Optimal Experimental Design for Parameter Identification and Model Selection

Dissertation

zur Erlangung des akademischen Grades

Doktoringenieur
(Dr.-Ing.)

von Dipl.-Ing. René Schenkendorf
geb. am 23. April 1982 in Magdeburg

genehmigt durch die Fakultät für Elektrotechnik und Informationstechnik
der Otto-von-Guericke-Universität Magdeburg

Gutachter:

Prof. Dr. Michael Mangold
Prof. Dr. Achim Kienle
Prof. Dr. Rudibert King

Promotionskolloquium am 16.05.2014

Abstract

The main focus of this thesis lies on elementary steps of Optimal Experimental Design (OED). In general, OED aims to determine operating conditions which are expected to provide informative measurement data. Here, the term *informative* depends on the intended task. In case of parameter identification problems, informative data correlate with precise parameter estimates and reliable simulation results, respectively. On the other hand, another objective in the framework of modelling is to select the most plausible model candidate from a pool of various model candidates/hypotheses. In this field, informative data are associated with measurements which facilitate the actual model selection process.

Indeed, these two strategies might be different in their outcomes, but both depend critically on the credibility of the applied algorithm for uncertainty propagation. Therefore, the Unscented Transformation (UT) approach as an alternative to standard approaches of uncertainty propagation is reviewed in detail. It is demonstrated that the UT method outperforms the linearisation concept in precision while utilising a low level of computational load compared to Monte Carlo simulations. In practice, when applied to OED problems for parameter identification the UT approach contributes to the overall performance beneficially.

Moreover, in case of model selection issues the UT method as part of the Unscented Kalman Filter enables an online model selection routine. That means, in parallel to the experimental run the operating conditions are optimised simultaneously. In doing so, the process of model selection becomes more robust against a potential poor initial guess of initial conditions and/or estimates of model parameters.

Finally, the concept of the flat input based parameter identification is introduced. It is shown, that by evaluating cost functions based on flat inputs instead of simulation results the parameter identification routine can

be speeded up significantly. This effect is achieved by replacing the cpu-intensive numerical integration algorithms for solving the underlying model equations by a less computationally cumbersome differentiation of surrogate functions. By analysing the associated cost functions it is illustrated that the flat input based expressions are likely to be more suitable candidates for a proper parameter identification, i.e., they may possess less local minima in comparison to the standard approach and, additionally, they are independent of the initial conditions. The general relation of the flat input concept to OED is given by a closer look at parameter sensitivities.

Kurzfassung

In der vorliegenden Arbeit werden wesentliche Aspekte aus dem Bereich der sogenannten Optimalen Versuchsplanung betrachtet und weiterentwickelt. Im Allgemeinen zielt die Optimale Versuchsplanung darauf ab, experimentelle Operationsbedingungen zu bestimmen, von denen man erwartet, informative Messdaten zu erzeugen. Hierbei lassen sich folgende Fälle unterscheiden. Im Bereich der Parameterbestimmung werden Messdaten als informativ angesehen, wenn sie möglichst präzise Parameterschätzungen ermöglichen. Auf der anderen Seite kann die Optimale Versuchsplanung auch darauf abzielen, den geeignetsten Modellkandidaten aus einem ganzen Satz möglicher Modellansätze/-hypothesen zu wählen. Folglich werden gewonnene Messdaten als informativ betrachtet, wenn sie die Modellselektion erleichtern.

Die Ziele als auch die ermittelten Operationsbedingungen beider Varianten sind somit durchaus verschieden. Beiden Ansätzen ist jedoch gemein, dass der Erfolg der Optimierung maßgeblich von der zuverlässigen Beschreibung und dem genauen Propagieren von (Mess)Unsicherheiten abhängt. Um dieser Anforderung gerecht zu werden, wurde das Konzept der "Unscented Transformation" (UT) auf ihre Eignung für die Optimalen Versuchsplanung geprüft. Es wird gezeigt, dass der UT Ansatz dem Standardkonzept basierend auf Linearisierung in puncto Genauigkeit überlegen ist. Weiterhin wird demonstriert, dass der UT Ansatz im Vergleich zur Monte Carlo Simulation mit einem geringeren Rechenaufwand auskommt. Somit liefert die UT Methode im Bereich der Optimalen Versuchsplanung eine praktikable als auch genauere Berücksichtigung von (Mess)Unsicherheiten. Die resultierenden Optimalen Versuchsplanungen sind somit eher in der Lage, informative Messdaten zu generieren.

Darüber hinaus wird aufgezeigt, wie der UT Ansatz als integraler Bestandteil des Unscented Kalman Filters genutzt werden kann, um einen onlinefähigen Algorithmus zur Modellselektion zu implementieren. In der Praxis

bedeutet dies, dass parallel zur eigentlichen Versuchsdurchführung die Operationsbedingungen optimiert werden. Hierdurch wird der Selektionsprozess robuster gegenüber ungenauen Anfangsbedingungen bzw. schlecht bestimmten Modellparametern.
Des Weiteren wird das Konzept der sogenannten flachen Eingänge zum Zweck der Parameterbestimmung implementiert. Gütefunktionale, welche mit geeigneten Parameteroptimierungsverfahren ausgewertet werden, sind hierbei Funktionen der flachen Eingänge. Im Vergleich zu dem Standardansatz, d.h. dem Auswerten des Gütefunktionals basierend auf Simulationsergebnissen, kann eine deutliche Rechenzeitreduktion festgestellt werden, da der rechenaufwendige Integrationsschritt der Modellgleichungen durch einen weniger aufwendigen Differentiationsschritt von Ersatzfunktionen ersetzt wird. Eine genauere Analyse der Gütefunktionale zeigt darüber hinaus, dass der vorgestellte Ansatz im Bereich der Parameteridentifizierung weitere vorteilhafte Eigenschaften liefert. So ist an einem Beispiel verdeutlicht, dass die Auswertung des Gütefunktionals basierend auf den flachen Eingängen im Vergleich zum Standardansatz zu weniger lokalen Minima führen kann. Den Zusammenhang zwischen der Verwendung von flachen Eingängen und der Optimalen Versuchsplanung wird anhand von Parametersensitivitäten demonstriert.

Acknowledgements

Above all I want to thank my supervisor Prof. Dr. Michael Mangold. He has taught me to get not lost in the vast, (for me) unexplored ocean of scientific literature in the field of optimal experimental design. His guidance, understanding, well considered ideas, and patience during my time as a research assistant at the Max Planck Institute for Dynamics of Complex Technical Systems in Magdeburg are the essentials for finishing my thesis successfully. He has been a true mentor - not only for my professional life.
I am also very grateful to Prof. Dr. Achim Kienle who cordially welcomed me as a member in his group, Process Synthesis and Process Dynamics, at the MPI. He has proved a recipe for the excellent working conditions at the MPI as well as for the fruitful connection to the Otto-von-Guericke-University Magdeburg.
I owe a very important debt to Prof. Dr. Rudibert King for acting as a referee for this thesis. The thesis greatly benefited from his detailed comments.
Moreover, many people have contributed to my personal and professional time at the MPI. Here, I would like to specially thank André Franz, Katharina Holstein, Robert Flassig, Anke Ryll, Philipp Erdrich, and Andreas Kremling for their support and encouragement.
Finally, I want to thank my family for their constant support. I am deeply grateful to Beatrice, who provided me continual motivation and encouragement during all these years. I dedicate this thesis to my two little girls Annabell and Pauline, for reminding me in the best of possible ways that scientific work is important - but not everything in life.

Contents

List of Figures

List of Tables

Glossary

Θ	domain of θ
$\hat{\theta}$	estimated model parameters
$\hat{\mathcal{S}}_j$	model structure of j'th model candidate
θ	model parameters
$(\cdot)$	place holder
α	scale factor of UT
$\bar{(\cdot)}$	expected value of $(\cdot)$
Δ_i	difference of AIC_i^c and AIC_{min}^c
ϵ	white noise
η	random vector
η_i	sample of random vector space
$\gamma(x(t))$	input vector field
$\hat{(\cdot)}$	approximation / estimation of $(\cdot)$
$\hat{x}_{k+1}^+$	corrected state variables
$\hat{x}_{k+1}^-$	predicted state variables
κ	scale factor of UT
λ	scale factor of UT
$\mathcal{D}^i$	differential operator
$\mathcal{Q}(\cdot)$	observability matrix
$\mathcal{S}$	system (physical process)
$\mathcal{W}$	Akaike weights
Ω	integration domain
$\Phi(t)$	basis functions
Π	model probabilities
$\sigma_{(\cdot)}^2$	variance of a quantity $(\cdot)$
Ψ_i	basis functions for PCE
υ	weighting parameter
$\vartheta(\cdot)$	scaling function
ξ	random vector
ξ_i	sample of random vector space
A^*	state adjacency matrix
AIC^c	corrected Akaike Information Criterion
AIC_{min}^c	value of the least complex candidate
Bi	bias
c	coefficients of surrogate functions
C^*	output adjacency matrix
$C_{(\cdot)}$	(co)variance matrix
CI	99% confidence interval of y^{data}
D	cost function of online OED
$E[(\cdot)]$	expected value of a quantity $(\cdot)$
$g(\cdot)$	(non)linear function
$GF[\cdot]$	generator function
$J_{(\cdot)}$	cost function for $(\cdot)$
K_I	Kullback's total measure of information
K_{k+1}	correction matrix
MSE	mean square error matrix
$O_{(\cdot)}$	overlap
P_{k+1}^+	corrected (co)variance matrix of state variables
P_{k+1}^-	predicted (co)variance matrix of state variables
Q	positive definite, diagonal matrix of process noise
R	positive definite, diagonal matrix of measurement noise
r_k	measurement residuals
S_i	Sobol' indices of first order
SE	Shannon's entropy
SM	sensitivity matrix
SR_{k+1}	(co)variance matrix of residuals

GLOSSARY

SSE	sum of squared errors
t	continuous time
t^{cpu}	cpu-time
t_k	discrete time
$TS1$	First-order Taylor Series Expansion
$TS2$	Second-order Taylor Series Expansion
u	input / stimulus
$U(\cdot)$	uncertainty of $(\cdot)$
$u^{flat/fict}$	flat inputs that are pure fictitious
$u^{flat/real}$	flat inputs that exist in practice
u^{flat}	flat inputs
w_i	weights of UT
x	state variables
y^{data}	measured quantities
y^{sim}	simulation result
y^{surr}	surrogate function
DDE	delay differential equation
EKF	extended Kalman filter
FIM	Fisher information matrix
KPI	key performance index
MC	Monte Carlo
ODE	ordinary differential equation
OED	optimal experimental design
PCE	polynomial chaos expansion
PI	parameter identification
TGF-β	transforming growth factor-beta
TS1	1. order Taylor Series Expansion
TS2	2. order Taylor Series Expansion
UKF	unscented Kalman filter
UT	unscented transformation
UT3	UT taking account of monomials of degree 3 or less
UT5	UT taking account of monomials of degree 5 or less

1

Introduction

1.1 Research Motivation

The utilisation of mathematical models to analyse complex systems is quite common in various research fields. One essential step in the process of model development is the parameter identification (PI) routine, i.e., the determination of unknown model parameters by minimising the difference between simulation results and measurement data. However, even by assuming the quite unrealistic case of perfect measurement data, i.e., the measurements are provided continuously without any measurement noise, the actual parameter identification process might be challenging. For example, different parameter configurations may result in similar simulation outcomes. That is, a unique set of parameter estimates cannot be warranted in principle. In this case, the underlying parameter identification problem is ill-posed and may only be solved properly by model reformulation or by adding different measurement quantities. In literature, this problem is known as theoretical/structural parameter identifiability analysis [WP97]. The situation gets worse by addressing measurement imperfections. In practice, measurement data are limited to discrete sample time points and, additionally, the data are corrupted by measurement noise. Indeed, these flaws might be compensated for by sophisticated measurement devices - but not entirely:

> “Because I had worked in the closest possible way with physicists and engineers, I knew that our data can never be precise.
> *Norbert Wiener (November 26, 1894 - March 18, 1964) [Lev66]* ”

Hence, the measurement data samples have to be treated as random variables instead of deterministic quantities. In consequence, the uncertainty about the measurement data has a serious impact on the identified model parameters. In the worst case, the

parameter identification routine may strongly amplify the uncertainty of the measurement data to some parameter estimates. Thus, even in case that the parameters are identifiable in principle the estimates might be useless due to an extreme parameter uncertainty. A resulting mathematical model may suffer in prediction. That means, although the model captures the essential features of the measurement data which have been applied to the parameter identification step, the parametrised model diverges at operating conditions which have not been part of the preceding parameter identification step. In short, the parametrised model cannot be generalised credibly.

The situation gets worse additionally by taking into account the model structure uncertainty. In general, any model imitates the described process approximately by modelling the most essential steps of interest but ignoring unimportant details. The universal dilemma in modelling can be summarised by the quotation given below.

> “Since all models are wrong the scientist cannot obtain a “correct” one by excessive elaboration. On the contrary following William of Occam he should seek an economical description of natural phenomena.
> *George E. P. Box (October 18, 1919 - March 28, 2013) [Box76]* ”

By reading the previous sentences one might get the impression that modelling is not worth worrying about, because any derived model will be just a crude approximation of the real process and potential simulation results are likely to suffer in precision, too. Nonetheless, in practice the last decades have shown: Modelling pays! Here, the rapidly evolving research field of system biology confirms this statement impressively. In spite of the potential pitfalls in modelling, mathematical models have demonstrated to be a versatile tool in modern biology. The following research achievement is selected to demonstrate the success of model-based systems biology substitutionally. In [CCFCO12], a dynamical model has revealed a plausible explanation to the so-called TGF-β paradox. The TGF-β, a cell process regulator, is known to be a *tumour suppressor* in mammalian cells. Measurement data analyses, however, have indicated a strong correlation between the severity of cancer and an increased level of TGF-β, i.e., cancer patients with the worst prognosis have the highest level of TGF-β. Thus, TGF-β has been considered as a potential cause of cancer, i.e., to be a *tumour promoter* in cancer cells - an obvious contradiction to its tumour suppressor role. The derived dynamical model reveals a more plausible explanation: tumour cells become gradually insensitive to TGF-β. To “steer” a cancer cell, therefore, needs an increased level of TGF-β. By

implication, TGF-β is not the cause but an effect of cancer - a conclusion which can be hardly drawn by pure data analysis. The result of this study can be generalised as follows: the dynamic interplay of essential elements in living systems can be captured at best by the use of dynamic models.

As a consequence, advanced branches of systems biology, e.g., synthetic biology [GvO09, CL12, RMS12] and systems medicine [WAJ+13, CS13], employ mathematical models to a large extend, too. Thus, to ensure meaningful model-based inferences the imperfect measurement data have to be utilised in the most efficient way. An experimental run has to be performed at operating conditions which provide the most informative data by evaluating Optimal Experimental Design strategies. In doing so, there is a fair chance that the uncertainty of model parameters and simulation results can be reduced to a reasonable level. In conclusion, Optimal Experimental Design is an essential component in the framework of model development and the major topic of this thesis.

1.2 Optimal Experimental Design

In the field of systems biology mathematics is of vital importance. For instance, mathematical algorithms are put in operation to prepare raw measurement data (data preprocessing), to test different hypotheses statistically, or to describe/explain analysed processes by dynamical models. Thus, mathematical concepts aim to enrich our understanding of real world phenomena:

> “People who wish to analyze nature without using mathematics must settle for a reduced understanding.
> *Richard P. Feynman (May 11, 1918 - February 15, 1988) [Fey13]* ”

In this thesis, the focus is on dynamical models solely. In detail, models based on ordinary differential equations (ODEs) are evaluated. In systems biology literature, ODEs are widely applied to extend or to confirm our knowledge of metabolic, genetic, and/or signalling processes in biological systems. Clearly, model-based inferences are only meaningful in case of predictive models. Thus, the overall credibility of simulation results has to be assessed systematically, and has to be improved by Optimal Experimental Design (OED) in case of need. Generally, one distinguishes between OED for parameter identification and OED for model selection. Both strategies are explained in more detail below.

1.2.1 OED for Parameter Identification

In practice, measurement data are affected by measurement noise which leads to some uncertainty of parameter estimates [WP97]. This fact has a strong impact on the model quality, because only parameter estimates with small variances are likely to ensure simulation results with a highly predictive power, see Fig. 1.1 for illustration purposes. Therefore, to avoid non-predictive / non-meaningful models, informative measurement data have to be utilised in the process of parameter identification. Here, Optimal Experimental Design (OED) comes into play. OED for parameter identification aims to identify operating conditions of experiments which are expected to ensure the most informative data. The general basics of OED in the field of system biology can be found in [PP07, Pro08, FM08, KT09, AWWT10] and references therein.

In general terms, OED consists of two sequential steps: (i) to determine the statistics of parameter estimates which are the result of a parameter identification algorithm by evaluating available experimental data, and (ii) to provide strategies to minimise these parameter uncertainty. That is, to adapt future experiments for the purpose of informative data. These steps have to be reiterated until a sufficient parameter accuracy is achieved. The overall OED performance depends on the quality of the determined parameter statistics. For instance, a frequently applied method in parameter statistics is based on the inverse of the Fisher Information Matrix (FIM) and the Cramer-Raó inequality [Kay93, Pro08, FM08]. In case of Gaussian noise and linear model parameters the FIM provides correct results of the parameter variances [Kay93]. Most applied models in systems biology, however, are non-linear in their parameters. This, in turn, can cause that the FIM may lead to poor approximations of the parameter statistics. To overcome this limitation, various methods have been developed to improve the calculation of the associated parameter uncertainties. In the majority of cases these approaches are based on Monte Carlo simulations [JSMK06]. This kind of implementation, however, has an increased computational load, which tends to prohibit their use in an iterative process like OED. That is, the parameter statistics has to be determined several times as sub-part of an optimisation routine. In the field of OED there is a strong need of reliable uncertainty approximation approaches which ensure a manageable computational load additionally. For this purpose, the Unscented Transformation (UT) is presented as a promising and versatile alternative. Additionally, it is demonstrated how the UT approach can be applied for the purpose of a parameter sensitivity analysis. That is, to figure out those parameters which have the strongest impact to

simulation results. Moreover, the UT method is put in operation to determine the uncertainty of simulation results explicitly. In doing so, the simulation uncertainties can be incorporated in OED in a more credible fashion.

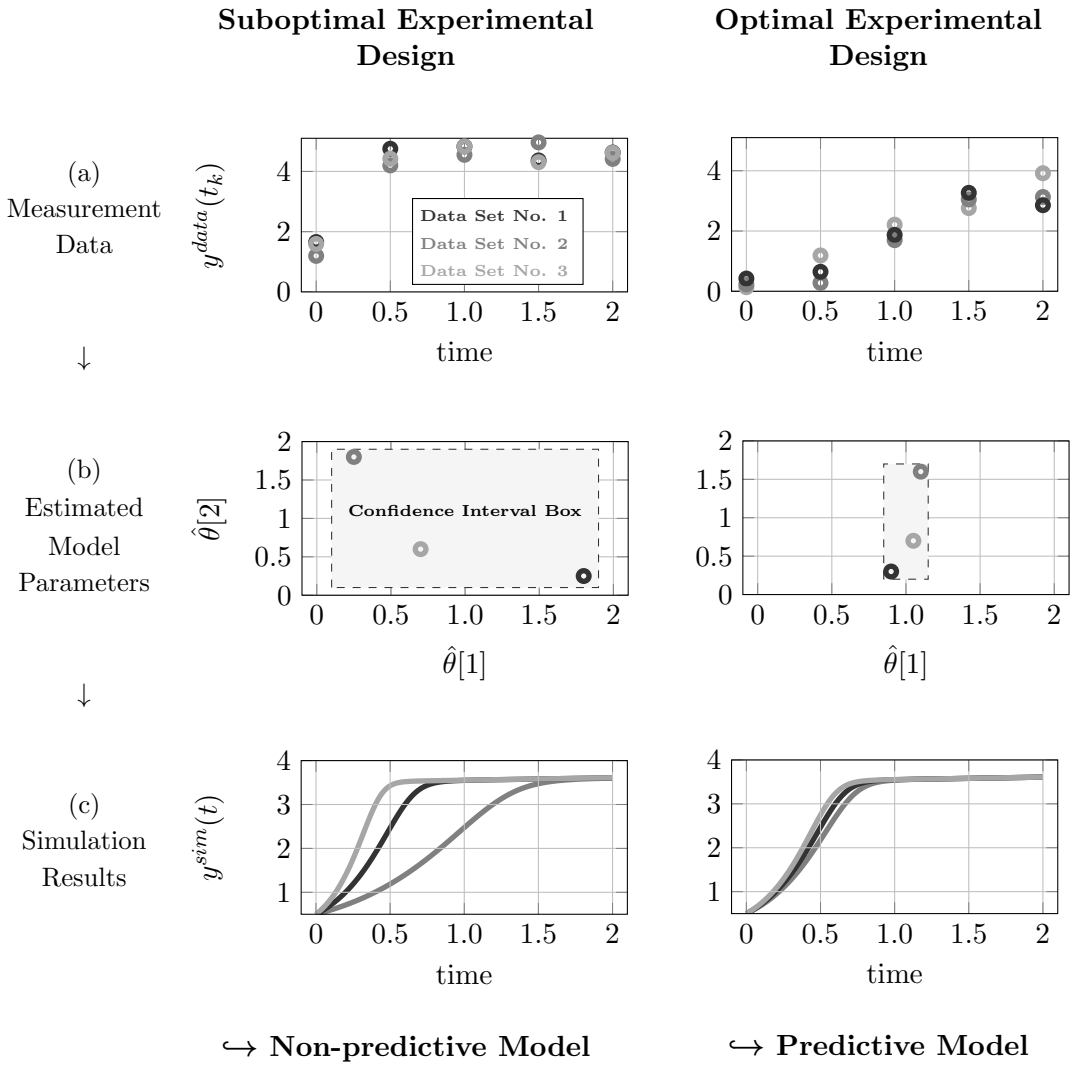


Figure 1.1: Suboptimal Experimental Design vs. Optimal Experimental Design. From top to the bottom, the flowchart of model parametrisation is illustrated according to: (a) the generation of measurement data by experimental runs, (b) the parameter identification step, and (c) the associated simulation results. In (a) the uncertainty of measurements is obvious. For example, experimental runs for a Suboptimal Experimental Design provide slightly different measurement data samples. The same is true for the Optimal Experimental Design. At suboptimal operating conditions, however, the data uncertainty is strongly amplified to the parameter estimates (b) as indicated by the large confidence interval box. Subsequently, the parameter uncertainty may result in a severe variation of simulation results (c) at operating conditions that have not been part of the former parameter identification step. In case of optimal operating conditions, the amplification of uncertainties can be reduced to an acceptable level. That is, measurement data provide the most precise parameter estimates for those parameter candidates which have the strongest impact to simulation results. In general, there is no need to identify all parameters at the same accuracy, see the (b)-row of the Optimal Experimental Design column.

1.2.2 OED for Model Selection

Due to the complexity of the analysed processes mathematical models capture only the essential features of interest. This approximate representation, which is usually combined with a vague knowledge of basic processes, leads in many cases to a variety of potential model candidates which describe the real process almost equally well. To determine the most plausible model candidate is the objective of model selection or model discrimination methods. If at given operating conditions, however, no sufficient discrimination can be achieved, Optimal Experimental Designs (OED) for model selection may provide a remedy.

