carrying out a GSA, with a few drawbacks. For all three cases of overpotential, the SG achieves high accuracy with a modest amount of points. The number of points might be not crucial in the case of an analytical model, like the targeted problem, as the computational costs are negligible compared to the costs of the DFT calculations. Nevertheless, if we are dealing with complex simulation-based models, like kinetic Monte Carlo, a single evaluation might be expensive, and the number of evaluations can be a major bottleneck [23]. For the efficient performance of the dimension and LA, the intrinsic dimensionality is significant. As discussed in Sec. 4.2, the model has a higher complexity for $\eta = 0.7\text{ V}$ and the SG method requires more points than for the other cases, due to the higher intrinsic dimensionality. So the biggest benefit of the adaptive SG will be achieved with a high true dimensionality and a low intrinsic dimensionality, as it can be seen for the $\eta = 0.4\text{ V}$ case, where we need significantly fewer points for the 20 dimensional model with an intrinsic dimensionality of 5. The drawback of the SG in general is the risk of premature termination. The harsh refinement strategy can overlook local effects, but concerning the features of the

models, this can be compensated with some preprocessing. In our case, we address this by applying a disproportionately low threshold and by limiting the maximal resolution for the initial grid. This method may not generally be applied to every model. For dealing with a complex model, such high accuracy might, for example, not be affordable.

Our results indicate that, using GSA, we can still extract information about the reaction mechanism from the uncertain model, even though we are not able to predict whether the catalyst is active or not. The GSA method successfully identifies only a few important parameters, even though we assume the worst case scenario of an equally distributed error for the PDF. Since we are dealing with microkinetic modelling and, in particular, with the efficiency of a catalytic surface, the GSA identifies key atomistic aspects which control the reactivity. In all three examples, the GSA identifies only a fraction of all input parameters that influence the model output, meaning that variations in the respective parameters cause a higher variation in the model outcome, compared to parameters with a vanishing TSI. In other words, the parameters with a high TSI are more kinetically relevant and control the efficiency of the catalyst. For the model itself, the TSI results mean that the physical details of the corresponding intermediate states are highly responsible for the reactivity behaviour. These results can be drawn, although we are dealing with this sizable error. Furthermore, the distinction between the different overpotentials has shown that the complexity and the important energies change with the applied overpotential and without a fixed potential, that a general conclusion on which atomistic indicators control the reactivity is not possible. If there is a previously known operational constraint, like a fixed window of the overpotential, important indicators can be identified, as the previous discussion showed.

In the discussion of SA, our results in Sec. 4.2.1 established that the LSA method draws erroneous conclusions by underestimating the number of important parameters and overestimating the importance of the identified parameters. As the LSA is the most common method to use for SA the erroneous analysis can be fatal when targeting at improving the model accuracy. As it can be seen for the case of $\eta = 0.7$ V, the LSA fails to identify all of the important parameters and strongly overestimates the importance of the detected parameters. If we use these results as a base to recalculate the respective DFT energies with a more accurate methodology, the potential of variance reduction would be limited. Additionally, the higher accurate DFT-numbers are obtained from significantly more expensive methods, which illustrates why a correct set of parameters is important. These more expensive methods are higher-level electronic structure methods, like domain-based local pair natural orbital based singles and doubles coupled-cluster method (DLPNO+CCSD(T)) or Constrained -Orbital -DFT[31, 32] As the LSA misses some important parameters, we would have a limited number of energies to recalculate. Based on the corresponding estimation of the variance, the latter

η [V]	reduced energy	LSA		true variance	
		$\text{Var}_{\text{org}}(y)$	$\text{Var}_{\text{red},3}$	$\text{Var}_{\text{org}}(y)$	$\text{Var}_{\text{red},3}$
0.4	$\Delta E_{\text{int},0}, \Delta E_{\text{int},4}, \Delta E_{\text{act},4}$	22.84	0.91	20.93	2.73
0.7	$\Delta E_{\text{int},2}, \Delta E_{\text{int},4}, \Delta E_{\text{act},4}$	17.18	0.97	9.50	4.58
1.0	$\Delta E_{\text{act},12}$	8.72	0.49	11.21	4.88

Table 4.4: Calculation of the total variance $\text{Var}_{\text{org}}(y)$ according to Eq. (2.10) for the LSA and Eq. (2.9) for the true variance and the reduced variance $\text{Var}_{\text{red},3}$ if the most important LSA energies contain a decreased error of $\Delta\Delta E_{\text{int/act},i} = 0.06$ eV for all three cases $\eta \in \{0.4, 0.7, 1.0\}$ V. This describes the recalculation of the most important energies with a higher accurate method.

more accurate energies would produce an artificially low variance for Eq. (2.10). This would lead to a false conclusion of an 'accurate' model. For a short demonstration, we assume for now, that the LSA results are correct and we can apply a higher level treatment for the energies and achieve an error of $\Delta\Delta E = 0.06$ eV, so a fifth of our initially employed error $\Delta\Delta E = 0.3$ eV. Then we take all three overpotential cases and apply the more accurate energies, assuming that the nominal values stay the same out of simplicity. Table 4.4 displays the variance (Eq. (2.10) and Eq. (2.9)) and the variance reduction based on refining the DFT-energies according to the LSA.

Since the SI for the LSA differ strongly between the most dominant ones and the small detection of less important parameters, we could concentrate on recalculating the three parameters with the most dominant SI. That means we recalculate $\Delta E_{\text{int},0}, \Delta E_{\text{int},4}, \Delta E_{\text{act},4}$ in the case of $\eta = 0.4$ V, $\Delta E_{\text{int},2}, \Delta E_{\text{int},4}, \Delta E_{\text{act},4}$ in the case of $\eta = 0.7$ V and $\Delta E_{\text{act},12}$ in the case of $\eta = 1.0$ V. To begin with, the variance of the LSA calculated by equation (2.10) differs from the true variance. After improving the energies the recalculated LSA variance shows the same trend in all three cases. The variance drops below 1, which would be an improvement factor of around 22 ($\eta = 0.4$ V), 17 ($\eta = 0.7$ V) and 16 ($\eta = 1.0$ V). However the true variance shows different results if we consider a reduced error for the same three LSA energies. While the variance for $\eta = 0.4$ V is also drastically reduced by a factor of around 8, the variance for the other two cases is approximately halved. This factor of 2 means we improve the nominal TOF by a factor less than one order of magnitude instead of three ($\eta = 0.7$ V) or two ($\eta = 1.0$ V) orders of magnitude as expected from the LSA results. For $\eta = 1.0$ V this is the final variance reduction, since the LSA does not indicate a dependency of the model outcome on another parameter. If we also include the fourth most important parameters in the other two cases the true variance can be decreased to 1.86 ($\eta = 0.4$ V) and 3.56 ($\eta = 0.7$ V), which is still very different from the variance reduction indicated by the results of the LSA. If we now base the

η [V]	all reduced energy	$\text{Var}_{\text{red},3}$	$\text{Var}_{\text{red},4}$	$\text{Var}_{\text{red, all}}$
0.4	$\Delta E_{\text{int},4}, \Delta E_{\text{act},4}, \Delta E_{\text{int},0},$ $\Delta E_{\text{int},1}, \Delta E_{\text{int},2}$	2.73	1.85	0.88
0.7	$\Delta E_{\text{act},4}, \Delta E_{\text{act},6}, \Delta E_{\text{int},4},$ $\Delta E_{\text{int},6}, \Delta E_{\text{int},2}, \Delta E_{\text{int},8},$ $\Delta E_{\text{act},8}, \Delta E_{\text{int},3}, \Delta E_{\text{int},5}$	3.45	3.11	0.57
1.0	$\Delta E_{\text{act},12}, \Delta E_{\text{int},12}, \Delta E_{\text{int},13}, \Delta E_{\text{act},10},$ $\Delta E_{\text{int},10}, \Delta E_{\text{int},14}, \Delta E_{\text{int},11}$	1.29	1.12	0.38

Table 4.5: Calculation of the total reduced variance if the three, four, or all GSA energies contain a decreased error of $\Delta\Delta E_{\text{int/act},i} = 0.06$ eV for all three cases $\eta \in \{0.4, 0.7, 1.0\}$ V. This describes the recalculation of the most important energies with a higher accurate method.

recalculation on the SI of the GSA and also consider at first only the three most important parameters to have the same costs as before, we get the results presented in table 4.5. It lists all energy parameters with a non-vanishing SI, starting with the most important one. For $\eta = 0.4$ V, the results are the same as for the LSA recalculation before, due to the same set of important parameters. For the most complex case, $\eta = 0.7$ V, the variance reduces to a third of the original variance, which indicates that the GSA identifies better the right parameters. For $\eta = 1.0$ V we have a reduction factor of around 10, which is caused by the extended amount of important parameters, compared to the one parameter identified by the LSA.

If we now extend the amount to four important parameters, as also documented in table 4.5, we achieve a slight improvement of the variance in all cases. Finally, taking all the important parameters into account for the recalculation, we achieve the target reduction with an actual variance of under 1 for all three cases.

This shows that the GSA has the potential to quantify the dependence of the model output on its input parameters. It is important to note here that such reduced kinetic models are often only valid for a certain set of parameters. So now the question arises, how robust the GSA results are against extending the model. As we already discussed, for this particular model, we excluded the reaction from odd-numbered states since, for nominal values, the intermediate state would not be sufficiently populated [133, 134]. However, if we consider the uncertainties in the energies, this may not be the case. An extension would be to include all the reaction pathways for reaction 4.1 from oxidation states between $i = 4$ and $i = 13$ and extend the model to 25 dimensions. Figure 4.9 shows the results for the LSA and GSA for the most complex case of $\eta = 0.7$ V. Reactions from intermediate states below 4 are not possible

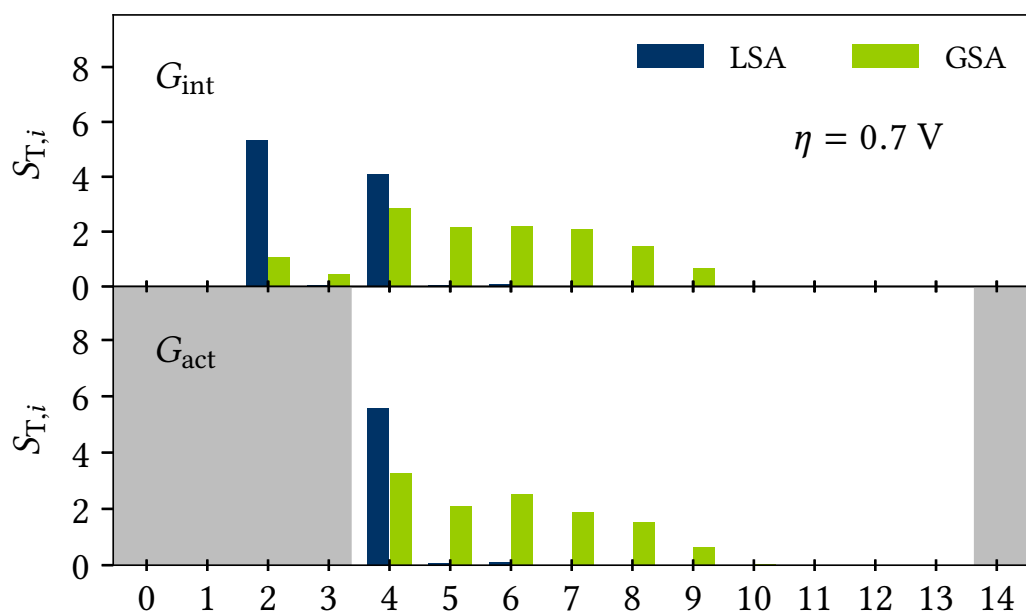


Figure 4.8: Local and global sensitivity indices $S_{T,i}$ for the extended model of the OER on the (110)-A surface for an overpotential $\eta = 0.7 \text{ V}$. The gray-shaded fields describe intermediate states that are not considered for a reaction in the model. Shown are the sensitivities of $\log_{10}(\text{TOF} \times s)$ with respect to the errors in the input parameters. Reproduced from [23], with the permission of AIP Publishing.

and are still excluded. For the energies of the new reactions, we assume a constant value of $\Delta\Delta E_{\text{act},5/7/9/11/13} = 0.6 \text{ eV}$ close to the other nominal energy values and an uncertainty of $\Delta\Delta E_{\text{act},5/7/9/11/13} = 0.3 \text{ eV}$.

The LSA numbers are nearly identical to the results of the not extended model in the middle panel of Fig. 4.6. For the GSA, the results are totally different. The GSA results shown in the Fig. 4.9 are performed on the SG results of $\epsilon = 0.01$. These results are different from the original case, although the TSIs for the even-numbered states up to 6 are nearly identical. However, for the extended model, the GSA shows a high dependency on the intermediate state and the corresponding activation energies of the odd states 5, 7 and 9. This rise of importance could have been expected from the results of the original model. The free energies of the odd-numbered intermediate states at $\eta = 0.7 \text{ V}$ are similar to the even states free energies. Furthermore, the barriers seem to be independent of the oxidation state so that there is no significant difference between even- and odd-states. A bigger difference occurs for the intermediate state of 8. While the original model only shows a minor dependency on state 8, the GSA shows now a significant TSI for the intermediate and activation energies. This shows that the GSA can guide chemical intuition when it comes to extending the model. As the initial result for the OER on the (110)-A surface showed a high dependency on the fourth and the sixth state for $\eta = 0.7 \text{ V}$, it is highly likely that the fifth and the seventh state would also be important. As the discussion about the results in figure 4.9 underlines that this would largely resemble the entire case, the first step would be to include the fifth and the seventh state.

Here it is important to note that the GSA for the full case shows significant dependencies ($S_{T,i} > 0.5$) on more than 14 parameters compared to 5 in the reduced model. This higher intrinsic dimensionality has a major impact on the SG performance. While the reduced model only need around 14000 points for a threshold of $\epsilon = 0.001$, it now needs approximately 70000 points.

This analysis shows the potential of the GSA to reduce the computational cost in microkinetic modelling. Based on the results, resources only need to be invested for the parameters in which uncertainties significantly impact the model outcome. This might make the GSA an essential feature for hierarchical modelling strategies. Instead of calculating all energies with an already expensive semi-local DFT method, the GSA allows starting with an even cheaper and less accurate method, e.g. machine learning potentials or semi-empirical Density Functional based Tight Binding [145, 146]. Then the GSA would identify the parameters that need to be recalculated with the next higher accurate method. In the next step, a new GSA can then determine which parameter to improve next and so on until the required accuracy is reached or the methods become too computationally expensive. In that context, a LSA

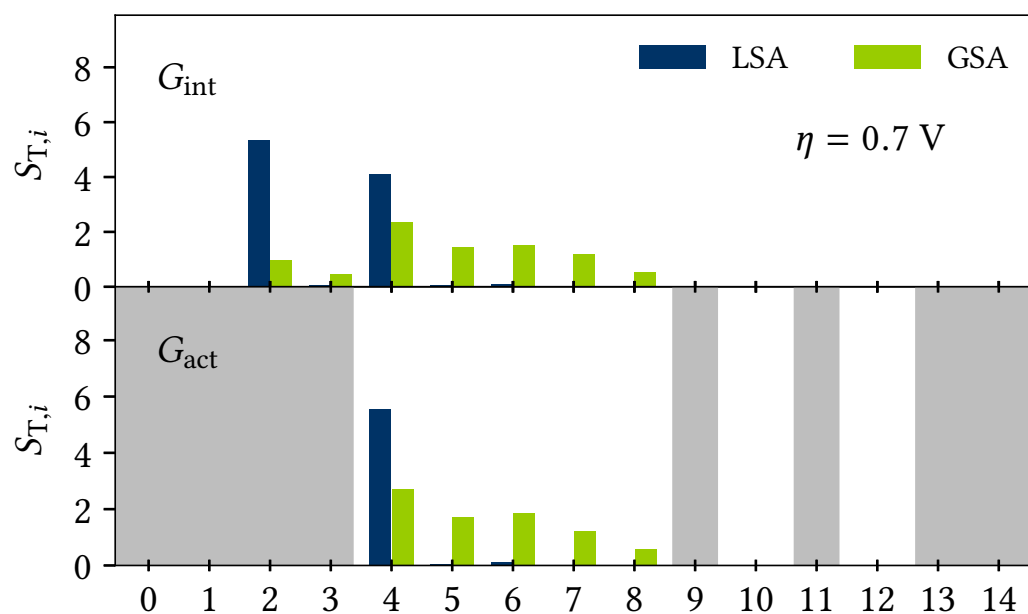


Figure 4.9: Global sensitivities indices and local sensitivities indices of the TOF for $\eta = 0.7 \text{ V}$ with respect to the energies for an extended model where the reactions from the intermediate states 5 and 7 are considered. Reproduced from the supplementary material of Ref. [23], with the permission of AIP Publishing.

method is questionable since the importance of non-linearities rises with increasing errors. So for such a hierarchical modelling structure, the LSA could be misleading and produce more computational costs than needed.

As an important note, the SA results are dependent on the employed error model and the corresponding choice of probability density function PDF. In this work, we utilize the worst case scenario, a uniformly distributed error with only bound constraints on the single error. In reality, the DFT errors may be correlated, but without more detailed information, except for the energy uncertainty of 0.3 V, we have to maximize the entropy for the PDF. For our test model, the OER on the Co_3O_4 , the errors in the intermediate energies result from an inaccurate description of the transition of Co(III) to Co(IV) by the DFT+U method. We, therefore, have to expect a strong correlation for the intermediate states. This means that they are either over- or underestimated, but definitely not independent as we assumed. The similar values of the barriers imply that they are independent of the degree of oxidation so that the errors are presumably identical for all barriers. Furthermore, the results already show some connection between intermediate and the corresponding activation barrier, which is backed up by the established Brønsted-Evans-Polanyi theory, which states that the barriers are correlated with the intermediate energies [6]. For such a correlated PDF we would assume a lower uncertainty than the one we employed, also causing changes in the SA indices. However, these changes are unpredictable because they strongly depend on the PDF and for a proper variance based analysis, we would have to employ an ANOVA decomposition for dependent parameters [94]. The GSA, as we introduced it, serves as a good starting point for identifying important parameters and their correlation that should be investigated in more detail [27, 147]. This can also be incorporated in the hierarchical method with firstly determining the correlation on a low accurate level and how it impacts the sensitivities. Based on that, the refining can continue with higher accurate methods.

This is the main application of the SA approach, but there are also possible applications in the field of material screening [6] or parametrization. In material screening, where the desire to cheaply examine a large class of materials often the variability causes the parameters uncertainties, the GSA and SG can serve as a tool to identify indicators of the catalytic function. Furthermore, in the case of comparison of different parametrisation methods (linear scalings[6], UBI-QEP [148], machine learning [149, 150]), the GSA results can help to choose the cheapest one. It helps to avoid the methods where important parameters are very expensive, but choose the parametrization where expensive parameters have a vanishing impact to avoid calculating them.

4.4 Conclusion

In this chapter, we have presented the most important individual parameters errors on the Quantity of Interest (QoI) for the Oxygen evolution model on Co_3O_4 . For that purpose, we performed a variance based Global Sensitivity Analysis (GSA) and compared the results to the common derivative based Local Sensitivity Analysis (LSA). Because of the need to efficiently estimating the numerical quadrature of high-dimensional integrals for determining the GSA indices, the approach utilizes a local- and dimension-adaptive Sparse Grid (SG) approach, introduced in chapter 3. This allows for an interpolation of the underlying function with a relatively low amount of points due to the efficient adaptation of the local features and the lower intrinsic dimensionality of the function. We employed the method on the Oxygen evolution on a Co_3O_4 (110)-A surface at room temperature. Here the lower amount of model evaluations have a limited impact on the computational costs, but for more complex models, like kinetic Monte Carlo [53] or methods for addressing the immediate solution of the Master Equation (ME) [30, 55], the number of model evaluations can be a significant drawback for the computational costs. However, the method's efficiency diminishes if the exploited characteristics, like low intrinsic dimensionality or local features, are not so pronounced.

For the employed model, we find uncertainties of two orders of magnitude or higher for the Turnover Frequency (TOF), which makes it impossible to assess whether the catalyst is active or not. It should be emphasized that the results and the uncertainty depend strongly on the parametrization and the employed error model. Both display an extreme test case with minimal pre-known information content about the model and the uncertainties of the parameters. In reality, correlations between the input parameters exist that might result in reduced uncertainty. However, even with the minimal information, the GSA can identify key atomistic features that are driving or hindering the catalytic performance [23]. In contrast, severe shortcomings have been found with the LSA to identify these parameters. The strong non-linearities of the model in the typical error bounds causes the LSA to identify a too small number of relevant parameters and to overestimate their importance.

This approach has a high potential to support the construction of higher accurate first principles-based models without tremendous costs by first identifying the important parameters, which then can be investigated more deeply by calculating the energies with higher-level methods [31, 32], by dedicated experiments or by incorporating previously ignored aspects, e.g. solvation effects [151]. Furthermore, the GSA can help to identify suitable indicators for catalytic performances in the context of material screening.

5 Multilevel Refinement

We already discussed in chapter 3.2, the advantages of the Sparse Grid (SG) approach in the context of numerical quadrature in high dimensions to overcome the *curse of dimensionality*. The key feature of adaptive refinement allows displaying function characteristics effectively without producing too many points. However, we showed only deterministic examples, although in many problem settings, the integrand has to be derived from a simulation algorithm, like Monte Carlo (MC), and single point evaluations can be costly. Furthermore, the simulation result will always carry a finite sampling error, which is inversely proportional to the computational cost. Even though the introduced adaptive Sparse Grid (ASG) method reduces the number of evaluations, the method might be unfeasible as a single evaluation can still be very expensive. Without more profound analysis, the SG and the ASG do not provide the information on how accurately single points have to be calculated and whether we have to spend the same computational costs on every grid point. In the past, multilevel MC have proven their ability to decrease computational costs by balancing the simulation with the numerical discretisation error [41, 152]. While these approaches mainly concentrated on estimating the expected value, we want to perform a Global Sensitivity Analysis (GSA) to extract more information from the MC model. Therefore we need a method to construct an accurate surrogate model, and in order to tackle our problem, we want to incorporate the multilevel idea in our SG approach to lower the costs.

So, if we want to employ such stochastic models for the GSA, we need an efficient approach to decrease the overall computational cost for approximating the integral and therefore the variance and the expected value of the target function $f(\mathbf{x})$ over a parameter domain Ω . Unlike common approaches, like Multilevel Monte Carlo (MLMC) [41] or Multilevel Sparse Grid (MLSG) [45, 46, 153], we will not utilize a hierarchy of functions with a different numerical approximations of the underlying model and approximate these functions independently. We construct a single surrogate model and exploit the intrinsic hierarchical structure of the SG.

As a starting point, we consider integrands implicitly given by a general MC simulation and where a finite sampling error causes deviations of the function value estimates. Parametric dependency of the output of such a model is necessary for the uncertainty quantification of chemical stochastic models [24, 52, 99]. However, the general method for the integrals can be applied to other popular fields like finance [100, 154] or molecular simulations [101].

In detail, there are many different methods for various applications, but often the efficiency is based on a high amount of simulations that can be performed on a low accuracy level and a

decreasing amount of simulations for higher accuracy levels.

The most common method is the MLMC approach with its first appearance in the late nineties [152, 155, 156]. The MLMC and later the Multilevel Quasi Monte Carlo (MLQMC) method has been applied to various problems, but in particular to uncertainty quantification, e.g for the treatment of partial differential equation (PDE)s with random input [43, 44]. In later years, the multilevel idea was also adapted for stochastic collocation with SG, MLSG methods [157]. As an extension to the multilevel idea, the multi-index idea was recently developed. The latter approach utilises a hierarchy of different general spatial approximations and different spatial approximations for different directions. The method produces different discretisation levels over the parametric domain along different axes. In the context of the SG, this utilises only the mixed derivatives that are most effective. The common Multi-Index Monte Carlo (MIMC) and Multi-Index Stochastic Collocation (MISC) can be seen as a combination of the SG approach and MC or Stochastic Collocation (SC) [47, 48]. These methods are used for problems controlled by multiple parameters computing the solution of a PDE with random data. Before we present our Multilevel Adaptive Sparse Grid (MLASG) approach, we will briefly introduce the common multilevel idea and discuss why this approach does not work for our problem.

5.1 Parametric Monte Carlo Models

Analog to the cases before, we consider a function $f : \Omega \rightarrow \mathbb{R}$ with $\Omega \in \mathbb{R}^D$, where we are interested in an accurate surrogate model to perform Sensitivity Analysis (SA). Again we restrict to those cases, where the domain equals a hypercube $\Omega = [0, 1]^D$, but here we assume that the function f is only implicitly given by a simulation code, some kind of Monte Carlo simulation.

To provide the integrands, the simulation code uses x as the input parameter and the Monte Carlo code provides an approximation of the function $f(x)$. Due to the random sampling, the approximation error correlates with the finite number of samples. Starting with a fixed x , the code draws samples from a probability distribution which is parametrically dependent on the input parameter x , with the outcome of a random variable Y_x . We suppose that the function value $f(x)$, is given by the expected value of Y_x , $E(Y_x)$ which is estimated from M_x samples and equals the function evaluation

$$f(x) = E(Y_x) \approx M_x^{-1} \sum_{i=1}^{M_x} y_{x,i} =: \overline{Y_{xM}} \quad (5.1)$$

where $y_{x,i}$ are the different sample outcomes according to the parameter x . We assume that the simulation code provides independently distributed samples for different values of x . For a fixed value of x all samples for $y_{x,i}$ are identical and independently distributed, but the distribution itself might depend on x . Then, the variance and the covariance of the estimates obey

$$\text{Var}(\overline{Y_{xM}}) = C_x M_x^{-1} \quad (5.2)$$

$$\text{Cov}(Y_{xM}, Y_{x'M}) = 0 \text{ for } x \neq x' \quad (5.3)$$

with $C_x = \text{Var}(Y_x)$. Eq. (5.2) follows from the assumption to have identical and independently distributed samples and a covariance of zero follows from the statistical independence for different values of x and x' . Therefore, with increasing sampling size M_x the variance decreases and with $M_x \rightarrow \infty$ the code would provide an accurate estimate. Unfortunately the computational costs increases with the M_x and $M_x \rightarrow \infty$ is infeasible. Therefore, our function value estimates will always carry some finite random noise, which will be propagated into the surrogate model. To practically estimate the variance we use the empirical variance formula

$$\text{Var}(\overline{Y_{xM}}) \approx (M_x(M_x - 1))^{-1} \sum_{i=1}^{M_x} (y_{x,i} - \overline{Y_{xM}})^2. \quad (5.4)$$

For the variance of the function evaluation we assume, that there exists an upper bound C^* , such that $C_x < C^*, \forall x \in \Omega$. Therefore we restrict to Monte Carlo models, which produce samples with finite variance irrespective of the value of the input parameter.

5.2 Multilevel idea

In this chapter, we address a similar problem setting as Heinrich and want to point out the differences to our *MLASG* method. One application by Heinrich uses the multilevel approach for approximating a function whose function evaluations are approximated by a parametric *MC* integration [49].