In detail, OED for the purpose of model selection searches for operating conditions which facilitate the overall selection process. Frequently, the underlying algorithms of OED are based on statistics and/or information theory [LBS94, BA02, MC04, KFG$^+$04, DBMM$^+$11]. These approaches have in common that the measurement data are evaluated in a batch mode. Thus, the measurement data are only processed after the entire experimental run has been finished. From the perspective of robust OED strategy, however, the immediate utilisation of measurement data might be desirable [KAG09]. To address this issue, an online model selection framework is presented in this thesis. Here, the Unscented Kalman Filter (UKF) provides statistical information which is used to assign probability values to every model candidate. These probability values are immediately updated as soon as new measurement data become available. Moreover, during the experimental run the process is steered in a fashion which maximises the differences in these candidates by evaluating a suitable cost function in real time. To overcome limitations caused by parameter uncertainties the most sensitive model parameters are simultaneously estimated in the course of the model selection framework. The combined application of the online framework and the joint estimation of sensitive model parameters provides an optimal usage of measurement data , i.e., the overall number of experiments might be reduced significantly.

1.3 Structure of the Thesis

The structure of the remaining chapters of the thesis is outlined as follows, see Fig. 1.2 for the illustration of the main topics.

Chapter 2 - Quantification and Propagation of Uncertainty

The general problem of uncertainty quantification and propagation is introduced. A short review about the most frequently applied uncertainty propagation methods is the starting point for a detailed look on the Unscented Transformation. That means, the origin of Unscented Transformation, its mathematical basics, as well as the associated assets and drawbacks for the intended task of Optimal Experimental Design are reviewed.

Chapter 3 - Optimal Experimental Design for Parameter Identification

First, the basics about Optimal Experimental Design for the purpose of parameter identification are summarised. It is stressed that the overall OED performance depends on the uncertainty approximation of the quantity of interest, e.g., model parameter and simulation results. Thus, the focus is on the influence of approximation errors on OED results. In this context, the Unscented Transformation approach demonstrates its superior approximation power while keeping the computation demand on a feasible level.

Chapter 4 - Optimal Experimental Design for Model Selection

The challenging problem of model selection is addressed. Statistics about simulation results are of essential relevance, too. Thus, the Unscented Transformation has a key role for the proposed model selection algorithm. Moreover, the measurement data samples are processed in a online fashion. That is, in parallel to the experimental run the operating conditions are optimised simultaneously. In doing so, the overall framework of model selection becomes more robust against a potential poor initial guess of initial conditions and/or estimates of model parameters. In cases where no online measurements can be provided, the Overlap approach is presented as an alternative. Here, too, the Unscented Transformation is a primary ingredient.

Chapter 5 - Flatness Approach for Parameter Identification

Some serious shortcomings of the traditional parameter identification philosophy are discussed. For example, by providing simulation results for any parameter combination which is explored by the parameter optimisation algorithm turns the actual parameter identification step into a computational demanding process. Thus, the concept of flat inputs is introduced to overcome potential flaws of the traditional approach - at least

to a certain extent. On that account, a short recall about the basics of differential flatness is given which provides the link to the concept of flat inputs and their merits for parameter identification issues. Moreover, a comprehensive literature review demonstrates the similarity, as well as the differences to parameter identification approaches which have been derived over the last few decades. The general relation to OED is given by a closer look on the parameter sensitivities. It is shown, that by applying the flatness based approach, there is a change in the spectrum of parameter sensitivities. That is, although the same measurement data - as provided for the traditional approach - are utilised there is a chance that previously insensitive parameters become sensitive by evaluating the flatness concept.

Chapter 6 - Conclusion and Future Work

Chapter 6 provides a summary of the work presented in the thesis and outlines the conclusions that can be drawn, with ideas for future work also being presented.

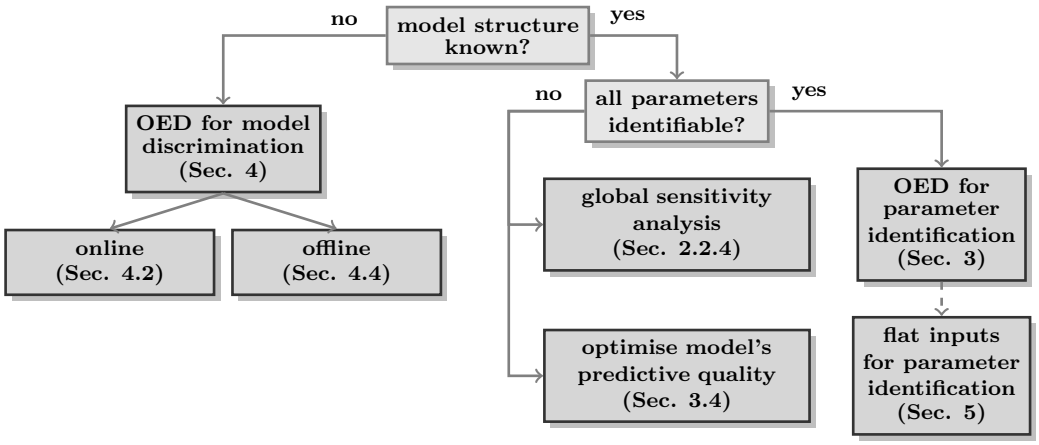


Figure 1.2: The figure shows different aspects of Optimal Experimental Design addressed in this thesis.

1.4 Publications

A part of the work contained in this thesis has already been published in the peer-reviewed literature and presented at numerous conferences. Here is a list of those publications:

1.4.1 Journal Articles

- R. Schenkendorf, A. Kremling, and M. Mangold. Optimal experimental design with the sigma point method. *IET Systems Biology*, 3:10–23, 2009
- R. Schenkendorf, A. Kremling, and M. Mangold. Influence of non-linearity to the optimal experimental design demonstrated by a biological system. *Mathematical and Computer Modelling of Dynamical Systems*, 18:413–426, 2012
- R. Schenkendorf and M. Mangold. Online model selection approach based on unscented kalman filtering. *Journal of Process Control*, 23:44–47, 2013
- R. Schenkendorf and M. Mangold. Parameter identification for ordinary and delay differential equations by using flat inputs. *Theoretical Foundations of Chemical Engineering*, accepted, 2014

1.4.2 Conferences / Proceedings

- R. Schenkendorf, A. Kremling, and M. Mangold. Application of sigma points to the optimal experimental design of a biological system. In *The 20th International Symposium on Chemical Reaction Engineering- Green Chemical Reaction Engineering for a Sustainable Future*, 2008
- R. Schenkendorf, A. Kremling, and M. Mangold. Optimal experimental design and model selection by a sigma point approach. In *Mathmod 2009 - 6th Vienna International Conference on Mathematical Modelling, Vienna, Austria*, 2009
- R. Schenkendorf and M. Mangold. Qualitative and quantitative optimal experimental design for parameter identification. In *18th IFAC World Congress Milano (Italy)*, 2011
- R. Schenkendorf and M. Mangold. Challenges of parameter identification for nonlinear biological and chemical systems. In *SIAM Conference on Optimization, Darmstadt*, 2011
- M. Mangold and R. Schenkendorf. Methods for optimal experimental design of nonlinear systems using sigma points. In *Dechema-Regionalkolloquium: "Design of Optimal Experiments", Magdeburg*, 2012
- R. Schenkendorf, U. Reichl, and M. Mangold. Parameter identification of time-delay systems: A flatness based approach. In *MATHMOD 2012 - Full Paper Preprint Volume*, 2012

2

Quantification and Propagation of Uncertainties

Mathematical models are the workhorse in systems biology. In what follows, ordinary differential equation (ODE) systems are of current interest. In its generic form ODE systems can be expressed by

$$\dot{x}(t) = f(x(t), u(t), \theta) \quad ; x \in \mathbb{R}^n \, , u \in \mathbb{R}^s \, , \theta \in \mathbb{R}^l \tag{2.1}$$

$$y^{sim}(t) = h(x(t)) \quad ; y^{sim} \in \mathbb{R}^m \tag{2.2}$$

$$x(t=0) = x_0 \quad ; x_0 \in \mathbb{R}^n, \tag{2.3}$$

where $x(t)$ is the state vector, $u(t)$ is the input vector, θ is the parameter vector, y^{sim} the output vector, and x_0 the initial conditions. The two vector functions, $f(\cdot)$ and $h(\cdot)$, are known as the state and the output function, respectively. At the very first step in model building a feasible model structure as well as model parameters have to be derived with the help of measurement data, $y^{data}(t_k)$. Thus, all potential benefits of model-based approaches are likely to fail in the light of non-informative data. To provide predictive models the available measurement data sets have to be utilised in the most informative way. In the absence of informative data at all, it might be necessary to run new optimally designed experiments.

Naturally, the question comes up how to quantify the information content of measurement data properly. Starting from a pure data analysis perspective the signal to noise ratio is of crucial interest, i.e., at best the measurement data should be corrupted by a minimum of measurement noise. Generally, the signal to noise ratio is determined by the measurement equipment, i.e., the measurement principle as well as the measurement range might influence the quality of measurement data.

In the context of modelling, however, the measurement data are the driving force of model parametrisation. That means, the data, $y^{data}(t_k)$, are not of primary interest but the identified model parameters, $\hat{\theta}$. Obviously, a low signal to noise ration of measurement data does not automatically ensure informative data for the purpose of parameter identification. Thus, data which provide precise parameter estimates are considered as informative.

As modelling aims to provide reliable inferences, even model parameters are not the ultimate goal. Generally, simulation results are of primary interest and have to be predictive as in any way possible. Consequently, by assuming a proper model structure measurement data are labelled as informative if they ensure a low uncertainty of simulation results.

For the sake of completeness, in some model-based applications even the simulation results are not analysed in detail. For instance, one might be interested in an overall performance measure of the process at hand. Thus, a so-called key performance index (KPI) might be of primary interest. In this particular case, measurement data are considered as informative if these data provide a low uncertainty about the KPI [KAGSW08].

In summary, after a first quantification of the present uncertainty about measurement data, this uncertainty is propagated sequentially up to the quantity of interest as illustrated in Fig. 2.1. For this purpose, probability theory provides a comprehensive framework which, however, may suffer in practicability and/or precision in the presence of non-linearity. Thus, in the following subsections suitable methods of uncertainty quantification, as well as propagation are presented and applied to overcome urgent problems in uncertainty analysis of non-linear systems. The proper use and interpretation of the resulting statistical quantities cannot be addressed comprehensively in this work. Therefore, the reader is referred to [SG01, CF05, Lan05] and references therein.

2.1 Quantification of Data Uncertainty

To a certain extent, variability is present in almost all physical systems. The quantification of this uncertainty, as well as its proper representation might be challenging by itself [OO04b]. When focusing on measurement data, $y^{data}(t_k)$, by way of example, exact values of measured quantities cannot be derived in practice, because of limitations in the measurement equipment or because of the inherent variability (process noise)

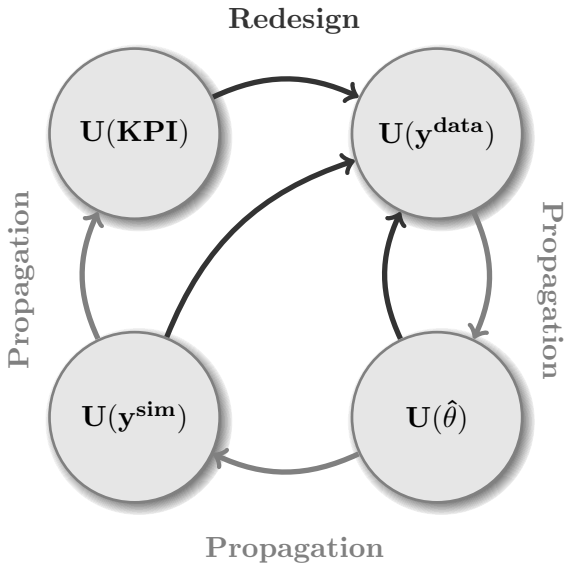


Figure 2.1: Quantification and Propagation of Uncertainties: By starting with the measurement data the associated measurement uncertainty, $U(y^{data})$, is sequentially propagated ($\longrightarrow$) to the model parameters, $U(\hat{\theta})$, the simulation results, $U(y^{sim})$, and the key performance index, $U(KPI)$. A credible quantification of these uncertainties is mandatory for the success of succeeding OED strategies, as the determined uncertainties give feedback, ($\longrightarrow$), for new informative measurement data.

[Ste94] of the system under study. Consequently, the uncertainty in measurements has to be addressed explicitly and its effect to the estimated model parameters, $\hat{\theta}$, and/or simulation results, $y^{sim}(t)$, has to be investigated adequately. Hence, a measurement data sample is not treated as a deterministic scalar value, but considered as a random variable instead. In this case, the variability of the measurement data can be described by a probability density function (PDF). In what follows, the probabilistic description of uncertainty is exclusively applied. Alternative techniques such as interval mathematics or fuzzy set theory are not subject of this work, but might be an alternative in case of missing descriptive PDFs. The presented methods of uncertainty propagation, however, are applicable in hybrid approaches as well, i.e., the joined analysis of probability distributions in combination with so-called possibility distributions [BD06, BPZ$^+$11] can be addressed in principle.

The immediate incorporation of the statistical information given by associated PDFs leads to meaningful results in modelling from the very first beginning [RvdSHV07]. In doing so, uncertainty analysis becomes an integral part in the modelling framework ensuring...

> “...that uncertainty assessment is not just something to be added after the completion of the modelling work. Instead uncertainty should be seen as a red thread throughout the modelling study starting from the very beginning.
> *J. C. Refsgaard [RvdSHV07]* ”

Thus, a main objective of this thesis is devoted to provide a general approach of uncertainty analysis. But before starting with the problem of uncertainty propagation, the question has to be addressed, how the probability density function of the measurement data can be derived. For the sake of simplicity, additive normally distributed uncorrelated noise is assumed at first. That means, the Gaussian probability density function describes the uncertainty of a measurement data sample by its two parameters uniquely, i.e., the mean value, $\bar{y}^{data}(t_k)$, and its associated variance, $\sigma^2_{y^{data}}(t_k)$, are utilised according to

$$y^{data}(t_k) \sim \mathcal{N}(\bar{y}^{data}(t_k), \sigma^2_{y^{data}}(t_k)) \tag{2.4}$$

In addition, the uncertainty of sets of data samples is considered as identical and independent. Thus, the uncertainty analysis is limited to independent and identically distributed (iid) random variables. The extension to more general random variables is presented in Sec. 2.2.3.2 separately. In practice, for a credible quantification of $\bar{y}^{data}(t_k)$ and $\sigma^2_{y^{data}}(t_k)$ several measurement data sets have to be generated, i.e., an experiment has to be repeated several times at the same operating condition. By taking monetary and time limitations into account the number of experimental runs has to be low. Therefore, re-sampling approaches such as the jackknife method or the bootstrap approach [Efr82, ET94] provide an alternative in comparison to an increased number of costly experimental reruns.
The actual identification of $\bar{y}^{data}(t_k)$ and $\sigma^2_{y^{data}}(t_k)$, however, is not subject of this work. In what follows, artificial measurement data are used instead, i.e., the Gaussian PDF of measurement data is known by definition. Thus, after the measurement uncertainty is given, the emphasis is much more on the propagation of uncertainty to the estimated model parameters, $\hat{\theta}$, and/or to simulation results, $y^{sim}(t)$, as presented in the next subsection.

2.2 Mapping of Uncertainty

This section aims to provide an overview of methods which are frequently applied in uncertainty propagation. To be more specific, the applicability of different approaches

for the intended task of Optimal Experimental Design is assessed. In doing so, it is assumed that a plausible model structure, $\hat{S}$, has already been derived, i.e., there is no uncertainty about the model structure itself. Thus, the uncertainty analysis is performed on noisy measurement data, $y^{data}(t_k)$, which induce variability in the estimated model parameters, $\hat{\theta}$, and in simulation results, $y^{sim}(t)$, respectively. Hence, $\hat{\theta}$ and $y^{sim}(t)$ can be considered as random variables as well [Bre70], i.e., the problem of uncertainty propagation exclusively acts in the probabilistic framework.
Consequently, there is a keen demand in a reliable determination of PDFs which are associated to model parameters, pdf_θ, and/or simulations results, $pdf_{y^{sim}}$. Moreover, as an inherent part of OED and for that reason part of an optimisation routine, the applied methods have to be low in computational cost. In general, the approaches of uncertainty propagation can be distinguished in analytical and in approximate methods, respectively. Though the analytical approach might be suitable to illustrate the general problem of uncertainty propagation for well chosen problems, it suffers from practical applicability. Thus, efficient as well as reliable approximate methods are of vital importance in OED as illustrated in Sec. 3. Firstly, however, the focus is solely on uncertainty propagation.

2.2.1 Analytical Expressions

In general, the uncertainty propagation describes how a random variable, ξ, is transferred by a (non)linear function, $g(\cdot)$, to the quantity of interest, η, according to

$$\eta = g(\xi) \tag{2.5}$$

One possible way to represent the uncertainty about η consists in calculating the associated probability density function, pdf_η. Assuming a monotonic function, $g(\cdot)$, an analytical solution of the resulting PDF can be derived immediately [Bre70, HMGB03b]

$$pdf_\eta = pdf_\xi \left(g^{-1}(\eta)\right) \left| \frac{dg^{-1}(\eta)}{d\eta} \right| \tag{2.6}$$

Any non-monotonic function has to be split up into monotonic sub-parts that are transferred separately [Bre70, HMGB03b].
Another point of interest might be in characteristic quantities of the associated PDF, i.e, statistical moments of pdf_η can be used as an alternative to characterise the induced uncertainty about η [Kay93, HMGB03b]. For instance, the mean, $E\left[g(\xi)\right]$, and the

related variance, σ_η^2, are frequently analysed and can be determined by

$$E\left[g(\xi)\right] = \int_\Omega g(\xi) pdf_\xi d\xi \tag{2.7}$$

$$\sigma_\eta^2 = \int_\Omega \left(g(\xi) - E\left[g(\xi)\right]\right)^2 pdf_\xi d\xi \tag{2.8}$$

Here, Ω represents the integration domain, i.e., in case of probability theory it is equivalent to the sample space [MK10]. Throughout this work, also higher central moments are applied, e.g., the skewness, μ_3, and the kurtosis, μ_4, are considered as well and expressed by

$$\mu_3 = \int_\Omega \left(g(\xi) - E\left[g(\xi)\right]\right)^3 pdf_\xi d\xi \tag{2.9}$$

$$\mu_4 = \int_\Omega \left(g(\xi) - E\left[g(\xi)\right]\right)^4 pdf_\xi d\xi \tag{2.10}$$

At this point it has to be stressed that the presented analytical solutions of the PDF and/or central moments of η can be solved only for a very limited number of uncertainty propagation problems [Bre70, HMGB03b]. In most practical cases, however, approximate methods based on Taylor series expansion, Gaussian quadrature, or Monte Carlo simulations are implemented and reviewed subsequently.

2.2.2 Basic Approaches in Approximate Methods

Commonly, the complexity of $g(\cdot)$ - if at all available explicitly - prohibits results in closed-form. Thus, the approximate methods aim: (1) to replace $g(\cdot)$ by handy surrogates, $\hat{g}(\cdot)$, which facilitate closed-form solutions of Eq. (2.6)-(2.10). Or alternatively (2), to solve these integral expressions by numerical routines approximately. In literature a vast number of approximate methods exist. Here, by being aware of the intended application in OED the merits and flaws of frequently used approaches are highlighted.

2.2.2.1 Taylor Series Expansion

To solve equations similar to Eq. (2.6)-(2.10) in closed-form the mapping function, $g(\cdot)$, is approximated by a surrogate function, $\hat{g}(\cdot)$, first. Here, the most common approach is known as the Taylor series expansion. Under the assumption that $g(\cdot)$ is sufficiently differentiable, the uncertainty propagation function can be expressed by a superposition of Taylor terms:

$$\eta \approx \hat{g}(\xi) = \sum_{i=0}^{N} \left.\frac{\partial^i g}{\partial \xi^i}\right|_{\xi=E[\xi]} \frac{(\xi - E[\xi])^i}{i!} \tag{2.11}$$

Generally, this sum is limited to a certain extent, $N << \infty$, which may introduce an approximation error but ensures a speedy computation. In the field of uncertainty propagation, therefore, the first-order Taylor expansion is applied as a standard approach. For instance, first-order Taylor principles are part of the Fisher-Information matrix (FIM) in the field of parameter statistics, as well as an inherent element in the Extended Kalman Filter (EKF) approach performing state reconstruction. Details about FIM as well as EKF are given later on. Firstly, however, the general calculation of statistical moments via the Taylor series is addressed.

According to Eq. (2.5), the first-order Taylor series approximation is expanded at $\overline{\xi} = E[\xi]$ as shown below

$$\eta \approx \hat{\eta} = g(\overline{\xi}) + \left.\frac{\partial g}{\partial \xi}\right|_{\xi=\overline{\xi}} (\xi - \overline{\xi}) \tag{2.12}$$

Here, the resulting function, $\hat{\eta}$, acts as a surrogate of the original function, η. Now, by evaluating $\hat{\eta}$ instead of η, the determination of statistical moments can be performed easily. For instance, the resulting mean $E[\hat{\eta}]$ is expressed by

$$E[\hat{\eta}] = g(E[\xi]) \tag{2.13}$$

In addition, still assuming that ξ represents a Gaussian random variable, the expectation of the squared difference of Eqs. (2.12) and (2.13) results into the variance expression of $\hat{\eta}$ according to

$$\sigma_{\hat{\eta}}^2 = \left(\left.\frac{\partial g}{\partial \xi}\right|_{\xi=\overline{\xi}} \right)^2 \sigma_{\xi}^2 \tag{2.14}$$

Obviously, the statistics about η is approximated by a linearisation scheme. Thus, the following question comes up naturally: When are these approximations good ones? In [Bre70] this question is addressed qualitatively by the statement given below.

> “The Taylor series will be a good approximation if $g(\cdot)$ is not too far from linear within the region that is within one standard deviation of the mean.”
> *A. M. Breipohl [Bre70]*

Consequently, as commonly known, the approximation quality depends strongly on two factors: (i) on the non-linearity of $g(\cdot)$ and (ii) on the scatter of ξ. Nevertheless, even under the assumption that the dispersion of ξ is known, the degree of non-linearity about

$g(\cdot)$, however, is rarely considered in most practical applications. Here, frequently, one just trusts that the linearised function represents the actual problem adequately. In many cases, therefore, strong deviating results are produced, which may lead to a sub-optimal performance of OED. Any approach which exceeds the linearisation concept in precision is of vital interests in the field of uncertainty propagation and OED, respectively.

Naturally, the utilisation of higher-order terms in the Taylor series expansion improves the accuracy gradually. For instance, it has been shown that even an incorporation of a moderate number of higher-order terms leads to a significant improvement in accuracy [XM12]. The determination of these higher-order terms, however, is accompanied by the calculation of an increased number of partial derivatives. Therefore, applying higher-order approaches is faced with the following dilemma.

> “Linearisation is widely recognised to be inadequate, but the alternatives incur substantial costs in terms of derivation and computational complexity.”
> *S. J. Julier & J. K. Uhlmann [JU04]*

This might be one reason, that higher-order approaches are still in the minority in the framework of uncertainty propagation which is confirmed by the quotation given below.

> “In practice, even the second order approximation is not commonly used and higher order approximations are almost never used.”
> *U. N. Lerner [Ler02]*

So far, the precision of the Taylor series has been the subject of consideration. For the purpose of practical applicability, however, it is also important that an approximate method has to be easy in implementation, and that the method copes well for a broad class of problems. As stated above, to increase the precision of the Taylor series more complex formulas have to be involved. The same is true, in case of non-Gaussian distributions and/or correlated random variables, respectively [Kay93, Zha06, MV08, And11, MALF12].