To explain the method, we start with a simple example that is close to our original problem and want to approximate the underlying function $u(x)$. With a basic parametric example

$$u(x) = \int_{\Omega_t} f(x, t) dt \quad (5.5)$$

with $f(x, t)$ as a function of $x \in \Omega_x$ and $\Omega_x \subseteq \mathbb{R}^{D_1}$, the parameter domain and $t \in \Omega_t$, with $\Omega_t \subseteq \mathbb{R}^{D_2}$, the integration domain. To approximate the integral we could apply a standard MC approach and fix a spatial approximation to a resolution of R with $x_i = \frac{i}{R}$, $i = 0, \dots, R$. Therefore, we estimate a single grid point evaluation with

$$u(x_i) = \frac{1}{M} \sum_{j=1}^M f(x_i, \omega_j). \quad (5.6)$$

ω_j are independently and identically distributed samples and the same set of ω_j is used for every x_i . This causes a high correlation between two approximations, which implies smooth approximation curves and avoids statistical fluctuation [49]. Based on the spatial (N_r) and the stochastic approximation the full function can be constructed by interpolation

$$u(x) \approx (P_R u)(x) = \sum_{i=0}^{N_R} u(x_i) \varphi(x) \quad (5.7)$$

$$\approx \frac{1}{M} \sum_{j=0}^M (P_R f(\cdot, \omega_j))(x), \quad (5.8)$$

where P_R describes the interpolation operator with a resolution of R . For a simple case, we deal with a piecewise linear interpolation and therefore with piecewise linear basis function φ , but the used interpolation operator depends on the problem [49]. The accuracy and the correlated computational cost of the approximation $u(x)$ is split in contributions of the spatial error and the stochastic error, assuming that with a finer spatial resolution $N \rightarrow \infty$ and increasing number of MC points $M \rightarrow \infty$ we achieve the exact solution. Achieving a high accuracy in a singular approach would cause high computational cost, which are not feasible for some problems. Therefore, the multilevel approach tries to balance the accuracy and the computational cost, to get the highest accuracy for minimal total cost.

For this purpose we sample not only from one fine spatial approximation, but from several ones, such that there is a sequence of grids with a hierarchy of spatial resolutions r , $x_{ri} = \frac{i}{r}$, $i = 0, \dots, r$, ($r = 0, \dots, R$). The key for multilevel approaches is to avoid approximating P_r directly on level r , but estimating the correction towards the next lower level $r - 1$, $\Delta P_r := P_r - P_{r-1}$, resulting in the telescoping sum

$$u_R(x)^{ML} = \sum_{r=0}^R (P_r u - P_{r-1} u) \approx \sum_{r=0}^R \frac{1}{M_r} \sum_{j=0}^{M_r} (P_r - P_{r-1}) f(\cdot, \omega_{rj})(x) \quad (5.9)$$

The approximation on the highest level R equals the lowest resolution level with additional corrections of the difference between two consecutive levels. Furthermore, the multilevel idea implies independently estimate each of these differences to lower the computational cost at a given variance. So instead of estimating all integrals $u(x_{ri})$ with a fixed M the multilevel adapts the number of MC points of every level and ω_{rj} , $r = 0, \dots, R$, $j = 0, \dots, M_r$ are independent and uniformly distributed over Ω_t . It turns out that often the most MC points have to be spent on the lowest spatial approximation, and with increasing resolution, the amount of MC points can be decreased [44, 158]. Although Heinrich's targeted model is very close to our problem, we cannot apply it. Heinrich's underlying assumption of strong coupling through using the same set of samples is not true for a kinetic Monte Carlo (kMC) model. Studies have shown that for a kMC simulation, the coupling is very weak, especially for strong trajectories. Furthermore, we need to use a SG approach to break the *curse of dimensionality* and use adaptivity to decrease the number of points. A different approach, which is very common, concentrates on approximating the expected value of the model with parametric uncertainty [43, 44]. For this case the multilevel approach is applied for the sampling accuracy of the numerical approximation and not for the spatial approximation. Regarding this strategy many methods have been developed including a SG approach to determine the sampling accuracy, but all of these examples deal with deterministic models and not stochastic models [157]. When transferred it to our problem we want to approximate the target function and the multilevel approaches decompose the function (or its integral) using a telescoping sum to approximate f on level R

$$f \approx f_R = f_0 + \sum_{r>0}^R f_r - f_{r-1} = f_0 + \sum_{r>0}^r \Delta f_r. \quad (5.10)$$

The function f_0 corresponds to the lowest level of evaluation accuracy and f_r corresponds to increasing level accuracy $r = [1, R]$. With $R \rightarrow \infty$, f_R approaches f . Now we need to discretise each difference independently, but because the terms will become smaller with increasing r , only coarse discretisations are needed. For our Monte Carlo model, the accuracy level would correspond to the drawn samples M per function evaluation. If we now apply the adaptive SG approach, we would produce random refinement due to the random error at low levels caused by the finite sampling. The method would not benefit from the adaptivity or, in the worst case, draw to wrong results due to too early terminations of the refinement. To enable the multilevel idea for an adaptive quadrature, we do not work with multiple sparse grid approximations but on a single SG and employ the intrinsic multilevel structure of the SG itself.

5.3 Sparse Grid

We employ the same *SG* structure as described in detail in section 3.2. In this section we want to briefly recap the important features of the *SG* approach for the multilevel extension. We work with piecewise linear basis function (Eq. (3.9)) to construct the *SG* surrogate model to approximate the function $f(\mathbf{x})$

$$f(\mathbf{x}) \approx u_L(\mathbf{x}) = \sum_{|\mathbb{I}_1| \leq L, \mathbf{i} \in \mathcal{I}_1} v_{\mathbf{1},\mathbf{i}} \varphi_{\mathbf{1},\mathbf{i}}(\mathbf{x}) \quad (5.11)$$

with the hierarchical indices sets \mathcal{I}_1 , defined by Eq. (3.20). The hierarchical surplus $v_{\mathbf{1},\mathbf{i}}$ can be calculated through the recursive formula, implied by the interpolation requirement Eq. (3.7)

$$v_{\mathbf{1},\mathbf{i}} = f(\mathbf{x}_{\mathbf{1},\mathbf{i}}) - u_{|\mathbb{I}_1|-1}(\mathbf{x}_{\mathbf{1},\mathbf{i}}). \quad (5.12)$$

It defines the hierarchical increments between two successive levels and characterizes the interpolation error of the $|\mathbb{I}_1| - 1$ approximation, making it a key characteristic for the refinement of the sparse grid.

The second important feature is the definition of the quadrature. The integral is approximated with (3.29) which in detail gives

$$\sum_{|\mathbb{I}_1| \leq L, \mathbf{i} \in \mathcal{I}_1} v_{\mathbf{1},\mathbf{i}} w_{\mathbf{1},\mathbf{i}} = \sum_{|\mathbb{I}_1| \leq L, \mathbf{i} \in \mathcal{I}_1} v_{\mathbf{1},\mathbf{i}} \prod_{d=1}^D w_{l_d, i_d} \quad (5.13)$$

$$\text{with } w_{l,i} = \int_0^1 \varphi_{l,i}(x) dx = \begin{cases} 2^{-l} & \text{for } l = 0 \\ 2^{-2l} & \text{for } l = 1 \\ 2^{-l} & \text{for } l > 1 \end{cases} \quad (5.14)$$

with the values for the one-dimensional basis function (*BF*)'s weights $w_{l,i}$. For this choice of *BF* the weight is the L_1 -norm of the *BF* $\varphi_{l,i}$. Therefore, $|v_{\mathbf{1},\mathbf{i}} w_{\mathbf{1},\mathbf{i}}|$ can be counted as the contribution by the *BF*.

5.3.1 Adaptivity

To decrease the number of points, compared to a normal *SG*, we use the local-adaptive refinement described in Sec. 3.3.3. With Eq. (3.46) we are able to neglect points with a small contribution $v_{\mathbf{1},\mathbf{i}} \varphi_{\mathbf{1},\mathbf{i}}(\mathbf{x})$ to the surrogate model. As a refinement indicator we still employ the L_1 -

norm of the contribution $\gamma_{1,i}$ (Eq. (3.40)), which shows a reasonable decrease with increasing refinement levels for localized rapid changes. According to Eq. (3.46), the refinement set \mathcal{S}_{loc} includes the children $\mathcal{C}_{1,i}$ (Eq. (3.14)) of the BFs in $|\mathbf{I}|_1 = L$, with L being the currently highest refinement level in the SG, whose indicator $\gamma_{1,i}$ exceed the predefined threshold ϵ .

To use the sparse grid's multilevel structure, later on, we consider only the local refinement, as it also has an implicitly given dimension adaptivity as points of unimportant dimensions have a quickly diminishing error indicator. Additionally to the classical local refinement, we include also all ancestors $\mathcal{A}_{1,i}$ (Eq. (3.16)) of the new set points in every refinement step. This procedure is repeated until no point shows an indicator higher than the predefined threshold. With this strategy and by equation (5.12) we ensure identical surpluses of the adaptive strategy compared to a full sparse grid, except for the points not included. This is a crucial necessity to avoid an optimization within every refinement step to minimize the numbers of sample for $f(\mathbf{x}_{1,i})$. Therefore the multilevel approach in the next section becomes less complex.

5.4 Multilevel Adaptive Sparse Grid

The MLASG method does not follow the outlined multilevel strategy but rather employs the multilevel structure of the SG which is based on the hierarchical structure. Higher-level BFs lead to higher resolution of the grid but it also means that their contribution $v_{1,i}w_{1,i}$ has a diminishing impact on the integral value. The weights as defined in Eq. (5.14) decay exponentially with the level, and the surplus should asymptotically decay, at least for bound mixed derivatives [104]. Furthermore, the discussion on the construction of the surrogate model and the numerical integration in sec. 5.1 shows that the accuracy increases by adding points, regardless of the sampling variance. Because of the decreasing impact of a single sparse grid point, we assume that the estimations' accuracy can decay with increasing refinement level, and we have to spend less computational effort for higher-order estimations. For the construction of the sparse grid, we include all the BF that increase the accuracy of the surrogate model, so that the error indicator $\overline{\gamma}_{1,i}$ are below a predefined accuracy ϵ

$$\overline{\gamma}_{1,i} = |v_{1,i} \cdot w_{1,i}| \leq \epsilon. \quad (5.15)$$

The idea of MLASG is to limit the sampling effort for the estimates $\overline{Y}_{1,i}$ with respect to the refinement level, such that the adaptive refinement is not drastically effected. In order to

avoid random refinement, we control the variance of the estimated error indicator $\gamma_{\mathbf{l},\mathbf{i}}$

$$\begin{aligned} \text{Var}(\overline{\gamma_{\mathbf{l},\mathbf{i}}}) &\leq c \cdot \epsilon^2 \\ \Leftrightarrow \text{Var}(\overline{v_{\mathbf{l},\mathbf{i}}}) &\leq c \cdot \epsilon^2 \cdot w_{\mathbf{l},\mathbf{i}}^{-2}, \end{aligned} \quad (5.16)$$

with a user defined constant c , whose use will be explained later. Because we use piecewise linear basis function we can limit the weight of a BF to $w_{\mathbf{l},\mathbf{i}} \leq 2^{-|\mathbf{l},\mathbf{i}|}$ (Eq. (5.14)) Together with Eq. (5.16), this means that we can increase the variance of the surplus $v_{\mathbf{l},\mathbf{i}}$ by a factor of four in every refinement step. Unfortunately, this relation cannot be directly propagated to the function estimation $\overline{Y_{\mathbf{l},\mathbf{i}}}$ whose MC sampling accuracy we can control. According to Eq. (5.12), the surplus are linear combinations of the estimates $\overline{Y_{\mathbf{l},\mathbf{i}}}$

$$v_{\mathbf{l},\mathbf{i}} = f(\mathbf{x}_{\mathbf{l},\mathbf{i}}) - u_{|\mathbf{l},\mathbf{i}|-1}(\mathbf{x}_{\mathbf{l},\mathbf{i}}) = \overline{Y_{\mathbf{l},\mathbf{i}}} - u_{|\mathbf{l},\mathbf{i}|-1}(\mathbf{x}_{\mathbf{l},\mathbf{i}}). \quad (5.17)$$

As the estimates are statistically independent we define a correlation between the surplus variance to the estimate variance

$$c w_{\mathbf{l},\mathbf{i}}^{-2} \epsilon^2 \geq \sum_{\mathbf{m},\mathbf{j}} A_{\mathbf{l},\mathbf{i};\mathbf{m},\mathbf{j}}^2 \cdot \text{Var}(\overline{Y_{\mathbf{m},\mathbf{j}}}), \quad (5.18)$$

where A results from the recursive formula Eq. (5.17) and maps function values to the surplus with \mathbf{m} including all levels up to \mathbf{l} . This means the inverse of A consists of the BFs evaluations at the grid points $\mathbf{x}_{\mathbf{i}}$, $A_{\mathbf{l},\mathbf{i};\mathbf{m},\mathbf{j}}^{-1} = \{\varphi_{\mathbf{m},\mathbf{j}}(\mathbf{x}_{\mathbf{l},\mathbf{i}})\}$. $A_{\mathbf{l},\mathbf{i};\mathbf{m},\mathbf{j}}$ is a sparse matrix with nonzero values for identical indices $(\mathbf{m},\mathbf{j}) = (\mathbf{l},\mathbf{i})$ or if (\mathbf{m},\mathbf{j}) is an ancestor of (\mathbf{l},\mathbf{i}) . The entries of the matrix only depend on the index pair and not on the function nor the threshold ϵ , so that it can therefore be calculated beforehand for (\mathbf{l},\mathbf{i}) if we include all ancestors. To be able to apply Eq. (5.17) for all points, the refinement strategy includes all ancestors of the node in each refinement step. This means that for fixed (\mathbf{l},\mathbf{i}) we have the same coefficients $A_{\mathbf{l},\mathbf{i};\mathbf{m},\mathbf{j}}^{-1}$ for the adaptive strategy as we would have for no adaptive refinement.

With this we want to choose the sampling effort of $\overline{Y_{\mathbf{l},\mathbf{i}}}$ beforehand so that the inequality (5.18) is fulfilled. The sampling effort of $\overline{Y_{\mathbf{l},\mathbf{i}}}$ depends on the number $M_{\mathbf{l},\mathbf{i}}$ of MC samples which have been drawn for the parameter set $\mathbf{x}_{\mathbf{l},\mathbf{i}}$. $M_{\mathbf{l},\mathbf{i}}$ displays the computational effort spent on that grid point and $\Delta M_{\mathbf{l},\mathbf{i}}$ describes the additional sampling to fulfill the inequality. If now new points are added in a refinement step, the total additional sampling effort $\sum_{\mathbf{l},\mathbf{i}} \Delta M_{\mathbf{l},\mathbf{i}}$ has to be minimized with the subject to the constrained Eq. (5.18) utilizing Eq. (5.4) or Eq. (5.2) if the variance of $\overline{Y_{\mathbf{l},\mathbf{i}}}$ or a bound C^* is known. Since we want to avoid to solve the integer minimization problem along the simulation, we assume that we can find a $B > 0$, such that

choosing

$$\text{Var}(\overline{Y_{\mathbf{l},\mathbf{i}}}) \leq c\epsilon^2 B^{|\mathbf{l},\mathbf{i}|}, \quad (5.19)$$

fulfills the inequality (5.18). This propagates into the inequality

$$c 2^{2|\mathbf{l},\mathbf{i}|} \epsilon^2 \geq \sum_{\mathbf{m},\mathbf{j}} A_{\mathbf{l},\mathbf{i};\mathbf{m},\mathbf{j}}^2 c \epsilon^2 B^{|\mathbf{m},\mathbf{j}|} \quad \text{with } w_{\mathbf{l},\mathbf{i}} \leq 2^{|\mathbf{l},\mathbf{i}|} \quad (5.20)$$

$$4^{|\mathbf{l},\mathbf{i}|} \geq \sum_{\mathbf{m},\mathbf{j}} A_{\mathbf{l},\mathbf{i};\mathbf{m},\mathbf{j}}^2 B^{|\mathbf{m},\mathbf{j}|}. \quad (5.21)$$

With our refinement strategy that includes all ancestors we can exploit that $A_{\mathbf{l},\mathbf{i};\mathbf{m},\mathbf{j}}^2$ for a fixed (\mathbf{l}, \mathbf{i}) contains the same entry as for a non adaptive approach. For points of a higher level $|\mathbf{m}|_1 > |\mathbf{l}|_1$ holds $A_{\mathbf{l},\mathbf{i};\mathbf{m},\mathbf{j}}^2 = 0$ by equation (5.17). To find the right B , we tested different values for B for various dimensionalities, using $A_{\mathbf{l},\mathbf{i};\mathbf{m},\mathbf{j}}^2$ from a non adaptive SG approach. As we want to apply the method to a 7 dimensional problem later on, we tested the method in this range of dimensionality and found that $B = 2$ fulfills the inequality of Eq. (5.21).

Using the upper bound C^* for the variance of a single Monte Carlo sample, we can now estimate the computational cost \mathcal{C}_{all} spent for Monte Carlo simulation for the final sparse grid with $|\mathbf{l}|_1 = l$ in terms of the total number of samples

$$\mathcal{C}_{all} \leq \frac{C^*}{c\epsilon^2} \sum_{l=0}^{l_{max}} 2^{-l} n(l). \quad (5.22)$$

where $n(l)$ is the number of SG nodes in the final adaptive grid. This is in contrast to a single level ASG where we estimate all points with the same number of MC samples and which has a complexity $\sim \sum_{l=0}^{l_{max}} n(l)$. If MLASG and ASG have roughly the same number of points in each level, MLASG therefore produces significantly lower costs.

The pseudocode for the MLASG approach is provided in algorithm 4. Starting with a level zero SG, which is the most accurate and therefore the most expensive point calculation of the SG, we have an initial kMC calculation for M samples. Here, \mathcal{S} is the set of points in the adaptive sparse grid, whereas \mathcal{S}_l is the set of points with $|\mathbf{l}|_1 = l$ added in the l -th refinement step. \mathcal{S}_{loc} is the total set of points added in each refinement step and $\mathcal{C}_{\mathbf{l},\mathbf{i}}$ and $\mathcal{A}_{\mathbf{l},\mathbf{i}}$ are the children and the ancestors of \mathbf{l}, \mathbf{i} , respectively. Of course, the above strategy works also using a higher level nonadaptive SG as starting point as long as the surplus obeys Eq. (5.16). The detailed pseudocode is shown in algorithm 4.

There are many different options for the choice of c , but the most promising seems to be $c = 1$. Much lower values would not improve the surrogate model's accuracy concerning the increasing computational effort. Choosing C much larger than one $c > 1$ increases the

Algorithm 4: Pseudocode for MLASG

```

Choose threshold  $\epsilon$ 
 $l = 0$ 
 $\mathcal{S} = \emptyset$ 
 $Y_{0,0} = \overline{Y_{x_{0,0}M}}$  using  $M$  samples  $\{y_{x_{0,0},i}\}_{i=1}^M$ . Choose  $M$  such that
 $(M(M-1))^{-1} \sum_{i=1}^M (y_{x_{0,0},i} - \overline{Y_{x_{0,0}M}})^2 \leq c\epsilon$ 
 $v_{0,0} = Y_{0,0}$ 
 $\mathcal{S}_0 = \mathcal{S} \cup \{(0,0)\}$ 
 $\mathcal{S} = \mathcal{S}_0$  while  $\mathcal{S}_0 \neq \emptyset$  do
   $R = \emptyset$ 
   $\mathcal{S}_{l+1} = \emptyset$ 
  for  $(l,i) \in \mathcal{S}_l$  do
    if  $v_{l,i} w_{l,i} \geq \epsilon$  then
       $\mathcal{R} = \mathcal{R} \cup \mathcal{C}_{l,i}$ 
       $\mathcal{S}_{l+1} = \mathcal{S}_{l+1} \cup \mathcal{C}_{l,i}$ 
      for  $(m,j) \in \mathcal{C}_{l,i}$  do
         $\mathcal{R} = \mathcal{R} \cup \mathcal{S}_{m,j} \setminus \mathcal{S}$ 
  for  $(m,j) \in \mathcal{R}$  do
     $Y_{m,j} = \overline{Y_{x_{m,j}M}}$  using  $M$  samples  $\{y_{x_{m,j},i}\}_{i=1}^M$ . Choose  $M$  such that
     $(M(M-1))^{-1} \sum_{i=1}^M (y_{x_{m,j},i} - \overline{Y_{x_{m,j}M}})^2 \leq c\epsilon B^l$ 
   $\mathcal{S} = \mathcal{S} \cup \mathcal{R}$ 
  for  $n=0:l+1$  do
    for  $(m,j) \in \mathcal{R}$  with  $|m|_1 = n$  do
      Calculate  $v_{m,j}$  using Eq. (5.12) from the nodes in  $\mathcal{S}$  and corresponding
      function estimates.
   $l = l + 1$ 

```

possibility of random refinement as the noise increases. With $c = 1$ the standard deviation of the lowest level point, the most accurate equals the threshold ϵ . For error indicators larger than ϵ , the corresponding variance is sufficient to avoid false refinement. Only if the surplus is in the same order as the ϵ the refinement strategy could pick up the noise of the surplus and falsely refine the points. However, $\gamma_{l,i}$ is a rough indicator. So possible errors due to this indicator have the same effect as if we erroneously refine points due to noise of order ϵ . The relatively wider choice of c is possible due to the conservative choice of B and the corresponding level variance ratio in Eq. (5.19). This implies that the error indicators might have a significantly lower variance than ϵ^2 .

So MLASG seems to be able to accelerate the construction of surrogates for Monte Carlo

models, i.e. a similar setting as Heinrich has treated initially in his pioneering work on MLMC [152, 155]. Before the method is tested at a CO oxidation model, a first principle kMC model, we display the method's performance at two examples, a low-dimensional 2D problem and a medium dimensional 7D problem.

5.5 An Illustrative model

To effectively test the method for a Monte Carlo model, the test model should reflect the same characteristics as the MC model. So we will test the method initially for the a function in $\Omega = [-0.5, 0.5]^D$ with similar characteristics as kinetic data

$$f(\mathbf{x}) = g\left(\sqrt{\sum_i^D (x_i + 0.5)^2}\right), \quad (5.23)$$

with

$$g(r) = \begin{cases} g_1(r) = 10 \cdot \left(\exp\left(\frac{-r + 0.35}{0.086}\right) + 1\right)^{-1} & , \text{for } r < 0.6 \\ g_2(r) = 0.005^r \cdot g_1(0.6) \frac{1}{(0.005^{0.6})} & , \text{for } r \geq 0.6 \end{cases} \quad (5.24)$$

which has a kink for $\sqrt{\sum_i^D (x_i + 0.5)^2} = 0.6$, describing the local non linear behavior of kinetic models [28]. The function for the $D = 2$ case is displayed in figure 5.1. As this model is deterministic, we have to add artificial noise to the function evaluations to mimic the costs for the MC model output. We use normally distributed, zero mean noise, whose variance increases with increasing level. The values $Y_{\mathbf{l};i}$ used for the sparse grid construction

$$Y_{\mathbf{l};i} = f(\mathbf{x}_{\mathbf{l};i}) + s_{\mathbf{l};i} \quad (5.25)$$

consists of the function evaluation and an additional noise $s_{\mathbf{l};i}$, whose variance equals

$$\text{Var}(s_{\mathbf{l};i}) = c \epsilon^2 2^{|\mathbf{l}|_1}. \quad (5.26)$$

So with every level increment, the variance of the noise doubles. In connection with the MC model we introduced the variance, connected to the computational cost by equation (5.2), so

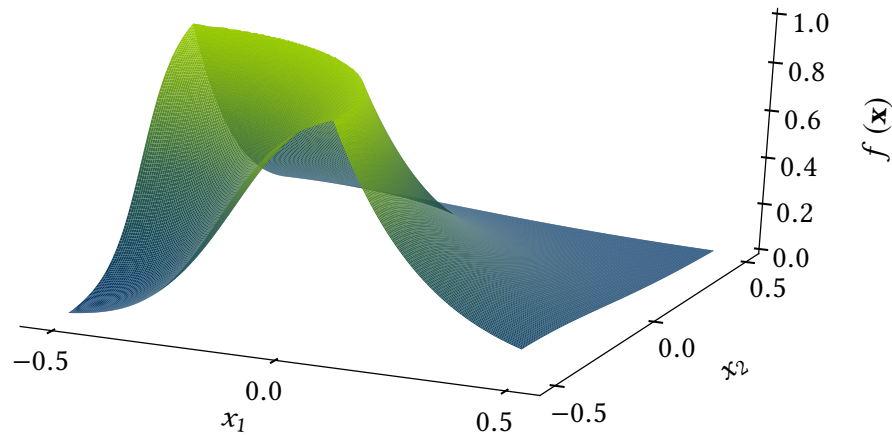


Figure 5.1: An illustrative function for the **MLASG** method which has the similar characteristics as the kinetic data. Kinetic data has rather sharp changes depending on more than one dimension. In the 2 dimensional figure the changes are displayed as a kink at a radial value of $r = \sqrt{\sum_i^D (x_i + 0.5)^2} = 0.6$.

that we assign a cost ratio of $2^{-|l|_1}$ to the sample $Y_{\mathbf{l};i}$ to mimic the cost of a parametric **MC** model. For analysing the **MLASG** approach, we are going to compare it to the classical **ASG** approach with local adaptivity.

5.5.1 2D-case

Before we start discussing the benefits of the multilevel approach in terms of computational costs, we will illustrate how much the multilevel idea influences adaptive refinement. Therefore we choose the test-model with $\epsilon = 10^{-4}$ for $D = 2$, in order to visualize the results more easily. As discussed before, we set $c = 1$ and the standard deviation σ_0 of the lowest level point $|\mathbf{l}|_1 = 0$ to ϵ . Figure 5.2 displays the sparse grids constructed by the **MLASG**, in the left figure, and the **ASG**, in the right figure. The **ASG** refinement is based on the function values $f(\mathbf{x}_{\mathbf{l};i})$ and the **MLASG** refinement on the estimates of Eq. (5.25). As the underlying function has the characteristics of a local kink with a locally non-linear behaviour, both sparse grids show an increased density of points toward this kink. For the areas of less local rapid changes, the grids show a more sparse pattern due to the rather linear behaviour in these areas. However,

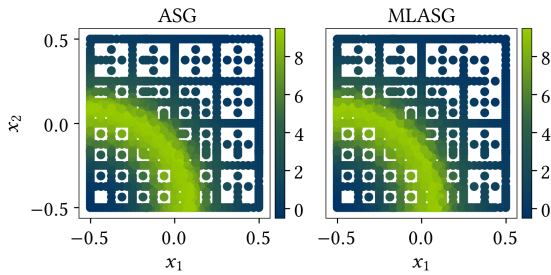


Figure 5.2: Comparison of the adaptive grids for $\epsilon = 10^{-4}$ for *ASG* without noise (*left figure*) and *MLASG* (*right figure*) with an initial standard deviation $\sigma_0 = 10^{-4}$, i.e. $c = 1$. Both final sparse grids show the same results for adapting the kink of the function with only a minimal difference in the number of points.

NoP_{ASG}	1117
$\text{NoP}_{\text{MLASG}}$	1118
$\text{NoP}_{\text{ASG}} \Delta \text{NoP}_{\text{MLASG}}$	11

Table 5.1: The number of points (NoP) for the final grids of both methods. The first row displays the results for *ASG* and the second for *MLASG*. The third row displays the actual difference in the number of points for both final grids.

both grids contain almost the same grid points and differ only in 1% of the points. Table 5.1 documents the number of points (number of points (NoP)) for both adaptive strategies and the symmetrical difference between the point sets, which are roughly 1% of the total amount of points. Both refinement methods increase the density of points in the local kink area so that the multilevel approach does not strongly influence the adaptive strategy. But how does it affect the overall results for the L_1 -norm of the approximation and the integral?

As our goal is to construct an accurate surrogate model, we will mostly concentrate on the L_1 -norm of the approximation error. Still, we will also inspect the quadrature error as it is a good indicator for the *GSA*, later on. Figure 5.3 displays the L_1 -norm (*upper row*) and quadrature error (*lower row*) of both *SG* approaches, the classical *ASG* and the *MLASG*. These do not show the convergence of the final results after refinement for different values of ϵ , but the behaviour during refinement for a fixed value of $\epsilon = 10^{-4}$ and $c = 1$.

The *MLASG* and the *ASG* show essentially the same convergence behaviour for the L_1 -norm during the refinement as both result lines overlap in the upper left figure for the *NoP* plot. Similar results can be seen for the quadrature error in the lower left figure. Only small differences occur for small quadrature errors due to a very fine resolution of the grid and a refinement criteria in the same order of magnitude as the applied noise. Despite the difference, both methods show almost the identical trend and even for the proceeded refinement, it is not definite which method shows a better performance. To discuss the benefits of the multilevel approach, we have to define a measure for the *CPU* savings. For this deterministic model, the savings are, similar to the noise, artificial, but gives a good impression of how much potential the method has.