Additionally, the Taylor series is limited to problems of differentiable transfer functions, $g(\cdot)$. At first, that means, the transfer function has to be known explicitly. Therefore, black-box type functions cannot be addressed immediately. Secondly, even in case of

explicit expressions, functions might be non-differential at all, e.g, the maximum function belongs to those terms. Hence, the Taylor series is likely to suffer in precision as well as in applicability. There is a strong need for alternatives.

2.2.2.2 Gaussian Quadrature

Instead of using surrogates of the original transfer function, $g(\cdot)$, to calculate the mean and the variance of η, an alternative is to solve the underlying integration problem numerically. In doing so, different approaches can be found in literature. For instance, the Gaussian quadrature might be put in operation to approximate the integral equations by discrete function evaluations and associated weights. For the 1-dimensional case, $\Omega \subset \mathbb{R}$, the approximation reads

$$\int_{\Omega} pdf_{\xi} g(\xi) d\xi \approx \sum_{i=1}^{l_{gq}} w_i g(\xi_i) \tag{2.15}$$

The basic idea is to choose l_{gq} points, ξ_i, and associated weights, w_i, in a way that the approximation is exact for polynomials of degree $2l_{gq} - 1$ or less [DR07]. In general, a non-linear equation system can be derived which is solved for an appropriate set of sample points and weights, $S(\xi_i, w_i)$, respectively. Subsequently, the transfer function, $g(\cdot)$, is evaluated at these discrete points. In the final step, the resulting function evaluations, $g(\xi_i)$, are summed up according to the weights, w_i. Obviously, in the Gaussian quadrature framework there is no need of time derivatives at all. Thus, even black-box type, as well as non-differentiable functions can be tackled natively.

If the original function, however, differs substantially in comparison to a general polynomial of degree $2l_{gq} - 1$, this approximate method suffers in precision, too. Obviously, the accuracy can be improved technically by evaluating an increased number of sample points, ξ_i. Due to the related computational load, only a low sample number is frequently used in practice, $l_{gq} \leq 5$. But even in this case, when applied to n-dimensional integration problems, the computational load increases exponentially. Gaussian quadrature, therefore, is subject of the *curse of dimensionality* [Bel66]. In detail, the set of points and weights, $S(\xi_i, w_i)$, which has been derived for the one-dimensional problem, is extended to n dimensions, $\mathbb{R}^n$, by tensor products

$$S^{\otimes n} = \underbrace{S \otimes S \otimes \ldots \otimes S}_{n \text{ times}} \tag{2.16}$$

In consequence, the function $g(\cdot)$ has to be evaluated for a total number of l_{gq}^n points. Therefore, the native Gaussian quadrature approach becomes prohibitive for many practical multi-dimension integration problems, including OED as well. Here, a remedy might be to apply partial tensor products, e.g., (adaptive) sparse grid methods, instead of the full tensor products. In doing so, the need of evaluation points can be reduced considerably. Usually, however, the resulting sample number is still too high for an efficient and general application in the field of OED [DBR10b].

2.2.2.3 Monte Carlo Simulation

Monte Carlo (MC) simulations are frequently applied as an approximate method. Similarly to Gaussian quadrature, discrete samples, ξ_i, are used to solve n-dimensional integration problems, e.g., Eq. (2.6)-(2.10). On the other side, MC differs in the sample generation fundamentally. According to the associated PDF of the random variable, ξ, under study, MC is based on random sampling. Subsequently, the resulting N realisations, ξ_i, are applied to the transfer function, $g(\cdot)$. Finally, the produced set of function evaluations, η_i, is utilised to determine the statistics of the random output variable, η, approximately. For instance, the mean, $E_{MC}[\eta]$, and the variance, $\sigma_{MC}^2(\eta)$, can be calculated via

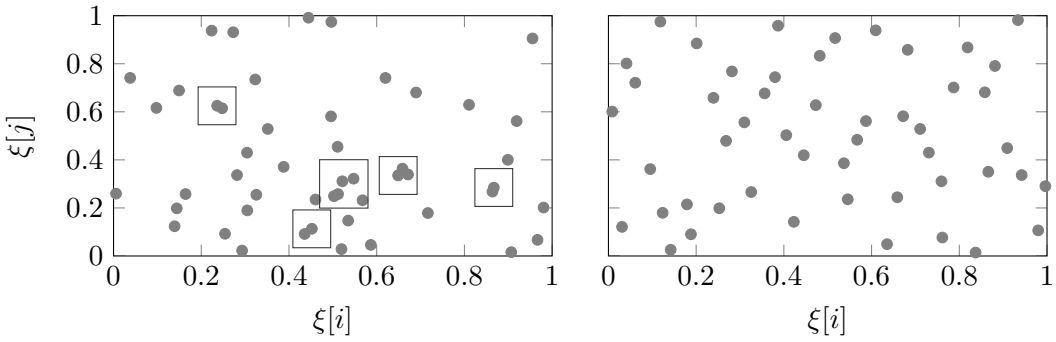
$$E_{MC}[\eta] = \frac{1}{N}\sum_{i=1}^{N}\eta_i \tag{2.17}$$

$$\sigma_{MC}^2(\eta) = \frac{1}{N-1}\sum_{i=1}^{N}(\eta_i - E_{MC}[\eta])^2 \tag{2.18}$$

The popularity of MC is based on its easy handling, i.e., the basics are readily understandable and straightforward in implementation. Additionally, MC does not suffer from the curse of dimensionality. The overall sample number, N, is not related to the dimension of the integration problem. Thus, this approach is tailor-made to address problems involving a large number of random input variables.

On the other side, however, the total sample number of realisations, N, has to sufficiently large to ensure reliable results. In fact, the convergence rate of estimates correlates to $N^{1/2}$ [Jam80]. That means, to achieve a noticeable improvement in the precision a substantial increased number of samples points has to be evaluated. Therefore, the computational load of the native MC approach is prohibitive in many real-life applications. Improved sampling strategies are available in order to accelerate the statistical convergence of MC, e.g., quasi-Monte Carlo methods or Latin hypercube. In

general, these concepts aim to avoid undesired clustering effects which may show up in the native MC implementation, see Fig. 2.2 for illustration. Usually, however, even an improved MC approach remains prohibitively expensive in the field of OED [JSMK06].



(a) Random sampling (b) Latin hypercube sampling

Figure 2.2: Random vs. Latin hypercube sampling: In the native MC approach sample points, •, are generated randomly. Therefore, sample points might be produced which are very close to each other as indicated in sub-figure (a) by □. This phenomena is known as clustering and reduces the performance of MC, because nearly the same information is provided by clustered points. Alternatively, performing Latin hypercube sampling avoids undesired clustering effects by combining random and grid point principles, see (b).

2.2.2.4 Polynomial Chaos Expansion

In uncertainty analysis, the concept of Polynomial Chaos Expansion (PCE) has become quite popular in the last two decades. Thus, for the purpose of completeness, the basics of PCE are presented in what follows. To represent the random variable, η, correctly, a weighted superposition of an infinity number of basis functions, $\Psi_i(\cdot)$, is needed [MK10]

$$\eta = g(\xi) = \sum_{i=0}^{\infty} a_i \Psi_i(\xi) \tag{2.19}$$

Similar to the Taylor series, however, one has to account for computational feasibility. Therefore, the expansion in Eq. (2.19) has to be implemented in a truncated form

$$\hat{\eta} = \sum_{i=0}^{l_{pce}} a_i \Psi_i(\xi) \tag{2.20}$$

By a proper choice of basis functions, $\Psi_i(\cdot)$, the determination of the unknown coefficients, a_i, can be simplified. In particular, different sets of orthogonal basis functions are applied depending on the associated PDF of the random input variable, ξ. For instance,

Hermite polynomials are utilised in case of a Gaussian distribution. In literature, different approaches are known to determine the coefficients, a_i, see [Tem09, MK10] and references therein. Here, the focus is on the least-square approach solely. In practical implementation, a residual, $r(\xi)$, emerges due to the truncation of PCE terms, $l_{pce} << \infty$,

$$r(\xi) = g(\xi) - \sum_{i=0}^{l_{pce}} a_i \Psi_i(\xi) \tag{2.21}$$

Now, the expected sum of squared errors can be defined as a suitable cost function

$$J_{PCE} = \int_\Omega [r(\xi)]^2 pdf_\xi d\xi \tag{2.22}$$

The additivity of the expectation operator enables the following reordering

$$J_{PCE} = \int_\Omega g(\xi)^2 pdf_\xi d\xi - 2\int_\Omega g(\xi) \sum_{i=0}^{l_{pce}} a_i \Psi_i(\xi) pdf_\xi d\xi + \int_\Omega \left(\sum_{i=0}^{l_{pce}} a_i \Psi_i(\xi) \right)^2 pdf_\xi d\xi \tag{2.23}$$

The minimum of this cost function can be found by differentiation of Eq. (2.23) with respect to a_i, and by setting the resulting derivative equal to zero. Here, due the orthogonality of Ψ_i the mathematical expression results in

$$\frac{\partial J_{PCE}}{\partial a_i} = -2\int_\Omega g(\xi)\Psi_i(\xi) pdf_\xi d\xi + 2a_i \int_\Omega \Psi_i(\xi)^2 pdf_\xi d\xi \stackrel{!}{=} 0 \tag{2.24}$$

Therefore, the i^{th} coefficient can be calculated according to

$$a_i = \frac{\int_\Omega g(\xi)\ \Psi_i\ pdf_\xi\ d\xi}{\int_\Omega \Psi_i(\xi)^2\ pdf_\xi\ d\xi} \tag{2.25}$$

In case of Hermite polynomials (still assuming a standard normal distribution, pdf_ξ), the denominator can be determined immediately, see [MK10], via

$$\int_\Omega \Psi_i(\xi)^2\ pdf_\xi\ d\xi = i! \tag{2.26}$$

The numerator of Eq. (2.25), however, has to be derived numerically. Obviously, instead of solving one of the original integrals, Eq. (2.6)-(2.10), a modified integration problem has to be tackled. First of all, the quantification of the coefficients, a_i, ensures a parametrisation of PCE, Eq. (2.20). Subsequently, associated moments of $\hat{\eta}$ can be

calculated analytically, e.g., the mean and the variance are determined by

$$E[\hat{\eta}] = a_0 \tag{2.27}$$

$$\sigma_{\hat{\eta}}^2 = \sum_{i=1}^{l_{pce}} a_i^2 \int_{\Omega} \Psi_i(\xi)^2 pdf_{\xi} d\xi = \sum_{i=1}^{l_{pce}} a_i^2 \cdot i! \tag{2.28}$$

In addition, a PDF approximation of $\hat{\eta}$ can be derived in combination with Monte Carlo simulations (Sec. 2.2.2.3) and standard Kernel density estimation algorithm which are available in standard computation/statistic tools, e.g., routines available in MATLAB® or in *R!*. Please bear in mind that $\hat{\eta}$ is an algebraic expression of ξ, Eq. (2.20). Therefore, MC simulations based on $\hat{\eta}$ can be performed at low computational costs. In summary, PCE benefits from its versatility and its good convergence behaviour, see [MK10] for additional details.

In case of OED, however, there is no special need of PDF approximations, because a very limited order of moments is sufficient as demonstrated in Sec. 3. Hence potential shortcomings of PCE may dominate in situations of an exclusive approximation of statistical moments. In general, the determined coefficients, a_i, are the result of an approximation of an approximation, i.e., the truncation error follows a numerical integration error, subsequently. Obviously, the incorporation of higher order terms in Eq. (2.20) increases the non-linearity of the integrand gradually, see Fig. 2.3. Thus, it is likely that the numerical error of coefficients, a_i, which are associated to higher order terms increases as well. Therefore, a balance of truncation and numerical integration error has to be found, which might be challenging. Additionally, the total number of coefficients, $\#(a_i)$, in n-dimensional problems increases rapidly as stated by the following expression which can be found in [MK10]

$$\#(a_i) = \frac{(n + l_{pce})!}{n! l_{pce}!} - 1 \tag{2.29}$$

For instance, assuming an input dimension of $n = 15$ and a order of PCE equal to $l_{pce} = 3$ ($l_{pce} = 5$) the overall number of unknown coefficients, a_i, is 816 (15504). Therefore, the versatility of PCE is paid by an increased computational load which might be prohibitive in the field of OED.

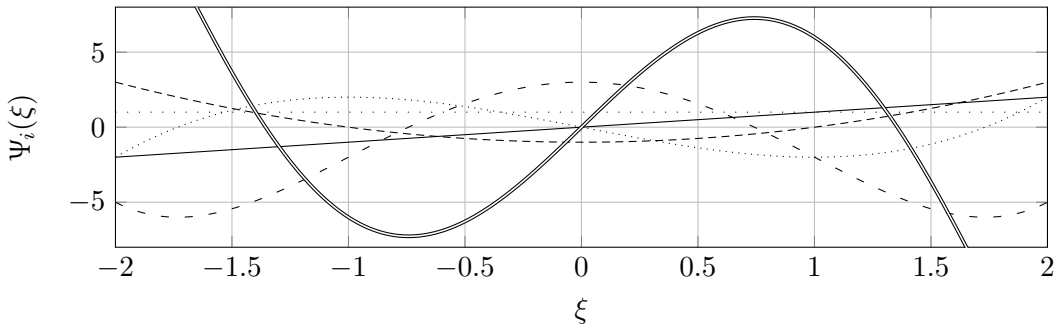


Figure 2.3: The first six Hermite Polynomials are shown: ····· $\Psi_0 = 1$; —— $\Psi_1 = \xi$; - - - - $\Psi_2 = \xi^2 - 1$; ····· $\Psi_3 = \xi^3 - 3\xi$; - - $\Psi_4 = \xi^4 - 6\xi^2 + 3$; ═══ $\Psi_5 = \xi^5 - 10\xi^3 + 15\xi$. Obviously, in higher order terms the degree of non-linearity increases gradually. Thus, as part of an integrand, proper/robust numerical integration routines have to put in operation to ensure reliable results of the related integration problem.

2.2.3 Unscented Transformation

Here, similar to Gaussian quadrature methods (Sec. 2.2.2.2) and Monte Carlo simulations (Sec. 2.2.2.3) one is interested in generating sample points, ξ_i, and associated weights, w_i, which are used to solve an n-dimensional integration problem approximately. The essential differences are: (i) the sample points are not chosen randomly but deterministically, and (ii) these points are generated directly in $\mathbb{R}^n$ instead of "tensoring" sample points which are chosen in $\mathbb{R}^1$. But before starting with the mathematical details, a short historical review of the Unscented Transformation is given.

The method of Unscented Transformation (UT), which has been introduced by Julier and Uhlmann in 1994 [JU94], has become quite popular in non-linear filter theory over the last two decades. Applied as an inherent part of the Unscented Kalman Filter (UKF) it has gradually superseded the standard approach in non-linear filtering, the Extended Kalman Filter (EKF). The guiding theme of UT has been the notion that

> "... it is easier to approximate a probability distribution than it is to approximate an arbitrary nonlinear function or transformation."
> *S. J. Julier & J. K. Uhlmann [JU04]*

The mathematical basics of UT, however, date back approximately 60 years in time [Tyl53]. At this time, the focus had been on efficient numerical integration routines for multi-dimensional integration problems. In detail, formulas had been derived which are intended to solve integration problems over symmetrical regions. Due to this symmetry,

numerical integration techniques have been developed which at best scale linearly to an n-dimensional integration problem.

After the first considered symmetric functions had no specific meaning, these functions have been associated to symmetric probability distributions shortly afterwards. In doing so, these formulas have been successfully applied to solve problems in uncertainty analysis where they are commonly known as Point Estimate Methods (PEMs). Here, the very first articles had been devoted to statistical tolerance calculation [Eva67, Eva74], i.e., tolerances of industrially manufactured products had been quantified. In the following years, however, PEMs have been implemented to uncertainty analysis in various disciplines ranging from complex technical systems up to environmental/biochemical modelling, see [MWL02, TF05, MPR07, FTM12, PK12, LL13] and references therein.

Naturally, the Point Estimate Methods have been subject of ongoing research for the last 50 years. In particular the increased computational power has enabled to implement less restrictive methods. That means, approaches that scale polynomially to n-dimensional integration problems have become applicable resulting in more precise calculation [Ler02]. Additionally, also asymmetric distributions have been become practical in the framework of PEM [Li92]. The general basics, however, are unchanged and shortly summarised in what follows based on the notations given in [Eva67, Ler02].

In Point Estimate Methods, the fundamental idea is to choose sample points, ξ_i, and associated weights, w_i, in relation to the first raw moments of the random input variable, ξ. Here, the so-called Generator Function, $GF[\cdot]$, [Tyl53, Ler02] is of vital importance. A GF describes how sample points are directly determined in $\mathbb{R}^n$ by permutation and the change of sign-combinations. According to that, the first three GFs operate in the following way: (i) $GF[0]$ changes no element of a given vector, (ii) $GF[\pm\vartheta]$ permutes one element of a given vector to all potential combinations, and (iii) $GF[\pm\vartheta, \pm\vartheta]$ permutes two elements of a given vector to all potential combinations. For instance, the proposed Generator Functions are illustrated with a problem in $\mathbb{R}^2$:

$$GF[0] = \{(0,0)^T\} \tag{2.30}$$

$$GF[\pm\vartheta] = \{(\vartheta,0)^T,\ (-\vartheta,0)^T,\ (0,\vartheta)^T,\ (0,-\vartheta)^T\} \tag{2.31}$$

$$GF[\pm\vartheta, \pm\vartheta] = \{(\vartheta, \vartheta)^T,\ (-\vartheta, -\vartheta)^T,\ (\vartheta, -\vartheta)^T,\ (-\vartheta, \vartheta)^T\} \tag{2.32}$$

Here, the scalar parameter, ϑ, controls the spread of the sample points, ξ_i, in $\mathbb{R}^n$. Generally, for the purpose of solving an n-dimensional integration problem, the idea is to use a weighted superposition of function evaluations at GF-based sample points, $g(\xi_i)$, according to

$$\int_\Omega g(\xi) pdf_\xi d\xi \approx w_0 g(GF[0]) + w_1 \sum g(GF[\pm\vartheta]) + w_2 \sum g(GF[\pm\vartheta, \pm\vartheta]) \tag{2.33}$$

As only a finite number of raw moments of the input random variable, ξ, is considered, the transfer function, $g(\cdot)$, is approximated by monomials of finite degree [Eva67, Ler02]. For instance, by taking account for the first two non-zero raw moments of ξ, the related monomials of the transfer function, $g(\cdot)$, are $g(\xi) = 1$ and $g(\xi) = \xi[i]^2$ (any $i \in \{1, \ldots, n\}$ might be used due to symmetry). Thus, the transfer function is approximated exactly for monomials of order three. The Unscented Transformation in relation to monomials of precision 3 is labelled as UT3. Remember that any odd power term is zero in association to Gaussian distributions. In this particular case, only the first two Generator Functions, $GF[0] \cap GF[\pm\vartheta]$, can be parametrised by solving the following equation system

$$w_0 + 2nw_1 = \int_\Omega 1 pdf_\xi d\xi = 1 \tag{2.34}$$

$$2w_1\vartheta^2 = \int_\Omega \xi[i]^2 pdf_\xi d\xi = 1 \tag{2.35}$$

In consequence, for $\vartheta \neq 0$, the related weights can be calculated via

$$w_0 = 1 - \frac{n}{\vartheta^2} \tag{2.36}$$

$$w_1 = \frac{1}{2\vartheta^2} \tag{2.37}$$

As shown in [JU04] higher-order moments of the analysed PDF can be used for the quantification of ϑ additionally. For instance, considering the 4'th raw moment of the standard Gaussian distribution leads to

$$2w_1\vartheta^4 = \int_\Omega \xi[i]^4 pdf_\xi d\xi = 3 \tag{2.38}$$

Therefore, applying $\vartheta = \sqrt{3}$ might be an optimal choice in case that the probability distribution of η is close to the normal distribution ...

> “…, but in general different values of ϑ can lead to better or worse approximations depending on the function $g(\cdot)$.
> *U. N. Lerner [Ler02]* ”

Obviously, there is some uncertainty about ϑ which might be expressed by an associated probability distribution, pdf_ϑ. To compensate for this uncertainty Generator Functions can be averaged in principle via

$$\int_\Omega (GF[\pm\vartheta] pdf_\vartheta d\vartheta \tag{2.39}$$

Here, too, the resulting 1-dimensional integration problem can be solved with one of the previously presented approaches. For instance, a nested UT approach might be implemented. That means, points and weights related to pdf_ϑ are generated at first. These quantities are applied in a second step to the original approximation scheme given in Eq. (2.33). The total cost of the modified approach correlates to $3 \cdot (2n+1)$ points which have to be evaluated by $g(\cdot)$, i.e., the computational effort still scales linearly to the dimension, n, of the original integration problem.

After a proper selection of points, $\eta_i = g(\xi_i)$, and associated weights, w_0 & w_1, the mean and the (co)variance of η can be determined by approximation according to

$$E[\eta] \approx \overline{\eta} = w_0\eta_0 + w_1 \sum_{i=1}^{2n} \eta_i \tag{2.40}$$

$$\sigma^2(\eta) \approx w_0(\eta_0 - \overline{\eta})(\eta_0 - \overline{\eta})^T + w_1 \sum_{i=1}^{2n} (\eta_i - \overline{\eta})(\eta_i - \overline{\eta})^T \tag{2.41}$$

As shown in [JU04] the incorporation of a (co)variance correction term, β, can improve the precision of $\sigma^2(\eta)$ by

$$\sigma^2(\eta) \approx (w_0 + \beta)(\eta_0 - \overline{\eta})(\eta_0 - \overline{\eta})^T + w_1 \sum_{i=1}^{2n} (\eta_i - \overline{\eta})(\eta_i - \overline{\eta})^T \tag{2.42}$$

Similar to the previous quantification of ϑ, different values of β provide better or worse approximations depending on the function $g(\cdot)$. Some general comments about the precision of $\sigma^2(\eta)$ are given in Sec. 2.2.3.1. At this point, however, it is stressed that in the particular case of a linear mapping, $\eta = c_1\xi + c_0$, the correction term, β, does not influence the outcome of $\sigma^2(\eta)$. Remember that a linear mapping of a Gaussian

distribution is bias-free, $\eta_0 - \overline{\eta} = 0$. Therefore, the (co)variance of the resulting Gaussian distribution of η is still determined correctly.

In the same manner also higher order moments of η can be approximated according to

$$\mu_3 \approx w_0(\eta_0 - \overline{\eta})(\eta_0 - \overline{\eta})^T(\eta_0 - \overline{\eta}) + w_1 \sum_{i=1}^{2n}(\eta_i - \overline{\eta})(\eta_i - \overline{\eta})^T(\eta_0 - \overline{\eta}) \tag{2.43}$$

$$\begin{aligned}\mu_4 \approx\ & w_0(\eta_0 - \overline{\eta})(\eta_0 - \overline{\eta})^T(\eta_0 - \overline{\eta})(\eta_0 - \overline{\eta})^T + \\ & w_1 \sum_{i=1}^{2n}(\eta_i - \overline{\eta})(\eta_i - \overline{\eta})^T(\eta_0 - \overline{\eta})(\eta_0 - \overline{\eta})^T\end{aligned} \tag{2.44}$$

Here, too, additional correction factors might be applied to improve the accuracy.

Naturally, the general precision of the UT approach can be increased gradually by considering higher order raw moments of ξ. For instance, an approximation scheme can be applied which represents monomials of $g(\cdot)$ correctly up to the precision of 5 via

$$E[g(\xi)] = \int_\Omega g(\xi) pdf_\xi d\xi \approx w_0 g(GF[0]) + w_1 g(GF[\pm\vartheta]) + w_2 g(GF[\pm\vartheta, \pm\vartheta]) \tag{2.45}$$

The Unscented Transformation in relation to monomials of precision 5 is labelled as UT5. In this case, the number of generated sample points, ξ_i, correlates to $2n^2 + 1$ for an n-dimensional integration problem. Here, for the purpose of parametrisation of w_i and ϑ an equation system can be derived taking into account monomials of degree 5 or less

$$w_0 + 2nw_1 + 2n(n-1)w_2 = \int 1 pdf_\xi d\xi = 1 \tag{2.46}$$

$$2w_1\vartheta^2 + 4(n-1)w_2\vartheta^2 = \int \xi[i]^2 pdf_\xi d\xi = 1 \tag{2.47}$$

$$2w_1\vartheta^4 + 4(n-1)w_2\vartheta^4 = \int \xi[i]^4 pdf_\xi d\xi = 3 \tag{2.48}$$

$$4w_2\vartheta^4 = \int \xi[i]^2 \xi[j \neq i]^2 pdf_\xi d\xi = 1 \tag{2.49}$$

Therefore, the four unknowns can be uniquely determined by the previous equation

system as

$$\vartheta = \sqrt{3} \tag{2.50}$$

$$w_0 = 1 + \frac{n^2 - 7n}{18} \tag{2.51}$$

$$w_1 = \frac{4 - n}{18} \tag{2.52}$$

$$w_2 = \frac{1}{36} \tag{2.53}$$

2.2.3.1 Computational Effort and Approximation Power

In the previous section, it is shown how the number of applied sample points, ξ_i, correlates to the precision power of the Unscented Transformation. For instance, the application of $2n + 1$ or $2n^2 + 1$ sample points ensures a correct approximation of monomials of precision 3 and 5 respectively. Therefore, the computational load scales linearly or quadratically in relation to the n-dimensions. In comparison to the Gaussian quadrature approach of Sec. 2.2.2.2, this might be a dramatic reduction in the total number of necessary sample points as shown in Fig. 2.4.

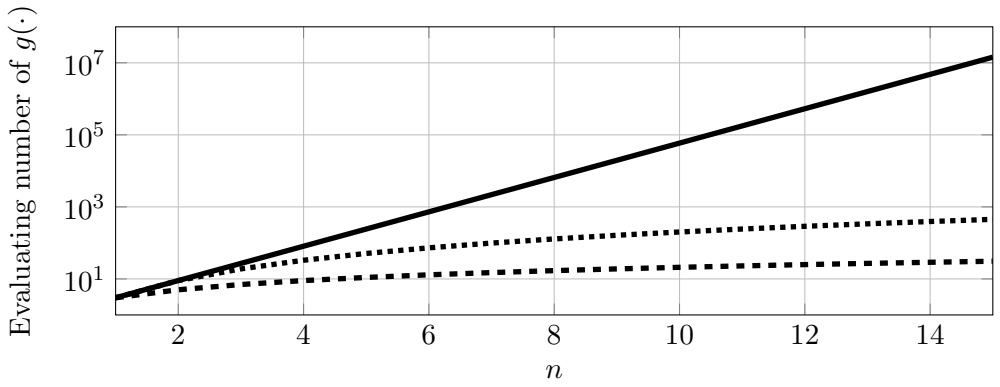
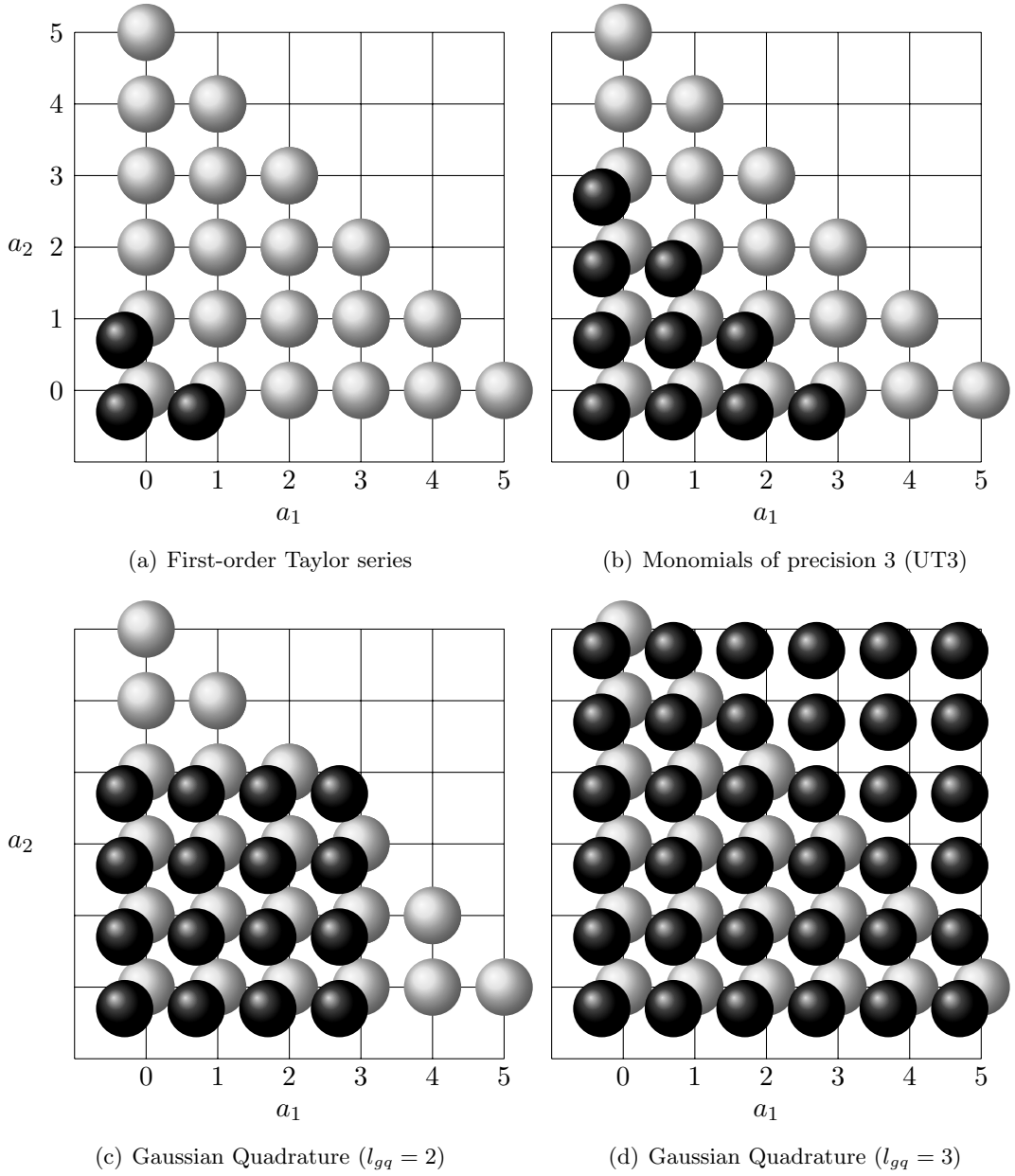


Figure 2.4: Benchmark of computational effort: ▪▪▪ Unscented Transformation (UT3), ▪▪▪▪ Unscented Transformation (UT5), ━━ Gaussian quadrature

On the other side, however, the precision of the UT is very sensitive to input interactions. Due the inherent relation to monomials of low degree, only low order interaction can be approximated appropriately [Ler02]. Consequently, higher order input interactions lead to a poor approximation performance of UT. For instance, by implementing UT for monomials of precision 5, $\prod_{i=1}^{n} \xi[i]^{a_i}$, the summation of the coefficients, a_i, is limited to 5 or less, $\sum_{i=1}^{n} a_i \leq 5$. That means, expressions similar to $g(\xi) = \xi[1]^2 \cdot \xi[2]^2$



(a) First-order Taylor series

(b) Monomials of precision 3 (UT3)

(c) Gaussian Quadrature ($l_{gq} = 2$)

(d) Gaussian Quadrature ($l_{gq} = 3$)

Figure 2.5: The exact approximation of $E[\eta] = g(\xi) = \xi[1]^{a_1} \cdot \xi[2]^{a_2}$ is illustrated for the proposed approximate methods. In detail, monomials of precision 5 (UT5), ●, in relation to alternative approaches, ●, is shown. Obviously, the first-order Taylor series approach causes approximation errors as soon as nonlinear terms show up, $a_1 + a_2 > 1$. UT3 and UT5 ensure correct results for $a_1 + a_2 \leq 3$ and $a_1 + a_2 \leq 5$, respectively. The superior approximation power of the Gaussian Quadrature is accompanied by an excessive computational load for real-life applications. UT3 and UT5 provide a fair compromise on an adequate approximation power and low computational costs.

can be approximated correctly, whereas $g(\xi) = \xi[1]^2 \cdot \xi[2]^2 \cdot \xi[3]^2$ causes already an approximation error. In this case, the Gaussian quadrature benefits from the tensor product. If the 1-dimensional Gaussian quadrature approach is of precision l, any product equal to $\prod_{i=1}^{n} \xi[i]^{a_i}$, where $a_i \leq l$, is approximated correctly. The approximation power of different approaches is illustrated in Fig. 2.5.

Generally, there seems to be some obscurities in literature about the approximation performance of the UT approach. Here, the following problem is discussed to overcome common misinterpretations: Why determines the UT3 approach the variance, σ_η^2, about $g(\xi) = \xi_1^2$ correctly but fails in case of $g(\xi) = \xi[1]^2 + \xi[2]^2$? Associated numerical results are presented in Sec. 2.2.3.3. Here, the theoretical framework is given. In doing so, the Eq. (2.41) is reformulated according to

$$\sigma_\eta^2 \approx \left(w_0 g(\xi_0)^2 + w_1 \sum_{i=1}^{2n} g(\xi_i)^2 \right) - \overline{g}(\xi)^2 \tag{2.54}$$

$$\sigma_\eta^2 \approx \overline{g(\xi)^2} - \overline{g}(\xi)^2 \tag{2.55}$$

Obviously, by calculating the variance, σ_η^2, the UT approach has to be not only a good approximation of $g(\cdot)$ but also of $g(\cdot)^2$, too. In general, the same is true for Gaussian quadrature methods as well as for Monte Carlo simulations. In relation to the previous examples, the expressions $g(\xi)^2 = \xi[1]^4$ and $g(\xi)^2 = \xi[1]^4 + 2\xi[1]^2\xi[2]^2 + \xi[2]^4$, respectively, have to be approximated properly. In both cases, monomials of order 4 show up. Thus, the application of UT3, which is correct up to monomials of order 3, should suffer in precision in either case, i.e., an approximation error is expected to emerge. But why, this being the case, determines the UT3 approach the variance about $g(\xi) = \xi^2;\ \xi \in \mathbb{R}^1$ exactly?

Here, for the special case of 1-dimensional integration problems, the UT3 approach, the UT5 version, as well as the Gaussian quadrature of precision 5 ensure the same accuracy. In detail, all three methods are perfect approximation schemes for monomials of degree 5 or less. This might be one reason why the performance of UT3 is frequently overrated in literature as it is done in [JU94, JU04, vdM04] to name but a few.
As shown in Sec. 2.2.2.1, the overall performance of the Taylor series approach depends on the precision of the resulting surrogate transfer function, $\hat{g}(\xi)$. That means, in case that the transfer function, $g(\cdot)$, is represented exactly by the surrogate, the mean, $E[\eta]$, and the variance, σ_η^2, are calculated correctly as well. Hence, there is no

need to approximate the squared transfer function, $g(\cdot)^2$, directly as it is done by the UT approach. Therefore, to ensure that UT has the same accuracy as the second-order Taylor series approach, monomials of order 4 and less have to be represented correctly. Consequently, in case of the variance, σ_η^2, only the UT5 approach outperforms the second-order Taylor series in precision.

Nevertheless, the UT3 leads in many practical applications to a significant improvement in the determined statistical quantities in comparison to the standard approach which is based on linearisation [Ler02, vdM04, Sim06]. Therefore, UT3 fills the gap of low computational demands and accuracy. In consequence, the UT3 approach is applied to a great extent in what follows.

2.2.3.2 Incorporation of General Probability Density Functions

Up to this point, the uncertainty about input variables, ξ, has been assumed to follow independent standard Gaussian distributions. According to this assumption the previous algorithms of uncertainty propagation have been explored. In most applications, however, quite different kinds of PDFs appear and have to be handled appropriately.

In general, the previously presented methods of uncertainty propagation can be adapted to non-Gaussian distributions immediately. For instance, there is a generalised Polynomial Chaos Expansion (gPCE) approach in literature which is applicable to a wide range of different distributions. In gPCE the essential idea is to apply specific types of orthogonal polynomials depending on the PDF under study. In detail, the Askey-scheme [MK10] provides a set of orthogonal polynomials which are optimal in relation to certain distributions [XK02].

Following the same basic idea, the Unscented Transformation method can be generalised by an appropriate incorporation of distribution-specific raw moments into the equation system given above, Eq. (2.36)-(2.38). In the framework of UT, however, one is limited to symmetric distributions, i.e., the odd statistical moments have to be zero [BD87, KK00, HMGB03b]. Nonetheless, a various number of distributions can be treated under this restriction. To name but a few this includes the following distributions:

- General Gaussian distribution
- Uniform distribution
- Student's t-distribution
- Laplace distribution

- Symmetric Triangle distribution
- Logistic distribution

Generally, the propagation of asymmetric distributions can be performed in an extended framework of Point Estimate Methods. Different algorithms can be found in literature for this purpose. For instance, the sampling scheme introduced in [Li92] might be applied. By assuming an asymmetric distribution asymmetric weights have to be evaluated. In this case, however, the overall number of sample points, ξ_i, correlates to $(n^2 + 3n + 2)/2$ for an n-dimensional problem. Clearly, the gain in flexibility, which means here the propagation of non-Gaussian distributions, has to get paid by an increased number of function evaluations.

In most practical applications, however, one is usually interested in an easy to implement, robust, as well as efficient algorithm. Therefore, a more practicable framework might be desirable for the purpose of general uncertainty propagation. Alternatively, instead of adapting the weights, w_i, and sample points, ξ_i, according to the distribution at hand, $pdf_{\xi'}$, one can derive a (non)linear transfer function, $q(\cdot)$, which renders a standard Gaussian distribution into the desired distribution, $\xi' = q(\xi)$. In detail, by utilising the inverse Rosenblatt transformation [LC07] given PDFs associated to ξ' are transformed into a set of independent random variables of standard Gaussian distributions. For instance, the transformation applied to the 1-dimensional problem is expressed by

$$\xi' = q(\xi) = F^{-1}\left(\Phi(\xi)\right) \tag{2.56}$$

Here, $F^{-1}(\cdot)$ represents the inverse of the cumulative distribution function (CDF) of the desired random variable ξ', and $\Phi(\cdot)$ denotes the CDF of the standard Gaussian random variable ξ. In the same manner even correlated random variables can be transformed into independent standard Gaussian representatives [MB12]. In conclusion, the Unscented Transformation becomes applicable for correlated non-Gaussian random variables, too. For example, in Tab. 2.1 some resulting transformation functions are given for frequently used PDFs. Additional transformation formulas can be found in [Isu99].

Obviously, in most cases, the transformation function, $q(\cdot)$, is a non-linear expression. Hence, as an inherent part of the original uncertainty propagation problem, $\eta = g(q(\xi))$, the overall non-linearity may become more severe. That means, the PEM/UT methods may suffer in precision to a certain extent. In many practical applications, however, this precision flaw might be acceptable in the light of the easiness in implementation.

Type of $pdf_{\xi'}$	Transformation: $q(\xi) =$
Normal(μ, σ)	$\mu + \sigma\xi$
Uniform(a, b)	$a + (b-a)\left(\frac{1}{2} + \frac{1}{2}\text{erf}(\xi\sqrt{2})\right)$
Log-normal(μ, σ)	$\exp(\mu + \sigma\xi)$
Gamma(a, b)	$ab\left(\xi\sqrt{\frac{1}{9a}} + 1 - \frac{1}{9a}\right)^3$
Exponential(λ)	$-\frac{1}{\lambda}\log\left(\frac{1}{2} + \frac{1}{2}\text{erf}\left(\frac{\xi}{\sqrt{2}}\right)\right)$

Table 2.1: Probability density function transformation formulas extracted and adapted from [Isu99]. Here, the term *erf* means the error function.

Moreover, numerical results in Sec. 2.2.3.3 confirm the usefulness of the transformation approach additionally.

As shown above, the framework of the Unscented Transformation can be extended to deal with different distribution types of random input vectors, ξ, properly. In addition, the problem of an adequate representation of the resulting output uncertainty is addressed in what follows. According to Eq. (2.40)-(2.41) approximations of the mean, $E[\eta]$, and the (co)variance, σ_η^2, are determined by UT, respectively. Usually, these two quantities are associated to a Gaussian distribution as well. For example, a related PDF might be parametrised or confidence intervals are determined. In cases, however, where the actual distribution of η diverges strongly in comparison to a Gaussian distribution misleading inferences are likely. Here, the additional information of higher order moments about η, e.g. the skewness and the kurtosis, might be used as correction factors. As demonstrated, the UT approach enables a straightforward calculation of these higher moments, see Eq. (2.43) and (2.44). Thus, one of the following concepts might be put in operation to take account for non-Gaussian output distributions:

(a) Instead to parametrise a Gaussian PDF via $E[\eta]$ and σ_η^2, alternative PDFs might be parametrised as well. In a subsequent step, the approximations about higher-order moments are utilised to select the most plausible PDF candidate, see Sec. 2.2.3.3 for illustration.

(b) A similar idea in comparison to the first point but performed more systematically, is to apply the Pearson system [AKM05]. Here, the Pearson system encompasses different types of distributions, e.g., Gaussian distribution, the beta distribution, the exponential distribution, and the uniform distribution. In detail, the first four statistical moments obtained for η, determine which of the enclosed distribution type is selected.

(c) Alternatively, the Gaussian distribution might be reshaped by correction factors which are correlated to the skewness and the kurtosis of η. In literature, different correction schemes can be found. For example, the Cornish-Fisher expansion [SA11] might be applied.

(d) Finally, the Unscented Transformation approach can be applied to approximate the integration expression of the Polynomial Chaos Expansion which has been presented in Sec. 2.2.2.4.

(e) Moreover, the UT concept in combination with a Gaussian Mixture density approximation strategy may perform well for non-Gaussian problems, too, but demands an increased sample number [RHK10].

2.2.3.3 Practical Implementation

This sub-section aims to present the most frequently applied algorithm of the Unscented Transformation in relation to monomials of precision 3 (UT3). In this framework, the Generator Function as well as the transformation step to any desired Gaussian distribution are performed inherently, see [JU94, JU04] for details. The generation of sample points is performed according to

$$\xi_0 = \quad 0 \tag{2.57}$$

$$\xi_i = \quad +\sqrt{(n+\lambda)}\left(\sqrt{\sigma_\xi^2}\right)_i \quad ; \quad i = 1,\ldots,n \tag{2.58}$$

$$\xi_i = \quad -\sqrt{(n+\lambda)}\left(\sqrt{\sigma_\xi^2}\right)_i \quad ; \quad i = n+1,\ldots,2n \tag{2.59}$$

where $\left(\sqrt{\sigma_\xi^2}\right)_i$ is the ith column of the matrix square root and $\lambda = \alpha^2 \cdot (n+\kappa) - n$.

Now, every sample point, ξ_i, is evaluated by the transfer function

$$\eta_i = g(\xi_i); \quad \forall i = 1,\ldots,(2\cdot n+1) \tag{2.60}$$

The resulting set of transformed points, η_i, is used subsequently to determine an approximation of the mean, $E[\eta]$, and the (co)variance matrix, σ_η^2, by the following weighted

superposition

$$E[\eta] = \sum_{i=0}^{2\cdot n} w_i^m \eta_i \tag{2.61}$$

$$\sigma_\eta^2 = \sum_{i=0}^{2\cdot n} w_i^c (\eta_i - \overline{\eta})(\eta_i - \overline{\eta})^T, \tag{2.62}$$

Here, the associated weights, w_i, are given by

$$w_0^m = \frac{\lambda}{n+\lambda} \tag{2.63}$$

$$w_0^c = \frac{\lambda}{n+\lambda} + 1 - \alpha^2 + \beta \tag{2.64}$$

$$w_i^m = w_i^c = \frac{1}{2\cdot(n+\lambda)}; \quad \forall i = 1, \ldots, 2n. \tag{2.65}$$

In comparison to the presented Point Estimate Method in Sec. 2.2.3, a modified notation is used and an additional tuning parameter, α, is introduced for the purpose of numerical robustness [JU04]. However, both approaches coincide for the following setting, $\alpha = 1$, $\kappa = 3 - n$, and $\beta = 0$.

In this work, the Unscented Transformation is applied to various non-linear uncertainty propagation problems that are given in the following.
a) relation between measurement noise and Akaike weights, $\mathcal{W}(\hat{\mathcal{S}}_i)$

- ξ is the vector of all available measurement data. If K is the number of measurement time points, ξ has the dimension $n = m \cdot K$ and is given by:

$$\xi = [y_1(t_1), \ldots, y_1(t_K), \ldots, y_m(t_1), \ldots, y_m(t_K)]^T.$$

- $g(\cdot)$ stands for the determination of Akaike weights, $\mathcal{W}(\mathcal{S}_i)$, i.e., the complete parameter identification process, as well as the calculation of AIC^c (Eq. (4.6)) are part of the transformation process.

- η is the resulting Akaike weight, $\mathcal{W}(\hat{\mathcal{S}}_i)$, of model candidate $\hat{\mathcal{S}}_i$, i.e., $\eta = \mathcal{W}(\hat{\mathcal{S}}_i)$.

b) relation between state vector, $x(t_k)$, and the next time instance, $x(t_{k+1})$

- ξ is the vector of all states, i.e., quantities that are described by the ODE system at time point t_k, i.e., $\xi = x(t_k)$.

- $g(\cdot)$ stands for the simulation process, i.e., the numerical solution of the underlying ODE system in the time interval $\Delta t = t_{k+1} - t_k$.
- η is the resulting vector of states at time point t_{k+1}, i.e., $\eta = x(t_{k+1})$.

c) relation between parameter perturbation and the simulation results

- ξ is the vector of all imprecisely known model parameters, $\xi = \theta$.
- $g(\cdot)$ stands for the simulation process, i.e., the numerical solution of the underlying ODE system.
- η is the resulting vector of the output function at a certain time point, i.e., $\eta(t_k) = h(t_k)$.

The last application is an essential part in the field of Global Sensitivity Analysis that is described in more detail in the next subsection.