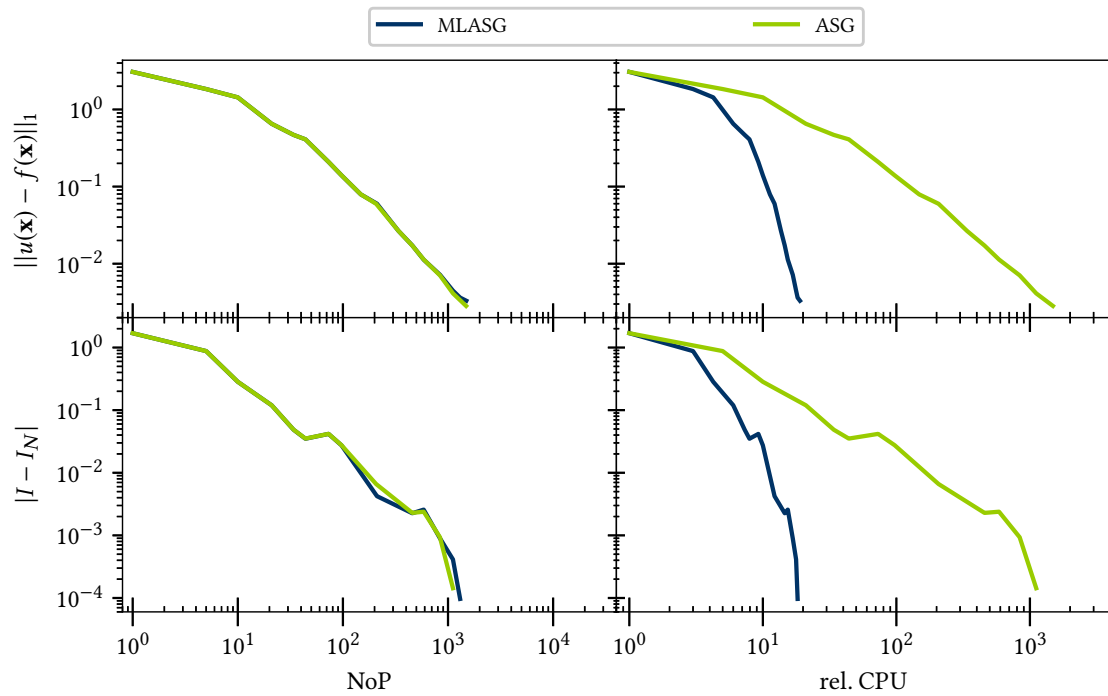


Figure 5.3: Results for the 2D case of the illustrative function (Eq. (5.24)) for the MLASG and the ASG method over the NoP (left column) and the relative CPU (right column). Upper row: Results for the L_1 -norm of the interpolation error $\|u(\mathbf{x}) - f(\mathbf{x})\|_1$ during the refinement. Lower row: Results for the quadrature error $|I - I_N|$ for the two methods during the refinement.

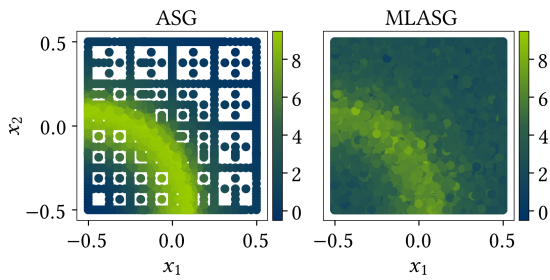


Figure 5.4: Comparison of the adaptive grids for $\epsilon = 10^{-4}$ for ASG (left figure) without noise and MLASG (right figure) with an initial standard deviation $\sigma_0 = 10^{-2}$, i.e. $c = 100$.

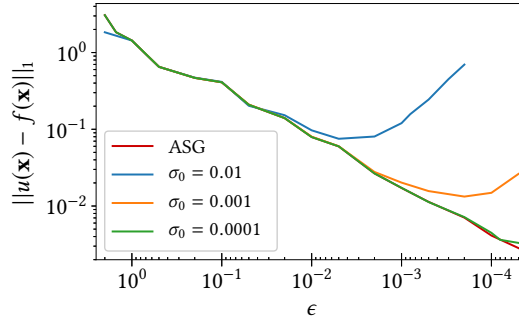


Figure 5.5: L_1 -Norm of the interpolation error for different initial noise values $\sigma_0 = [10^{-2}, 10^{-3}, 10^{-4}]$ compared to the ASG method without noise.

As a reference value, we define a single point calculation of the ASG, i.e. with the highest accuracy. Therefore the CPU results of the ASG are identical to the NoP results. For the multilevel approach we calculate the computational costs of one SG point according to $\mathcal{C} = 2^{-|l|_1}$ (Eq. (5.22)) which describes the roughly the factor of CPU savings compared to the ASG method. If the CPU time equals the NoP for the ASG we would have a CPU time for the MLASG of $\frac{\text{NoP(ASG)}}{2^{|l|_1}}$ for the MLASG.

The results of the two methods for the relative CPU, shown in the right column of figure 5.3, indicate a rapid drop of the interpolation and the quadrature error for the MLASG approach compared to the ASG. As both methods start with the same initial grid (5 points) for $|l|_1 = 1$ and the same accuracy for the first centre point of the sparse grids, the initial cost is similar for both sparse grids. During the refinement of the sparse grid, the computational savings of the MLASG increase, but also at very coarse sparse grids, the effect of the savings are around half a magnitude. Since both approaches show the same behaviour for the NoP, the MLASG approach can decrease the cost ratio by two orders of magnitude without affecting the accuracy of the quadrature.

Before we discuss the multilevel effect in a higher dimensional case, we want to discuss the limits of the MLASG and the effect of c , which is responsible for the ratio between the variance and the refinement tolerance ϵ .

The right column of figure 5.4 shows the grid results if we consider an initial increased standard deviation of $\sigma_0 = 10^{-2}$ for the lowest level point for the MLASG method. Compared to that, the left column displays the results for the ASG with exact function values. For both we exploit a threshold of $\epsilon = 10^{-4}$, thus the difference to the earlier results is an increased $c = 100$ instead of $c = 1$ (as in figure 5.2). For the increased value of c , the MLASG refinement does not follow the function's characteristics but randomly refines points to spread over the

domain. Due to the value of c , the refinement criteria' noise is a factor hundred bigger than the threshold ϵ . As a consequence, the adaptive refinement falsely refines the noise.

In order to analyze the limits of the **MLASG** method according to different c values, we tested the performance for varying thresholds $\epsilon = [10^{-4}, 2]$ for three zero-level standard deviations $\sigma_0 = [10^{-2}, 10^{-3}, 10^{-4}]$.

Figure 5.5 shows the interpolation error of the three cases as a function of the threshold ϵ , where we included the ideal **ASG** case for comparison. Starting at the largest values for ϵ , all three curves follow the ideal **ASG** until the threshold ϵ becomes close to the respective value of σ_0 with $c = 1$. A further refinement would lead to points containing a variance bigger than the threshold ϵ , which causes random refinement. This is indicated by the flatted curves and the increase of the interpolation error towards smaller ϵ . This underlines that the accuracy of the interpolant highly depends on the initial noise applied to the function. We lose the benefit of the adaptive refinement and interpolation accuracy for much larger values of c than $c \approx 1$. This means, for high accuracy, the initial noise has to be chosen adequately low, while for a lower accuracy, the benefit of the multilevel approach can be enhanced with an increased c .

5.5.2 7D-case

As a higher dimensional case, we consider the function Eq. (5.23) for 7 dimension. As in the previous section, we employ a refinement tolerance of $\epsilon = 10^{-4}$ and $c = 1$. Similar to the 2D-case in subsection 5.5.1, figure 5.6 shows the convergence during the refinement of the two different methods, **ASG**, **MLASG**, with respect to the L_1 -norm of the interpolation error and the quadrature error. In the left column both quantities of interest are displayed as a function of points and of the relative **CPU** in the right column. As expected, the curves for the number of points for both errors do not differ much between the **ASG** and the **MLASG** method for, except for a very high accuracy, where the **MLASG** outperforms the **ASG** regarding the quadrature error. This phenomenon is rather random, since the error indicator for the refinement is very coarse compared to the little deviations in the indicator, as it is described in sec. 5.4. For the interpolation error (*upper left panel*), the difference for a higher accuracy is not as pronounced but still visible. Regarding the results for the **CPU**-savings for both Quantity of Interest (**QoI**)s, the multilevel method, **MLASG**, saves two orders of magnitude compared to **ASG**.

All in all, the **MLASG** achieves the higher accuracy with two orders of magnitude in cost savings compared to the **ASG** method. This example already shows the possibilities of the

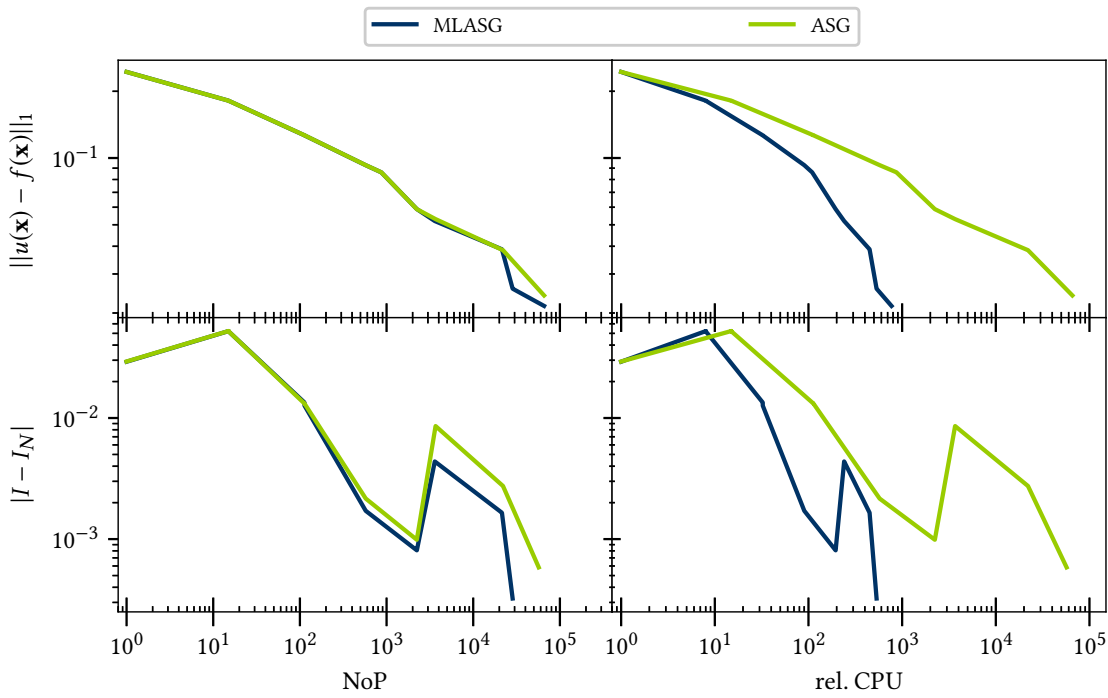


Figure 5.6: Results for the 7D case of the illustrative function (Eq. (5.24)) for the MLASG and the ASG method over the NoP (left column) and the relative CPU(right column). Upper row : Results for the L_1 -norm of the interpolation error $\|u(\mathbf{x}) - f(\mathbf{x})\|_1$ during the refinement. Lower row: Results for the quadrature error $|I - I_N|$ for the two methods during the refinement.

MLASG. Due to the local refinement strategy, the impact of excluding points is strongly visible in the $7D$ case, but still, the implicitly given dimension adaptivity is not shown. Since the characteristics depend on every dimension, the adaptivity has to refine in every dimension and towards higher levels. In the kinetic model, only a fraction of the parameters are considered important for the surrogate model's construct. There the adaptivity is even stricter with a bigger impact on the computational costs so that it is very interesting whether the **MLASG** still behaves similar to the **ASG**.

5.5.3 7D-2D-case

To test the case of lower intrinsic dimensionality, we consider a $7D$ domain and the function Eq. (5.23) is only defined in $2D$, i.e. we include 5 dummy dimensions. Figure 5.7 is equivalent to the results figure of $2D$ and $7D$, which shows the convergence during refinement of the two methods for the L_1 -norm of the interpolation error and quadrature error, regarding the **NoP** (*left column*) and the relative **CPU** (*right column*). The value for c and the initial error is identical to the recent examples ($\sigma_0 = 10^{-4}$, $c = 1$). As we expected for the **NoP**, the **ASG** and the **MLASG** curves mostly agree for the interpolation and the quadrature error. Thus, the adaptation of the lower dimensional function in $2D$ works well for the **MLASG** method.

The multilevel approach in figure 5.7 achieves a cost saving of around two orders of magnitude compared to the **ASG**. These three examples present the multilevel approach's potential and that the performance of the multilevel approaches depends on the function itself. All in all, we can save 1 to 2 orders of magnitude with a multilevel extension of the adaptive **SG**. Since these are only toy models, we will address in the next sections 5.6 and 5.7 the **MLASG** approach for a number test functions and a realistic **kMC** model, respectively.

5.6 Test examples

We have now established that the multilevel method works well for functions with similar characteristics as the target kinetic data. However, we want to know how the **MLASG** method performs for general examples, and where the possible limits are for the method. Genz [131, 132] publishes a set of examples in 1984 from which we are going to choose the oscillatory example in two different modifications and the discontinuous function to show the limits and

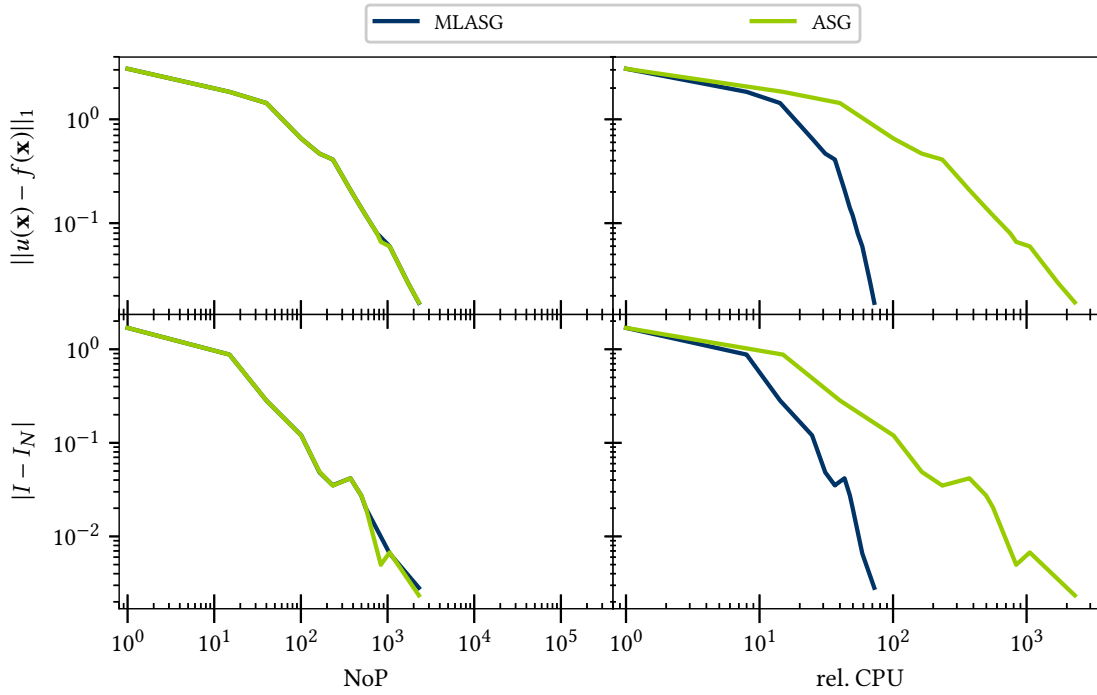


Figure 5.7: Results for the $7D - 2D$ case of the illustrative function (Eq. (5.24)) for the MLASG and the ASG method over the NoP (left column) and the relative CPU (right column). The intrinsic dimensionality is $D = 2$ and 5 dummy dimensions are added, that do not have an impact on the function outcome.

Upper row : Results for the L_1 -norm of the interpolation error $\|u(\mathbf{x}) - f(\mathbf{x})\|_1$ during the refinement.
 Lower row: Results for the quadrature error $|I - I_N|$ for the two methods during the refinement.

which key characteristics can be in favour for the [MLASG](#) approach. More examples of the Genz publication can be found in the section 7.3 in the Appendix. Particularly, we consider the functions

$$f_{\text{osc}}(\mathbf{x}) = \cos\left(\sum_i^D a_i(x_i + 0.5) + \pi\right) \quad a_i = 4.0; D = 2 \quad (5.27)$$

$$f_{\text{osc dec}}(\mathbf{x}) = \cos\left(\sum_i^D \frac{a_i}{2^i}(x_i + 0.5) + \pi\right) \quad a_i = 8.0; D = 14 \quad (5.28)$$

$$f_{\text{dis}}(\mathbf{x}) = \begin{cases} \exp\sum_i^2(a_i(x_i + 0.5)) & \text{if } (x_i \geq 0.0) \forall i \in [1, 2] \\ 0 & \text{else} \end{cases} \quad a_i = 4.0; D = 12 \quad (5.29)$$

We already introduced the oscillatory function as an example in the section 3.4. The results established that an adaptive [SG](#) approach shows a limited gain for this function due to local features that are spread over the whole domain. Since the standard oscillatory function (Eq. (5.27)) is equally dependent on all dimensions, and the adaptivity effect is limited, we assume that the performance. For the second example, we adapt this example for a more realistic approach for a kinetic model, and we would not deal with an equally strong dependence on all dimensions. Therefore we decreased the importance of increasing dimension so that the method can benefit from adaptation. As a last example function we use the discontinuous function, which describes a difficult characteristics for a [SG](#) to approximate. This function can cause a lot of refinement of the [SG](#) at the sharp jump and is, therefore, unfeasible for the adaptive sparse grid approach. We find that if we increase the dimensionality $D > 3$, the [ASG](#) and the [MLASG](#) needs a high number of points to achieve the lowest accuracy. So we restrict the intrinsic dimensionality to $D = 2$. If you are only interested in the quadrature, then we suggest to use methods like [MLQMC](#) (Section 7.4) for a discontinuity in higher dimensions. In the previous section, we essentially assumed that the noise per [MC](#) sample is so high that we can consider the number of samples to be drawn at a certain grid point a quasi-continuous. In this section, we now address the more realistic scenario that a single [MC](#) sample $Y_{1,i}$ has a finite standard deviation ([STD](#)) of σ . So the [STD](#) for our estimates $\overline{Y_{1,i}}$ is never below the value of σ . That means with $c = 1$ for the refinement that both refinement strategies [MLASG](#) and [ASG](#) will only produce one sample grid point as long as the threshold ϵ is bigger than σ of a single point. Only for smaller values of ϵ , the number of samples will increase, and [MLASG](#) has to draw fewer samples than the [ASG](#). For that, we now test three different noises $\sigma = [1, 0.1, 0.01]$ and display the final results for decreasing tolerances and not the results along the refinement process. This means the following figures show the final points of the previous examples just for various tolerances.

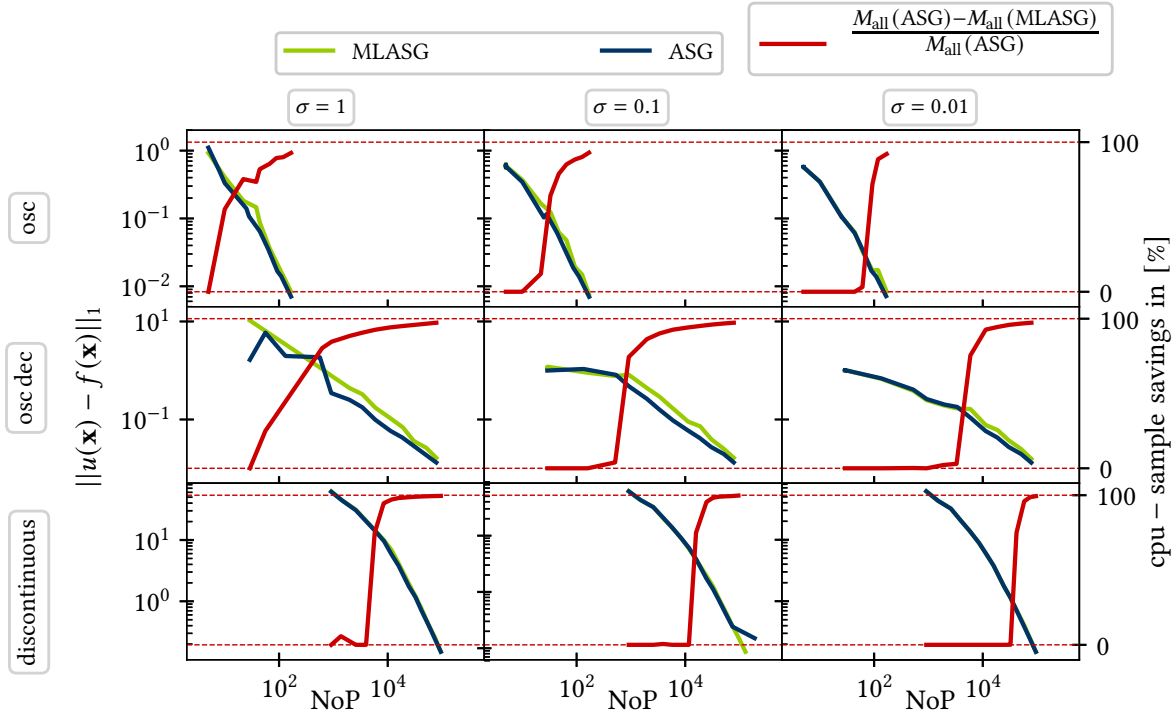


Figure 5.8: Results of the two refinement methods, ASG (—) and MLASG (—) for L_1 -norm of the interpolation error. Displayed are three different noises $\sigma = [1, 0.1, 0.01]$ and the CPU savings (—) for the MLASG compared to the ASG.

Upper row : Performance for the function f_{osc} in $D = 2$, which is a classical example for quadrature calculations.

Middle row : Performance for the function $f_{\text{osc dec}}$ in $D = 14$ with a decreasing hierarchy of dimensions.

Lower row : Performance for the function f_{dis} in $D = 12$ with a intrinsic dimensionality of two.

For all three examples we analyse the performance of the L_1 -norm of the interpolation error, displayed in figure 5.8, and the quadrature error, displayed in figure 5.9, but only over the NoP and not the relative CPU. Instead we additionally display the CPU savings over the NoP for the MLASG compared to the ASG. The CPU savings by the MLASG method are quantified by the difference in the total number of drawn samples for all sparse grid points in case of the ASG compared to the MLASG, normalized with the value for ASG. and divided by the samples for all ASG points

$$\frac{M_{\text{all}}(\text{ASG}) - M_{\text{all}}(\text{MLASG})}{M_{\text{all}}(\text{ASG})} \quad (5.30)$$

This equals a savings of 0 % if both sparse grids a constructed with the same amount of samples and reaches the 100% as the impact of the MLASG grows and the total numbers of samples of MLASG are vanishing small compared to the ASG samples.

The upper panel of figure 5.8 displays the results of classical low dimensional case f_{osc} . For the smallest noise $\sigma = 0.01$ (right column) both methods show an identical performance with

only a small CPU saving for the multilevel method. While for increasing noise $\sigma = [1, 0.1]$ the MLASG and ASG still perform very similar, the CPU-savings shift toward smaller NoP. So with increasing noise already at a coarse accuracy level, the MLASG is able to save 70-80% of the CPU- effort.

For the second function, $f_{\text{osc dec}}$, the two adaptive approaches show an almost identical behaviour, regarding the NoP, with only a small deviation for increasing accuracy. However, the CPU savings show that the MLASG can save up to almost 100% of the computational costs. The convergence is not as effective as in the classic oscillatory example, which is caused by the increasing dimensionality of the function. The local adaptivity of the SG approaches can intrinsically adapt important dimensions, but with a limited effect as all dimensions must be sampled to a certain accuracy level, and the adaptivity can not exclude dimensions in an early refinement state. Furthermore, no feature is localised but spread over the whole domain. The adaptive method is more effective if localised features need a higher density of points.

The f_{dis} example is favourable for an adaptive approach since 10 of the dimensions are dummy dimensions that can be excluded regarding the refinement. In the lower panel of figure 5.8, the discontinuity, defined in 2 dimensions, is well approximated by both refinement strategies, with only little deviations for $\sigma = 0.1$. Regarding the total performance for the L_1 -norm, both methods are able to decrease the interpolation error by two orders of magnitude. Due to the higher dimension of 12, the NoP is shifted to higher numbers in comparison to the examples before, and we assume with smaller thresholds, the convergence rate would slow down. If the threshold causes a very fine grid, the error caused by the overshooting of the SG for a discontinuity causes multiple additional refinements. For higher dimensions, this effect occurs earlier in the refinement procedure so that the cost becomes unfeasible. Regarding the computational savings, the multilevel approach can save almost 98% with 10^4 points in case of the highest noise. So even for an example that is less suited for the SG approach, the MLASG performs equally to the ASG with fewer costs.

Regarding the overall results of the two SG approaches, we see no significant difference in the results for all three examples. Although we tested both methods for different characteristics of a function, the MLASG performs nearly identical to the ASG method. We see only a minimal discrepancy between the MLASG and the ASG, due to higher noise at coarse refinement level, which produces more points. These result in a constant shift toward a higher NoP. So with a smaller noise of $\sigma = 0.01$, the shift is smaller. Taken the CPU-savings into account the MLASG is cheaper than the ASG approach, although it produces a higher amount of points for the same accuracy of the interpolation.

Next to the L_1 -norm, we analyse the performance of both methods regarding the quadrature error, displayed in figure 5.9. For the f_{osc} and the highest noise, we detect a deviation in

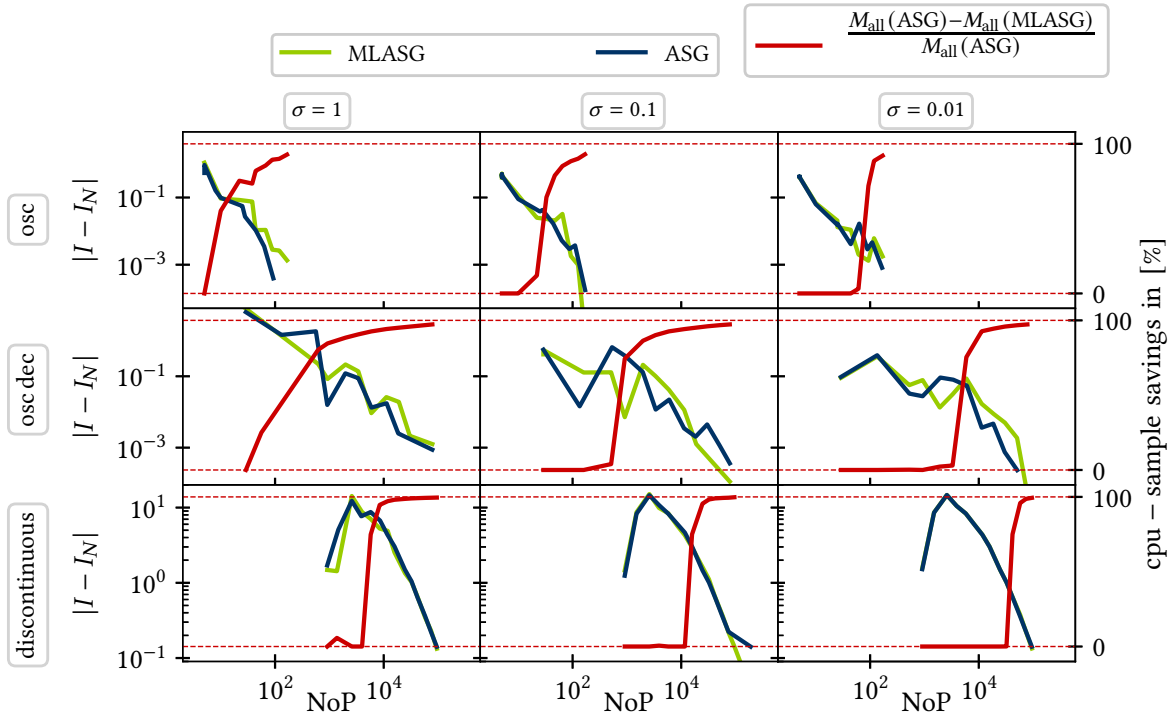


Figure 5.9: Results of the two refinement methods, ASG (—) and MLASG (—) for quadrature error. Displayed are three different noises $\sigma = [1, 0.1, 0.01]$ and the CPU savings (—) for the MLASG compared to the ASG.