2.2.4 Global Sensitivity Analysis

Especially for models in systems biology, the influence of model parameters, θ, on the model output varies strongly. On the one hand there are parameters, $\theta_l \subset \theta$, that can be changed by orders of magnitude without notable influence on the dynamic behaviour and on the other hand a slight change of certain parameters, $\theta_h \subset \theta$, leads to a strong output variation. Evidently, in the framework of parameter identification it is much more easier to identify the latter parameter subset, θ_h.
To determine the influence of model parameters θ on $y^{sim}(t)$ related parameter sensitivities have to be calculated. If identified parameters provide tight confidence regions, i.e., their values are almost certainly known, then the sensitivities can be determined by a local approach evaluating the Sensitivity Matrix (SM)

$$SM(t_k) = \left.\frac{\partial y^{sim}(t_k)}{\partial \theta}\right|_{\overline{\theta}} \tag{2.66}$$

Usually, this is not the case and global methods taking parameter uncertainties explicitly into account have to be applied. These requirements are automatically fulfilled by variance-based approaches. Treating parameters, θ, and the output, $y^{sim}(t)$, as random variables, the amount of variance that each parameter, $\theta[i]$, contributes to the variance of the output, $\sigma^2(y^{sim}(t))$, is determined.
The ranking of a parameter $\theta[i]$ is done by the amount of output variance that would vanish, if this parameter $\theta[i]$ is assumed to be known. Formally, for every assumed

known parameter $\theta[i]$ a conditional variance, $\underset{-i}{\sigma}^2(y^{sim}|\theta[i])$, can be determined. The subscript $-i$ indicates that the variance is taken over all parameters other than $\theta[i]$. As $\theta[i]$ itself is a random variable in reality, the expected value of the conditional variance $\underset{i}{E}\left[\underset{-i}{\sigma}^2(y^{sim}|\theta[i])\right]$ has to be determined, where the subscript $\underset{i}{E}$ indicates that the expected value is only taken over the parameter $\theta[i]$. Now, the output variance, $\sigma^2(y^{sim})$, can be separated [SRTC05] into the following two additive terms.

$$\sigma^2(y^{sim}) = \underset{i}{\sigma}^2(\underset{-i}{E}[y^{sim}|\theta[i]]) + \underset{i}{E}[\underset{-i}{\sigma}^2(y^{sim}|\theta[i])] \tag{2.67}$$

The variance of the conditional expectation, $\underset{i}{\sigma}^2(\underset{-i}{E}[y^{sim}|\theta[i]])$, represents the contribution of parameter $\theta[i]$ to the variance $\sigma^2(y^{sim})$ indicating the importance of this parameter. The normalised expression in Eq. 2.68 is known as the first order sensitivity index [Sob93] and is used in the following for parameter sensitivity analysis.

$$S_i^y = \frac{\underset{i}{\sigma}^2(\underset{-i}{E}[y^{sim}|\theta[i]])}{\sigma^2(y^{sim})} \tag{2.68}$$

Usually, the integrals associated to $\sigma^2(y^{sim})$, $\underset{-i}{E}[y^{sim}|\theta[i]]$, and $\sigma^2(y^{sim}|\theta[i]))$ are evaluated by Monte Carlo simulations [Sob01], which have been presented in Sec. 2.2.2.3. As previously stated, MC correlates with a high computational effort. Thus, to reduce the computation load the Unscented Transformation is put in operation instead. Firstly, the overall variance, $\sigma^2(y^{sim})$, is determined by the Unscented Transformation related to monomials of precision 5. A total number of $2n^2+1$ sample points have to be evaluated and analysed. Subsequently, the evaluated samples can be reused to calculate the variance of the conditional expectation, $\underset{i}{\sigma}^2(\underset{-i}{E}[y^{sim}|\theta[i]])$, immediately. That means, the total number of function evaluations correlates to $2n^2+1$. In comparison to standard Monte Carlos methods, this indicates are significant reduction in computational load for many practical applications. By implementing the proposed strategy, precision demands are fulfilled automatically, i.e., determined variances are related to monomials of precision 5, whereas the expectations are associated to monomials of precision 3. Details can be found in App. A.2.

In summary, the Unscented Transformation renders the Global Sensitivity Analysis into a feasible approach which can be applied with manageable computational effort in many cases. Thus, the Global Sensitivity Analysis can be implemented not only for parameter ranking but also for more general approaches as well. For example,

Global Sensitivity Analysis might be used for model reduction purposes [QDSM11] or to provide a better insight into (bio)chemical processes [RWL+12].

2.2.5 Numerical Results

This section aims to demonstrate the proposed methods of uncertainty propagation. Starting with a single input single output problem the complexity is increased gradually up to multi-input multi-output expressions. In general, the performance of the Unscented Transformation is analysed in detail.

2.2.5.1 Sigmoid Function

Here, a sigmoid transfer function is reviewed, which transfers the variability about one random variable, ξ, to another one, η, according to

$$\eta = \frac{1}{1 + e^{-\xi}} \tag{2.69}$$

As illustrated in Fig. 2.6, the uncertainty propagation is subject of non-linearity. The proposed transfer function, even thought to represent an academic example, correlates to essential problems in modelling as well. For instance, a constrained parameter identification problem shows a similar behaviour. In this case, ξ represents some measurement data, y^{data}, and η is thought as an unknown parameter, θ, which has lower and upper constraints. Hence, the uncertainty propagation step describes a non-linear mapping of the uncertain data sample to the unknown parameter by an optimisation routine which has incorporated some penalty terms [Yen05] to account for the parameter constraints. Another important effect, especially in systems biology, is known as ultra sensitivity [HF96, Kho00]. That means, there is a tiny interval at which a variation of a factor, θ, leads to a strong variation in a related output, y^{sim}. Beyond this interval, an additional variation of θ has almost no impact on y^{sim}, i.e., a saturation emerges. In relation to the expression of Eq. (2.69), the random variable ξ correlates to θ, whereas y^{sim} is associated to η.

In a first step, the precision of the Unscented Transformation (UT5) is analysed in comparison to results based on linearisation as well as Monte Carlo simulation. The mean, $E[\eta]$, and variance, σ_η^2, are determined at different levels of uncertainty about ξ. As demonstrated in Tab. 2.2, the Unscented Transformation outperforms the Taylor series approach in precision while utilising a minimum number of sample points in comparison to Monte Carlo simulations.

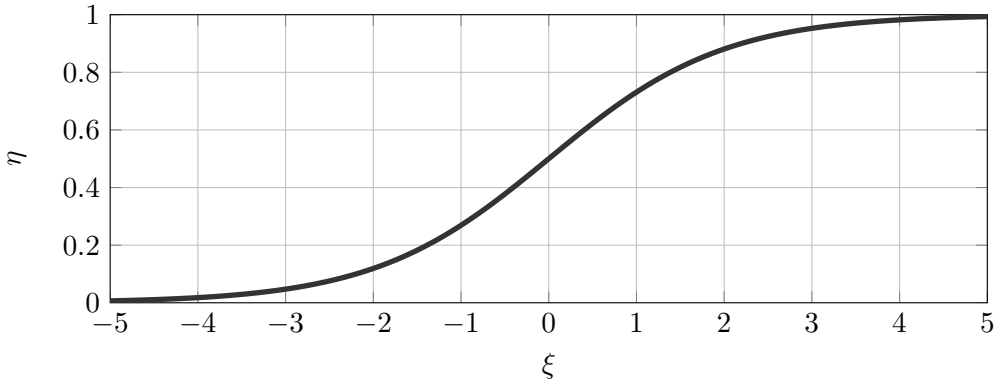


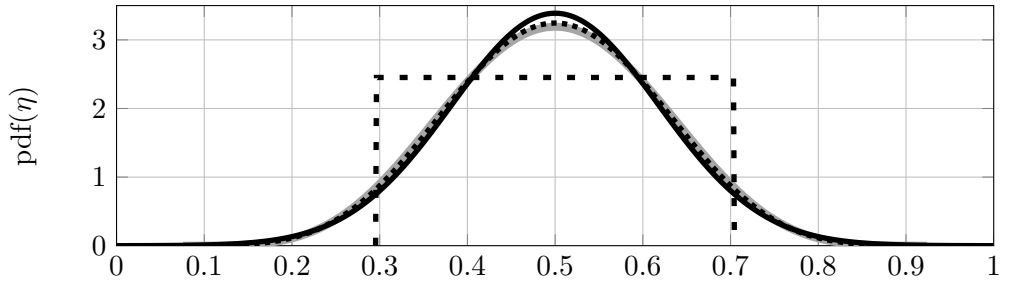
Figure 2.6: Illustration of the test case study: a sigmoid function

	Linearisation		UT5 (3 Samples)		MC (10^5 Samples)	
$\xi \sim$	$E[\eta]$	σ_η^2	$E[\eta]$	σ_η^2	$E[\eta]$	σ_η^2
$\mathcal{N}(0, 0.50)$	0.50	$1.56 \cdot 10^{-2}$	0.50	$1.39 \cdot 10^{-2}$	0.50	$1.40 \cdot 10^{-2}$
$\mathcal{N}(0, 1.50)$	0.50	$1.41 \cdot 10^{-1}$	0.50	$0.62 \cdot 10^{-1}$	0.50	$0.73 \cdot 10^{-1}$
$\mathcal{N}(0, 2.00)$	0.50	$2.50 \cdot 10^{-1}$	0.50	$0.74 \cdot 10^{-1}$	0.50	$0.98 \cdot 10^{-1}$
$\mathcal{N}(1, 1.25)$	0.73	$6.04 \cdot 10^{-2}$	0.69	$4.73 \cdot 10^{-2}$	0.68	$4.76 \cdot 10^{-2}$

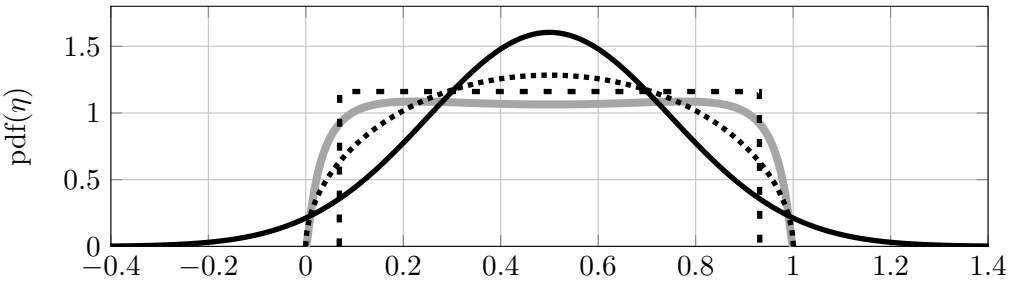
Table 2.2: Benchmark study of approximation methods: Linearisation vs. Unscented Transformation at different input uncertainty configurations, ξ. Monte Carlo simulations are used for assessment purposes. Obviously, the Unscented Transformation (UT5) provides credible results by evaluating a very low sample number.

Obviously, the mean and variance of η can be determined satisfactorily. A Gaussian distribution based on these two quantities, $\eta \sim \mathcal{N}(E[\eta], \sigma_\eta^2)$, might be poor representative of the actual distribution, see Fig. 2.7(a)-(d). Here, different types of PDF which are parametrised by $E[\eta]$ and σ_η^2 are more suitable candidates. As previously mentioned, the approximated skewness and kurtosis might be used to select the most plausible PDF candidate.

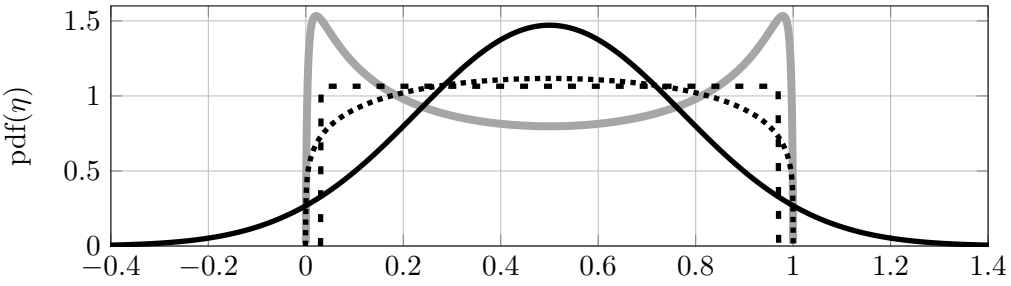
Naturally, the assumption of a Gaussian distribution in relation to η is likely to introduce errors in confidence intervals, too. In the particular case of Eq. (2.69), η is limited to $\eta \in [0\ 1]$ by definition. Therefore, any plausible confidence interval should not exceed these bounds. The assumption of a Gaussian distribution, however, leads to a confidence interval of $CI = (0.03, 1.34)$ for $\xi \sim \mathcal{N}(1, 1.25)$. As an alternative, the Cornish-Fisher expansion [SA11] is applied to account for the approximately determined skewness and kurtosis. In doing so, the confidence interval is corrected to $CI = (0.10, 0.95)$, which is in much better agreement to the PDF shown in Fig. 2.7(d).



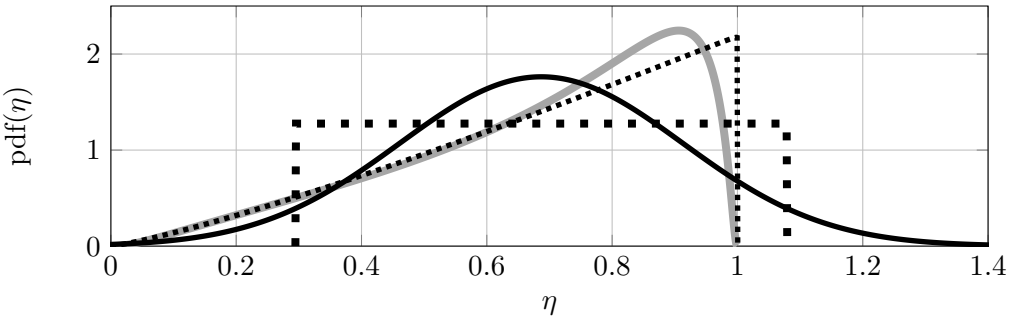
(a) $\xi \sim \mathcal{N}(0, 0.50)$



(b) $\xi \sim \mathcal{N}(0, 1.50)$



(c) $\xi \sim \mathcal{N}(0, 2.00)$



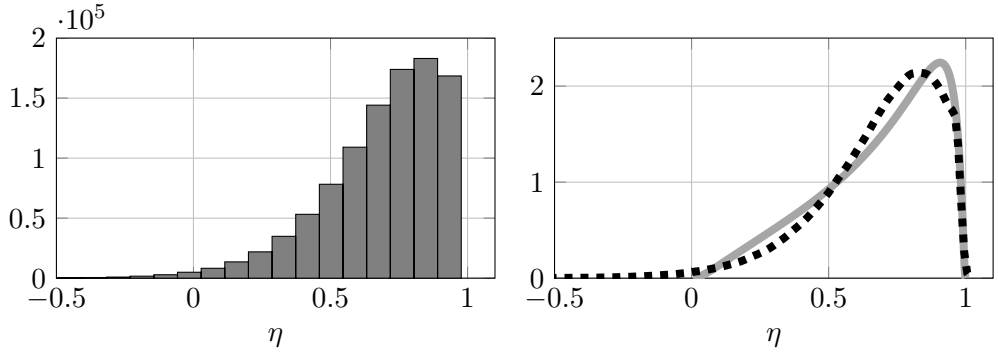
(d) $\xi \sim \mathcal{N}(1, 1.25)$

Figure 2.7: Illustration example of an UT-based PDF approximation: For different input configurations, $\xi \sim \mathcal{N}(E[\xi], \sigma_\xi)$, the UT3 approach is applied to propagate the associated mean and variance, respectively. Subsequently, different PDF candidates are parametrised by these two statistical quantities, i.e., (——) Normal PDF, (▪ ▪) Uniform PDF, and (▪▪▪▪▪) Beta PDF. The true PDF is given by (——).

	a_0	a_1	a_2	a_3
UT5	0.6869	0.2084	-0.0883	0.0000
MC	0.6834	0.2133	-0.0640	-0.0581
$\frac{UT5-MC}{MC}$	0.51%	-1.37 %	37.96%	-100%

Table 2.3: Precision of the determined PCE coefficients: Unscented Transformation (UT5) vs. Monte Carlo simulation (MC).

In cases, however, where the primary interest is in good approximation of pdf_η exclusively, one might implement the framework of Polynomial Chaos Expansion directly. As shown in Sec. 2.2.2.4, the most crucial part in PCE is to derive reliable coefficients, a_i, via numerical integration routines. Assuming $l_{pce} = 3$, an overall number of four integral expressions according to Eq. (2.25) have to be determined. Hence, the UT5 approach can be put in operation for this purpose as well. In Tab. 2.3, the outcome of this procedure is given. The accuracy of the determined coefficients, a_i, decreases for coefficients of increasing order gradually, because the nonlinear characteristic increases as well. Nevertheless, the associated surrogate function based on PCE provides a good approximation of the actual PDF as shown in Fig. 2.8.



(a) Histogram (b) Probability density function

Figure 2.8: Probability density approximation via the PCE approach: Once the PCE-based surrogate function is determined it can be "cheaply" evaluated by Monte Carlo simulations resulting in the shown histogram (a). Subsequently, Kernel-based approaches might be put in operation to determine the associated probability density function, pdf_η. In (b), the resulting PDF (▪▪▪) is shown in relation to true PDF (▬).

2.2.5.2 An n-Dimensional Input Problem

In this subsection, a test case example which has been presented in [GH12] is revisited. In detail, the following non-linear mapping is analysed

$$\eta = g(\xi) = \xi^T \xi, \quad \xi \sim \mathcal{N}(0_n, I_{n\times n}) \tag{2.70}$$

Here, $I_{n\times n}$ represents the identity matrix. Thus, η is defined as a superposition of n squared standard Gaussian random variables. In this particular case, the statics about η follows a χ^2-distribution [HMGB03b]. Thus, the mean, $E[\eta]$, and the variance, σ_η^2, can be determined analytically [HMGB03b] in accordance to

$$E[\eta] = n \tag{2.71}$$

$$\sigma_\eta^2 = 2n \tag{2.72}$$

As shown in Sec. 2.2.3, for the purpose of calculating the variance, σ_η^2, via sample based approaches also $g(\xi)^2$ has to be approximated properly. In doing so, monomials up to order 4 have to be addressed correctly. Thus, applying the UT3 approach leads to an approximation error about σ_η^2, whereas the mean, $E[\eta]$, is determined correctly, because only monomials of order 2 are involved. To ensure a correct result of the variance, too, the UT5 approach has to be put in operation as demonstrated in Tab. 2.4.

	True	$TS1$	$TS2$	$UT3(\beta=0)$	$UT3(\beta=2)$	$UT3(\beta=1)$	$UT5$
$E[\eta]$	n	0	n	n	n	n	n
σ_η^2	$2n$	0	$2n$	$3n-n^2$	$3n+n^2$	$3n$	$2n$

Table 2.4: Analytical results about $E[\eta]$ and σ_η^2, respectively, for different approximation schemes.

By comparing different correction factors, β, it is obvious that the approximation error of UT3 can be reduced significantly. For example, by applying no correction factor, $\beta = 0$, or alternatively $\beta = 2$ as recommended in [JU04] the approximation error scales quadratically in relation to the dimension n, see App. A.1 for more details. By implementing $\beta = 1$, however, a good approximation performance can be achieved even for UT3 as shown in Fig. 2.9. Thus, a proper choice of β is an essential component to ensure reliable results in the calculation of variances by the UT3 approach.

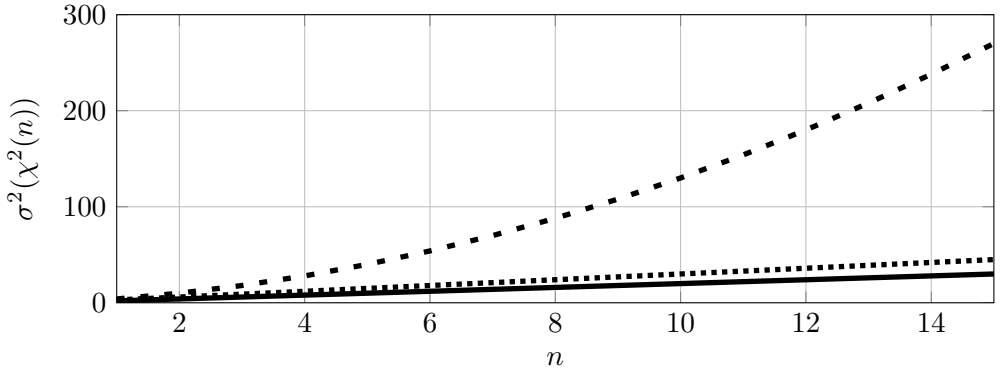


Figure 2.9: Variance of $\chi^2(n)$ for an increased number of dimension, n: The associated variances, $\sigma^2(\chi^2(n))$, are approximated via (- -) UT3, (·····) UT3e, and (——) UT5. Here, the UT5 approach provides the correct result.

2.2.5.3 Gompertz Function

The intention of this example is to demonstrate the performance of the Unscented Transformation approach in the presence of non-Gaussian distributions. As stated in Sec. 2.2.3, distribution-specific transfer functions, $q(\xi)$, might be utilised for this purpose. Here, the induced uncertainty about the time-dependent outcome of the so-called Gompertz function is analysed

$$\eta(\xi', t) = \xi'[1]e^{-\xi'[2]e^{-\xi'[3]t}} \tag{2.73}$$

The elements of the random vector, $\xi' \in \mathbb{R}^3$, are associated to a non-standard Gaussian, a Uniform, and Log-normal distribution, respectively. The detailed specifications of the applied distributions are given by

$$\xi'[1] \sim \mathcal{N}(1, 0.05) \tag{2.74}$$

$$\xi'[2] \sim \mathcal{U}(3, 9) \tag{2.75}$$

$$\xi'[3] \sim \ln\mathcal{N}(0.05, 0.3) \tag{2.76}$$

In a subsequent step, the UT3 approach in combination with the appropriate transfer function, $q(\xi)$, given in Tab. 2.1 are applied to determine the mean, $E[\eta(t)]$, and the variance, $\sigma_\eta^2(t)$. The numerical results are illustrated in Fig. 2.10. In comparison to Monte Carlo simulations, the proposed concept provides working results, i.e, the uncertainty about $\eta(t)$ is approximated properly by a minimum of computational load.

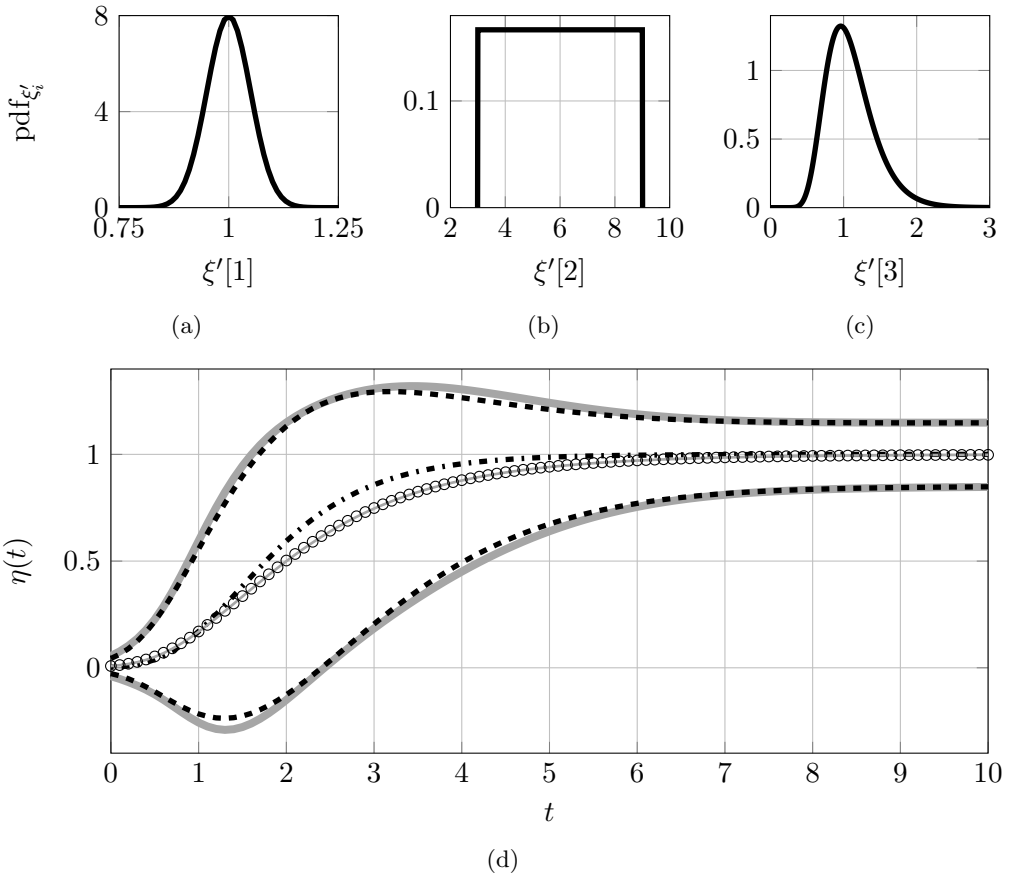


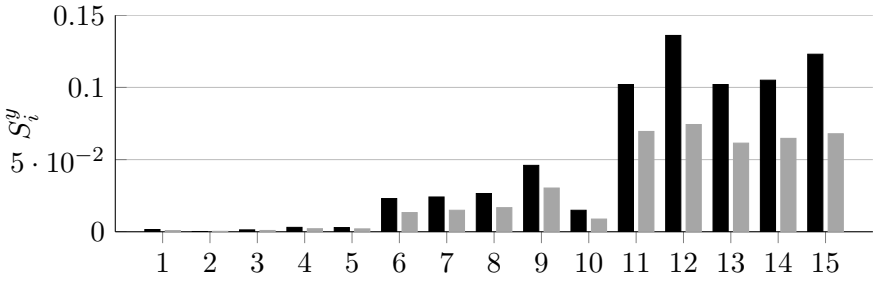
Figure 2.10: Uncertainty about the Gompertz function: In the first row, (a)-(c), the independent distributions of elements of the random vector, ξ, are presented. In (d) the simulation result is shown. Here, (·-·-·) relates to the deterministic outcome given by $\eta(\overline{\xi'}, t)$. More adequate inferences can be done taking the dispersion of ξ' into account explicitly as demonstrate by the mean, $E[\eta]$, and the associated 99%-confidence interval (CI). In detail, MC results are related to: $E_{MC}[\eta]$ (——) and CI_{MC} (▬). Where quantities determined by UT3 are indicated by: $E_{UT3}[\eta]$ (○) and CI_{UT3} (- - -), respectively.

2.2.5.4 Case Study of the Global Sensitivity Analysis

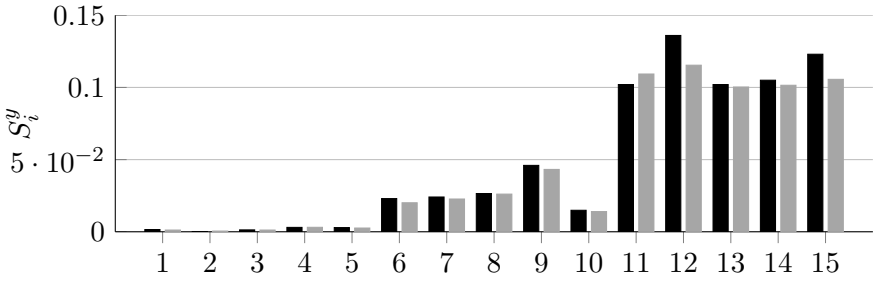
Finally, the usefulness of the Unscented Transformation in the framework of Global Sensitivity Analysis is presented. For this purpose the so-called O'Hagan & Oakley function is implemented according to

$$\eta = a_1^T \xi + a_2^T \cos(\xi) + a_3^T \sin(\xi) + \xi^T M \xi, \quad \xi \in \mathbb{R}^{15} \tag{2.77}$$

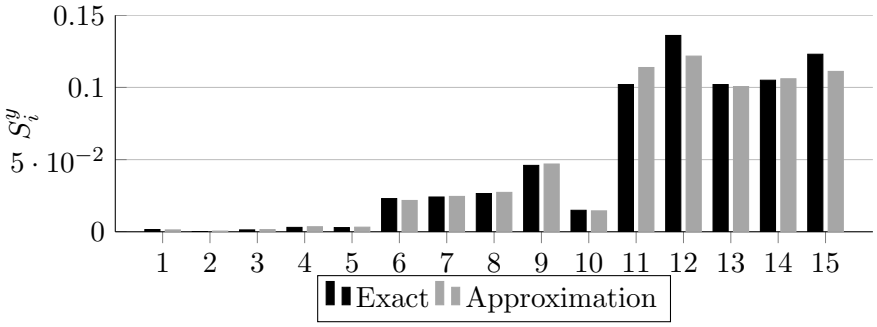
The numerical values associated to the parameter vectors, a_i, as well as to the matrix M are given in App. A.2. In addition, the elements of the random vector, ξ, are assumed to be independent and to follow a standard Gaussian distribution, $\xi \sim \mathcal{N}(0_{15}, I_{15\times 15})$. The O'Hagan & Oakley function is frequently used as a benchmark problem for methods in global sensitivity analysis, because the Sobol' indices can be derived analytically [OO04a]. Moreover, this function imitates a behaviour which can be found in various models in the field of systems biology, namely the sloppiness of model parameters [GWC+07]. This feature characterises the omnipresence of three different groups of parameter sensitivities, i.e., one can find a group of parameters with low ($\xi[1] - \xi[5]$), medium ($\xi[6] - \xi[10]$), and high ($\xi[11] - \xi[15]$) sensitivity, respectively. To ensure predictive simulations, parameters of high sensitivity are of special interest in parameter identification. Therefore, any helpful method in sensitivity analysis has to classify the model parameters into the mentioned groups reliably, whereas the exact numerical value of an individual sensitivity index might be of less importance. As demonstrated in Fig. 2.11 the proposed concept of Sec. 2.2.4 fulfils this desired property satisfactorily, i.e., the elements of ξ a correctly grouped. Moreover, the determined Sobol' indices are in good agreement in comparison to the analytical results. In detail, the utilisation of a total number of $2 \cdot 15^2 + 1 = 451$ sample points provides a good approximation while keeping the computation load to a minimum level. For example, generally more than 1000 sample points are applied in the native Monte-Carlo simulation framework to determine the Sobol' indices of first order for only one input factor. Therefore, the total number of function evaluations exceeds the value 15,000 in this particular case easily. In conclusion, this academic example demonstrates impressively that the Unscented Transformation is a promising approach to increase the applicability of Global Sensitivity Analysis. Thus, GSA might become more competitive in comparison to the standard approach of local sensitivity analysis which is based on linearisation, but still dominating the sensitivity analysis in systems biology [DBR10a].



(a) Applied total sample number: 451



(b) Applied total sample number: 451



(c) Applied total sample number: 12241

Figure 2.11: The approximated Sobol' indices, S_i^y, are compared to the exact results. Different approximation schemes are used: In (a), $\sigma^2(y^{sim})$ by UT3, $\underset{i}{\sigma}^2(\cdot)$ by UT5, and $\underset{-i}{E}[y^{sim}|\theta[i]]$ by UT3. In (b), $\sigma^2(y^{sim})$ by UT5, $\underset{i}{\sigma}^2(\cdot)$ by UT5, and $\underset{-i}{E}[y^{sim}|\theta[i]]$ by UT3. In (c), $\sigma^2(y^{sim})$ by UT5, $\underset{i}{\sigma}^2(\cdot)$ by UT5, and $\underset{-i}{E}[y^{sim}|\theta[i]]$ by UT5. Thus, the configuration of (b) shows the best ratio of approximation power and computational load.

2.3 Chapter Summary

In this chapter, the problem of uncertainty quantification as well as propagation is addressed. The focus is on various methods of uncertainty propagation techniques, including standard approximate concepts, e.g., Taylor series, Gaussian quadrature, Monte Carlos simulation, and Polynomial Chaos expansion. Additionally, the Unscented Transformation approach for the purpose of uncertainty propagation is examined in more detail.

It is demonstrated, how to incorporate non-Gaussian and/or dependent distributions in the framework of UT. Here the transformation concept, i.e., the introduction of an additional transfer function which maps a standard Gaussian distribution to any desired distribution, is a versatile tool for a broad class of real-life problems.

For the purpose of demonstration, comparative studies are performed using illustrative examples. Although a limited number of examples is analysed, some general conclusions can be given. It is obvious that UT provides good approximations in comparison to Taylor series of first order, but utilising a minimum level of computation load in relation to Monte Carlo simulations or Gaussian quadrature. Therefore, the Unscented Transformation is likely to fill the gap of accuracy demands and computational efficiency in a number of practical applications.

Apart from the mathematical details and benchmark studies, some comments about the origin of the Unscented Transformation are given. Here the so-called Point Estimated Methods are the direct precursors of the Unscented Transformation, but rarely or never addressed in recent works. As a consequence, some redundancy can be found in the literature, i.e., results well known in the field of PEMs are rediscovered in the field of UT. Thus, a proper renaming of UT might be considered even though this involves a great loss to the figurative language:

> “For nonlinear filtering problems where the nonlinearity is severe compared to the prior state information, the classical extended Kalman filter (EKF) “stinks” compared to the unscented Kalman filter (UKF), which has been concluded in a large number of application.
> *F. Gustafsson & G. Hendeby [GH12]* ”

3

OED for Parameter Identification

After the uncertainty of a quantity of interest has been determined appropriately, the Optimal Experimental Design (OED) for parameter identification aims to minimise this uncertainty in case of need, i.e., when the associated uncertainty prohibits any meaningful inferences. The origin of OED dates back to the pioneer work of R. A. Fisher [Fis35]. At the very first beginning of OED the focus had been on empirical models, i.e., data-driven model concepts had been subject of research. In recent decades the principles of OED have been applied for mechanistic model approaches as well. For example, in the field of systems biology a vast number OED applications can be found [BSSR94, BCRFB07, FM08, KT09, DBI10]. In most cases, however, OED applied for nonlinear problems is based on linearisation without any proof of credibility. Thus, the focus in this thesis is on the influence of approximation errors to OED results. Before test case studies are presented some basics about OED are given below.

First of all, the term "experimental design" may need some explanation:

> "A schedule of experimental conditions is referred to as an experimental design."
> *G. E. P. Box & N. R. Draper [BD87]*

Hence the schedule may cover measurement specifications:

- When to measure?
- What to measure?
- Where to measure?
- How to measure?,

and general operating conditions as well:

- How to choose initial conditions?
- How to steer the system?

To give appropriate answers to these problems, a proper cost function has to be defined. Generally, this cost function represents the uncertainty about the quantity of interest. In this context, the experimental condition which provides the lowest value of the cost function is labelled as the "optimal experimental design". In conclusion, a suitable cost function is a key element in OED and needs some more detailed examination.

3.1 Definition of Cost Functions

In most practical cases, the Fisher Information Matrix (FIM) is put in operation to quantify the uncertainty about estimated model parameters, $\hat{\theta} \in \mathbb{R}^l$. For a set of measurement sample time points, $y^{data}(t_k),\ \forall k = 1, \ldots, K$, and the associated measurement (co)variance matrix, $C_{y(t_k)}$, the FIM is defined by

$$FIM = \sum_{t_k}^{K} SM_{t_k}^T \cdot C_{y(t_k)}^{-1} \cdot SM_{t_k} \tag{3.1}$$

Here, the sensitivity matrix, SM_{t_k}, reads

$$SM_{t_k} = \begin{bmatrix} \left.\frac{\partial y_1(t_k)}{\partial \theta_1}\right|_{\hat{\theta}_1} & \left.\frac{\partial y_1(t_k)}{\partial \theta_2}\right|_{\hat{\theta}_2} & \cdots & \left.\frac{\partial y_1(t_k)}{\partial \theta_l}\right|_{\hat{\theta}_l} \\ \left.\frac{\partial y_2(t_k)}{\partial \theta_1}\right|_{\hat{\theta}_1} & \left.\frac{\partial y_2(t_k)}{\partial \theta_2}\right|_{\hat{\theta}_2} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \left.\frac{\partial y_m(t_k)}{\partial \theta_1}\right|_{\hat{\theta}_1} & \cdots & \left.\frac{\partial y_m(t_k)}{\partial \theta_{l-1}}\right|_{\hat{\theta}_{l-1}} & \left.\frac{\partial y_m(t_k)}{\partial \theta_l}\right|_{\hat{\theta}_l} \end{bmatrix} \tag{3.2}$$

In case of ODEs, a matrix differential equation system for the sensitivities has to be solved in parallel to the original model

$$\dot{SM} = \frac{\partial f}{\partial y} \cdot SM + \frac{\partial f}{\partial \theta} \quad ; \quad SM(0) = 0_{m \times l} \tag{3.3}$$

Subsequently, the inverse of FIM in relation to the Cramer-Raó inequality [Kay93] provides a lower bound of the parameter covariance matrix, $C_{\hat{\theta}}$, according to

$$C_{\hat{\theta}} \geq \frac{\partial g\left(\theta\right)}{\partial \theta} FIM^{-1} \frac{\partial g\left(\theta\right)^T}{\partial \theta}, \tag{3.4}$$

where $g\left(\theta\right) = \theta + Bi(\hat{\theta})$. Here, the bias term, $Bi(\hat{\theta}) = E[\hat{\theta}] - \theta$, describes a systematic deviation of the estimated to the true parameter values. The equality only holds, if (i) the measurement errors are additive, and (ii) the model is linear in its parameters.

Moreover, the FIM does not give any information on $E[\hat{\theta}]$. Therefore, in many cases it is additionally assumed that (iii) the estimates are unbiased, $E[\hat{\theta}] = \theta \;\rightarrow\; Bi(\hat{\theta}) = 0$. Consequently, the parameter covariance matrix is approximated by

$$C_{\hat{\theta}} \approx FIM^{-1} \tag{3.5}$$

In practical applications, however, the FIM approach may lead to poor approximations of $C_{\hat{\theta}}$. Obviously, the Fisher Information Matrix is based on the same linearisation principles as shown in Sec. 2.2.2.1. Here, too, linear surrogates of non-linear problems might be poor representatives, i.e., results which are based on linearisation are likely to fail. Moreover, the derivatives in FIM have to be evaluated at parameter values, θ, which are actually unknown. A remedy might be the following procedure which is frequently implemented in practice:

> “... a prior guess $\hat{\theta}^0$ for θ is used to design the experiment, with the hope that the local optimal design for $\hat{\theta}^0$ will be close to the optimal one for the unknown θ. When the alternation of estimation and design phases is possible, sequential design permits to progressively adapt the experiment to an estimated value of θ that (hopefully) converges to its unknown true value.” *A. Pázman & L. Pronzato [PP07]*

To avoid, or at least to reduce, costly reiterations of parameter estimation and design phases, robust OED strategies based on min-max optimisation principles have been derived [KKBS04, TLDI12a, TLDI12b]. Here the parameter uncertainty is addressed explicitly, i.e., the parameters are characterised by confidence intervals. Hence, the FIM is not evaluated at a single parameter vector of estimates exclusively. Instead, a bounded parameters space is explored in parallel searching for parameter values at which OED has the worst performance in spite of optimal operating conditions. That means, the worst-case scenario in relation to the unknown model parameters, $\theta \in \Theta$, is solved for OED.

In the particular case of FIM-based OED, however, even a proper choice of model parameters is no guarantee of optimally designed experiments for non-linear problems. As shown in Sec. 3.2, the approximation error due to linearisation may provide sub-optimal results as well, see [BW04, JSMK06, VG07] for confirming references. Thus, an improvement in the approximation accuracy is likely to provide more suitable operating conditions, i.e., to provide more informative data. Nevertheless, in the field of

systems biology, even today, the implementation of the Fisher Information Matrix is state-of-the-art in OED.

Obviously, the Unscented Transformation approach is an appropriate alternative to determine the uncertainty about the estimated model parameters, $\hat{\theta}$. As demonstrated in Sec. 2.2.3, the UT method provides approximations about the mean, $E[\hat{\theta}]$, and the covariance matrix, $C_{\hat{\theta}}$, respectively. Thus, the mean square error matrix of the estimated parameters, $MSE_{\hat{\theta}}$ [Kay93], can be determined via

$$MSE_{\hat{\theta}} = E[(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T] \approx C_{\hat{\theta}} + Bi(\hat{\theta})Bi(\hat{\theta})^T \tag{3.6}$$

Consequently, instead of analysing the covariance matrix exclusively, $MSE_{\hat{\theta}}$ is processed in OED. But independently of the analysed uncertainty matrix, $C_{\hat{\theta}}$ or $MSE_{\hat{\theta}}$, numerical optimisation routines require an adequate representation of these matrices.

Thus, for the purpose of numerical implementation, a scalar utility function of the approximated parameter uncertainty matrix has to be derived. Instead of solving a multi-objective optimisation problem which aims to minimise every element of $C_{\hat{\theta}}$ or $MSE_{\hat{\theta}}$, respectively, a one-dimensional compromise function is used. Therefore, well known optimality criteria exist in literature and are frequently applied in practice [WP97], e.g.,

$$A - optimal\ design \quad \Phi_A(C_{\hat{\theta}}) = trace(C_{\hat{\theta}}) \tag{3.7}$$

$$D - optimal\ design \quad \Phi_D(C_{\hat{\theta}}) = det(C_{\hat{\theta}}) \tag{3.8}$$

$$E^* - optimal\ design \quad \Phi_{E^*}(C_{\hat{\theta}}) = \frac{\lambda_{max}(C_{\hat{\theta}})}{\lambda_{min}(C_{\hat{\theta}})} \tag{3.9}$$

with λ_{max} (λ_{min}) as the maximum (minimum) eigenvalue of $C_{\hat{\theta}}$. Please note that $C_{\hat{\theta}}$ might be replaced by $MSE_{\hat{\theta}}$ in case of the UT approach. In general, however, the choice of the design criterion influences the outcome of OED and it is not clear in advance which criterion will produce the best result, i.e., which criterion will produce the most precise parameter estimates. Naturally, at this point, it raises the question whether it is really necessary to identify all model parameters with the same accuracy.

As stated previously, the objective of most mathematical models is to provide reliable simulation results, i.e., to provide meaningful model-based inferences. Therefore, a cost function which is based on the parameter precision exclusively is usually not an ideal

option for non-linear models. Admittedly, in the special case of linear problems there might be no differences between these two objectives. Both are equivalent objectives under stringent conditions, see [KW60, Won94, Won95] and references therein. For non-linear problems, however, this equivalence is not any longer a valid assumption [BHC+04, GWC+07]. Only a subset of the unknown model parameters and combinations thereof determine the qualitative behaviour of the model. Thus, when the main interest lies in obtaining a predictive model but not in identifying certain parameter values, it is not necessary to reduce all parameter uncertainties by the same amount.

In consequence, OED should provide models with tight confidence intervals of simulation results, which are indirectly influenced by the imperfect parameter identification process. Here, the universal concept of the UT method provides also an appropriate and elegant way to take these considerations into account, see Sec. 2.2.3.3. In detail, the UT approach provides the mean and the (co)variance matrix of the states of the simulated time interval. A suitable cost function taking the uncertainty about $x(t)$ into account is

$$\Phi_{MSE_x}^{UT} = trace\left(\int_{t_0}^{t_{end}} MSE(\mathbf{x}(t,\hat{\theta},u))dt\right). \tag{3.10}$$

Here, the mean square error matrix of the states is approximated according to

$$MSE(x(t,\hat{\theta},u)) = E[(x(t,\hat{\theta},u)-x(t,\theta,u))(x(t,\hat{\theta},u)-x(t,\theta,u))^T] \approx C_{\hat{x}(t)}^{UT}+Bi_{\hat{x}(t)}Bi_{\hat{x}(t)}^T \tag{3.11}$$

In doing so, the $MSE(x(t,\hat{\theta},u))$ is evaluated at validation conditions, i.e., conditions which have not been part of a former parameter identification process. Hence, by minimising $\Phi_{MSE_x}^{UT}$ the uncertainty about the most sensitive parameters and parameter combinations is reduced automatically. The correlation of global parameter sensitivities and the mean square error of prediction has been analysed empirically in [LML+09, LMM10].

3.2 Demonstration of FIM-based OED Inaccuracy

The following academic example demonstrates potential pitfalls of the Fisher Information Matrix in the field of Optimal Experimental Design. In detail, the test case

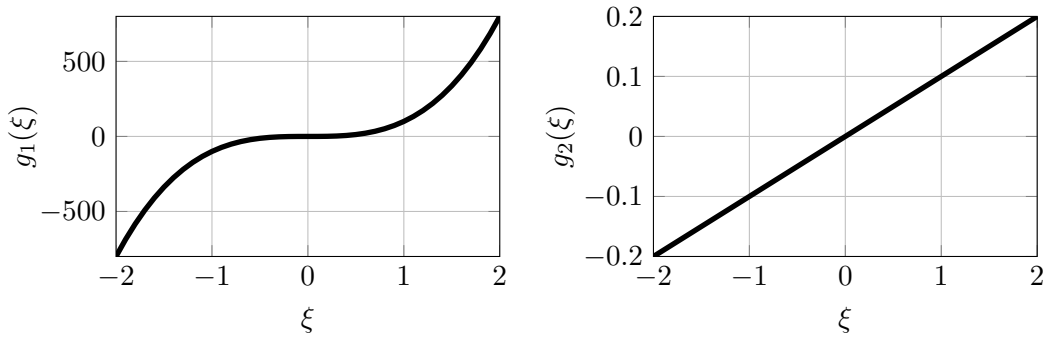
problem is defined as

$$g_1(\xi) = a \cdot \xi^3 \tag{3.12}$$

$$g_2(\xi) = b \cdot \xi \tag{3.13}$$

$$\eta(\xi) = w \cdot g_1(\xi) + (1 - w) \cdot g_2(\xi) \tag{3.14}$$

Hence, the function of interest, $\eta(\xi)$, is a weighted combination of a nonlinear function, $g_1(\xi)$, and a linear function, $g_2(\xi)$, respectively, see Fig. 3.1.



(a) (b)

Figure 3.1: Illustration of the test case function, $\eta(\xi) = w \cdot g_1(\xi) + (1-w) \cdot g_2(\xi)$, separated in its nonlinear (a) and its linear (b) component, respectively.

Additionally, it is assumed that the random input variable, ξ, follows a standard Gaussian distribution. As mentioned previously, OED aims to minimise the uncertainty about, $\eta(\xi)$, by an optimal choice of available design variables, ζ. Applied to the example given in Eq. (3.14) the objective of OED is to find the most appropriate value of the weighting factor, w, which provides the lowest uncertainty about $\eta(\xi)$, i.e., to figure out the optimal design value, $\zeta = w \in [0\ 1]$, which minimises the variance, σ_η^2. Therefore, an optimisation problem is defined according to

$$\arg\min_{\zeta} \ \sigma_\eta^2(\zeta); \quad 0 \leq \zeta \leq 1 \tag{3.15}$$

Naturally, the outcome of this optimisation problem depends on the precision of the approximated variance, σ_η^2, critically. A worse approximation of the variance, σ_η^2, might cause sub-optimal or even adverse optimisation results. For the purpose of demonstration, the optimisation problem of Eq. (3.15) is solved for different approximate methods, i.e., the Fisher Information Matrix, the Unscented Transformation, as well as Monte Carlo simulations are put in operation. In addition, different test case scenarios are analysed. In detail, the function parameters, a & b, are modified to provide a low,

a medium, and a significant impact of the nonlinearity to $\eta(\xi)$, respectively.