Upper row : Performance for the function f_{osc} in $D = 2$, which is a classical example for quadrature calculations.

Middle row : Performance for the function $f_{\text{osc dec}}$ in $D = 14$ with a decreasing hierarchy of dimensions.

Lower row : Performance for the function f_{dis} in $D = 12$ with a intrinsic dimensionality of two.

the NoP. The MLASG needs more points to reach the same accuracy as the ASG. Taking the cpu-savings into account, the MLASG is still cheaper than the non multilevel approach, although it requires more points. For decreasing noise the MLASG and the ASG present similar results with only small fluctuations for high tolerances. Overall both methods are able to decrease the quadrature error over two orders of magnitude.

For the $f_{\text{osc dec}}$ function, we can see more fluctuations but an overall decreasing trend for the three noise cases. Although the performances are not identical for the interpolation error, the convergence trend is still similar, and both decrease the error by two orders of magnitude. For the CPU-savings, the MLASG outperforms the ASG and is able to save up to 97% of the computational costs.

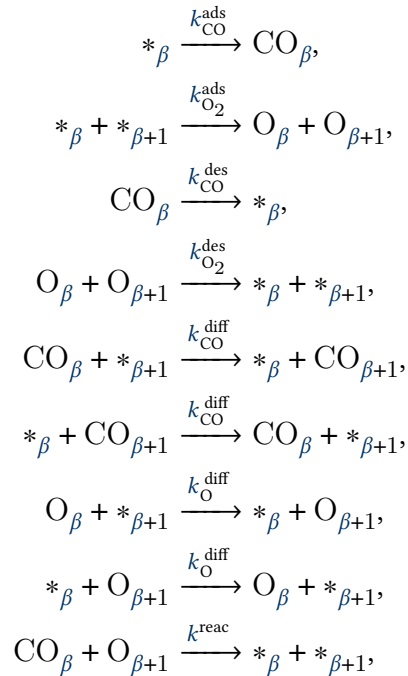
In the case of the f_{dis} , both methods draw a similar picture to the examples before, except for the increasing error at the beginning of the refinement. For the first two refinement steps, the error increases by a factor of almost one order of magnitude due to the coarse grid that does not adapt to the discontinuity yet. During further refinement, the ASG and MLASG results show a decreasing trend with smaller fluctuations for a smaller noise and greater fluctuations for increasing noise. All in all, both methods manage to increase the interpolation's accuracy by two orders of magnitude, with the MLASG saving up to 98% of the computational cost.

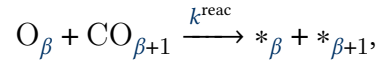
The comparison between the ASG and the MLASG outlines that their results are not identical for the quadrature error, whereas their error in terms of the norm is almost identical. This is due to the quadrature error being smaller than the L_1 -norm of the interpolation error. Even a coarse approximation results in a smaller quadrature error than the interpolation error so that small errors have a bigger effect of fluctuations on the quadrature error than the interpolation error. The smallest noise gives the closest aligned results for the two refinement strategies over the NoP. Still also for the bigger value, the performances do not differentiate strongly. Considering the CPU-savings, the MLASG performs at lower costs, and even with more points for the MLASG, the method is cheaper than the ASG. Furthermore, the MLASG can also achieve a reduction of the CPU savings at coarser refinement levels, so that even at a coarse level, we save computational effort with the MLASG

So for all three cases, it shows that although the model is not the most favourable for a SG approach, both approaches produce a good accuracy at a reasonable number of points. Furthermore, the CPU savings underline that the multilevel approach also works for a difficult example and can still decrease the computational cost.

5.7 Parametric Monte Carlo Model: CO oxidation model

As a realistic parametric model, on which the test model was based on, we use a reduced version of the well established CO oxidation on Rutheniumdioxide (RuO₂(110)). This is a heterogeneous catalyst model, whose original quantum chemistry model was introduced by Reuter and Scheffler [13]. This kMC model describes the chemical kinetics of the CO oxidation as Markov jump processes on a chain of the catalytic surface, consisting of two different adsorption sites, *cus* and *bridge*. Each of these jumps correspond then to one of the allowed elementary reactions, which include the adsorption/desorption of CO and O₂, the diffusion of O and CO to neighboring sites and the reaction of the educts to gaseous CO₂. In total, the model consists in total of 22 elementary steps, considering two adsorption sites. Later on, the model was reduced to seven dimensions[29, 30], since the reaction is mainly controlled by processes on one adsorption site, the *cus* site [24, 34, 35] and therefore all processes on *bridge* sites were excluded. Since we are working on a lattice model we can describe the state of the surface as a chain of singular sites. Each *cus* site on the catalytic surface, enumerated by its position β in the chain, can be in one of three states: I) empty ($*_{\beta}$), II) CO covered (CO_{β}), or III) oxygen covered (O_{β}). Since all of the processes are happening on the *cus* sites, we omit the site indexing for the rates. These states of the individual sites, that are documented in a vector, can be changed by elementary processes





where $k_{\text{CO}}^{\text{ads}}$ etc. are the rates, with which the respective process occurs. For further details on the model, we refer to references [29, 30]. Compared to the previous study [30], we choose a different parametrisation for the uncertainty analysis. Instead of using the adsorption and desorption rates directly, we use the equilibrium constant (EC) of the adsorption of CO and O₂ ($K_{\text{CO}}, K_{\text{O}_2}$) and the adsorption rates themselves of CO and oxygen on the *cus* site ($k_{\text{CO}}^{\text{ads}}, k_{\text{O}_2}^{\text{ads}}$). The corresponding desorption rates are implicitly given by the equilibrium constant and the adsorption rate by the reaction $k_{\text{CO/O}_2}^{\text{des}} = k_{\text{CO/O}_2}^{\text{ads}} \times K_{\text{CO/O}_2}^{-1}$. Furthermore, we include the diffusion rates ($k_{\text{CO}}^{\text{diff}}, k_{\text{O}_2}^{\text{diff}}$) and the actual reaction (k^{reac}).

Although it is a reduced model, it is still too high dimensional to be solved with classical numerical methods, so we use the lattice kinetic Monte Carlo code *KMOS* [159] to simulate the stationary expected values for the *QoI*.

The process rates, whose activation energies have been derived from Density Functional Theory (DFT), are used as an input parameter for the model. As previously stated, the rate constants carry a finite uncertainty due to the approximation of the density functional for the barrier calculation. We assume now that the input parameters' logarithm is independent and uniformly distributed in a specific range of values dominated by the errors in the activation energies that enter the input exponentially [13]. These ranges and the default values are displayed in table 5.2. Out of multiple outputs to employ, but we choose the logarithm of

Parameters	Default value [1/s]	Range [1/s]
K_{CO}	$\frac{2.0}{9.2} \times 10^2$	$\frac{2.0}{9.2} \times 10^0 - \frac{2.0}{9.2} \times 10^4$
K_{O_2}	$\frac{9.7}{2.8} \times 10^6$	$\frac{9.7}{2.8} \times 10^2 - \frac{9.7}{2.8} \times 10^6$
$k_{\text{CO}}^{\text{ads}}$	2.0×10^8	$1.0 \times 10^8 - 4.0 \times 10^8$
$k_{\text{O}_2}^{\text{ads}}$	9.7×10^7	$4.85 \times 10^7 - 1.94 \times 10^8$
$k_{\text{CO}}^{\text{diff}}$	5.0×10^{-1}	$5.0 \times 10^{-3} - 5.0 \times 10^1$
$k_{\text{O}_2}^{\text{diff}}$	6.6×10^{-2}	$6.6 \times 10^{-4} - 6.6 \times 10^0$
k^{reac}	1.7×10^5	$1.7 \times 10^3 - 1.7 \times 10^7$

Table 5.2: List of input parameter for the reduced model for CO₂ oxidation on at RuO₂ with its default values at $T = 600$ K, $p_{\text{CO}} = p_{\text{O}_2} = 1$ bar. Included are the range of variation for each parameter, spanning our domain.

the expected value of the Turnover Frequency (TOF) $y = \log_{10}(\text{E}(\text{TOF}))$ as the *QoI*. Since

we want to perform a **GSA** with the results, we use the logarithm of the expected **TOF**. Using the logarithm allows the **GSA** to be comparable to other established Local Sensitivity Analysis (**LSA**) and makes the interpolation easier. Without logarithmic settings, we cannot expect a linear dependency of the **TOF** on the kinetic parameter for most of the interpolation domain, which leaves us with only narrow regions with non-linear behaviour. For this setup linear basis functions are suitable to solve the problem. In similar research, we have used the CO coverage as a **QoI** to extract information about the dependencies of dominant regimes of the surface coverage on the input parameter [42].

The explored parameterisation and the parameter range has to advantage to be comparable to other **SA** indices, e.g. the Degree of Rate Control (**DRC**), that requires the same parameterisation for a comparison. Additionally, we can exploit the rather small uncertainty for the adsorption rates constants, for which extensive studies with different density functionals indicate that the respective barrier is very close to zero [160]. With the coordinate transformation to the equilibrium constant, it is easy to bring the required integral with the parameter distribution and the parameter-dependent **TOF** into a form (5.5). In the explored parameter range, the integral of the $\log_{10}(\mathbf{E}(\text{TOF}))$ is approximately -3.591, calculated by 10^5 Quasi Monte Carlo (**QMC**) points, with each point averaged over 100 trajectories with 10^8 **kMC** steps., that are used as a reference value for the accuracy of the **SG** integrals.

In a very strictly manner, the **QoI** of the logarithm of the **TOF** is not suitable for the multilevel approach we are presenting. Since we are taking the logarithm of the **TOF** determined by M samples, we are producing a bias. Fortunately the method still works since the bias appears only in higher order. If we approximate the bias with a Taylor expansion it arrives at

$$\mathbf{E}(\log(\text{TOF})) - \log \mathbf{E}(\text{TOF}) = \mathbf{E}(\log(d\text{TOF} + \mathbf{E}(\text{TOF}))) - \log \mathbf{E}(\text{TOF}) \quad (5.31)$$

$$= \mathbf{E}(\log(\mathbf{E}(\text{TOF})) + \frac{d\text{TOF}}{\mathbf{E}(\text{TOF})} + \mathbf{E}(\mathcal{O}(\frac{d\text{TOF}^2}{\mathbf{E}(\text{TOF})^2}))) - \log \mathbf{E}(\text{TOF})$$

$$\text{with } d\text{TOF} = \text{TOF} - \mathbf{E}(\text{TOF}) \quad (5.32)$$

and since $\mathbf{E}(\log(\mathbf{E}(\text{TOF}))) = \log \mathbf{E}(\text{TOF})$ and the expected difference of the expected **TOF** and the real **TOF** equals zero ($\mathbf{E}(d\text{TOF}) = 0$) we can determine

$$\mathbf{E}(\log(\text{TOF})) - \log \mathbf{E}(\text{TOF}) = \mathbf{E} \left(\mathcal{O} \left(\frac{d\text{TOF}^2}{\mathbf{E}(\text{TOF})^2} \right) \right). \quad (5.33)$$

This shows a relative fluctuation of the bias in a second-order, which is negligible compared to the random error of the first order.

We perform the simulations on a surface of 20 sites and periodic boundary conditions. Since

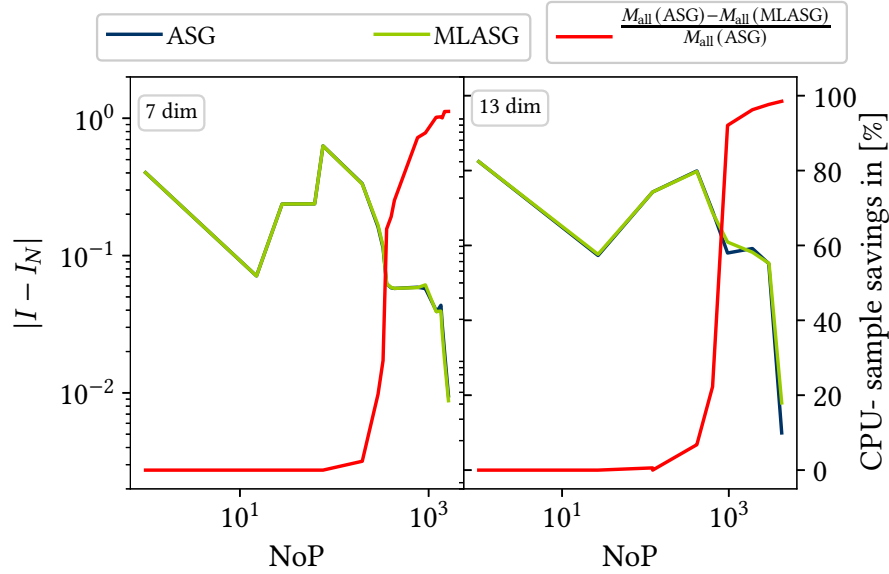


Figure 5.10: Comparison of the quadrature errors for the $\log_{10}(\mathbb{E}(\text{TOF}))$ of the **ASG** and **MLASG** for different tolerances ϵ and the **MLQ** method in the CO oxidation model. *Left figure:* The original 7 dimensional case. *Right figure:* Extension of the original model with 6 dummy parameters to a 13 dimensional model.

we are interested in estimating the stationary expected value, we perform the first 10^7 steps for relaxation to reach the steady-state and 10^7 points for estimating the expected value by time averaging over these. To achieve the target variance, we run these 2×10^7 long trajectories multiple times with different, hardware-generated random numbers as seeds for the kMC simulations. For the CO oxidation model we compare the final errors of the **MLASG** and the **ASG** for different target tolerances $\epsilon = [1, 5 \cdot 10^{-4}]$. Therefore we chose the initial standard variation to be equal to the applied tolerance $\sigma_0 = \epsilon$ for the **MLASG** and the **ASG** method, i.e. we employ $c = 1$ in Eq. (5.19). With increasing level the noise for the **MLASG** increase by $\sigma_l = \sigma_0 \cdot \sqrt{2^l}$ while the σ for the **ASG** stays the same. In practice we will use the empirical variance, since we do not know the upper bound of the variance, so that we draw M samples from the kMC points to fulfill

$$\text{Var}(\overline{Y_{1i}}) \leq \epsilon^2 \cdot 2^l \text{ for MLASG} \quad \text{Var}(\overline{Y_{1i}}) \leq \epsilon^2 \text{ for ASG} \quad (5.34)$$

with $\text{Var}(\log(\text{TOF})) = \frac{1}{\mathbb{E}(\text{TOF})^2} + \mathcal{O}\left(\frac{d\text{TOF}}{\mathbb{E}(\text{TOF})^3}\right)$.

For estimating the initial variance, we draw a minimal sample size of 20 for each point. The left column of figure 5.10 shows the quadrature error displayed over the **NoP** for both refinement methods. We display the **CPU-savings** of the multilevel approach compared to the standard **ASG**, calculated as the difference of all the drawn samples of the **ASG** method and the **MLASG**

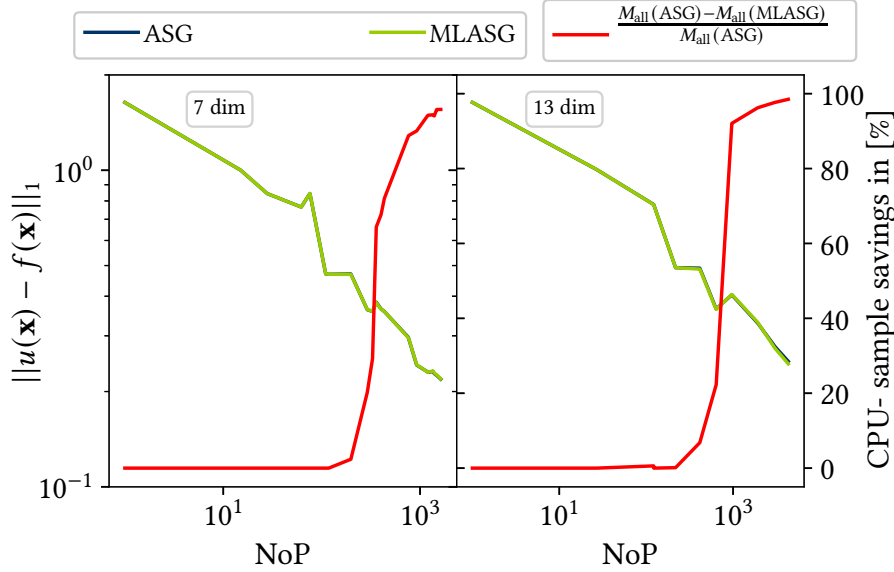


Figure 5.11: Comparison of the L_1 -norm for the $\log_{10}(\mathbb{E}(\text{TOF}))$ of the ASG and MLASG for different tolerances ϵ and the MLQ method in the CO oxidation model. *Left figure:* The original 7 dimensional case. *Right figure:* Extension of the original model with 6 dummy parameters to a 13 dimensional model.

with regard to all ASG samples (Eq. (5.30)) This gives a relative measurement of the CPU - savings caused by the multilevel extension.

With decreasing tolerances, the quadrature error shows some fluctuating behaviour till the refinement reaches smaller thresholds where both methods can decrease the error by a factor of 10^2 . The fluctuations appear for SG with the size up to 100 points. The SGs use for the finest grid around 1000 points to decrease the quadrature error to $|I - I_N| = 0.0088$. As we are mostly interested in the accuracy of the surrogate model to extract further information, e.g. for the sensitivity analysis, we also analyse the error of the L_1 -norm for both methods. The results are displayed in figure 5.11. The convergence of the interpolation error for the 7-dimensional case in the left figure is not as rapid as the one for the quadrature error for both adaptive methods. The refinement strategies are able to decrease the L_1 -norm by one order of magnitude with decreasing threshold and a maximum size of around 1000 points. The minimal interpolation error for the tolerance range of $\epsilon = [1, 5 \cdot 10^{-4}]$ is 0.22. The slow convergence implies that the underlying model is not an optimal model for the SG method. The results underline previous findings of a very complex structure of the RuO₂ model. The previous studies indicate that we are mostly dealing with a second order phase transition for the default setting of the model, which includes inconsistent derivatives [30, 35]. This is also underlined by the fluctuations seen in the quadrature errors. The general behaviour of the SG of an increasing quadrature error for the first refinement steps indicates that the

kinetic model structure is rather complex, and the initial SG is a good guess. Still, not all kinetic features are mapped with the initial grid. With further refinement, the quadrature error shows a plateau caused by regions of sharp transitions requiring a higher density of points to be approximated correctly.

For all tolerances, both SG refinement strategies show the same behaviour with respect to the NoP. Although the MLASG increases the variance with every refinement step of the sparse grid, the strategy avoids randomly refining noise but shows the same refinement pattern as the ASG with more accurate points. As we expected, the MLASG outperforms the standard ASG in regard to the CPU-savings. For the lower accuracy and NoP below 100, we cannot see the multilevel approach's impact since the points' variances are below the target threshold ϵ . Lowering the threshold causes the ASG to strongly increase the number of samples M compared to the multilevel approach. Therefore, the CPU-savings increase already at coarser refinement steps. At NoP of around 1000, we can save $\sim 90\%$ of the CPU-time without lowering the accuracy of the quadrature integral. For further refinement, the CPU-savings slope decreases so that for the minimal error of the quadrature, we can save roughly 95% or a factor of 20 less than for the ASG.

As the toy model results imply, the MLASG method, as well as ASG, can exploit low intrinsic dimensionality. To demonstrate this, we added three dummy species Z^i , that can be adsorbed/desorbed from the surface but do not take place in any reaction paths. This extends the system with three more states on the *cus* site Z_{cus}^i $i \in 1, 2, 3$, resulting in a total of six states ($\text{CO}_{cus}, \text{O}_{cus}, \text{e}_{cus}$). To the kMC model, we add the adsorb and desorb process



with $k_{Z^i}^{\text{des}} = k_{Z^i}^{\text{ads}} K_{Z^i}^{-1}$ and $k_{Z^i}^{\text{ads}} \in [1, 10^5]$ and $K_{Z^i} \in [1, 10^4]$, thereby increasing the dimensionality to 13. This could reflect a situation where trace gases are present in the gas phase above the catalyst, which can be adsorbed but do not take part in the reaction. The other parameters stay the same, and the right column of figure 5.10 shows the results of the 13-dimensional model. As before, we plot for both methods the quadrature error over the NoP in the right figure of figure 5.10 and the interpolation error in the right figure of figure 5.11. For both QoIs, we see a similar behaviour as for the 7 dimensional case. Regarding the quadrature error, the results show less strong fluctuations and a non-decreasing error for coarser sparse grids. Towards smaller threshold, both SG methods are able to decrease the quadrature error to $|I - I_N| = 0.0023$. The L_1 -norm of the interpolation error in the 13-dimensional case is very similar to the 7 dimensional case, except that the final error is minimally higher. The right figure shows the same slow convergence rate for the model.

As we expected, the sparse grid performances are not much influenced by the higher dimensionality. Compared to the 7-dimensional case, the adaptive methods produce more points due to the local adaptation, which also produces points in irrelevant dimensions. There is a small difference in the CPU-savings behaviour. Also, here we analyse tolerances $\epsilon = [1, 5 \cdot 10^{-4}]$ but we can save $\sim 98,5\%$ of the CPU-costs with the multilevel approach. This is more than in the 7-dimensional case, which implies that the effect of the multilevel approach might shift to higher tolerances.

These results show the positive impact of the multilevel idea on the ASG approach and underline the computational savings without a lower accurate surrogate model. To examine how much information we can extract from the MLASG surrogate model, we want to perform a GSA and analyse the important input parameters of the CO oxidation model.

5.8 Global Sensitivity Analysis

After having tested the L_1 -norm convergence of ASG and MLASG on a realistic kMC model with realistic error bounds, we now turn to an application in terms of global sensitivity analysis (GSA), which has been introduced in subsection 2.2.2. We are only going to perform it on the 7-dimensional case, since the dummy dimension in the 13-dimensional case, should not have an influence on the GSA results. The final results for the Total Sensitivity Index (TSI) are displayed in the figure 5.12 for three different tolerances ϵ . We choose three accuracy levels $\epsilon = [0.5, 0.008, 5 \cdot 10^{-4}]$ of the model for different convergence states and compare the performance of the ASG and the MLASG. The upper panel of figure 5.12 displays the coarsest approximation with no impact of the multilevel approach, resulting in a CPU-saving of 0%. For the applied threshold of $\epsilon = 0.5$ the sparse grid does not refine the initial grid that much further so that the threshold is bigger than the sample variance of the grid points. Based on this, we expect no difference between the ASG and the MLASG. The Sensitivity Indices for the coarse grid show a strong dependency on the ECs K_{CO} , K_{O_2} and the reaction rate k_{reac} . In relative comparison, the system depends on the K_{O_2} the strongest, followed by the K_{CO} , and k_{reac} . If we then increase the accuracy to $\epsilon = 0.008$ (*middel panel*), the strongest dependency shifts to the EC of CO, followed by the EC of O_2 . Additionally to the three dependencies of the coarsest results, the system also show a small dependency on both adsorption rates. Regarding the difference between the ASG and MLASG results, we detect a only minor differences in the TSI, but all below 10^{-3} . Both methods identify the same dependencies, qualitatively and quantitatively, but the MLASG is able to save almost 30% of the computational cost compared to ASG. This effect is even stronger for the highest accuracy $\epsilon = 5 \cdot 10^{-4}$. Both methods show

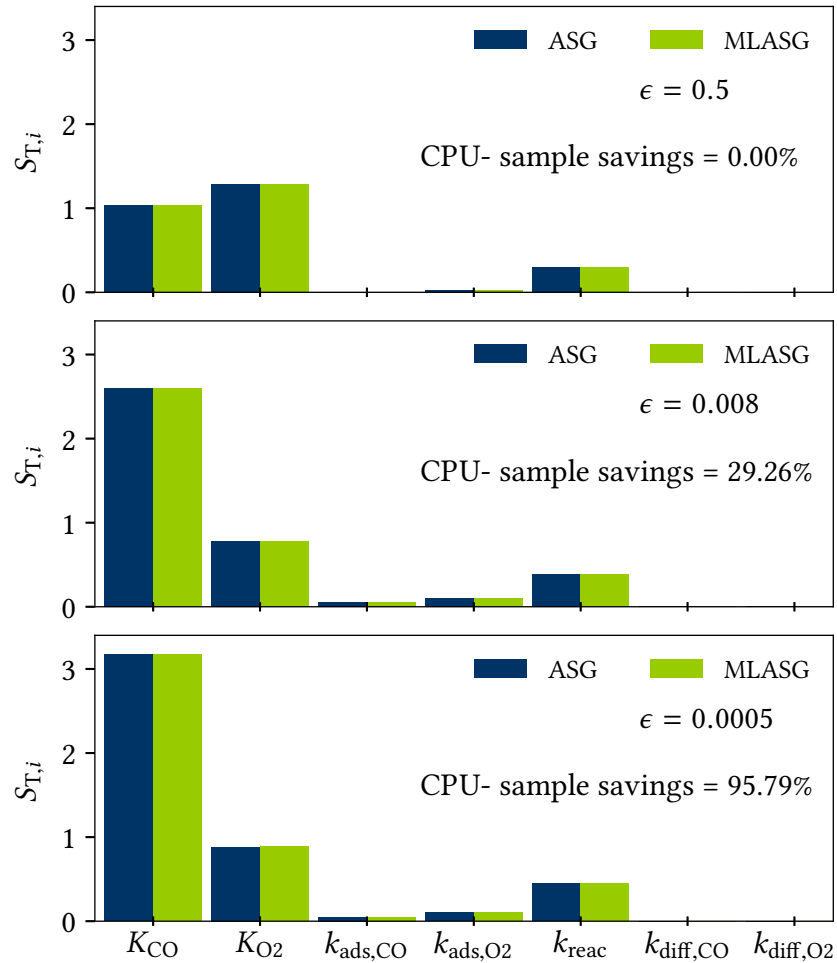


Figure 5.12: GSA of the 7- dimensional CO oxidation model for three different accuracies $\epsilon = [0.5, 0.008, 5 \cdot 10^{-4}]$. We analysed the $\log_{10}(E(\text{TOF}))$ and compared the GSA-results of the ASG and the MLASG surrogate model.

identical results for the TSI except that for this high accuracy the MLASG saves almost 96 % of the computational costs. Although the TSI are not totally converged at $\epsilon = 5 \cdot 10^{-4}$ there are only minor differences compared to $\epsilon = 0.008$, which shows a good approximation of the TSIs for our case. The impact of the K_{CO} increases even more, similar to the K_{O_2} , but the relative order stays the same.

Previous work on that model has shown similar results for the sensitivity on the parameters [24, 34] of this model, except for the missing strong dependency on the adsorption rates. We do not find this strong dependency due to the coordinate transformation and the smaller domain range we chose. We can decrease the domain range for the adsorption rates since the DFT calculations have shown no activation barrier for these processes. Since they are then only correlated to the sticking coefficient, these are also the only source of uncertainty [160]. Therefore, we can expect a smaller uncertainty for the rate constant [28].

Regarding the sensitivity results, we have to take into account what previous studies already revealed about the model [34]. A detailed analysis has shown that we have three dominant regimes: I) CO covered II) O covered and III) a mixed regime. For the CO covered regime, we expect an equilibrium between the $k_{\text{ads,CO}}$ and $k_{\text{des,CO}}$, while the adsorption of the oxygen is rate-limiting. For a CO covered surface, the biggest challenge is to get O_2 on the surface to enable a reaction. This explains the high dependency on the EC of CO and the low dependency on the $k_{\text{ads,CO}}$. Furthermore, the importance of $k_{\text{ads,CO}}$ is negligibly small in all regimes in contrast to $k_{\text{ads,O}_2}$, which shows high importance for the first regime. Since this regime is only a part of the whole domain, the importance is comparably small but still bigger than the $k_{\text{ads,CO}}$.

The results with the same settings from the LSA in paper [34] imply that there is no dependency on the adsorption rate in the second regime since they are both in equilibrium. This underlines the dependency on the EC, but not the adsorption rates. For the mixed regime, we expect a high dependency on both adsorption rates to enable a reaction. This regime displays a very complex reaction structure, where we can not imply equilibrium between the adsorption and desorption rates. Similar results are also presented with the DRC of paper [34]. The importance of the adsorption constant is still so insignificant due to the smaller size of the regime. Furthermore, the dependency on the adsorption rates is rather limited if the uncertainty of the domain is decreased.

All in all, these results show that the CO oxidation model on RuO_2 depends strongly on the ratio between the adsorption and desorption rates of both educts, symbolised by the EC. This implies that the occupation of the surface is a rate-limiting factor. As one species dominantly covers the surface, and to enable the reaction, we need to bring the minority educt on the surface. The subsequent reaction is the only reaction channel the model shows the third

strongest dependency on the reaction constant.

5.9 Conclusion

In this work, we introduced a multilevel sparse grid approach for parametric *MC* models to address the problem of high computational effort for estimating parametric numerical quadrature and approximation. To this end, we exploit the sparse grid model's intrinsic multilevel structure. It allows us to control the sampling variance of a single grid point such that the noise does not influence the adaptivity of the sparse grid in the *MC* simulation results. The upper bound of the variance assures a low computational effort without compromising neither the refinement strategy nor the accuracy of the *SG* approximation. By analysing the *SG* structure and testing the new method on an artificially constructed and a realistic model, we found that the sampling variance can be doubled in every refinement step, resulting in a computational effort with a factor of two orders of magnitude lower compared to non-multilevel *ASG*. Naturally, these results depend on the kind of problem whose integral should be estimated.

If we would just consider the quadrature and not the surrogate model, the classical telescopic sum approach, the *MLQ* would outperform the adaptive sparse grid approach, as described in the appendix 7.4. Regarding our purpose to perform a *SA*, we can not assume that the *MLQ* method outperforms the *MLASG*. Therefore, the *MLQ* needs to control the *MC* sampling error. If we then consider the functionals we need to approximate for the *GSA*, the *MLQ* would need to control the noise, and the bias and the easy approach with one sample per sparse grid point would not work anymore.

Beyond the quadrature results, we also included the interpolation error for the model to underline that the surrogate results can be used for further analysis, e.g. gradient-based *GSA*. For this *QoI* the multilevel approach performed as accurate as of the more expensive *ASG* method, and the impact on the *TSI* results was negligible.

All in all, we presented the potential of the *MLASG* method for producing an accurate surrogate model without reducing the efficiency of the refinement method. This can be very useful for multiscale models that have a coupling between low-fidelity and high-fidelity data. Our *MLASG* model is able to construct a surrogate, whose construction and, therefore, the evaluation of the function is cheaper than the evaluation of the underlying model.

6 Conclusion and Outlook

This work dealt with the error propagation in first-principles models in the field of heterogeneous (electro-) catalysis. Such models often have multiple error-prone input parameters that span a high-dimensional parameter domain that needs to be analysed. Furthermore, the parameters are often drawn from electronic structure theory, whose intrinsic approximation cause a very large error in the parameters. These errors, propagated into the first-principle models, can cause uncertainty of 2 or more orders of magnitude for the catalytic activity. Therefore it is not possible to predict whether the catalyst is active or not. We could have either tried to decrease the uncertainty in the input parameters or tried to extract the most information out of the given data to tackle this problem.

Based on the selectiveness of catalysed reaction, we expected only a few of the parameters to influence the reaction and, therefore, the activity of the model. On that basis, we quantified the importance of single input parameters with a Global Sensitivity Analysis (GSA). We preferred the GSA over a Local Sensitivity Analysis (LSA) due to the variance-based approach that takes the significant parameter errors and the non-linearities into account. This Sensitivity Analysis (SA) gave us better qualitative insights that, in fact, only a fraction of input parameters influence the model outcome. We concentrated purely on the input parameters and not on the system parameters (temperature, pressure...). However, we could see the different importance of the input parameter depended on the applied overpotential for the Oxygen evolution on Cobalt. Depending on the applied overpotential to the electro-catalytic model, the cobalt-surface can achieve different oxidation states, determining the reaction path. Such a model is complex due to the multiple possible reaction paths with possible dependencies between the parameters. The GSA gave us not only the overall most important parameters but also revealed important dependencies between parameters. Thus, we can determine the relevant oxidation state of the surface and dominant reaction paths. Our analysis also included a LSA to compare the results with the GSA and underline why the GSA is a better option for such models. Due to the mentioned characteristics of non-linear behaviour and large parameter domains, the gradient-based LSA underestimated the number of important parameters and simultaneously overestimated the impact of a single parameter.

To perform the GSA we needed a surrogate model for the non-linear functionals of the GSA. As mentioned before, we were working on high-dimensional parameter domains, where we expected only a fraction of the dimensions to be important. Therefore, we concentrated on Sparse Grid (SG) approach, that did not suffer from the *curse of dimensionality*. A general full SG approach would have meant to produce a lot of points in dimensions that would

be unnecessary because the outcome did not show a dependence on them. So we used a dimension-adaptive method to only refine the relevant dimensions and minimise the points for unimportant dimensions. Furthermore, we knew that the kinetic data mostly show a linear behaviour on a logarithmic scale and behaves non-linearly only in minor areas of the parameter domain. Based on this, we utilised a piecewise linear basis function with a local adaptivity that produces a higher density of points in non-linear areas. Similar methods were already content of earlier studies, but we adapted the methods to perform refinement steps according to a refinement level instead of singular points. While other methods concentrated on the *SG* points with the highest error and choose the refinement of these points as a refinement step, we considered all points above a predefined accuracy in one refinement step. All these things considered, the method allows us to deal with very high-dimensional models and construct an accurate surrogate model with a modest amount of *SG* points.

In the case of models, where the integrand is derived from a simulation algorithm, like Monte Carlo (*MC*) models, the presented *SG* would minimise the number of points but still be unfeasible as a single evaluation of the model is already very expensive. These simulation results carry a finite sampling error, which is inversely proportional to the computational cost. So to increase the computational savings, we needed to control the accuracy of a single point. Since the *SG* approach does not provide the information on how accurate one point has to be calculated, we needed a method that combines adaptive refinement and a sampling size adaptation. Therefore we employed the hierarchical structure of the *SG*, with a decreasing impact of the points along with the refinement steps and found out that with every refinement, we could increase the variance by two without compromising the refinement strategy. Together with the choice of basis function, this allows us to predefine the variance of points according to the preset threshold and avoid an on-the-fly optimisation of the *SG*, which makes the use uncomplicated. All in all, the threshold/accuracy ϵ is the single point of control for the sparse grid construction as it controls the refinement strategy as well as the accuracy of the *SG* points.

During our studies, we also identified some limitations of the *SG* approach. Next to the common difficulties, like choosing the proper basis function for effective construction of the surrogate model, we dealt with early termination problems and quasi-discontinuous function. If the initial *SG* is too coarse and is not able to approximate important local characteristics the refinement strategy terminates too early in the refinement procedure. This can be avoided with some coarse inexpensive precalculation that gives a rough qualitative approximation as a base for the *SG* or with a very accurate initial *SG* construction. In this case, we mean an initial *SG* with a limited maximum refinement level but with a very low threshold. Therefore we avoid a very dense sparse grid but include the smallest variation in the evaluation of

points. Based on this initial grid, we can then run the **SG** method for higher refinement levels but with a much reduced risk to overlook important areas of the domain. Combining the preprocessing and the adaptive sparse grid would make it even suitable for high-dimensional complex kinetic models, such as kinetic Monte Carlo (**kMC**) models [53].

Methods like a coordinate transformation prior to the sparse grid approach can effectively tackle the early determination. Estimating the variance with random data points of poor accuracy can identify the areas of high variances and areas that can be excluded *a priori*. Transforming the coordinates of the **SG** toward the high variance dimensions will increase the performance of the **SG**. Similar methods have already been established for least-squares algorithms for high-dimensional problems [161]. Especially for complex kinetic models as the CO-oxidation model in 5.1, this can be very beneficial to further understanding on the microscopic level.

Furthermore, the results of the CO oxidation model for the Multilevel Adaptive Sparse Grid (**MLASG**) indicated the problems with discontinuous functions. Some of these functions can be suitable for the sparse grid with a coordinate transformation. However, it often includes a major amount of precalculation or cannot be effectively approximated in a practical mesh-spacing.

This work shows the potential of the presented methods for various applications in the future. Regarding the **GSA**, we can see the advantage of a surrogate model over other methods, particularly for **kMC**-models. An accurately constructed surrogate model is easier and cheaper to evaluate, especially when it comes to gradient-based **GSA**, where for other methods, the gradient calculation causes high computational cost because of multiple evaluations of the underlying model. The results of the **GSA** for both chemical models have shown how robust the results are against the quality of input data or the choice of model. On that basis, we can expect that we could also apply a hierarchical approach to the **SA**. First, starting with very poor and cheap input data and let the **GSA** define the most important one to be calculated with a more accurate method to limit the computational cost. This can be extended to a one-by-one approach that only improves one parameter at a time. Therefore we could individually decide which method is the best to improve the data and lower the computational cost even further. Approaches like these will include a complex accuracy vs cost analysis based on different quantum-mechanical approaches. This would not only mean a higher accurate first-principles model outcome but also an impact for the underlying electronic structure theories that can concentrate on the important parameter and reduce the computational costs.

In reality, the worst-case scenario of uncorrelated input parameters is not true for chemical models. The Density Functional Theory (**DFT**)-energies and therefore the **DFT**-errors are correlated. Functionals like the Bayesian error estimation functional takes exchange-correlation

for dispersion interactions into account. The error correlation will be taken into account by adapting the probability density function (PDF). However, the SA performance will not be influenced by that, except that the higher-order sensitivity indices will display the correlation. In the aspect of material or catalyst screening, an effective GSA with a surrogate model provides a method to identify the important parameters of a material class with data of very poor accuracy data for different materials. This means we can easily scan an extensive parameter range of materials inexpensively. Regarding the construction of the surrogate model, the next step would be developing a first local and then dimensional adaptivity. Since the kinetic data often describes different regimes from non-active to active, the SA can effectively identify which parameter has the highest impact for these local regimes. This is beneficial to increasing the Turnover Frequency (TOF) in already active regimes or identifying potentially active regimes and characterising them.



We only used the SG approach for interpolation, but potential applications are manifold. They can also be used for a regression algorithm or Galerkin-projection for a parameter domain, where they have to be proven very effective for dynamic models with multiple parameters [113, 162]. In regards to the MLASG approach, we presented only a local-adaptive refinement, but we could also extend the refinement to dimension adaptivity or other refinement strategies. The critical requirement of the multilevel approach is to include all the parents of one point. Therefore we keep the hierarchical decreasing importance structure that allows us to increase the variance of a single evaluation. As we mentioned early, we could also construct the SG as a minimisation problem during the refinement. Instead of predefining the accuracy of the point, we minimise the computational cost on the fly with every refinement step. This would also allow for other basic functions. Furthermore, the future incorporation of integer programming approaches for an optimal distribution of the sampling effort will optimise the computational effort. This allows testing more refinement strategies and various basis functions. The possible application of the methods expands to various real-life problems. The next step would be to test the approach on different applications, e.g. from the field of molecular simulations, to achieve a thorough picture of the performances and possible improvements.






At last, we want to point out the good performance of the SG method in regards to kinetic data. Other studies have shown that Gaussian process regression is not ideal for extracting information of kinetic data [29]. The local rapid changes are often problematic, or the underlying low-dimensional model causes more computational cost than necessary. The refinement strategies of the SG show a high efficiency regarding the local features. Although we have seen some quantitative fluctuations for the GSA results for different accuracies of the SG, the qualitative results did not change for increasing accuracy. This shows how efficient the SG

can be even with a coarse refinement level and a small number of points. Other methods, like a random-forest tree mechanism or a spline approach, have been shown to work for [kMC](#) data but only for a small amount of dimension or smooth models. For higher-dimensional models it is left to be proven efficient to be an alternative for the [SG](#)[[163](#), [164](#)].

In the kinetic field, a method like our approach that can control the sampling accuracy and allows for only one threshold to define the construction of the sparse grid, is rarely used. This prevents higher computational costs for unnecessary production of testing points, as can be seen in other methods, e.g. for the construction of neural networks. To consider both aspects, the amount of points and the sampling accuracy is not very common yet. Most studies only focus on a minimal amount of points. They do not take the sampling accuracy into account, meaning most of the evaluations are more accurate than needed and include a waste of computational effort. At last, we want to point out the fast convergence of the [MLASG](#) method, although we double the variance with every refinement step. All the results show almost the same behaviour as the non-multilevel adaptive Sparse Grids and show a good conversion rate, even for a very complex model like the CO oxidation on RuO_2 .

List of Figures

-
- 2.1 Description of the four considered kinds of elementary reactions : I) Ad-/Desorption of reactants (blue and red balls) on the surface II) Diffusion of a reactant over the surface and III) Reaction of two reactants, including a desorption of the surface. 17
- 2.2 Energy diagram of a reaction from state \mathbf{i} to state \mathbf{j} , including the description of the Transition State (TS) as the saddlepoint of the Potential Energy Surface (PES). The system has to overcome the energy barrier ΔE_{ij} for the forward reaction and ΔE_{ji} for the backward reaction 20
- 2.3 Model output of the CO oxidation on RuO₂ model for $T = 298.15$ K. The conversion, measured as $\log_{10}(\text{TOF} \times s)$, is plotted over the CO adsorption rate constant $k_{\text{CO,ads}}$. For nominal values of all input parameters the conversion is displayed as the blue dots. The green shaded area describes the standard deviation of $\log_{10}(\text{TOF} \times s)$ by assuming uniformly distributed DFT-errors in range of $[-0.2, 0.2]$ eV. 21
- 3.1 This is the CO oxidation on Rutheniumdioxide(110) dependent on the adsorption rate of CO. The logarithm of the Turnover frequency (TOF) behavior plotted over the reaction coordinate, which spans the uncertainty of one system parameter. Along the reaction coordinate the TOF values highly variate and show areas of linear behaviour and also of non linear behavior . . . . 36
- 3.2 Sparse Grid Interpolation for locally discontinuous function $g(x)$  with splines of different orders for level $l = 2$. *Left panel* : Interpolation of $p = 2$ order splines, with significant overshoots towards the discontinuity. *Right panel* Interpolation of $p = 1$ order splines, with a closer adaptation of the characteristics of the underlying function. 36
- 3.3 Illustration of the hierarchical interpolation in one dimension with increasing level index. *Left Panel*: Basis functions added at the corresponding levels l . *Middle Panel*: The original function $f(x)$ and the hierarchical interpolation $u(x)$ *Right Panel*: Illustration of the contribution of the basis functions added at each level $v_{l,i}\phi_{l,i}$. The contributions decrease with increasing level, and, at the finest level, only those basis functions close to the sharp non-linearity still have significant contributions to the interpolation. 38
-

- 3.4 Hierarchical difference spaces for the 2D Sparse Grid Interpolation according to the refinement level. The top left and the top right row show the one dimensional basis function for each dimension and level. In the inner diamond grid, the 2D tensor products of the one dimensional functions are displayed. An interpolation constructed by the shown basis function has the maximum level of $L = 2$. For the SG, all hierarchical difference spaces above the orange dotted line  are included, thus it consists of spaces with an L_1 -norm $|l|_1 \leq 2$. In comparison to that, a full grid includes all spaces with a maximum norm $|l|_\infty \leq 2$, which includes also the spaces below the orange dotted line  42
- 3.5 Refinement procedure and interpolation of a target function with a jump characteristics with the basis functions for different refinement levels $l = [0, 1, 2, 3]$ 47
- 3.6 Scheme for the dimension adaptivity. *Upper panel:* All possible hierarchical difference spaces up to $|l|_\infty \leq 3$ in 2D and the corresponding grid points. Solid color symbols the refinement spaces, considered by the dimension adaptivity while the opaque color symbols the spaces that are neglected by the method. Thus all refinement spaces according to dimension 1 are included while the second dimension is only refined once. *Lower panel, left:* Target function of a sigmoid character that needs to be approximated by the Sparse Grid . *Lower panel, right:* Final Sparse Grid with dimension adaptivity. 50
- 3.7 Scheme for the local adaptivity. *Upper panel:* All possible hierarchical difference spaces up to $|l|_\infty \leq 3$ in 2D and the corresponding grid points. The solid darkblue points  and the red crosses  map out the points included by the local adaptivity. The opaque light blue  color symbols the points that are neglected by the method. The red crosses denote the points, which carry a high local error and need to be refined. Thus the method concentrates the points closest to non linear behavior. *Lower panel, left:* Target function of a sigmoid character, that needs to be approximated by the Sparse Grid . *Lower panel, right:* Final Sparse Grid with local adaptivity. 53

- 3.8 Combination technique of the dimension and local adaptivity. *Upper left panel:* Initial grid of $|\mathbf{l}|_1 = 1$ ● and all possible refinement points of $|\mathbf{l}|_1 = 2$ ✖. *Upper right panel:* Reduction of refinement points according to the dimension adaptivity; thus only points in the first dimension along the variation of the function are considered. *Lower left panel:* Reduction of refinement points according to the local adaptivity; thus only point in the area of the variation are considered. *Lower right panel:* The intersection of both refinement sets is added to the Sparse Grid construction. 56
- 3.9 Four test functions for the refinement strategies. *Upper panel:* Significant functions for kinetic data that show the characteristics of the TOF of kinetic data (spherical edge, *left figure*) and the coverage for one species (sigmoid, *right figure*). *Lower panel:* Classical examples for integration problems, introduced by Genz. *Left figure* shows the shifted peak of the continuous function and the *right figure* shows the oscillatory function. 58
- 3.10 Final Sparse Grid for the approximation of f_{SE} (spherical edge) with a threshold of $\epsilon = 0.01$. The lighter green dots indicate the kink of the function. *Left panel* Grid points for the dimension adaptivity (DA) refinement strategy which equals a SG without adaptivity. *Right panel* Grid points for the local adaptivity (LA) refinement strategy, which resolves in a higher resolution in the area of the non-linear kink. *Middle panel* Grid points for the combined refinement technique whose refinement is mostly dominated by the LA and results in a similar grid as LA. 59
- 3.11 Performance of the three refinement methods for the L_1 -norm of the interpolation error (*upper panel*) and the quadrature error (*lower panel*) in case of the spherical edge function. Displayed are three different dimensional cases. All of them have an intrinsic dimensionality of 2, and in the case of $D = 4$ two and the case of $D = 8$, six dummy dimensions are added. In all 6 cases, the combined technique $\mathcal{S}_{\text{dim,loc}}$ performs the best. The increasing dimensionality worsens the performance of the \mathcal{S}_{loc} method due to the increasing refinement points in all dimensions. Contrary to that, the DA method can not efficiently adapt the function's local features but gives a two-dimensional full sparse grid. 60

-
- 3.12 Final Sparse Grid for the approximation of the sigmoid function f_S with a threshold of $\epsilon = 0.01$. *Left panel* Grid points for the DA refinement strategy which equals a SG without adaptivity. *Right panel* Grid points for the LA refinement strategy, which resolves in a higher resolution in the area of the sharp changes and excludes the points in the constant areas. *Middle panel* Grid points for the combined refinement technique whose refinement is mostly dominated by the LA and results in a similar grid as LA. 62
- 3.13 Performance of the three refinement methods for the L_1 -norm of the interpolation error (*upper panel*) and the quadrature error (*lower panel*) in case of the sigmoid function. Displayed are three different dimensional cases in which all of them have an intrinsic dimensionality of 2. In the case of $D = 4$ (*middel panel*) 2 and in the case of $D = 8$ (*right panel*) 6 dummy dimension added. In all 6 cases, the combined technique $\mathcal{S}_{\text{dim,loc}}$ performs the best. The increasing dimensionality worsens the performance of the \mathcal{S}_{loc} method due to the increasing refinement points in all dimensions. In contrary to that, the DA method can not efficiently adapt the local features of the function, but the intrinsic dimensionality and therefore gives a two-dimensional full sparse grid for all the dimensional cases. This leads to DA and LA showing the same refinement behaviour along the accuracy for $D = 8$ for the expected value. 63
- 3.14 Final Sparse Grid for the approximation of f_{CF} (continuous function) with a threshold of $\epsilon = 0.01$. *Left panel* Grid points for the DA refinement strategy which equals a full SG for $D = 2$. *Right panel* Grid points for the LA refinement strategy which resolves in a higher resolution towards the non-linear peak at $[-0.2, -0.2]$. *Middle panel* Grid points for the combined refinement technique whose refinement is mostly dominated by the LA and results in a similar grid as LA, except for a smaller number of points. The combined technique uses 37% fewer points than the LA method. 65
-

- 3.15 Performance of the three refinement methods for the L_1 -norm of the interpolation error (*upper panel*) and the quadrature error (*lower panel*) in case of f_{CF} (continuous function). Displayed are two different dimensional cases in which all of them have an intrinsic dimensionality of 2. In the case of $D = 8$ (*right panel*) 6 added dummy dimension. In all 4 cases, the combined technique $\mathcal{S}_{dim,loc}$ performs the best. The increasing dimensionality worsens the performance of the \mathcal{S}_{loc} method due to the increasing refinement points in all dimensions. Contrary to that, the DA method can not efficiently adapt the local features, which causes a factor 5 of the number of points (NoP) to achieve the same accuracy as the combined technique for both measures. 66

- 3.16 Final Sparse Grid for the approximation of f_{OF} (oscillatory function) with a threshold of $\epsilon = 0.01$. *Left panel* Grid points for the DA refinement strategy which equals a full SG but results in less points than the LA grid. *Right panel* Grid points for the LA refinement strategy, which resolves in a slightly higher resolution in the curved area. *Middle panel* Grid points for the combined refinement technique, which shows a slightly higher density in the area of the dense curves like the LA but with fewer points in total. This is caused by the excluding feature of the dimension adaptivity feature of the combined technique. 68

- 3.17 Performance of the three refinement methods for the L_1 -norm of the interpolation error (*upper panel*) and the quadrature error (*lower panel*) in case of the oscillatory function. Displayed are two different dimensional cases in which both of them have an intrinsic dimensionality of 2 and in the case of $D = 8$ (*right panel*) 6 added dummy dimension. For the 2D case, all of the methods perform equally well. For the expected value, it appears that the methods have some fluctuations, but none of the methods can outperform the others. Only for the increasing dimensionality, the performance of the LA method worsens due to the increasing refinement points in all dimensions. 69

- 3.18 Results of the performance of the three refinement methods for the L_1 -norm of the interpolation error (*upper panel*) and the quadrature error (*lower panel*) in case of the $f_{2D\ sum}$ function in the *left coloumn* and the product function $f_{prod\ func}$ in the *right coloumn*. The $f_{2D\ sum}$ has a dimensionality of 6 and the product function a dimensionality of 4. Both functions have the characteristic of a parameter dependency that can be decomposed into smaller dimensional contributions. The combined method outperforms the other methods in all 4 cases. 71

- 4.1 The cyclic Oxygen evolution reaction on the 110A site of Co_3O_4 . The structure consists of two redox-active Cobalt atoms bridged by an Oxygen atom. Structure 1) has two bridging hydroxyl groups that will take part in the reaction, which are deprotonated in the first two steps (Structure 2/3) and results in a Cobalt atom with two bound oxos. We create a hydroxyl group for the oxo (Structure 4) through water addition, which is turned into a superoxo by another deprotonation (Structure 5). The superoxo ligand desorbs as an elementary Oxygen molecule (Structure 6), whose vacancies are filled by another water addition (Structure 1). 76
- 4.2 Adapted energy diagram for the oxygen evolution reaction (OER) reaction on Co_3O_4 . $\Delta E_{\text{int},i}$ indicates the free energy of the intermediate state A_i for the oxidation state i with respect to the reference state A_0 . The reaction can only happen from intermediate states A_i of even numbered states with an energy barrier $\Delta E_{\text{act},i}$ to arrive at the product state. 78
- 4.3 TOF as a function of the applied overpotential η for $T = 298.15$ K for the nominal energy values (blue line —). The green area \blacksquare depicts the standard deviation of the model output by assuming uniformly distributed DFT-errors in range of $[-0.3, 0.3]$ eV. Reproduced from [23], with the permission of AIP Publishing. 82
- 4.4 Two dimensional dependence of the TOF on the energy errors $\Delta\Delta E_{\text{int},4}$ and $\Delta\Delta E_{\text{act},4}$ at an overpotential of $\eta = 0.7$ V. (I) no dependence on $\Delta\Delta E_{\text{int},4}$ and $\Delta\Delta E_{\text{act},4}$, (II) only depending on $\Delta\Delta E_{\text{act},4}$, (III) depending on both, and (IV) only depending on $\Delta\Delta E_{\text{int},4}$. Reproduced from [23], with the permission of AIP Publishing. 84
- 4.5 Adaptive grid for three applied tolerances $\epsilon = 0.1$ (left) $\epsilon = 0.01$ (middle) and $\epsilon = 0.001$ (right). Dotted line - - - indicates borders between four different regimes with different dependencies on the parameters $\Delta\Delta E_{\text{int},4}$ and $\Delta\Delta E_{\text{act},4}$. Refinement primarily happens close to the origin, where the four regimes meet. Partially reproduced from [23], with the permission of AIP Publishing. 87
- 4.6 Local und global sensitivity indices $S_{T,i}$ for the 20 input parameters for the (110)-A structure of the Oxygen evolution reaction at three different overpotentials $\eta \in \{0.4, 0.7, 1.0\}$ V. The gray-shaded fields describe intermediate states that are not considered for a reaction in the model. Shown are the sensitivities of $\log_{10}(\text{TOF} \times s)$ with respect to the errors in $\Delta E_{\text{int},i}$ and $\Delta E_{\text{act},i}$. Reproduced from [23], with the permission of AIP Publishing. 89

4.7	Second order indices with respect to the input parameters of the OER model on the (110)- A surface for an overpotential $\eta \in \{0.4, 0.7, 1.0\}$ V. Darker areas correspond to as stronger interaction between the two parameters. Partially reproduced from [23], with the permission of AIP Publishing.	94
4.8	Local and global sensitivity indices $S_{T,i}$ for the extended model of the OER on the (110)-A surface for an overpotential $\eta = 0.7$ V. The gray-shaded fields describe intermediate states that are not considered for a reaction in the model. Shown are the sensitivities of $\log_{10}(\text{TOF} \times s)$ with respect to the errors in the input parameters. Reproduced from [23], with the permission of AIP Publishing.	99
4.9	Global sensitivities indices and local sensitivities indices of the TOF for $\eta = 0.7$ V with respect to the energies for an extended model where the reactions from the intermediate states 5 and 7 are considered. Reproduced from the supplementary material of Ref. [23], with the permission of AIP Publishing.	101
5.1	An illustrative function for the MLASG method which has the similar characteristics as the kinetic data. Kinetic data has rather sharp changes depending on more than one dimension. In the 2 dimensional figure the changes are displayed as a kink at a radial value of $r = \sqrt{\sum_i^D (x_i + 0.5)^2} = 0.6$	116
5.2	Comparison of the adaptive grids for $\epsilon = 10^{-4}$ for adaptive Sparse Grid (ASG) without noise (<i>left figure</i>) and MLASG (<i>right figure</i>) with an initial standard deviation $\sigma_0 = 10^{-4}$, i.e. $c = 1$. Both final sparse grids show the same results for adapting the kink of the function with only a minimal difference in the number of points.	117
5.3	Results for the 2D case of the illustrative function (Eq. (5.24)) for the MLASG and the ASG method over the NoP (<i>left coloumn</i>) and the relative Central Processing Unit (CPU)(<i>right coloumn</i>). <i>Upper row</i> : Results for the L_1 -norm of the interpolation error $\ u(\mathbf{x}) - f(\mathbf{x})\ _1$ during the refinement. <i>Lower row</i> : Results for the quadrature error $ I - I_N $ for the two methods during the refinement.	118
5.4	Comparison of the adaptive grids for $\epsilon = 10^{-4}$ for ASG (<i>left figure</i>) without noise and MLASG (<i>right figure</i>) with an initial standard deviation $\sigma_0 = 10^{-2}$, i.e. $c = 100$	119
5.5	L_1 -Norm of the interpolation error for different initial noise values $\sigma_0 = [10^{-2}, 10^{-3}, 10^{-4}]$ compared to the ASG method without noise.	119