The correlation of the variance, σ_η^2, and the design variable, ζ, at the three different scenarios are illustrated in Fig. 3.2. Obviously, in all three test cases the optimisation problem has an unique optimal solution, which should be detected by the applied approximate methods, too. In Tab. 3.1 the results of this benchmark study are summarised. For a low level of nonlinearity all three methods provide the same optimal value, $\zeta = 1$. In case of an increased nonlinear impact, the FIM which is based on linearisation principles, fails completely. That means, suboptimal/adverse design variables are calculated. The Unscented Transformation, however, provides for all cases suitable design variables with a minimum effort in computational load compared with the Monte Carlo approach which is applied as a reference method.

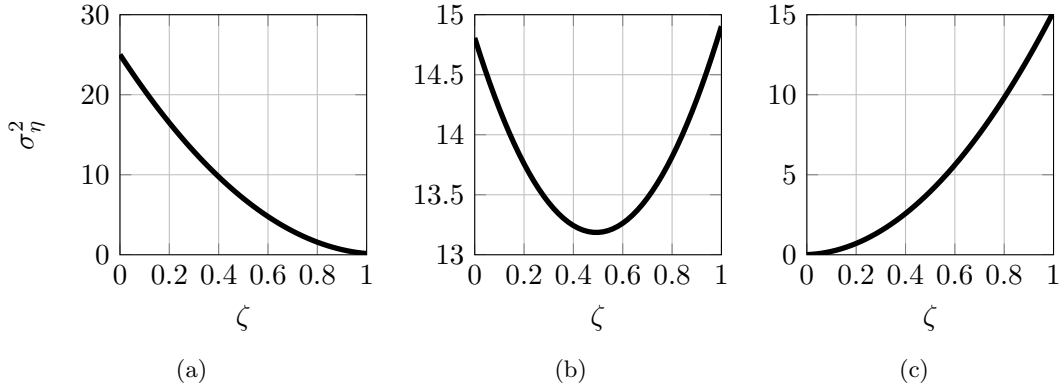


Figure 3.2: Outcome of the benchmark study: range of weighting/design factors, ζ, in comparison to the resulting variance, σ_η^2. Here, the lowest entry of σ_η^2 corresponds to the optimal choice of ζ. Three different scenarios are illustrated: In (a) the linear behaviour is dominating ($a = 0.10, b = 5.00$). In (b) the nonlinear impact is slightly increased ($a = 1.00, b = 3.85$). Finally, the nonlinearity dominates in (c) ($a = 1.00, b = 0.10$).

	MC		UT5		FIM	
(a, b)	ζ	σ_η^2	ζ	σ_η^2	ζ	σ_η^2
$(0.10, 5.00)$	1.00	0.15	1.00	0.15	1.00	0.15
$(1.00, 3.85)$	0.51	13.28	0.62	13.42	1.00	15.11
$(1.00, 0.10)$	0.00	0.01	0.00	0.01	1.00	14.82

Table 3.1: Outcome of the benchmark study : optimal weighting/design factor, ζ, in comparison to the resulting variance, σ_η^2.

3.3 Single-Substrate Uptake Model[1]

In this study a single-substrate uptake model is of special interest. In detail, the growth of biomass, $c_B[gl^{-1}]$, which is induced by a substrate uptake, $c_S[gl^{-1}]$, is simulated for a continuous stirred tank bio-reactor. By assuming that the inlet flow, $q_{in}[lh^{-1}]$, and the outlet flow, $q_{out}[lh^{-1}]$, of the reactor are equal, ($q_{in} = q_{out} =: q[lh^{-1}]$), the following ODE system can be derived

$$\dot{c}_B = \mu \cdot c_B - D \cdot c_B \tag{3.16}$$

$$\dot{c}_S = -\frac{1}{Y_{B|S}} \cdot \mu \cdot c_B + (c_{s,in} - c_S) \cdot D \tag{3.17}$$

Here, the dilution rate is defined as $D = \frac{q}{V}[h^{-1}]$, and the specific growth rate, μ, is determined by a Monod kinetics as

$$\mu = \frac{\mu_m \cdot c_S}{K_s + c_S}. \tag{3.18}$$

The simplicity of this unstructured growth model does not provide a significant insight into biological mechanisms, but it seems appropriate to demonstrate characteristic problems in determining and minimising the parameter covariance matrix for models that are non-linear in their parameters. Obviously, the model has three parameters: (i) $Y_{B|S}$ - the yield factor describes how much biomass is produced by the uptake of a certain amount of substrate, (ii) μ_m - the maximum growth rate is the upper limit of the growth rate μ, and (iii) K_s - the substrate affinity constant represents the substrate concentration at which the specific growth rate is half its maximum value. For reasons of simplification, the following assumptions are made: (i) only the concentration of biomass is measurable $y(t_k) = c_B(t_k)$, and (ii) $Y_{B|S}$ is known from literature, consequently only the parameters of the Monod kinetics, μ_m and K_s, have to be identified, which results in a two-dimensional parameter space.

3.3.1 Parameter Identifiability

Before starting the parameter identification, it seems reasonable to check whether the unknown parameters can be determined in principle for perfect measurement information, i.e., measurement data are continuous in time and not effected by measurement noise, respectively. In this very special case, the measurement data at time point zero,

[1] A large majority of results contained in this section have already been published in the peer-reviewed literature, [SKM09b].

$y^{data}(0)$, as well as associated time derivatives, $\dot{y}^{data}(0), \ddot{y}^{data}(0), \ldots$, up to some arbitrary order are available in theory [WP97]. Hence, by assuming $y^{data}(0) = c_x(0)$ the following equation system can be derived in relation to Eq. (3.16) & (3.17)

$$
\begin{aligned}
y^{data}(0) &= c_x(0) && (3.19)\\
\dot{y}^{data}(0) &= \dot{c}_B(0) = (\mu - D) \cdot c_{x_0} = (\mu - D) \cdot y^{data}(0) && (3.20)\\
\ddot{y}^{data}(0) &= (\mu - D)\dot{y}^{data}(0) + \frac{d\mu}{dc_S} \cdot \dot{c}_S(0) \cdot y^{data}(0) && (3.21)\\
&= \frac{\dot{y}^{data}(0)^2}{y^{data}(0)} + y^{data}(0)\frac{d\mu}{dc_S}\left(-\frac{1}{Y_{B|S}} \cdot \mu \cdot y^{data}(0) + D(c_{s,in} - c_S)\right) && (3.22)
\end{aligned}
$$

In addition, it is assumed that the initial substrate concentration, $c_S(0)$, is known. Hence, Eq. (3.20) and (3.22) can be rearranged to

$$
\begin{aligned}
\mu|_{t=0} &= D + \frac{\dot{y}^{data}(0)}{y^{data}(0)}\\
\left.\frac{d\mu}{dc_S}\right|_{t=0} &= \frac{\ddot{y}^{data}(0) - \frac{\dot{y}^{data}(0)^2}{y^{data}(0)}}{y^{data}(0)\left(-\frac{1}{Y_{B|S}} \cdot (D \cdot y^{data}(0) + \dot{y}^{data}(0)) + D(c_{s,in} - c_S(0))\right)} =: \mu'(0)
\end{aligned}
$$

Obviously, it is possible to express $\mu|_{t=0}$ and $\left.\frac{d\mu}{dc_S}\right|_{t=0}$ by the known quantities, $y(0)$, $\dot{y}(0)$, $\ddot{y}(0)$, and $c_S(0)$, analytically. Thus, from the definition of μ it follows that also K_s and μ_m can be expressed uniquely in the following way

$$
\begin{aligned}
K_s &= \frac{c_S(0)^2}{\frac{\mu(0)}{\mu'(0)} - c_S(0)} && (3.23)\\
\mu_m &= \frac{\mu(0) \cdot (K_s + c_S(0))}{c_S(0)} && (3.24)
\end{aligned}
$$

Therefore, at least in principle, it is possible to identify the unknown parameters, K_s and μ_m, by biomass measurements.

3.3.2 Optimal Experimental Design

Here, a benchmark study compares the outcome of the FIM-based OED in relation to Unscented Transformation method. Therefore, the following assumptions are made concerning the single-substrate uptake model:

- the measurement of the biomass concentration is taken at three time points $t_k =$ [0.5 1.0 1.5]h;

- artificial measurement data are used, i.e., data are obtained from a reference simulation and are corrupted by an artificial Gaussian noise additionally.

The reference values of the parameters are $K_s = 2$ and $\mu_m = 5$. Moreover, the variance of the measurement noise is set to $C_y = 10^{-4}[g^2 l^{-2}]$. Here, the small measurement error is chosen deliberately to demonstrate that the applied methods for OED provide different results in spite of small measurement errors. The computation is done in MATLAB, i.e., the ODE-solver *ode15s* and the optimiser *lsqnonlin* (Levenberg-Marquardt algorithm) are put in operation, respectively.

The OED problem can be solved numerically by using one of the above mentioned optimality criteria. As the correlation of μ_m and K_S complicates a proper identification, the E^*-criterion (Eq. (3.9)) is used. In doing so, the correlation of the parameters is addressed appropriately. The theoretical minimum of $\Phi_{E^*}(C_\theta)$ is 1 in case of $\lambda_{max}(C_\theta) = \lambda_{min}(C_\theta)$, where $\lambda(\cdot)$ means the corresponding eigenvalues. Moreover, for the purpose of OED, the inlet flow, $q(t)$, is defined as the design variable, ζ. That means, OED aims to find an optimal trajectory of $q(t)$ which maximises the measurement information content and minimises the parameter uncertainty, respectively. Usually, optimally determined simple inlet profiles are good candidates to reduce the parameter uncertainty [BSSR94]. Therefore, the inlet flow is restricted to follow a practicable profile equal to

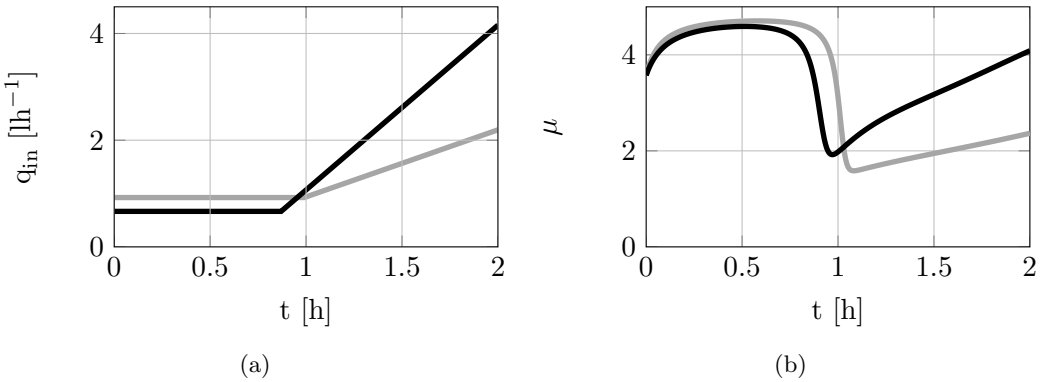
$$q(t) = \begin{cases} a, & t < c \\ a + b(t - c), & t \geq c \end{cases} \tag{3.25}$$

Thus, the OED problem consists in the optimal choice of the parameters a, b, and c, which minimises the associated objective function $\Phi_{E^*}(C_{\boldsymbol{\theta}})$. By running the benchmark study, the covariance matrix, C_θ, is computed either from the inverse of the FIM or directly by the UT approach. Because the (co)variances computed by FIM and UT differ, the two methods provide different optimal operating conditions, too, see Tab. (3.2). Thus, the inlet flow, $q_{in}(t)$, and the corresponding growth rate, $\mu(t)$, differ as well, see Fig. (3.3) for details. Here, the most obvious deviation concerns the slope parameter, b. The UT approach suggests an earlier and steeper increase of the inlet flow, $q(t)$, which may allow a more precise estimation of the substrate limiting constant, K_s. To assess which of the two experiments, the FIM as well as the UT based design, provides the most informative data, a Monte-Carlo simulation is performed at both operating conditions. The resulting scatter plot in Fig. (3.4) with 2,000 samples per

newly designed experiment shows that the UT approach generates a smaller and more roundish point cloud of the parameter estimates, $\hat{\theta}$. Obviously, the UT approach results into a reduction of the parameter uncertainties and their correlation in comparison to the FIM-based design. Especially the variance of K_S in the UT-based experiment is smaller than in the FIM-based experiment. The main reason for this seems to be that the FIM-based covariance underestimates the variance of K_s significantly. That means, in the FIM-based OED a too high accuracy of K_s is assumed erroneously. In consequence, no adequate effort is made to increase the associated uncertainty properly. Moreover, the scatter plot provides characteristic values of the parameter statistics which confirm the usefulness of UT in the field of OED, see Tab. (3.3).

	a	**b**	**c**
OED(FIM)	0.925	1.257	0.992
OED(UT3)	0.666	3.089	0.867

Table 3.2: Parameters of the optimal inlet flow function, $q(t)$, associated to the FIM-based and UT-based OED, respectively.



(a) (b)

Figure 3.3: Outcome of OED for the single substrate model associated to FIM approach (—) and the UT method (—), respectively. The inlet flow, q_{in}, is shown in (a). The corresponding growth rate, μ, is given in (b) additionally.

	$\sigma^2_{K_s}$	$\sigma^2_{\mu_m}$	$E[K_s]$	$E[\mu_m]$	ρ
OED(FIM)	0.9891×10^{-3}	0.0611×10^{-3}	2.0020	5.0005	0.9971
OED(UT3)	0.2130×10^{-5}	0.0452×10^{-5}	1.9999	4.9999	0.8584

Table 3.3: Resulting parameter statistics corresponding to the FIM-based and UT-based OED, respectively. Here, ρ represents the correlation of the two parameters, K_s and μ_m.

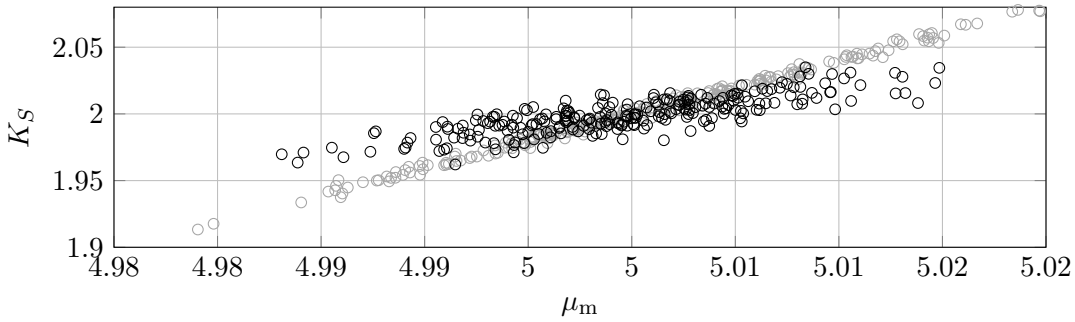


Figure 3.4: Scatter plot of the parameter estimates associated to OED: Results based on the FIM approach (○) show a strong correlation of the two model parameters, K_S and μ_m. In comparison to the UT method (○) also the final parameter uncertainty is increased, i.e., the FIM-based estimates are wider scattered.

3.4 Two-Substrate Uptake Model[1]

In this subsection it is demonstrated how the computational cost of Unscented Transformation based OED can be reduced by utilising interpolation techniques in parallel. Instead of evaluating the original cost function during the process of OED a surrogate cost function based on Kriging interpolation (see App. A.3) is explored. The general procedure is demonstrated by a two-substrate uptake model. Here, the carbohydrate uptake of the micro-organism *Escherichia coli* (*E.coli*) is of particular interest. The considered process mimics the uptake of glucose (Glc) and glucose 6-phosphate (G6p), where the latter is preferred by *E.coli.*, i.e., the uptake of glucose starts only when no G6p is at hand. Thus, in contrast to the so called "glucose effect" [TKA00] glucose is not able to inhibit the uptake of G6p when a mixture of these two substrates is available.

The associated mathematical model encompasses 9 ordinary differential equations and more then 30 model parameters, θ. As previously stated, the model addresses the growth of biomass B, the substrate uptake of $G6p$ and Glc, the evaluation of related uptake enzymes E_{G6p} and E_{Glc}, essential components of the glycolysis pathway, and $EIIA_P$ as an element of the bacterial phosphotransferase system (PTS). A detailed explanation of this model can be found in [KBG07]. In this thesis, however, the focus is on the identification of four related model parameters, K_1, K_{g6p}, K_2, and K_{eiiap}. As these parameters are part of the uptake- and enzyme synthesis rates, just the directly

[1] A large majority of results contained in this section have already been published in the peer-reviewed literature, [SKM12].

involved ODEs are shown. The complete set of ODEs is given in App. A.4.

$$\dot{B} = u \cdot B \tag{3.26}$$

$$\dot{G6p} = -mw_{g6p} \cdot rup_{g6p} \cdot B \tag{3.27}$$

$$\dot{Glc} = -mw_{glc} \cdot rup_{glc} \cdot B \tag{3.28}$$

$$\dot{E}_{G6p} = rsyn_{g6p} - (kd + u) \cdot E_{G6p} \tag{3.29}$$

$$\dot{E}_{Glc} = rsyn_{glc} - (kd + u) \cdot E_{Glc} \tag{3.30}$$

Here, the related uptake rates are

$$rup_{g6p} = \frac{1.5 \cdot kg6p \cdot (E_{G6p} \cdot G6p)}{K_{g6p} + G6p} \tag{3.31}$$

$$rup_{glc} = \frac{kptsup \cdot EIIA_P \cdot (E_{Glc} \cdot Glc)}{(K_{glc} + G6p) \cdot K_{eiia_p} \cdot x0 + EIIA_P \cdot (K_{glc} + Glc)} \tag{3.32}$$

The associated enzyme synthesis rates are defined as stated below

$$rsyn_{g6p} = k_1 \cdot \frac{kb + ksyn \cdot {EIIA_P}^6}{{EIIA_P}^6 + K^6} \cdot \frac{rup_{g6p}}{K_1 + rup_{g6p}} \tag{3.33}$$

$$rsyn_{glc} = k_2 \cdot \frac{KI}{E_{G6p} + KI} \cdot \frac{kb + ksyn \cdot {EIIA_P}^6}{{EIIA_P}^6 + K^6} \cdot \frac{rup_{glc}}{K_2 + rup_{glc}} \tag{3.34}$$

Obviously, K_1 and K_{g6p} determine the uptake of $G6p$, whereas K_2 in combination with K_{eiiap} influences the consumption of Glc. Thus, with the knowledge of the rate expressions, a statement about the parameter estimates to be expected is possible: K_1 and K_{g6p} as well as K_2 and K_{eiiap} are correlated, which makes a parameter identification difficult. A first parameter estimation is done with intuitively chosen initial concentrations, $G6p_{ini}$=2.00 μmol/gDW and Glc_{ini}=2.00 μmol/gDW. The measurement information is gathered by the following setting: samples of biomass, glucose 6-phosphate and glucose ($m = 3$) are taken at seven time points ($K = 7$), between 0h and 8h ([0.2 1.2 2.45 3.7 4.95 6.2 7.45]h). Here artificial (simulated) measurements with a constant variance, $C_{y,ii}$=0.1^2 [gDW^2,$\mu mol^2/gDW^2$], are applied. The MATLAB optimiser *fminsearch* is used to minimise the squared difference between the simulation, y^{sim}, and measurements, y^{data}, weighted by the inverse of the measurement covariance matrix, C_y. This has to be done for each of the 43 measurement sample vectors ($2 \cdot m \cdot K + 1$), see Sec. 2.2.3.3. The resulting set of 43 parameter vectors are used to approximate the mean square error matrix of the estimates, $MSE_{\hat{\theta}}$. In contrast to standard sample based approaches it is a tremendous reduction of computational effort, e.g., for a Monte Carlo approach hundreds or thousands of measurement samples are

usually used.

Almost estimated parameters exhibit a percentage bias over 1% (Tab. 3.4) even for small measurement noise indicating a non negligible non-linearity of the model with respect to its parameters [Rat83]. (Please note that the evaluated bias is based on UT as it outperforms the more traditional Box-bias concept, see App. A.5.) With this in mind it seems natural that the frequently used Fisher Information matrix (based on linearisation) might be insufficient to provide a proper quantification of parameter uncertainties.

$\sqrt{C_y}$	0.01	0.05	0.1	0.15	0.2	0.25
$\%Bi_{K_1}$	0.04	1.09	4.34	9.77	17.36	27.13
$\%Bi_{K_2}$	2.72	67.97	271.88	611.73	1087.52	1699.25
$\%Bi_{K_{g6p}}$	0.20	4.88	19.53	43.93	78.10	122.04
$\%Bi_{K_{eiiap}}$	0.72	18.00	72.01	162.01	288.07	450.04

Table 3.4: Percentage Bias (Eq. A.18) related to an increased measurement error for fixed initial concentrations of $G6p_{ini}$=2.00μmol/gDW and Glc_{ini}=2.00μmol/gDW.

The batch process possesses only few degrees of freedom for OED: the initial conditions and the time points of measurements are potential design variables. Here, the initial concentrations of the two carbohydrates ($G6p_{ini}$ and Glc_{ini}) are optimised to reduce the uncertainties of the parameters and the states, respectively.

	$\frac{E[K_1]}{K_1}$	L_{K_1}	$\frac{E[K_2]}{K_2}$	L_{K_2}	$\frac{E[K_{G6p}]}{K_{G6p}}$	$L_{K_{G6p}}$
Worst Case	13.50	133.60	86.81	899.11	0.39	50.99
Initial Point	1.04	2.51	3.71	33.12	0.68	9.83
Optimum	1.14	3.85	3.61	31.40	0.50	29.10

	$\frac{E[K_{eiiap}]}{K_{eiiap}}$	$L_{K_{eiiap}}$	$\frac{\Phi}{\Phi_{Opt}}$
Worst Case	8.61	82.37	520.08
Initial Point	1.71	10.26	1.38
Optimum	1.21	5.34	1.00

Table 3.5: Statistics of estimated parameters at the three different design points. Here, $L_{(\cdot)}$ represents the length of the associated 99%-Confidence Interval.

The universal concept of the UT method is applied to evaluate the design criterion for the purpose of OED. The UT method in combination with the Kriging interpolation is put in operation for the special case of the carbohydrate uptake model in order to

find optimal initial concentrations of the two carbohydrates, glucose and glucose 6-phosphate. Hence 50 sample points of the two dimensional design space are chosen by Latin Hypercube Sampling, see Sec. 2.2.2.3. These samples are evaluated by the UT approach to determine the uncertainty about simulation results at a validation point of $G6p_{ini}$=2.50 μmol/gDW and Glc_{ini}=2.50 μmol/gDW. The summed up MSE's of the model states B, G6p, and Glc (Eq. 3.10) at the 50 sample points are used as the collocation points for the Kriging interpolation. The resulting surrogate cost function can be seen in Fig. 3.5. Obviously, the intuitive choice of $G6p_{ini}$ and Glc_{ini} is not the worst case but also not an optimal point of the design space. Furthermore, the worst case for the parameter identification is determined by the absence of any substrate, i.e., no information on the uptake dynamics is available. As an experiment under this condition is meaningless an alternative point with a high cost function value in the presence of glucose and glucose 6-phosphate ($G6p_{ini}$=0.45 μmol/gDW and Glc_{ini}=0.94 μmol/gDW) is labelled as the Worst Case. Moreover, the surrogate cost function in Fig. 3.5 indicates that the initial concentration of both substrates should not drop below 2 μmol/gDW. In addition, the concentrations of $G6p_{ini}$=3.47 μmol/gDW and Glc_{ini}=4.36 μmol/gDW can be identified as optimal design values.