-
- 5.6 Results for the 7D case of the illustrative function (Eq. (5.24)) for the MLASG and the ASG method over the NoP (*left column*) and the relative CPU(*right column*). *Upper row* : Results for the L_1 -norm of the interpolation error $\|u(\mathbf{x}) - f(\mathbf{x})\|_1$ during the refinement. *Lower row*: Results for the quadrature error $|I - I_N|$ for the two methods during the refinement. 121
- 5.7 Results for the 7D – 2D case of the illustrative function (Eq. (5.24)) for the MLASG and the ASG method over the NoP (*left column*) and the relative CPU(*right column*). The intrinsic dimensionality is $D = 2$ and 5 dummy dimensions are added, that do not have an impact on the function outcome. *Upper row* : Results for the L_1 -norm of the interpolation error $\|u(\mathbf{x}) - f(\mathbf{x})\|_1$ during the refinement. *Lower row*: Results for the quadrature error $|I - I_N|$ for the two methods during the refinement. 123
- 5.8 Results of the two refinement methods, ASG(—) and MLASG (—) for L_1 -norm of the interpolation error. Displayed are three different noises $\sigma = [1, 0.1, 0.01]$ and the CPU savings (—) for the MLASG compared to the ASG. *Upper row* : Performance for the function f_{osc} in $D = 2$, which is a classical example for quadrature calculations. *Middle row* : Performance for the function $f_{\text{osc dec}}$ in $D = 14$ with a decreasing hierarchy of dimensions. *Lower row* : Performance for the function f_{dis} in $D = 12$ with a intrinsic dimensionality of two. 125
- 5.9 Results of the two refinement methods, ASG(—) and MLASG (—) for quadrature error. Displayed are three different noises $\sigma = [1, 0.1, 0.01]$ and the CPU savings (—) for the MLASG compared to the ASG. *Upper row* : Performance for the function f_{osc} in $D = 2$, which is a classical example for quadrature calculations. *Middle row* : Performance for the function $f_{\text{osc dec}}$ in $D = 14$ with a decreasing hierarchy of dimensions. *Lower row* : Performance for the function f_{dis} in $D = 12$ with a intrinsic dimensionality of two. 127
- 5.10 Comparison of the quadrature errors for the $\log_{10}(E(\text{TOF}))$ of the ASG and MLASG for different tolerances ϵ and the Multilevel Quadrature (MLQ) method in the CO oxidation model. *Left figure*: The original 7 dimensional case. *Right figure*: Extension of the original model with 6 dummy parameters to a 13 dimensional model. 132
- 5.11 Comparison of the L_1 -norm for the $\log_{10}(E(\text{TOF}))$ of the ASG and MLASG for different tolerances ϵ and the MLQ method in the CO oxidation model. *Left figure*: The original 7 dimensional case. *Right figure*: Extension of the original model with 6 dummy parameters to a 13 dimensional model. 133
-

5.12	GSA of the 7- dimensional CO oxidation model for three different accuracies $\epsilon = [0.5, 0.008, 5 \cdot 10^{-4}]$. We analysed the $\log_{10}(E(\text{TOF}))$ and compared the GSA-results of the ASG and the MLASG surrogate model.	136
7.1	Covergence plot of the Total Sensitivity Index (TSI) for the three different overpotentials $\eta \in \{0.4, 0.7, 1.0\}$ V. We display the global sensitivity indices $S_{T,i}$ for the 20 input parameters for the (110)-A structure of the oxygen evolution reaction. The results are shown for three decreasing tolereances $\epsilon = [0.1, 0.01, 0.001]$ of the sparse grid construction. The gray-shaded fields describe intermediate states that are not considered for a reaction in the model. We present the sensitivities of $\log_{10}(\text{TOF} \times s)$ with respect to the errors in $\Delta E_{\text{int},i}$ and $\Delta E_{\text{act},i}$	176
7.2	Results of ASG(—) and MLASG (—) for L_1 -norm of the interpolation error (<i>upper panel</i>) and the quadrature error (<i>lower panel</i>) for the three functions $f_{\text{sig}}, f_{\text{peak}}, f_{\text{cos kink}}$. Displayed are three different noises $\sigma = [1, 0.1, 0.01]$ and the CPU savings for the MLASG compared to the ASG (—). <i>Upper row</i> : Performance for the function f_{osc} in $D = 2$, which is a classical example for quadrature calculations. <i>Middle row</i> : Performance for the function $f_{\text{osc dec}}$ in $D = 14$ with a decreasing hierarchy of dimensions. <i>Lower row</i> : Performance for the function $f_{\text{cos kink}}$ in $D = 8$ with a intrinsic dimensionality of two. . . .	178
7.3	The sum of the squared quadrature weights for $\text{dim}=[1, 2, 3, 4]$ shows the convergence for increasing dimensionality over a common denominator, the refinement level of the Full Sparse Grid (FSG).The actual NoP for grid with refinement level increase with the dimensionality.	181
7.4	The quadrature error $ I - I_N $ results for the cosinus kink in 4D- function for the MLQ (Quasi Monte Carlo (QMC)) and the MLASG methods for different tolerances ϵ . Displayed are the results of three different initial noise cases $\sigma_0 = [1, 0.1, 0.01]$ over the sum of drawn samples M_{all}	183

List of Tables

2.1	Factor k_{error} , caused by the energy error $\Delta\Delta E$ $k \propto e^{\left(\frac{\Delta E_{\text{process}} \pm \Delta\Delta E}{k_B T}\right)}$ resulting a parameter variation of $k = \left[\frac{k_0}{k_{\text{error}}}, k_0 \cdot k_{\text{error}}\right]$ with $k_{\text{error}} = e^{\left(\frac{\Delta\Delta E}{k_B T}\right)}$. The effect of the error uncertainties lowers with increasing temperature, so high temperature models are less influenced by the effect. With most models performing between 500 K-600 K the effect on each rate constant is at least one order of magnitude difference.	21
3.1	Different refinement criteria for the refinement of the target function in figure 3.5. The middle column documents the hierarchical surplus, i.e., the maximum norm of the local error, which decreases very slowly with the increasing level. Compared to that, the right column shows the L_1 -norm of the local error with a constant decrease with the refinement.	47
4.1	Intermediate energies and activation barriers for the model for Oxygen evolution on the Co_3O_4 (110)-A surface, taken from Refs. [133] and [134]. Reproduced from [23], with the permission of AIP Publishing.	80
4.2	The SG results for the initial grid and the final grid for decreasing threshold $\epsilon = [0.1, 0.01, 0.001, 0.0001]$ and the estimates for the expected $E(\log_{10}(\text{TOF}))$ and the variance $E(\log_{10}(\text{TOF}))$ for the SG results. Partially reproduced from supplementary material [23], with the permission of AIP Publishing.	86
4.3	The total variance and the contribution of the first and second order decomposition, documented for three cases of overpotential η and displayed in the upper row of the table. The importance of either the first, second or the remaining residual R on the total variance are underlined by the percentage contribution are displayed in the lower row.	93
4.4	Calculation of the total variance $\text{Var}_{\text{org}}(y)$ according to Eq. (2.10) for the LSA and Eq. (2.9) for the true variance and the reduced variance $\text{Var}_{\text{red},3}$ if the most important LSA energies contain a decreased error of $\Delta\Delta E_{\text{int/act},i} = 0.06$ eV for all three cases $\eta \in \{0.4, 0.7, 1.0\}$ V. This describes the recalculation of the most important energies with a higher accurate method.	97
4.5	Calculation of the total reduced variance if the three, four, or all GSA energies contain a decreased error of $\Delta\Delta E_{\text{int/act},i} = 0.06$ eV for all three cases $\eta \in \{0.4, 0.7, 1.0\}$ V. This describes the recalculation of the most important energies with a higher accurate method.	98

5.1	The number of points (NoP) for the final grids of both methods. The first row displays the results for ASG and the second for MLASG. The third row displays the actual difference in the number of points for both final grids.	117
5.2	List of input parameter for the reduced model for CO ₂ oxidation on at RuO ₂ with its default values at $T = 600$ K, $p_{\text{CO}} = p_{\text{O}_2} = 1$ bar. Included are the range of variation for each parameter, spanning our domain.	130
7.1	Results for the final sparse grid, its L_1 -norm of the interpolation error an quadrature error for the kinetic model function.	173
7.2	Results for the final sparse grid, its L_1 -norm of the interpolation error an quadrature error for the sigmoid function	174
7.3	Results for the final sparse grid, its L_1 -norm of the interpolation error an quadrature error for the continuous function	175
7.4	Results for the final sparse grid, its L_1 -norm of the interpolation error an quadrature error for the oscillatory function	175

Bibliography

- [1] J. N. Armor, "So you think you may have a better process: how can you define the value?" *Catalysis today* **178**, 8 (2011) (cit. on p. 9).
- [2] F. Zaera, "New challenges in heterogeneous catalysis for the 21st century," *Catalysis letters* **142**, 501 (2012) (cit. on pp. 9, 10).
- [3] F. Zaera, "Outstanding mechanistic questions in heterogeneous catalysis," *The Journal of Physical Chemistry B* **106**, 4043 (2002) (cit. on p. 9).
- [4] F. Zaera, "Probing catalytic reactions at surfaces," *Progress in surface science* **69**, 1 (2001) (cit. on p. 9).
- [5] G. A. Somorjai and Y. Li, *Introduction to surface chemistry and catalysis* (John Wiley & Sons, 2010) (cit. on p. 9).
- [6] J. K. Nørskov, T. Bligaard, J. Rossmeisl, and C. H. Christensen, "Towards the computational design of solid catalysts," *Nat. Chem.* **1**, 37 (2009) (cit. on pp. 9, 21, 25, 102).
- [7] M. K. Sabbe, M.-F. Reyniers, and K. Reuter, "First-principles kinetic modeling in heterogeneous catalysis: an industrial perspective on best-practice, gaps and needs," *Catalysis Science & Technology* **2**, 2010 (2012) (cit. on pp. 10, 11, 16).
- [8] J. A. Dumesic, A. A. Trevino, B. A. Milligan, L. A. Greppi, V. R. Balse, K. T. Sarnowski, C. E. Beall, T. Kataoka, and D. F. Rudd, "A kinetic modeling approach to the design of catalysts: formulation of a catalyst design advisory program," *Industrial & engineering chemistry research* **26**, 1399 (1987) (cit. on p. 10).
- [9] S. Rangarajan, C. T. Maravelias, and M. Mavrikakis, "Sequential-optimization-based framework for robust modeling and design of heterogeneous catalytic systems," *The Journal of Physical Chemistry C* **121**, 25847 (2017) (cit. on p. 10).
- [10] A. J. Medford, M. R. Kunz, S. M. Ewing, T. Borders, and R. Fushimi, "Extracting knowledge from data through catalysis informatics," *ACS Catalysis* **8**, 7403 (2018) (cit. on p. 10).
- [11] S. Matera, W. F. Schneider, A. Heyden, and A. Savara, "Progress in accurate chemical kinetic modeling, simulations, and parameter estimation for heterogeneous catalysis," *ACS Catalysis* **9**, 6624 (2019) (cit. on p. 10).

-
- [12] M. Stamatakis and D. G. Vlachos, "Unraveling the complexity of catalytic reactions via kinetic monte carlo simulation: current status and frontiers," *Acs Catalysis* **2**, 2648 (2012) (cit. on p. 10).
- [13] K. Reuter and M. Scheffler, "First-principles kinetic monte carlo simulations for heterogeneous catalysis: application to the co oxidation at ru o 2 (110)," *Physical Review B* **73**, 045433 (2006) (cit. on pp. 10, 17, 19, 129, 130).
- [14] M. Stamatakis, "Kinetic modelling of heterogeneous catalytic systems," *Journal of Physics: Condensed Matter* **27**, 013001 (2014) (cit. on p. 10).
- [15] J. Bray and W. Schneider, "First-principles thermodynamic models in heterogeneous catalysis," *Computational Catalysis* **14**, 9781849734905 (2013) (cit. on p. 10).
- [16] H. Prats, F. Illas, and R. Sayós, "General concepts, assumptions, drawbacks, and misuses in kinetic monte carlo and microkinetic modeling simulations applied to computational heterogeneous catalysis," *International Journal of Quantum Chemistry* **118**, e25518 (2018) (cit. on p. 10).
- [17] K. Reuter, "First-principles kinetic monte carlo simulations for heterogeneous catalysis: concepts, status, and frontiers," *ChemInform* **43**, no (2012) (cit. on p. 10).
- [18] M. Saeys, M.-F. Reyniers, M. Neurock, and G. B. Marin, "Ab initio reaction path analysis of benzene hydrogenation to cyclohexane on pt(111)," *J. Phys. Chem. B* **109**, PMID: 16851197, 2064 (2005) (cit. on pp. 10, 16).
- [19] O. R. Inderwildi, S. J. Jenkins, and D. A. King, "Fischer-tropsch mechanism revisited: alternative pathways for the production of higher hydrocarbons from synthesis gas," *J. Phys. Chem. C* **112**, 1305 (2008) (cit. on pp. 10, 16).
- [20] A. A. Gokhale, J. A. Dumesic, and M. Mavrikakis, "On the mechanism of low-temperature water gas shift reaction on copper," *J. Am. Chem. Soc.* **130**, PMID: 18181624, 1402 (2008) (cit. on pp. 10, 16).
- [21] K. S. Exner, F. H. H. Over, and A. P. Seitsonen, "Combined experiment and theory approach in surface chemistry: stairway to heaven?" *Surf. Sci* **640**, 165 (2015) (cit. on pp. 10, 16).
- [22] S. Matera, S. Blomberg, M. J. Hoffmann, J. Zetterberg, J. Gustafson, E. Lundgren, and K. Reuter, "Evidence for the active phase of heterogeneous catalysts through in situ reaction product imaging and multiscale modeling," *ACS Catal.* **5**, 4514 (2015) (cit. on pp. 10, 16).

- [23] S. Döpking, C. P. Plaisance, D. Stobusch, K. Reuter, C. Scheurer, and S. Matera, “Addressing global uncertainty and sensitivity in first-principles based microkinetic models by an adaptive sparse grid approach,” *The Journal of Chemical Physics* **Vol 148**, 3, 034102 © 2018 Author(s). Published by AIP Publishing (2018) (cit. on pp. 11, 15, 16, 24, 26, 30, 31, 41, 45, 75, 77, 79, 80, 82, 84–89, 91, 92, 94, 95, 99, 101, 103).
- [24] S. Döpking and S. Matera, “Error propagation in first-principles kinetic monte carlo simulation,” *Chem. Phys. Lett.* **674**, 28 (2017) (cit. on pp. 11, 15, 16, 95, 105, 129, 137).
- [25] A. J. Medford, J. Wellendorff, A. Vojvodic, F. Studt, F. Abild-Pedersen, K. W. Jacobsen, T. Bligaard, and J. K. Nørskov, “Assessing the reliability of calculated catalytic ammonia synthesis rates,” *Science* **345**, 197 (2014) (cit. on pp. 11, 16).
- [26] J. E. Sutton, W. Guo, M. A. Katsoulakis, and D. G. Vlachos, “Effects of correlated parameters and uncertainty in electronic-structure-based chemical kinetic modelling,” *Nat. Chem.*, 331 (2016) (cit. on pp. 11, 15, 16, 27).
- [27] E. Walker, S. C. Ammal, G. A. Terejanu, and A. Heyden, “Uncertainty quantification framework applied to the water–gas shift reaction over pt-based catalysts,” *J. Phys. Chem. C* **120**, 10328 (2016) (cit. on pp. 11, 15, 16, 102).
- [28] K. Reuter and M. Scheffler, “First-principles kinetic monte carlo simulations for heterogeneous catalysis: application to the co oxidation at ruo₂(110),” *Phys. Rev. B* **73**, 045433 (2006) (cit. on pp. 11, 115, 137).
- [29] J. M. Lorenzi, T. Stecher, K. Reuter, and S. Matera, “Local-metrics error-based shepard interpolation as surrogate for highly non-linear material models in high dimensions,” *The Journal of chemical physics* **147**, 164106 (2017) (cit. on pp. 11, 21, 22, 129, 130, 142).
- [30] P. Gelß, S. Matera, and C. Schütte, “Solving the master equation without kinetic monte carlo: tensor train approximations for a co oxidation model,” *J. Comput. Phys.* **314**, 489 (2016) (cit. on pp. 11, 18, 21, 22, 103, 129, 130, 133).
- [31] A. Kubas, D. Berger, H. Oberhofer, D. Maganas, K. Reuter, and F. Neese, “Surface adsorption energetics studied with “gold standard” wave-function-based ab initio methods: small-molecule binding to tio₂(110),” *J. Phys. Chem. Lett.* **7**, 4207 (2016) (cit. on pp. 11, 96, 103).
- [32] C. P. Plaisance, R. A. van Santen, and K. Reuter, “Constrained-Orbital Density Functional Theory. Computational Method and Applications to Surface Chemical Processes,” *J. Chem. Theory Comput.*, **10** . 1021/acs.jctc.7b00362 (2017) (cit. on pp. 11, 12, 96, 103).

-
- [33] C. T. Campbell, “Future directions and industrial perspectives micro- and macro-kinetics: their relationship in heterogeneous catalysis,” English, *Top. Catal.* **1**, 353 (1994) (cit. on pp. 11, 25, 88).
- [34] H. Meskine, S. Matera, M. Scheffler, K. Reuter, and H. Metiu, “Examination of the concept of degree of rate control by first-principles kinetic monte carlo simulations,” *Surf. Sci.* **603**, 1724 (2009) (cit. on pp. 11, 25, 83, 85, 129, 137).
- [35] M. J. Hoffmann, F. Engelmann, and S. Matera, “A practical approach to the sensitivity analysis for kinetic monte carlo simulation of heterogeneous catalysis,” *J. Chem. Phys.* **146**, 044118 (2017) (cit. on pp. 11, 25, 27, 85, 129, 133).
- [36] G. E. B. Archer, A. Saltelli, and I. M. Sobol, “Sensitivity measures, anova-like techniques and the use of bootstrap,” *J. Stat. Comput. Simul.* **58**, 99 (1997) (cit. on pp. 11, 12, 27, 33).
- [37] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola, *Global sensitivity analysis. the primer* (Wiley VCH Verlag, 2008) (cit. on pp. 11, 27, 91, 92).
- [38] D. Strobusch and C. Scheurer, “Adaptive sparse grid expansions of the vibrational hamiltonian,” *J. Chem. Phys.* **140**, 074111 (2014) (cit. on p. 11).
- [39] K. Reuter, “Ab initio thermodynamics and first-principles microkinetics for surface catalysis,” *Catalysis Letters* **146**, 541 (2016) (cit. on p. 11).
- [40] J. D. Jakeman and S. G. Roberts, “Local and dimension adaptive sparse grid interpolation and quadrature,” *arXiv preprint arXiv:1110.0010* (2011) (cit. on pp. 12, 32, 39, 45, 46, 48, 52, 55, 59, 64).
- [41] M. B. Giles, “Multilevel monte carlo methods,” in *Monte carlo and quasi-monte carlo methods 2012*, edited by J. Dick, F. Y. Kuo, G. W. Peters, and I. H. Sloan (2013), pp. 83–103 (cit. on pp. 13, 105).
- [42] S. Döpking and S. Matera, “Multilevel adaptive sparse grid quadrature for monte carlo models,” *arXiv preprint arXiv:1810.00810* (2018) (cit. on pp. 13, 46, 131).
- [43] A. L. Teckentrup, R. Scheichl, M. B. Giles, and E. Ullmann, “Further analysis of multilevel monte carlo methods for elliptic pdes with random coefficients,” *Numerische Mathematik* **125**, 569 (2013) (cit. on pp. 13, 106, 109).
- [44] K. A. Cliffe, M. B. Giles, R. Scheichl, and A. L. Teckentrup, “Multilevel monte carlo methods and applications to elliptic pdes with random coefficients,” *Computing and Visualization in Science* **14**, 3 (2011) (cit. on pp. 13, 106, 109).

- [45] A. L. Teckentrup, P. Jantsch, C. G. Webster, and M. Gunzburger, “A multilevel stochastic collocation method for partial differential equations with random input data,” *SIAM/ASA J. Uncertainty Quantification* **3**, 1046 (2015) (cit. on pp. 13, 105).
- [46] H.-W. van Wyk, “Multilevel sparse grid methods for elliptic partial differential equations with random coefficients,” (2014) (cit. on pp. 13, 105).
- [47] A.-L. Haji-Ali, F. Nobile, L. Tamellini, and R. Tempone, “Multi-index stochastic collocation convergence rates for random pdes with parametric regularity,” *Foundations of Computational Mathematics* **16**, 1555 (2016) (cit. on pp. 13, 106).
- [48] A.-L. Haji-Ali, F. Nobile, and R. Tempone, “Multi-index monte carlo: when sparsity meets sampling,” *Numerische Mathematik* **132**, 767 (2016) (cit. on pp. 13, 106).
- [49] S. Heinrich, “Multilevel monte carlo methods,” in *International conference on large-scale scientific computing* (Springer, 2001), pp. 58–67 (cit. on pp. 13, 107, 108).
- [50] P. E. Heckman and G. G. Meyers, “The calculation of aggregate loss distributions from claim severity and claim count distributions,” in *Proceedings of the casualty actuarial society*, Vol. 70, 133-134 (Casualty Actuarial Society, 1983), pp. 49–66 (cit. on p. 15).
- [51] G. Meyers and N. Schenker, “Parameter uncertainty in the collective risk model,” *PCAS LXX*, 111 (1983) (cit. on p. 15).
- [52] A. Alexanderian, F. Rizzi, M. Rathinam, O. P. Le Maître, and O. M. Knio, “Preconditioned Bayesian regression for stochastic chemical kinetics,” *J. Sci. Comp.* **58**, 592 (2014) (cit. on pp. 15, 31, 105).
- [53] M. Stamatakis and D. G. Vlachos, “Unraveling the complexity of catalytic reactions via kinetic Monte Carlo simulation: current status and frontiers,” *ACS Catal.* **2**, 2648 (2012) (cit. on pp. 16, 103, 141).
- [54] C. W. Gardiner et al., *Handbook of stochastic methods*, Vol. 3 (springer Berlin, 1985) (cit. on p. 17).
- [55] G. J. Herschlag, S. Mitran, and G. Lin, “A consistent hierarchy of generalized kinetic equation approximations to the master equation applied to surface catalysis,” *J Chem. Phys.* **142**, 234703 (2015) (cit. on pp. 18, 21, 103).
- [56] A. Chatterjee and D. G. Vlachos, “An overview of spatial microscopic and accelerated kinetic monte carlo methods,” *Journal of computer-aided materials design* **14**, 253 (2007) (cit. on p. 18).
- [57] J. Rogal, K. Reuter, and M. Scheffler, “Co oxidation on pd (100) at technologically relevant pressure conditions: first-principles kinetic monte carlo study,” *Physical Review B* **77**, 155410 (2008) (cit. on pp. 19, 21).

-
- [58] K. Reuter, D. Frenkel, and M. Scheffler, “The steady state of heterogeneous catalysis, studied by first-principles statistical mechanics,” *Physical review letters* **93**, 116105 (2004) (cit. on p. 19).
- [59] R. I. Masel et al., *Chemical kinetics and catalysis*, Vol. 10 (Wiley-Interscience New York, 2001) (cit. on p. 19).
- [60] M. K. Sabbe, M.-F. Reyniers, and K. Reuter, “First-principles kinetic modeling in heterogeneous catalysis: an industrial perspective on best-practice, gaps and needs,” *Catal. Sci. Technol.* **2**, 2010 (2012) (cit. on pp. 21, 81, 83).
- [61] C. M. Nguyen, M.-F. Reyniers, and G. B. Marin, “Theoretical study of the adsorption of c1–c4 primary alcohols in h-zsm-5,” *Physical Chemistry Chemical Physics* **12**, 9481 (2010) (cit. on p. 21).
- [62] C. Morley, M. J. Pilling, R. Walker, C. Morley, S. H. Robertson, P. W. Seakins, D. Baulch, A. S. Tomlin, T. Turányi, S. Scott, et al., *Low-temperature combustion and autoignition*, 1997 (cit. on p. 22).
- [63] A. Saltelli, K. Chan, and E. Scott, *Sensitivity analysis. probability and statistics*, 2009 (cit. on p. 22).
- [64] D. Makowski, C. Naud, M.-H. Jeuffroy, A. Barbottin, and H. Monod, “Global sensitivity analysis for calculating the contribution of genetic parameters to the variance of crop model prediction,” *Reliability Engineering & System Safety* **91**, 1142 (2006) (cit. on p. 23).
- [65] B. Iooss and P. Lemaître, “A review on global sensitivity analysis methods,” in *Uncertainty management in simulation-optimization of complex systems* (Springer, 2015), pp. 101–122 (cit. on p. 23).
- [66] S. Lefebvre, A. Roblin, S. Varet, and G. Durand, “A methodological approach for statistical evaluation of aircraft infrared signature,” *Reliability Engineering & System Safety* **95**, 484 (2010) (cit. on p. 23).
- [67] R. Bolado-Lavin and A. C. Badea, “Review of sensitivity analysis methods and experience for geological disposal of radioactive waste and spent nuclear fuel,” JRC Scientific and Tech. Rep. EUR **23712** (2008) (cit. on p. 23).
- [68] E. de Rocquigny, N. Devictor, and S. Tarantola, *Uncertainty in industrial practice: a guide to quantitative uncertainty management* (John Wiley & Sons, 2008) (cit. on p. 23).
- [69] H. Christopher Frey and S. R. Patil, “Identification and review of sensitivity analysis methods,” *Risk analysis* **22**, 553 (2002) (cit. on p. 23).

- [70] J. C. Helton, J. D. Johnson, C. J. Sallaberry, and C. B. Storlie, "Survey of sampling-based methods for uncertainty and sensitivity analysis," *Reliability Engineering & System Safety* **91**, 1175 (2006) (cit. on p. 23).
- [71] J. P. Kleijnen, "Sensitivity analysis and related analyses: a review of some statistical techniques," *Journal of statistical computation and simulation* **57**, 111 (1997) (cit. on p. 23).
- [72] G. Chastaing, F. Gamboa, C. Prieur, et al., "Generalized hoeffding-sobol decomposition for dependent variables-application to sensitivity analysis," *Electronic Journal of Statistics* **6**, 2420 (2012) (cit. on p. 23).
- [73] S. Da Veiga, F. Wahl, and F. Gamboa, "Local polynomial estimation for sensitivity analysis on models with correlated inputs," *Technometrics* **51**, 452 (2009) (cit. on pp. 23, 28).
- [74] J.-P. Gauchi, S. Lehuta, and S. Mahévas, "Optimal sensitivity analysis under constraints: application to fisheries," *Procedia-Social and Behavioral Sciences* **2**, 7658 (2010) (cit. on p. 23).
- [75] J. Jacques, C. Lavergne, and N. Devictor, "Sensitivity analysis in presence of model uncertainty and correlated inputs," *Reliability Engineering & System Safety* **91**, 1126 (2006) (cit. on pp. 23, 28).
- [76] D. Kurowicka and R. M. Cooke, *Uncertainty analysis with high dimensional dependence modelling* (John Wiley & Sons, 2006) (cit. on p. 23).
- [77] G. Li, H. Rabitz, P. E. Yelvington, O. O. Oluwole, F. Bacon, C. E. Kolb, and J. Schoendorf, "Global sensitivity analysis for systems with independent and/or correlated inputs," *The journal of physical chemistry A* **114**, 6022 (2010) (cit. on p. 23).
- [78] A. Saltelli and S. Tarantola, "On the relative importance of input factors in mathematical models: safety assessment for nuclear waste disposal," *Journal of the American Statistical Association* **97**, 702 (2002) (cit. on pp. 23, 28).
- [79] C. Xu and G. Gertner, "Extending a global sensitivity analysis technique to models with correlated parameters," *Computational Statistics & Data Analysis* **51**, 5579 (2007) (cit. on p. 23).
- [80] M. D. Morris, "Factorial sampling plans for preliminary computational experiments," *Technometrics* **33**, 161 (1991) (cit. on pp. 23, 27).
- [81] W. Hoeffding, "A class of statistics with asymptotically normal distribution," in *Break-throughs in statistics* (Springer, 1992), pp. 308–334 (cit. on p. 23).

-
- [82] K. Campbell, M. D. McKay, and B. J. Williams, "Sensitivity analysis when model outputs are functions," *Reliability Engineering & System Safety* **91**, 1468 (2006) (cit. on p. 23).
- [83] F. Gamboa, A. Janon, T. Klein, and A. Lagnoux, "Sensitivity indices for multivariate outputs," arXiv preprint arXiv:1303.3574 (2013) (cit. on p. 23).
- [84] M. Lamboni, H. Monod, and D. Makowski, "Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models," *Reliability Engineering & System Safety* **96**, 450 (2011) (cit. on p. 23).
- [85] A. Marrel, B. Iooss, M. Jullien, B. Laurent, and E. Volkova, "Global sensitivity analysis for models with spatially dependent outputs," *Environmetrics* **22**, 383 (2011) (cit. on p. 23).
- [86] A. Saltelli, M. Ratto, S. Tarantola, and F. Campolongo, "Sensitivity analysis for chemical models," *Chem. Rev* **105**, 2811 (2005) (cit. on pp. 23, 25, 26).
- [87] I. M. Sobol, "Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates," *Mathematics and computers in simulation* **55**, 271 (2001) (cit. on p. 24).
- [88] R. Cukier, H. Levine, and K. Shuler, "Nonlinear sensitivity analysis of multiparameter model systems," *Journal of computational physics* **26**, 1 (1978) (cit. on p. 24).
- [89] M. D. McKay, "Nonparametric variance-based methods of assessing uncertainty importance," *Reliability engineering & system safety* **57**, 267 (1997) (cit. on p. 24).
- [90] Y. Chu, Z. Huang, and J. Hahn, "Global sensitivity analysis procedure accounting for effect of available experimental data," *Industrial & engineering chemistry research* **50**, 1294 (2011) (cit. on p. 24).
- [91] A. Saltelli, S. Tarantola, F. Campolongo, and M. Ratto, *Sensitivity analysis in practice: a guide to assessing scientific models*, Vol. 1 (Wiley Online Library, 2004) (cit. on pp. 26, 27).
- [92] I. M. Sobol', "On sensitivity estimation for nonlinear mathematical models," *Matematicheskoe modelirovanie* **2**, 112 (1990) (cit. on p. 28).
- [93] S. Kucherenko, S. Tarantola, and P. Annoni, "Estimation of global sensitivity indices for models with dependent variables," *Comput. Phys. Comm.* **183**, 937 (2012) (cit. on p. 28).
- [94] S. Rahman, "A generalized anova dimensional decomposition for dependent probability measures," *SIAM/ASA J. Uncertainty Quantification* **2**, 670 (2014) (cit. on pp. 28, 102).

- [95] C. Xu and G. Z. Gertner, “A general first-order global sensitivity analysis method,” *Reliability Engineering & System Safety* **93**, 1060 (2008) (cit. on p. 28).
- [96] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola, “Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index,” *Computer Physics Communications* **181**, 259 (2010) (cit. on p. 29).
- [97] A. Mood, F. Graybill, and D. Boes, *Introduction to the theory of statistics*, Vol. 69 (Dec. 1974) (cit. on p. 29).
- [98] T. Homma and A. Saltelli, “Importance measures in global sensitivity analysis of nonlinear models,” *Reliab. Eng. Syst. Saf.* **52**, 1 (1996) (cit. on pp. 30, 91, 92).
- [99] T. J. Sullivan, *Introduction to uncertainty quantification*, Vol. 63 (Springer, 2015) (cit. on pp. 31, 105).
- [100] M. Griebel and M. Holtz, “Dimension-wise integration of high-dimensional functions with applications to finance,” *Journal of Complexity* **26**, 455 (2010) (cit. on pp. 31, 45, 105).
- [101] D. Frenkel and B. Smit, *Understanding molecular simulation: from algorithms to applications*, Vol. 1 (Elsevier, 2001) (cit. on pp. 31, 105).
- [102] T. Gerstner and M. Griebel, “Dimension-adaptive tensor-product quadrature,” *Computing* **71**, 65 (2003) (cit. on pp. 32, 45, 47, 48, 64).
- [103] R. E. Bellman, *Adaptive control processes: a guided tour* (Princeton university press, 2015) (cit. on p. 32).
- [104] H.-J. Bungartz and M. Griebel, “Sparse grids,” *Act. Num.* **13**, 147 (2004) (cit. on pp. 32–35, 111).
- [105] P. J. Davis and P. Rabinowitz, *Methods of numerical integration* (Courier Corporation, 2007) (cit. on p. 32).
- [106] J. Hammersley and D. Handscomb, “Monte carlo methods, methuen & co,” Ltd., London **40** (1964) (cit. on p. 32).
- [107] M. Dror, P. L’ecuyer, and F. Szidarovszky, *Modeling uncertainty: an examination of stochastic theory, methods, and applications*, Vol. 46 (Springer Science & Business Media, 2002) (cit. on p. 32).
- [108] T. Gerstner, “Sparse grid quadrature methods for computational finance,” *Habilitation, University of Bonn* **77** (2007) (cit. on p. 32).
- [109] P. L’Ecuyer and C. Lemieux, “Recent advances in randomized quasi-monte carlo methods,” *Modeling uncertainty*, 419 (2002) (cit. on p. 32).

-
- [110] T. Gerstner and M. Griebel, “Numerical integration using sparse grids,” *Numer. algorithms* **18**, 209 (1998) (cit. on p. 32).
- [111] O. Knio and O. Le Maitre, “Uncertainty propagation in cfd using polynomial chaos decomposition,” *Fluid Dynamics Research* **38**, 616 (2006) (cit. on p. 32).
- [112] X. Ma and N. Zabarar, “An adaptive high-dimensional stochastic model representation technique for the solution of stochastic partial differential equations,” *Journal of Computational Physics* **229**, 3884 (2010) (cit. on pp. 32, 35, 52).
- [113] B. Ganapathysubramanian and N. Zabarar, “Sparse grid collocation schemes for stochastic natural convection problems,” *Journal of Computational Physics* **225**, 652 (2007) (cit. on pp. 32, 142).
- [114] C. Zenger, “Sparse grid,” in *Parallel algorithms for partial differential equations, proceedings of the 6th gamm-seminar kiel 1990*, edited by Hackbusch (Vieweg-Verlag, Jan. 1991), pp. 241–251 (cit. on p. 32).
- [115] M. Griebel, *A parallelizable and vectorizable multi-level-algorithm on sparse grids* (Technische Universität, 1990) (cit. on p. 32).
- [116] H.-J. Bungartz, *Dünne gitter und deren anwendung bei der adaptiven lösung der dreidimensionalen poisson-gleichung* (Technische Universität München, 1992) (cit. on p. 32).
- [117] S.A.Smolyak, “Quadrature and interpolation formulas for tensor products of certain classes of functions,” *Dokl. Akad. Nauk SSSR* **148**, 1042 (1963) (cit. on p. 32).
- [118] J. D. Jakeman and S. G. Roberts, “Local and dimension adaptive stochastic collocation for uncertainty quantification,” in *Sparse grids and applications* (Springer, 2012), pp. 181–203 (cit. on pp. 32, 37, 39, 45).
- [119] E. Plischke, E. Borgonovo, and C. L. Smith, “Global sensitivity measures from given data,” *Eur. J. Oper. Res.* **226**, 536 (2013) (cit. on p. 33).
- [120] K. Tang, P. M. Congedo, and A. R., “Sensitivity analysis using anchored anova expansion and high-order moments computation,” *Int. J. Numer. Methods Eng.* **102**, 1554 (2015) (cit. on p. 33).
- [121] J. Garcke, “Sparse grids in a nutshell,” in *Sparse grids and applications* (Springer, 2012), pp. 57–80 (cit. on p. 34).
- [122] D. M. Pflüger, “Spatially adaptive sparse grids for high-dimensional problems,” PhD thesis (Technische Universität München, 2010) (cit. on pp. 34, 39, 40).

- [123] F. Nobile, R. Tempone, and C. G. Webster, “A sparse grid stochastic collocation method for partial differential equations with random input data,” *SIAM Journal on Numerical Analysis* **46**, 2309 (2008) (cit. on p. 35).
- [124] A. Klimke and B. Wohlmuth, “Algorithm 847: spinterp: piecewise multilinear hierarchical sparse grid interpolation in matlab,” *ACM Trans. Math. Software* **31**, 561 (2005) (cit. on pp. 35, 39).
- [125] C. G. Webster, G. Zhang, and M. D. Gunzburger, *An adaptive wavelet stochastic collocation method for irregular solutions of stochastic partial differential equations*, tech. rep. (Oak Ridge National Laboratory (ORNL), 2012) (cit. on p. 35).
- [126] H. Yserentant, “Sparse grids, adaptivity, and symmetry,” *Computing* **78**, 195 (2006) (cit. on p. 37).
- [127] M. Griebel, “Adaptive sparse grid multilevel methods for elliptic pdes based on finite differences,” *Computing* **61**, 151 (1998) (cit. on pp. 37, 40, 45, 48).
- [128] H.-J. Bungartz and S. Dirnstorfer, “Multivariate quadrature on adaptive sparse grids,” *Computing* **71**, 89 (2003) (cit. on p. 45).
- [129] M. Stoyanov, “Hierarchy-direction selective approach for locally adaptive sparse grids,” Oak Ridge National Laboratory, Tennessee (2013) (cit. on p. 45).
- [130] M. Stoyanov, “Adaptive sparse grid construction in a context of local anisotropy and multiple hierarchical parents,” in *Sparse grids and applications-miami 2016* (Springer, 2018), pp. 175–199 (cit. on p. 45).
- [131] A. Genz, “Testing multidimensional integration routines,” in *Proc. of international conference on tools, methods and languages for scientific and engineering computation* (1984), pp. 81–94 (cit. on pp. 58, 64, 122).
- [132] A. Genz, “A package for testing multiple integration subroutines,” in *Numerical integration* (Springer, 1987), pp. 337–340 (cit. on pp. 64, 122).
- [133] C. P. Plaisance and R. A. van Santen, “Structure sensitivity of the oxygen evolution reaction catalyzed by cobalt (ii, iii) oxide,” *Journal of the American Chemical Society* **137**, 14660 (2015) (cit. on pp. 75, 80, 81, 98).
- [134] C. P. Plaisance, K. Reuter, and R. A. van Santen, “Quantum chemistry of the oxygen evolution reaction on cobalt(ii,iii) oxide – implications for designing the optimal catalyst,” *Faraday Discuss.* **188**, 199 (2016) (cit. on pp. 75, 80, 98).
- [135] M. G. Walter, E. L. Warren, J. R. McKone, S. W. Boettcher, Q. Mi, E. A. Santori, and N. S. Lewis, “Solar Water Splitting Cells,” *Chem. Rev.* **110**, 6446 (2010) (cit. on p. 75).

-
- [136] A. Marshall, B. Børresen, G. Hagen, M. Tsykin, and R. Tunold, “Hydrogen production by advanced proton exchange membrane (PEM) water electrolyzers—Reduced energy consumption by improved electrocatalysis,” *Energy* **32**, 431 (2007) (cit. on p. 75).
- [137] A. Roudgar, M. Eikerling, and R. van Santen, “Ab initio study of oxygen reduction mechanism at Pt(4) cluster,” *Phys. Chem. Chem. Phys.* **12**, 614 (2010) (cit. on p. 77).
- [138] M. Cococcioni and S. De Gironcoli, “Linear response approach to the calculation of the effective interaction parameters in the lda+ u method,” *Physical Review B* **71**, 035105 (2005) (cit. on p. 80).
- [139] H. J. Kulik, M. Cococcioni, D. A. Scherlis, and N. Marzari, “Density functional theory in transition-metal chemistry: a self-consistent hubbard u approach,” *Physical Review Letters* **97**, 103001 (2006) (cit. on p. 81).
- [140] V. I. Anisimov and O. Gunnarsson, “Density-functional calculation of effective Coulomb interactions in metals,” *Phys. Rev. B* **43**, 7570 (1991) (cit. on p. 81).
- [141] A. I. Liechtenstein, V. I. Anisimov, and J. Zaanen, “Density-functional theory and strong interactions: Orbital ordering in Mott-Hubbard insulators,” *Phys. Rev. B* **52**, R5467 (1995) (cit. on p. 81).
- [142] V. I. Anisimov, F. Aryasetiawan, and A. I. Lichtenstein, “First-principles calculations of the electronic structure and spectra of strongly correlated systems: the LDA + U method,” *J. Phys. Condens. Matter* **9**, 767 (1997) (cit. on p. 81).
- [143] A. J. Cohen, P. Mori-Sanchez, and W. Yang, “Insights into Current Limitations of Density Functional Theory,” *Science* **321**, 792 (2008) (cit. on p. 81).
- [144] E. Plischke, “An effective algorithm for computing global sensitivity indices (easi),” *Reliability Engineering & System Safety* **95**, 354 (2010) (cit. on p. 93).
- [145] J. Behler, “Perspective: machine learning potentials for atomistic simulations,” *The Journal of chemical physics* **145**, 170901 (2016) (cit. on p. 100).
- [146] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, and G. Seifert, “Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties,” *Physical Review B* **58**, 7260 (1998) (cit. on p. 100).
- [147] J. Wellendorff, K. T. Lundgaard, A. Møgelhøj, V. Petzold, D. D. Landis, J. K. Nørskov, T. Bligaard, and K. W. Jacobsen, “Density functionals for surface science: exchange-correlation model development with bayesian error estimation,” *Phys. Rev. B* **85**, 235149 (2012) (cit. on p. 102).

- [148] E. Shustorovich and H. Sellers, “The ubi-qep method: a practical theoretical approach to understanding chemistry on transition metal surfaces,” *Surf. Sci. Rep.* **31**, 1 (1998) (cit. on p. 102).
- [149] Z. W. Ulissi, A. J. Medford, T. Bligaard, and J. K. Nørskov, “To address surface reaction network complexity using scaling relations machine learning and DFT calculations,” *Nat. Comm.* **8**, 14621, 14621 (2017) (cit. on p. 102).
- [150] L. M. Ghiringhelli, J. Vybiral, E. Ahmetcik, R. Ouyang, S. V. Levchenko, C. Draxl, and M. Scheffler, “Learning physical descriptors for materials science by compressed sensing,” *New J. Phys.* **19**, 023017 (2017) (cit. on p. 102).
- [151] S. Ringe, H. Oberhofer, C. Hille, S. Matera, and K. Reuter, “Function-space-based solution scheme for the size-modified poisson–boltzmann equation in full-potential dft,” *Journal of chemical theory and computation* **12**, 4052 (2016) (cit. on p. 103).
- [152] S. Heinrich, “Monte carlo complexity of global solution of integral equations,” *Journal of Complexity* **14**, 151–175 (1998) (cit. on pp. 105, 106, 115).
- [153] L. W.-T. Ng and M. Eldred, “Multifidelity uncertainty quantification using non-intrusive polynomial chaos and stochastic collocation,” in 53rd aiaa/asme/asce/ahs/asc structures, structural dynamics and materials conference 20th aiaa/asme/ahs adaptive structures conference 14th aiaa (2012), p. 1852 (cit. on p. 105).
- [154] M. Holtz, *Sparse grid quadrature in high dimensions with applications in finance and insurance*, Vol. 77 (Springer Science & Business Media, 2010) (cit. on p. 105).
- [155] S. Heinrich, “Wavelet monte carlo methods for the global solution of integral equations,” *Proceedings of the Third International Conference on Monte Carlo and Quasi-Monte Carlo Methods*, Claremont 1998, 227 (1999) (cit. on pp. 106, 115).
- [156] S. Heinrich, “The multilevel method of dependent tests,” *Advances in Stochastic Simulation Methods*, 47 (2000) (cit. on p. 106).
- [157] A. L. Teckentrup, P. Jantsch, C. G. Webster, and M. Gunzburger, “A multilevel stochastic collocation method for partial differential equations with random input data,” *SIAM/ASA Journal on Uncertainty Quantification* **3**, 1046 (2015) (cit. on pp. 106, 109).
- [158] S. Heinrich, “Multilevel monte carlo methods,” in *Large-scale scientific computing*, edited by S. Margenov, J. Waśniewski, and P. Yalamov (2001), pp. 58–67 (cit. on p. 109).
- [159] M. J. Hoffmann, S. Matera, and K. Reuter, “Kmos: a lattice kinetic monte carlo framework,” *Computer Physics Communications* **185**, 2138 (2014) (cit. on p. 130).

- [160] H. Wang, T. Shen, and X. Xu, “A good prediction of the overall reaction rate may not mean a correct description of the reaction kinetics: a case study for co oxidation on ruo2 (110) surfaces,” *The Journal of Physical Chemistry C* **125**, 9169 (2021) (cit. on pp. 131, 137).
- [161] B. Bohn, M. Griebel, and J. Oettershagen, “Optimally rotated coordinate systems for adaptive least-squares regression on sparse grids,” in *Proceedings of the 2019 siam international conference on data mining*, Also available as INS Preprint No. 1812 (2019), pp. 163–171 (cit. on p. 141).
- [162] W. Guo and Y. Cheng, “A sparse grid discontinuous galerkin method for high-dimensional transport equations and its application to kinetic simulations,” *SIAM Journal on Scientific Computing* **38**, A3381 (2016) (cit. on p. 142).
- [163] M. Bracconi and M. Maestri, “Training set design for machine learning techniques applied to the approximation of computationally intensive first-principles kinetic models,” *Chemical Engineering Journal* **400**, 125469 (2020) (cit. on p. 143).
- [164] M. Klingenberger, O. Hirsch, and M. Votsmeier, “Efficient interpolation of precomputed kinetic data employing reduced multivariate hermite splines,” *Computers & Chemical Engineering* **98**, 21 (2017) (cit. on p. 143).

7 Appendix

7.1 Sparse Grid

The following section displays the detailed data for the SG examples of section 3.4 for (110)-A surface structure of the Co_3O_4 for the threshold of $\epsilon = 0.01$. We included all dimensional cases, 2D, 4D and 8D for all four examples with the results for the DA, LA and the combined technique with dimensional and local adaptivity. In all cases is the intrinsic dimensionality 2 but we extended the models of the 4D and 8D cases with 4 and respectively 8 dummy dimensions to underline the effect of the dimension adaptivity.

For the continuous and the oscillatory examples, we also included the 4D case compared to the figures

	refinement strategy	dimension	dimension-local	local
2D	\mathcal{A}_{2D}	3.85e+02	1.33e+02	1.41e+02
	$\ u(\mathbf{x}) - f(\mathbf{x})\ _1$	1.13e-01	2.36e-01	2.28e-01
	$ I - I_N $	1.33e-02	6.42e-02	6.36e-02
4D	\mathcal{A}_{4D}	3.89e+02	1.37e+02	3.53e+02
	$\ u(\mathbf{x}) - f(\mathbf{x})\ _1$	1.13e-01	2.34e-01	2.26e-01
	$ I - I_N $	1.33e-02	6.42e-02	6.36e-02
8D	\mathcal{A}_{8D}	3.97e+02	1.45e+02	7.77e+02
	$\ u(\mathbf{x}) - f(\mathbf{x})\ _1$	1.13e-01	2.34e-01	2.26e-01
	$ I - I_N $	1.33e-02	6.42e-02	6.36e-02

Table 7.1: Results for the final sparse grid, its L_1 -norm of the interpolation error an quadrature error for the kinetic model function.

3.15 and 3.17 to underline the strong effect of the dimension adaptivity for the dummy dimension. Table 7.3 and table 7.4 show in case of the DA that for the increasing dimensionality only points of the initial grid are added, resulting in 4 and 12 more points for 4D and 8D compared to the 2D example.

7.2 Oxygen Evolution Model

Based on the convergence of the sparse grid results, we expect similar results for the TSI of the oxygen evolution model. Figure 7.1 shows the TSI results for the three cases of overpotential η for three

refinement strategy		dimension	dimension-local	local
2D	\mathcal{A}_{2D}	2.25e+02	5.00e+01	5.80e+01
	$\ u(\mathbf{x}) - f(\mathbf{x})\ _1$	9.06e-03	1.07e-02	7.73e-03
	$ I - I_N $	4.03e-03	5.50e-03	2.78e-03
4D	\mathcal{A}_{4D}	2.29e+02	5.40e+01	1.34e+02
	$\ u(\mathbf{x}) - f(\mathbf{x})\ _1$	9.01e-03	1.07e-02	7.88e-03
	$ I - I_N $	4.03e-03	5.50e-03	2.78e-03
8D	\mathcal{A}_{8D}	2.37e+02	6.20e+01	2.86e+02
	$\ u(\mathbf{x}) - f(\mathbf{x})\ _1$	9.01e-03	1.07e-02	7.87e-03
	$ I - I_N $	4.03e-03	5.50e-03	2.78e-03

Table 7.2: Results for the final sparse grid, its L_1 -norm of the interpolation error an quadrature error for the sigmoid function

different tolerances, $\text{tol} = 0.1$ (dark blue), $\text{tol} = 0.01$ (green) and $\text{tol} = 0.001$ (orange). Already at the smallest threshold of $\text{tol} = 0.1$, the important parameters are properly identified. The variations caused by lowering the tolerances are minimal and do not change significantly after $\text{tol} = 0.001$. Furthermore, no variations are visibly for the identified parameters. Although the importance may vary slightly, the important underlying parameters are already identified with a very coarse sparse grid.

refinement strategy		dimension	dimension-local	local
2D	\mathcal{A}_{2D}	2.10e+01	1.70e+01	2.70e+01
	$\ u(\mathbf{x}) - f(\mathbf{x})\ _1$	4.96e-02	5.01e-02	3.02e-02
	$ I - I_N $	8.83e-03	1.35e-02	2.11e-02
4D	\mathcal{A}_{4D}	2.50e+01	2.10e+01	6.30e+01
	$\ u(\mathbf{x}) - f(\mathbf{x})\ _1$	4.96e-02	5.01e-02	3.01e-02
	$ I - I_N $	8.83e-03	1.35e-02	2.11e-02
8D	\mathcal{A}_{8D}	3.30e+01	2.90e+01	1.35e+02
	$\ u(\mathbf{x}) - f(\mathbf{x})\ _1$	4.96e-02	5.01e-02	3.01e-02
	$ I - I_N $	8.83e-03	1.35e-02	2.11e-02

Table 7.3: Results for the final sparse grid, its L_1 -norm of the interpolation error an quadrature error for the continuous function

refinement strategy		dimension	dimension-local	local
2D	\mathcal{A}_{2D}	8.10e+01	7.30e+01	8.90e+01
	$\ u(\mathbf{x}) - f(\mathbf{x})\ _1$	4.07e-02	5.25e-02	4.05e-02
	$ I - I_N $	2.13e-03	6.86e-03	6.91e-03
4D	\mathcal{A}_{4D}	8.50e+01	7.70e+01	2.37e+02
	$\ u(\mathbf{x}) - f(\mathbf{x})\ _1$	4.07e-02	5.25e-02	4.045e-02
	$ I - I_N $	2.13e-03	6.86e-03	6.91e-03
8D	\mathcal{A}_{8D}	9.30e+01	8.50e+01	5.33e+02
	$\ u(\mathbf{x}) - f(\mathbf{x})\ _1$	4.07e-02	5.25e-02	4.04e-02
	$ I - I_N $	2.13e-03	6.86e-03	6.91e-03

Table 7.4: Results for the final sparse grid, its L_1 -norm of the interpolation error an quadrature error for the oscillatory function

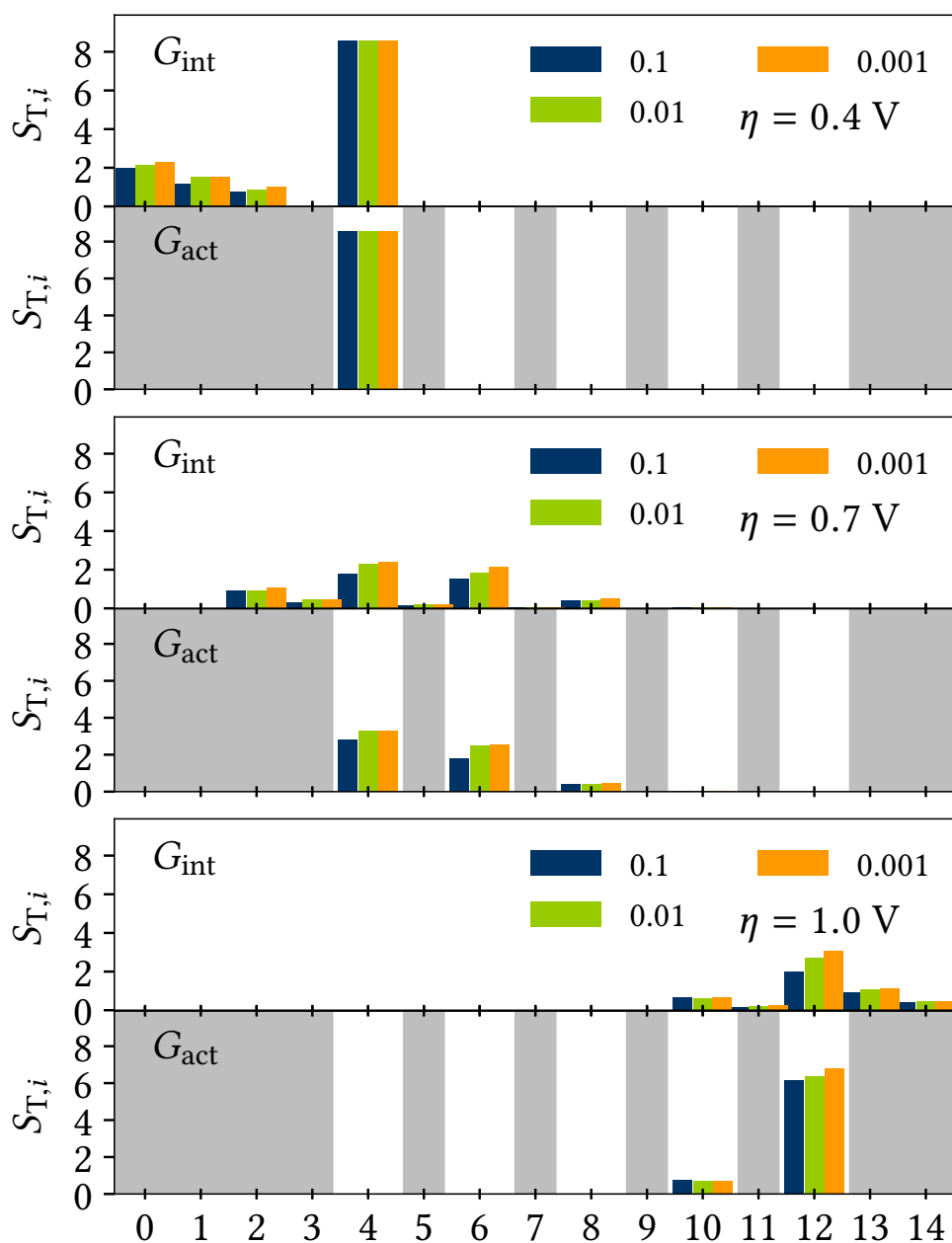


Figure 7.1: Convergence plot of the TSI for the three different overpotentials $\eta \in \{0.4, 0.7, 1.0\}$ V. We display the global sensitivity indices $S_{T,i}$ for the 20 input parameters for the (110)-A structure of the oxygen evolution reaction. The results are shown for three decreasing tolerances $\epsilon = [0.1, 0.01, 0.001]$ of the sparse grid construction. The gray-shaded fields describe intermediate states that are not considered for a reaction in the model. We present the sensitivities of $\log_{10}(\text{TOF} \times \text{s})$ with respect to the errors in $\Delta E_{\text{int},i}$ and $\Delta E_{\text{act},i}$.

7.3 Multilevel refinement

This chapter introduces three additional examples for multilevel refinement that can be of interest for chemical data. First we analyse the sigmoid function (Eq. (7.1)) that we also used as an example for the sparse grid refinement in Sec. 3.4. In that section, we explained why this function is useful for chemical data as it has the same characteristics as the coverage behaviour of one species. Here, we just increase the outer dimensionality to $D = 4$.

The second example is the sharper peak which is a sharper version of the continuous peak, introduced by Genz [1]. As we already explained, the adaptive SG with piecewise linear functions is not effective for functions with no locally restricted features. We modified the original example to a stronger peak. The same basic structure as the oscillatory case (Eq. (5.27)) of section 5.6 has the cosinus kink function (Eq. (7.3)), including an edge caused by the absolute value function. This characteristic with locally sharp edges and features are expected to profit from the local-adaptive sparse grid approach. Furthermore, we decrease the intrinsic dimensionality to 2, so the important dimensions are quartered. With this, we shape the general oscillatory function so that the MLASG method would be favourable and compare the results to the original, whether the multilevel impact is stronger. Similar to the previous example we test three different noises $\sigma = [1, 0.1, 0.001]$.

sigmoid

$$f_{\text{sig}}(\mathbf{x}) = \frac{1}{1 + e^{(x_1 - 0.3) \cdot 100}} \cdot \frac{1}{1 + e^{(-x_2 - 0.3) \cdot 100}} \quad D = 4 \quad (7.1)$$

sharper peak

$$f_{\text{peak}}(\mathbf{x}) = \exp(-a_i * \|\mathbf{x}\|_1) \quad a_i = 30.0; D = 6 \quad (7.2)$$

cosinus kink

$$f_{\text{cos kink}}(\mathbf{x}) = \left| \cos \left(\sum_i^2 a_i (x_i + 0.5) + \pi \right) \right| \quad a_i = 4.0; D = 8 \quad (7.3)$$

Figure 7.2 displays the performance of the two adaptive approaches for the L_1 -norm of the interpolation error and the quadrature error regarding the NoP, including the CPU savings using the multilevel approach compared to the standard ASG.

None of the examples shows a significant difference between the results of the ASG and the MLASG for the L_1 -norm of the interpolation error. Only for the sharper peak and the higher noises, both refinement methods show deviations for coarser approximations that vanish for increasing accuracy. Overall, the approximation for this function is not as effective as for the other examples. For the

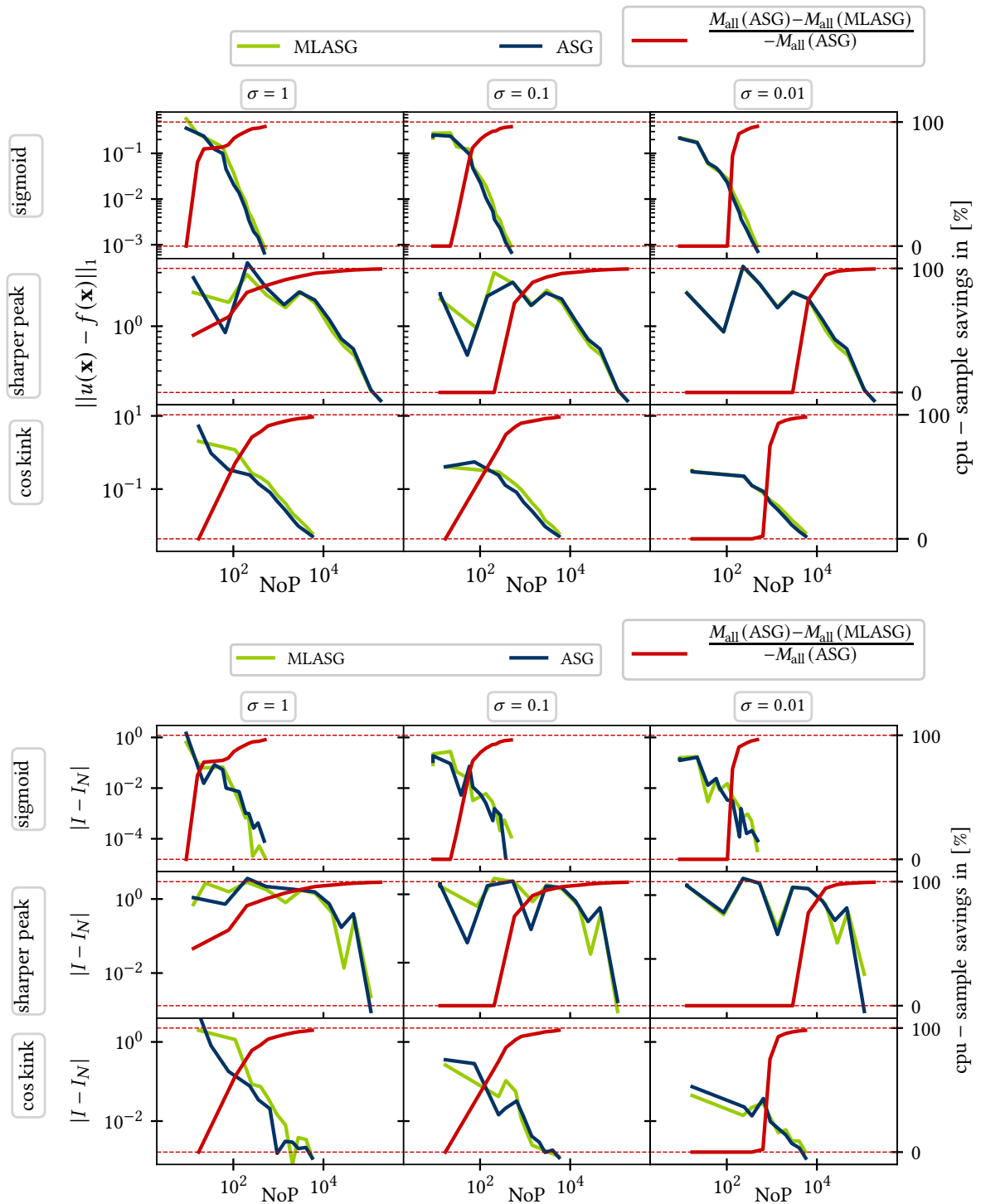


Figure 7.2: Results of ASG (—) and MLASG (—) for L_1 -norm of the interpolation error (*upper panel*) and the quadrature error (*lower panel*) for the three functions f_{sig} , f_{peak} , $f_{\text{cos kink}}$. Displayed are three different noises $\sigma = [1, 0.1, 0.01]$ and the CPU savings for the MLASG compared to the ASG (—).

Upper row : Performance for the function f_{osc} in $D = 2$, which is a classical example for quadrature calculations.

Middle row : Performance for the function $f_{\text{osc dec}}$ in $D = 14$ with a decreasing hierarchy of dimensions.

Lower row : Performance for the function $f_{\text{cos kink}}$ in $D = 8$ with a intrinsic dimensionality of two.

major amount of the refinement, the methods show no decreasing trend for the error. At already very high NoP the curves start to decrease with a slope of lower than 0.5, which is very small compared to the other examples whose slope is > 1 . Thus the adaptive SG methods do not provide a sufficient convergence. Regarding the CPU-savings, the MLASG can save up to $> 90\%$ computational cost for all cases, compared to the normal adaptive method. Considering the σ , major savings can be achieved with very coarse grids when the noise is high.

The $f_{\cos \text{ kink}}$ example is favourable for an adaptive approach since 6 of the dimensions are dummy dimensions that can be excluded regarding the refinement. The included artificial edge of the function gives a good local feature for the ASG and MLASG. The lower row of the L_1 -norm in figure 7.2 shows similar results again, except for a small constant deviation for the MLASG towards a higher NoP. The MLASG needs a small number of points more than the ASG to reach the same accuracy, but the CPU savings of almost a 100% illustrates the impact of the multilevel approach for the computational cost. So even though the MLASG needs more points, it can save most of the CPU-effort of the ASG.

For the quadrature error, the results show more fluctuations but also similar trends as before. We have a very slow convergence of both methods for the sharper peak, which is in line with the interpolation error results. We see an even stronger decrease for the other two results than for the L_1 -norm without a strong deviation between both methods. In the case of the $f_{\cos \text{ kink}}$, both methods draw a similar picture to the examples before. The ASG and MLASG results show a decreasing trend with smaller fluctuations for a smaller noise and greater fluctuations for increasing noise. All in all, both methods manage to increase the interpolation's accuracy, with the MLASG saving up to 92% of the computational cost. Overall the results show no negative impact of the multilevel approach. On the contrary, we can already reduce the computational cost for the coarse sparse grid without challenging the interpolation accuracy of the integral.

7.4 Multilevel Quadrature

The main work is concentrating on the case to have a surrogate model for further analysis like a GSA. If we would be only interested in the quadrature, there are alternatives to the SG approach. Especially MC or multilevel-MC methods are very efficient to approximate the quadrature. However, those are not applicable for, e.g. function approximation and therefore not a real alternative to our problem. Due to its high efficiency toward the quadrature, we want to introduce the general idea of a MLQ.

Assuming we would approximate the integral (Eq. (3.2)) of function f

$$I = \int_{\Omega} f(x) dx^D \tag{7.4}$$

with some kind of numerical quadrature

$$I = \lim_{N \rightarrow \infty} I_N \text{ with } I_N = \sum_{i=1}^N w_i f(x_i), \quad (7.5)$$

we would have a sum of w_i weighted function evaluation $f(x_i)$ at the quadrature nodes $\{x_i\}_{i=1}^N$ over N samples. With the classical Multilevel (ML) approach we decompose the intergal I using a telescoping sum

$$I_R = I_0 + \sum_{r>0}^R I_r - I_{r-1} = I_0 + \sum_{r>0}^R \Delta I_r \text{ with } I_r = \int_{\Omega} f_r(x) dx^D \quad (7.6)$$

where $f_r(\mathbf{x})$ is the approximation of $f(x)$ on the level r . For a parametric MC model, the r correlates with the accuracy of the evaluation, so the number of samples M_r per integration node drawn from the model. The Integrals I_0 and ΔI_r are numerically solved by independent quadrature rules. f_r are generated from a finite number of pointwise independent Monte Carlo samples and therefore they are discontinuous at every point in Ω . However, the samples are drawn independently, i.e. the errors should be statistically independent. Then, ΔI_r equals zero, $\Delta I_r = 0$, when employing suitable quadrature rules in Eq. (7.5), whose properties we will specify below. Thereby, the method equals a single numerical quadrature with drawing only a single MC sample per node.

$$I = I_{r=0} \quad (7.7)$$

In order to proof under which conditions the above statement holds true, we consider the numerical quadrature and the limit of infinitely many quadrature nodes. In the considered multilevel setting for MC models, the true function values at the nodes have to be replaced with the MC estimates $\overline{Y_{x_i, M}}$, (compare section 5.1), based on M Samples per node. We then approximate the integral I using

$$I \approx I_N \approx \overline{I_N} = \sum_{i=0}^N w_i \overline{Y_{x_i, M}}. \quad (7.8)$$

with $\mathbf{E}(\overline{Y_{x_i, M}}) = f(x_i)$ and $\text{Var}(\overline{Y_{x_i, M}}) \leq C^* M^{-1}$.

By the linearity of Eq. (7.8) the expected value of $\overline{I_N}$ equals the numerical quadrature I_N with the exact function values, $\mathbf{E}(\overline{I_N}) = I_N$ [2]. For the mean square error for the estimator we arrive at

$$\begin{aligned} \mathbf{E}(I - \overline{I_N})^2 &= \\ &= \mathbf{E}(|(I - I_N) - (\overline{I_N} - I_N)|^2) \\ &= \mathbf{E}(|I - I_N|^2) - 2\mathbf{E}((I - I_N)(\overline{I_N} - I_N)) + \mathbf{E}((\overline{I_N} - I_N)^2) \\ &= |I - I_N|^2 + (I - I_N)\mathbf{E}(\overline{I_N} - I_N) + \text{Var}(\overline{I_N}) \\ &= |I - I_N|^2 + \text{Var}(\overline{I_N}) \end{aligned}$$

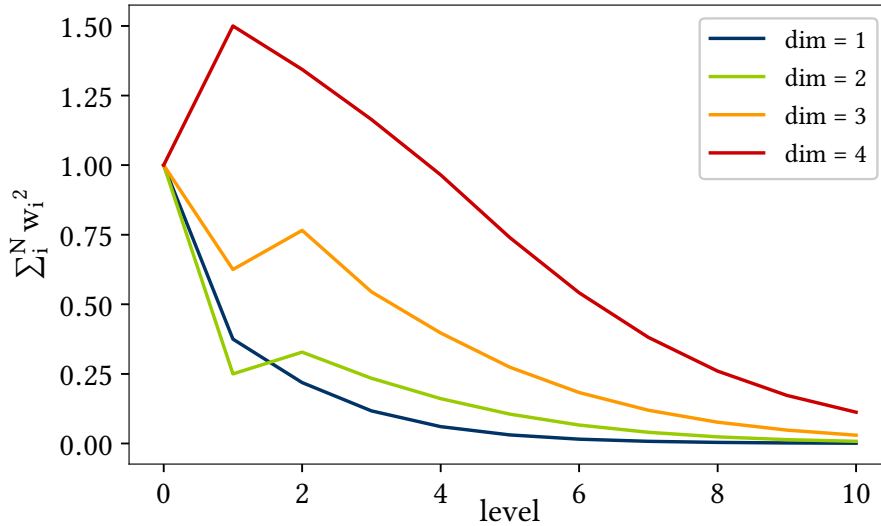


Figure 7.3: The sum of the squared quadrature weights for $\text{dim}=[1, 2, 3, 4]$ shows the convergence for increasing dimensionality over a common denominator, the refinement level of the FSG. The actual NoP for grid with refinement level increase with the dimensionality.

where, for the last equality, we employed that $\mathbb{E}(\overline{I_N}) = I_N$. The first term after the last equality is simply the quadrature error if we would use the exact function values. By Eq. (7.5), this converges to zero for $N \rightarrow \infty$. What remains is to show under which conditions $\lim_{N \rightarrow \infty} \text{Var}(\overline{I_N}) = 0$. Since the samples are drawn independently, the variance of the integral can be estimated by

$$\text{Var}(\overline{I_N}) = \sum_{i=1}^N w_i^2 \text{Var}(\overline{Y_{x_i, M}}) \leq C^* \sum_{i=1}^N w_i^2. \quad (7.9)$$

The variance $\text{Var}(\overline{I_N})$ converges to zero for $N \rightarrow \infty$, when the sum of the squared quadrature weights tends to zero. In this case, our approximate $\overline{I_N}$ converges against the true integral I irrespectively of the choice of the number samples M per node. In this sense, we can say $I_r = I_{r'} \forall r, r'$ and hence $\Delta I_r = 0$ and Eq (7.7) holds. This is the case for equal-weight quadrature rules (e.g. QMC), where the variance $\text{Var}(\overline{I_N})$ tends to zero as $\mathcal{O}(N^{-1})$ (which is the fastest rate, which one can expect for a MC model). Thus Multilevel quadrature boils down to a straightforward approach in the considered parametric Monte Carlo models. One draws a single sample per node and estimates the integral using this data.

We want to emphasise that this procedure can not be employed using arbitrary quadrature rules. For rules, where the sum of the squares of the quadrature weights does not converge to zero for $N \rightarrow \infty$, the variance will also not converge to zero. Even if it converges, this might be at a prolonged rate if the sum of the squared weights tends only slowly to zero. This can be the case for the SG weights. Figure 7.3 shows the behaviour of the squared weights for different dimensions for a FSG, and as can be seen, the convergence becomes slower with increasing dimension. The behaviour is displayed over

the levels to have a common comparison between the dimension. What needs to be considered is the difference in the number of points. While the grid of $l = 10$ for $dim = 1$ only includes 1025 points the grid for same level for $dim = 4$ includes 113409 points. This means that even though the method only uses one sample per point, the number of points can be increasingly high till the integral converges. Applying now an adaptive grid refinement, [ASG](#), would be of no substantial help because I) it will not alter the weights significantly, and II) the [MC](#) noise in the samples would prohibit a reasonably accurate estimation of the surplus resulting in random refinements. The large random noise of a single point would prohibit a reasonably accurate estimation of the surplus.

For a better understanding of the [MLQ](#) with [QMC](#) we include an example case for the quadrature error and compare the performance of the [MLASG](#) and the [MLQ](#). As an example we choose the cos kink function, introduced in section 7.3, except we will not 6 dummy dimensions just 2

cosinus kink in 4D

$$f_{\text{cos kink}}(\mathbf{x}) = \left| \cos \left(\sum_i^2 a_i(x_i + 0.5) + \pi \right) \right| \quad a_i = 4.0; D = 4. \quad (7.10)$$

Figure 7.4 shows the results for both methods for the cos kink in 4D- function for 3 different noises $\sigma_0 = [0.01, 0.1, 1]$. As we mentioned before we draw only one sample per node for the [MLQ \(QMC\)](#) method, so that we choose the all drawn sample M_{all} , which will equal the number of [QMC](#)-points for the [MLQ](#), as an indicator for the [CPU](#)-effort and display it on the x-axis. For the [MLASG](#) method, M_{all} equals the sum of all drawn samples to reach an accuracy of every point of $\epsilon = \sigma_0 * 2^{||l||_1}$. As it can be seen in the right pannel of figure 7.4, the [MLASG](#) performs worse than the [MLQ\(QMC\)](#) even for the smallest noise. For coarser accuracy the results are similar but with further refinement the [MLQ](#) is able to keep a higher decreasing slope compared to the [MLASG](#). Therefore, the [MLQ](#) needs around 3 orders of magnitude less M samples to arrive at the same accuracy for the quadrature error. Similar results are displayed for the higher noises in the middel pannel ($\sigma = 0.1$) and the left pannel ($\sigma = 1$) in figure 7.4. The [MLQ](#) is able to reach a high accuracy with at least 4 orders of magnitude less [CPU](#)-effort than the [MLASG](#) with increasing fluctuations toward higher noises.

All in all, the [MLQ](#) outperforms the [MLASG](#) for approximating the quadrature and is superior for this application. Nevertheless, the limited functionality of the [MLQ](#) is a significant disadvantage to [MLASG](#). If we want to apply this to the [GSA](#), we would have the problem that the errors are not cancelled out for the Sensitivity Index ([SI](#)), so that we need accurate values for the nodes. Furthermore, the [GSA](#) approximations are far more expensive due to the multiple integral approximations. So if the application is restricted to the expected value/ quadrature, the [MLQ](#) is a better solution than [MLASG](#). For more diverse applications, the surrogate model gives more flexibility.

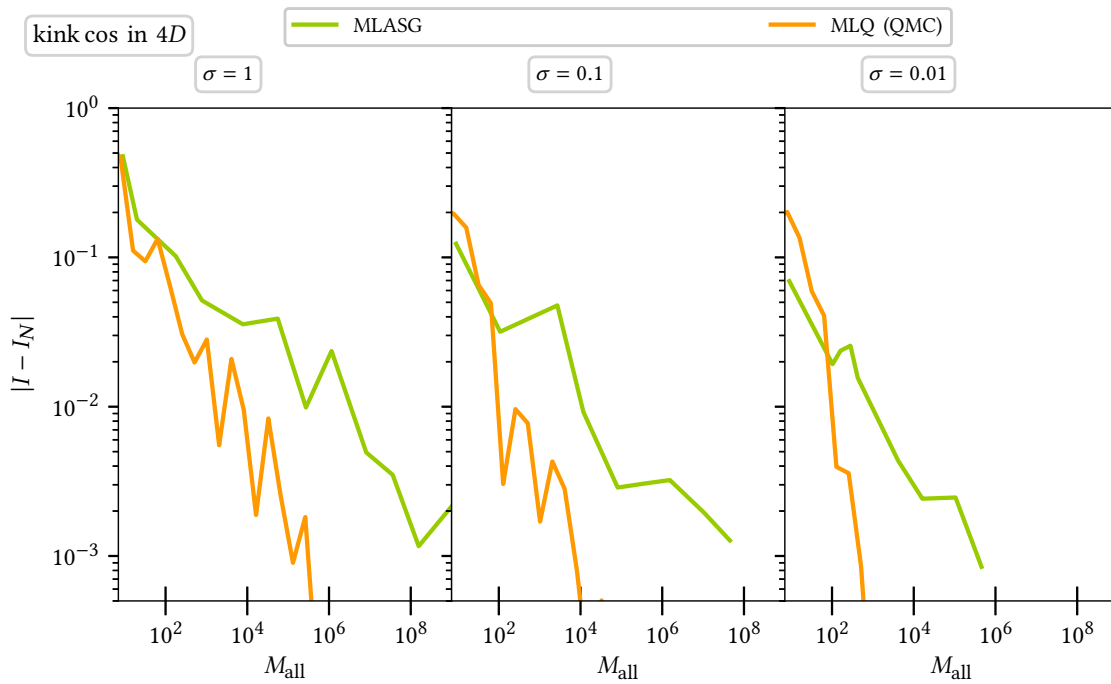


Figure 7.4: The quadrature error $|I - I_N|$ results for the cosine kink in 4D- function for the MLQ (QMC) and the MLASG methods for different tolerances ϵ . Displayed are the results of three different initial noise cases $\sigma_0 = [1, 0.1, 0.01]$ over the sum of drawn samples M_{all} .

Appendix Bibliography

- [1] A. Genz, “Testing multidimensional integration routines,” in Proc. of international conference on tools, methods and languages for scientific and engineering computation (1984), pp. 81–94 (cit. on p. 177).
- [2] S. Döpking and S. Matera, “Multilevel adaptive sparse grid quadrature for monte carlo models,” arXiv preprint arXiv:1810.00810 (2018) (cit. on p. 180).

Deutsche Zusammenfassung

Auf dem Gebiet der heterogenen Katalyse hat sich die First-Principle-basierte mikrokinetische Modellierung als wesentliches Werkzeug bewährt, um ein tieferes Verständnis der mikroskopischen Wechselwirkung zwischen Reaktionen zu ermöglichen. Leider basieren die katalytischen Modelle auf Informationen aus der elektronischen Strukturtheorie (z. B. Dichtefunktionaltheorie), die aufgrund intrinsischer Näherungen einen beträchtlichen Fehler enthalten. In dieser Arbeit werden wir analysieren wie signifikant die Auswirkungen dieser Fehler auf das Modellergebnis sein können. Dazu erklären wir zunächst, wie diese Fehler in ein Modellergebnis, wie z. B. Turnover-Frequency (TOF), übertragen werden. Des Weiteren quantifizieren wir die Auswirkung einzelner Fehler mittels einer lokalen und globalen Sensitivitätsanalyse und erklären die Unterschiede beider Methoden.

Der globale Sensitivitätsansatz erfordert das Lösen hochdimensionaler Integrale bzw. ein akkurates Ersatzmodell zum Auswerten, wofür wir einen lokalen und dimensions-adaptiven Sparse Grid-Ansatz benutzen. Sparse Grids (SG) haben sich für mitteldimensionale Probleme als sehr nützlich erwiesen, da ihre Adaptivitätsfunktion ein genaues Ersatzmodell mit einer kleinen Anzahl von Punkten ermöglicht. Trotz der hohen Dimensionalität der Modelle wird das Ergebnis meist von einem Bruchteil der Modellparameter dominiert, was eine hohe Verfeinerung in nur einem Bruchteil der Dimensionen erfordert (dimensionsadaptiv). Darüber hinaus zeigen die kinetischen Daten Charakteristiken scharfer Übergänge zwischen "nicht aktiven" und "aktiven" Bereichen, die eine höhere Verfeinerung (lokal-adaptiv) erfordern. Die Effizienz des adaptiven SG wird an verschiedenen Testmodellen und einem realistischen First-Principle-Modell, einschließlich der Sensitivitätsanalyse, getestet. Die Ergebnisse zeigen, dass für katalytische Modelle eine lokale Sensitivitätsanalyse auf Basis lokaler Ableitungen nur begrenzte Informationen liefert. Dagegen kann der globale Ansatz die wichtigen Parameter identifizieren und ermöglicht es, Informationen aus komplexeren Modellen detaillierter zu extrahieren.

Der Sparse Grid-Ansatz reduziert die Gesamtzahl an Punkten, aber was ist, wenn die Auswertung eines Punktes schon sehr teuer ist? Deswegen konzentriert sich der zweite Teil dieser Arbeit auf die Lösung hochdimensionaler Integrale für Modelle, deren Auswertungen nur implizit, z.B. durch ein Monte-Carlo-Modell, gegeben ist. Wir erweitern die SG-Methode um einen mehrstufigen Ansatz, der die Kosten senken soll. Im Gegensatz zu bestehenden Ansätzen werden wir nicht die Teleskopsumme verwenden, sondern die intrinsisch gegebene hierarchische Struktur des SG ausnutzen. Jede Funktionsauswertung enthält einen Fehler, aufgrund einer begrenzten Probenmenge, aber nicht alle SG-Punkte benötigen die gleiche Genauigkeit. Deswegen können wir bei jedem Verfeinerungsschritt die Varianz der Punkte verdoppeln und somit die Menge der gezogenen Stichproben halbieren und die Kosten minimieren. Wir demonstrieren die Methodik an verschiedenen Testmodellen und einem realistischen kinetischen Monte-Carlo-Modell. Dabei vergleichen wir den reinen adaptiven Sparse Grid (ASG) Ansatz mit dem Multilevel Adaptive Sparse Grid (MLASG). Die Ergebnisse zeigen, dass wir mit der mehrstufigen Erweiterung im Vergleich zur ASG, bis zu zwei Größenordnungen an CPU (Central Processing Unit)- Zeit einsparen können, ohne die Genauigkeit des Ersatzmodells zu beeinflussen.

Eidesstattliche Erklärung

Ich versichere hiermit an Eides statt, dass diese Arbeit von niemand anderem als meiner Person verfasst worden ist. Alle verwendeten Hilfsmittel wie Berichte, Bücher, Internetseiten oder ähnliches sind im Literaturverzeichnis angegeben. Zitate aus fremden Arbeiten sind als solche kenntlich gemacht. Die Arbeit wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfungskommission vorgelegt und auch nicht veröffentlicht.

Berlin, den 08. August 2021

Sandra Döpking

10 Acknowledgements

I would like to take this opportunity to express my gratitude to all those who encouraged me to write this thesis and supported me along the way. First and foremost, my thanks goes to my supervisor Sebastian Matera for his continuous support and guidance as well as giving me the opportunity to write this thesis. Second, I would like to direct my gratitude to Professor Klein and his group for the interesting discussion and the valuable input. I am specially indebted to my office colleagues, Gottfried Hastermann, Tom Dörfel and Patrick Gelß for the creative input and your warm, friendly and productive atmosphere. Without you the work would only have been halve the fun. I highly appreciate the invaluable effort of Ulrike Eikers to help me out and keep the wheel turning.

Special thanks should be given to Professor Reuter, his colleague Craig Plaisance and his group for providing the kinetic models, my work is based on, and the constant interactions and the valuable advice. Finally I would like to thank my family, especially my mother, and Markus Sinstein, who I am deeply grateful for. This work would have not been possible without their support, encouragement and honesty. Thank you for keeping me motivated.

This research has been funded by the Matheon, the Einstein Center and the math+ (DFG).