To assess the potential of the OED, the statistics of the estimated parameters are analysed at the three different points of the design space (see Tab. 3.5). Indeed the parameters identified under the Worst Case condition provide the highest uncertainties, i.e., the 99%-Confidence Intervals are large and the expected values deviate from the true values significantly. At the Initial Point the uncertainties are smaller but still unsatisfactory. Even the Optimum Point does not lead to credible estimates. The uncertainties are still high, but compared to the Worst Case they are reduced by order of magnitudes.

As described, the cost function is based on uncertainties of the states and not directly on parameter statistics. With this in mind, the Confidence Intervals of the model states are visualised in Fig. 3.6. A model calibrated with the measurement information at the Worst Case is not able to describe the uptake process in a sufficient manner. The Confidence Intervals of simulation results, which are related to parameter uncertainties at the Initial Point and at the Optimum could be reduced, but are still too high for meaningful inferences. Additional optimally designed experiments or more accurate measurements are necessary to guarantee models with an increased predictive power.

To offer an explanation for the benefit of OED, simulations of the biomass, glucose and

glucose 6-phosphate with their measurement samples are shown in Fig. 3.7. Under the condition of the Worst Case, i.e., low initial concentrations of the two substrates, the transient to the steady state settles more quickly than in the optimal case. In the Worst Case the last three measurements are taken at the steady state, hence the last measurements hardly bring new information. In the Optimal Case the amount of $G6p_{ini}$ and Glc_{ini} is sufficiently high that the transient behaviour covers the complete experimental time horizon and all measurements are taken at different system dynamics. At the determined optimum the values of the parameters and especially of the model states possess reduced uncertainties. Consequently, the UT approach and Kriging Interpolation are beneficial in the framework of OED.

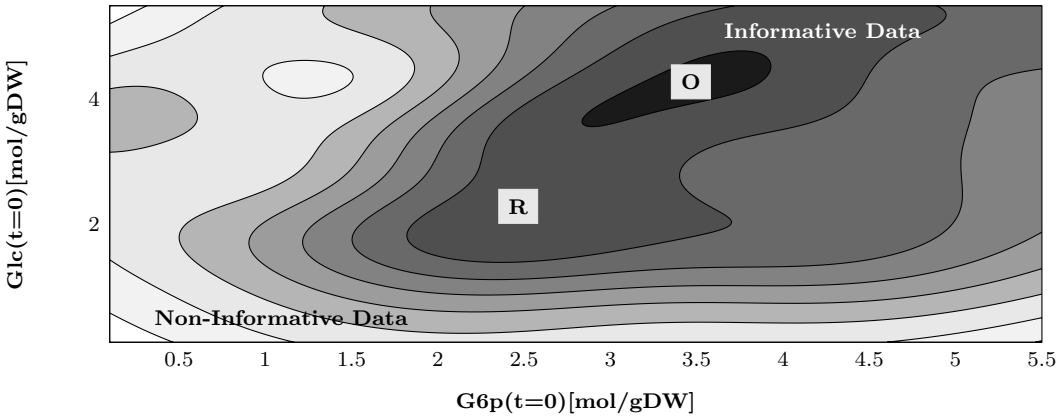


Figure 3.5: Contour plot of the logarithm of the cost function: 50 sample points are used as an input set for the Kriging interpolation. The resulting surrogate cost function can be easily evaluated by optimisation routines. Informative regions are shown dark grey to black, whereas non-informative regions are shown light grey to white. Hence, for a pre-defined reference point, **R**, the optimal initial substrate concentrations are given at **O**.

3.5 Chapter Summary

Compared to the Fisher Information Matrix and the Monte Carlo simulation, the Unscented Transformation approach has a number of advantages that make it highly attractive for Optimal Experimental Design:

- One does not only obtain a lower bound of C_θ as provided by FIM, but an approximation that is sufficiently accurate as long the measurement noise is not too large or the non-linearity is not too severe.

- While the FIM postulates an unbiased parameter estimator, the UT method also gives reliable information on the estimated mean values, $E[\hat{\theta}]$. Therefore, in the framework of OED it becomes possible not only to minimise the covariance but also the bias of the estimated parameters.
- The UT approach does not assume some idealised parameter identification procedure, but works with the numerical method that is actually applied to the identification problem. Therefore, all kinds of imperfections and special properties of the numerical optimiser are taken into account.
- In contrast to the FIM, the UT method does not require the calculation of gradients or Jacobians for the sensitivities. Therefore, it is applicable to a broad class of models, including e.g. black-box like models [HKK08].
- The implementation of the UT framework is very easy, the parallelization is straight-forward.
- The sample points in UT are not chosen randomly but deliberately. Therefore, the number of required samples will usually be much smaller than for the MC simulation. In addition, the necessary number of samples is clearly defined and not a matter of guessing, as it is to some extent for the MC simulation.

Thus, the Unscented Transformation for OED is a versatile concept which may help to overcome a long-lasting problem in non-linear OED:

> “In a nonlinear problem, the statistician can say to the experimenter: “You tell me the value of θ and I promise to design the best experiment for estimating θ.” If the experimenter replies, “Who needs you?,” this is natural but not helpful.
> *W. G. Cochran [Coc73]* ”

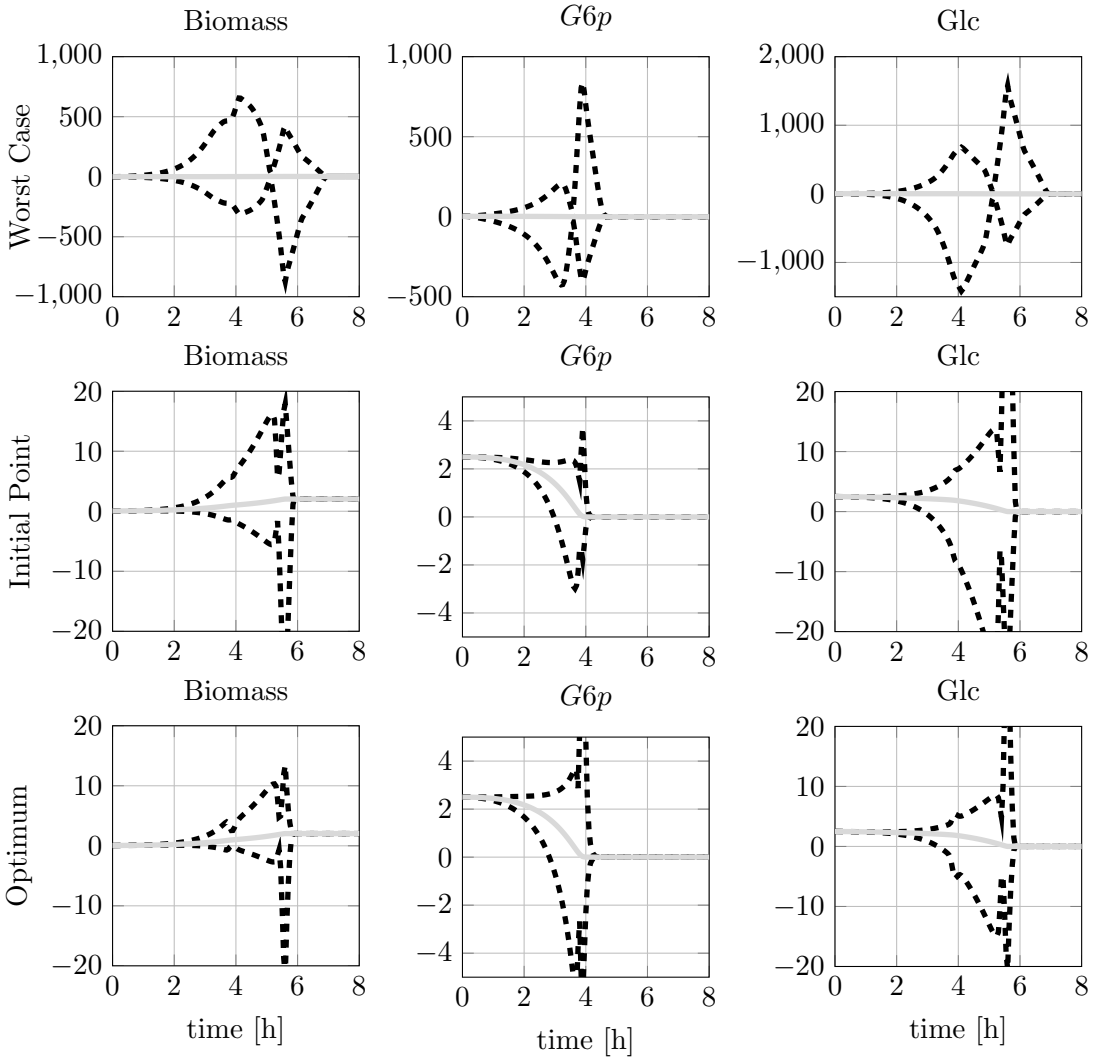


Figure 3.6: Statistics of a subset of model states. Here, the dashed lines, - - -, represent the 95% confidence intervals of biomass, glucose, and glucose 6-phosphate. All simulation results, ——, are performed at reference conditions, $G6p_{ini}$=2.50 μmol/gDW and Glc_{ini}=2.50 μmol/gDW.

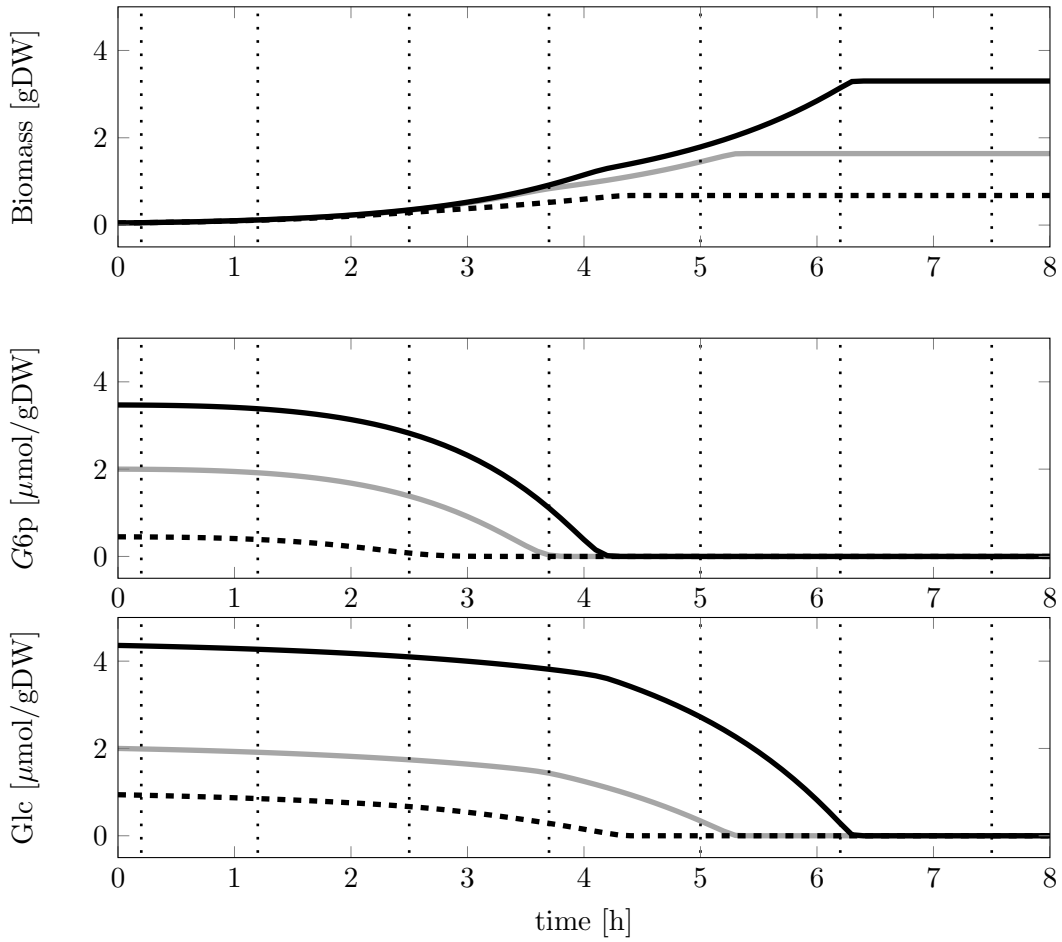


Figure 3.7: Simulation results of the two substrate model at three different operating conditions: (i) the initial substrate concentration, ——, (ii) the optimised substrate concentration, ——, and (iii) the worst case scenario, - - -. Obviously, the measurements at pre-defined time points (vertical dotted lines, · · · ·) are only efficiently used in the case of the optimised initial substrate concentrations, ——.

4

OED for Model Selection

4.1 State-of-the-Art in Model Selection[1]

As mentioned previously, essential processes in cell biology, e.g., signalling or metabolic processes, can be described and analysed using deterministic models. Hence, expert knowledge has to be transferred into a suitable model structure, $\hat{\mathbb{S}}$, which represents the interaction of model components. Depending on the intended use, the resulting mathematical model may help to explore unknown mechanisms in living systems or to validate competing hypotheses of elementary processes. As shown in the previous section, Sec.3, even with a correct model structure, the identification of related model parameters, $\theta \in \Theta \subset \mathbb{R}^l$, might be challenging. Moreover, the situation changes for the worse if there are uncertainties about the model structure itself and the kind of interplay of model components, e.g., kinetics. The generation of a whole bank of model candidates which represent competing hypotheses of the same basic processes, but describing measurement data, $y^{data}(t)$, similarly well, is a quite common situation in systems biology. Generally, suitable model candidates, $\hat{\mathbb{S}}_i$; $\forall i = 1, \ldots, M$, as well as associated model parameters, $\theta(\hat{\mathbb{S}}_i)$, are determined by minimising the mismatch between simulation results, $y^{sim}(\hat{\mathbb{S}}_i, t)$, and measurement data, $y^{data}(t)$. In the field of systems biology, however, one has to accomplish the feat to figure out the most plausible model structure, $\hat{\mathbb{S}}_i$, in conjunction with a large number of unknown model parameters, $\theta(\hat{\mathbb{S}}_i)$, from sparse data. Here, sparse data means that only a small subset of model components can be measured directly at a limited number of discrete time points, t_k, $\forall k = 1, \ldots, K$.

Various methods have been developed which are intended to support the modeller

[1] A large majority of results contained in this chapter has already been published in the peer-reviewed literature, [SKM09a, SM13].

in identifying the most likely or plausible model candidate. Usually, those methods are based on statistics and/or information theory [LBS94, BA02, MC04, KFG+04, DBMM+11]. In general, these two concepts might be different in their origin, as well as in their quality of results, but they share common features, too [LBS94]. For instance, measurement data, $y^{data}(t_k)$, are processed in a batch mode, i.e., data are collected until the experiment is finished and the whole set of data samples is subsequently used for the purpose of model validation/selection. Furthermore, parameter uncertainties are usually not explicitly considered in the framework of model selection. These shortcomings, however, might influence the result of the selection process strongly. In addition, there is no guarantee at all that the model candidates can be distinguished properly at given operating conditions by applying one of these methods. Here, Optimal Experimental Design (OED) comes into play [KT09, MSM10, SL10]. In detail, OED searches for operating conditions which are expected to facilitate the overall selection process. Subsequently, after the determination of suitable operating conditions by OED a new experiment has to be conducted, which is likely to provide new informative data. Unfortunately, optimally designed experiments are usually non-standard experiments, i.e., for the lab assistants those experiments are non-routine jobs and prone to error in consequence. Thus, in combination with some inherent uncertainties, e.g., uncertainties about initial conditions, environmental conditions, and/or model parameters, previously optimally designed experiments may become suboptimal in practice [ATE+08, KAG09]. An online approach which aims to compensate for these uncertainties is presented in what follows.

4.1.1 Preliminaries

Similarly to the previous sections, the considered mathematical models are given as a system of ordinary differential equations (ODEs) according to

$$\dot{x}(t) = f(x(t), \theta, u) \quad ; x \in \mathbb{R}^n \, , \theta \in \mathbb{R}^l \, , u \in \mathbb{R}^s \tag{4.1}$$

$$y^{sim}(t) = h(x(t)) \qquad ; y^{sim} \in \mathbb{R}^m \tag{4.2}$$

Here, the states, $x(t)$, describe the temporal evolution of the quantities of interest, e.g., concentration of enzymes. The output function, $h(\cdot)$, pinpoints the states, $x(t)$, which are directly or indirectly measurable. For the purpose of readability, the proposed

methodologies are introduced without loss of generality for 1-dimensional problems, i.e., $y^{sim} \in \mathbb{R}^1$. Generally, an online model selection strategy calls for outputs which can be measured online in parallel to the experimental run. In the field of systems biology, this might be a difficult task to undertake, but, an increased number of advanced sensor technologies are at hand in the field of (bio)chemical process monitoring, which includes the subsequently analysed MAP kinase as well [FdVN$^+$10]. On the other side, for a flexible OED operation there is a need to excite the analysed system in a desired way by a feasible system stimulus. In the particular case of the MAP kinase such a stimulus might be physically put in operation by ionising radiation [CAR$^+$98, NK04].

4.1.2 How to Separate the Wheat from the Chaff

Generally, in systems biology the analysed processes are quite complex and suffer in validated assumptions about (bio)chemical principles. Therefore, any inference which is based on a mathematical model has to be handled with utmost caution. For instance, it might be that the sum of squared errors, SSE (Eq. (4.3)), can be minimised satisfactorily although unrealistic interactions of model components are part of the model.

$$SSE = \sum_{k=1}^{K} (y^{data}(t_k) - y^{sim}(t_k))^2 \tag{4.3}$$

Naturally, this effect becomes more likely if (i) the dimension of the model parameter vector, $\mathbb{R}^l$, is high, (ii) the number of data points, K, is limited, and (iii) these data are disturbed by measurement noise. Here, too, it is assumed that the measurement noise, $\epsilon(t_k)$, is additive and that $\epsilon(t_k)$ is described by a Gaussian distribution, $\epsilon(t_k) \sim \mathcal{N}(0, \sigma^2_{y_k^{data}})$, leading to the following data generating function

$$y^{data}(t_k) = h(x(t_k)) + \epsilon(t_k) \tag{4.4}$$

An essential part in modelling is to figure out the most plausible model, $\hat{\mathcal{S}}_i$, from a pool/bank of M potential model candidates. A performance index, J_{GC}, which assesses each individual candidate, has to take at least two properties into account

$$J_{GC} = \Psi(\text{Goodness of Fit},\ \text{Complexity}) \tag{4.5}$$

In detail, these are the goodness of fit, i.e., agreement of model output and measurement data (Eq. (4.3)), as well as the model complexity, which is to some extent related to the dimension of the parameter dimension, $\theta \in \mathbb{R}^l$ [BA02].

For example, in statistics, the problem of model selection is frequently reformulated as a test of hypotheses [Fis71, BA02, MC04, HMGB03a]. By assuming two model candidates the simplest model is considered as the so called null hypothesis, H_0, whereas the second candidate is treated as the alternative hypothesis, H_1. Now, the measurement data, $y^{data}(t_k)$, are applied to reject H_0 with a certain probability. Thus, there are two possible outcomes of such a statistical approach, the given data fail to reject H_0, i.e., the simpler model is selected. Or else the data succeeds to reject H_0, that means the more complicated model (the alternative hypothesis, H_1) is the favourite. In its original version only two model candidates can be compared. Furthermore, the outcome of this test is a binary decision, i.e., H_0 fails or it does not fail, but there is no hint how clearly it fails or how well it is accepted.

A more flexible approach of model selection was derived in information theory. Here, no hypotheses are tested, but models are ranked explicitly by a trade-off of the goodness of fit and model complexity in accordance to Eq. (4.5). For instance, the Akaike Information Criterion (AIC) is the most widely used basic formula in this field of application [BA02, MC04]. Throughout this section, the so-called "corrected Akaike Information Criterion" (AIC^c) formula is used

$$AIC^c = K \cdot ln\left(\frac{SSE}{m \cdot K}\right) + 2(l+1) + \frac{2(l+1)(l+2)}{m \cdot K - l} \tag{4.6}$$

Remember, K is the number of measurement time points, and m indicates the number of measured quantities, $y^{data}(t_k) \in \mathbb{R}^m$.

As an individual AIC^c value is difficult to interpret, the AIC^c differences, Δ_i (Eq. (4.7)), are more appropriate for the purpose of model selection. Hence, all individual AIC_i^c values are compared to the least complex candidate, AIC_{min}^c, i.e., the model with the lowest number of parameters.

$$\Delta_i = AIC_i^c - AIC_{min}^c \tag{4.7}$$

By using these difference values, Δ_i, the associated likelihood of a model can be expressed by Akaike weights

$$\mathcal{W}(\hat{\mathcal{S}}_i) = \frac{exp\left(-\frac{1}{2}\Delta_i\right)}{\sum_{r=1}^{m} exp\left(-\frac{1}{2}\Delta_r\right)} \tag{4.8}$$

Furthermore, the Akaike weights, $\mathcal{W}(\hat{\mathcal{S}}_i)$, are equivalent to probabilities values, Π_i, of the potential model candidates and sum up to one

$$\sum_{i=1}^{M} \mathcal{W}(\hat{\mathcal{S}}_i) = \sum_{i=1}^{M} \Pi(\hat{\mathcal{S}}_i) = 1 \tag{4.9}$$

Naturally, the most desired outcome of model selection is to assign one model candidate with a probability value close to one, whereas the remaining candidates are assigned with probability values close to zero. Therefore, a proper distribution of these probabilities is essential in model selection and would be a suitable cost function for the purpose of OED as shown in the subsequent section. In comparison to the hypotheses test, this approach can be easily extended to a large number of potential model candidates. Moreover, all of these candidates are assessed qualitatively, i.e., the selection of the very best model is based on probability values instead of a binary rejection-acceptance decision [MC04]. Nevertheless, at this point it has to be stressed that the outcome of any model assessment strategy has to be handled with care. For instance, due to measurement imperfections calculated results may be affected by large uncertainties which prohibit any meaningful inferences as shown below, e.g., see Fig. 4.11.

4.1.3 Optimal Experimental Design for Model Selection

In the previous section, different approaches for model selection have been proposed. Ideally, by applying one of these methods the most plausible model candidate is selected for a given set of measurement data. In practice, however, one measurement data set derived from a single experiment might be insufficient to discriminate rival model candidates properly. More informative data have to be gathered in addition by new experimental runs. To save time and money these new experiments should be conducted at deliberate operating conditions that were previously determined by a model-based Optimal Experimental Design (OED) strategy. As stated previously,

potential operating conditions which can be adjusted in principle are (i) initial conditions, $x(t=0)$, (ii) the stimulus of the system, $u(t)$, and (iii) measurement sample time points, t_k. In what follows, only the stimulus, $u(t)$, is optimised.

Firstly, for the purpose of defining feasible operating conditions, a cost function, J_D, for OED has to take the differences of the analysed model candidates into account. For instance, a suitable measure of the expected model differences is based on the Kullback-Leibler distance (KLD) (Eq. (4.10)) which is closely related to AIC [BH67, BA02]. Originally, KLD determines the differences in probability density functions, $pdf(\hat{\mathbb{S}}_i)$, of the associated model candidates. By assuming Gaussian probability distributions ($\sigma^2_{\hat{\mathbb{S}}_i} = \sigma^2_{\hat{\mathbb{S}}_j} = c$) the KLD given in Eq. (4.10) simplifies to Eq. (4.11), details can be found in [BH67, RP02].

$$J_D^{KLD}(pdf(\hat{\mathbb{S}}_i), pdf(\hat{\mathbb{S}}_j), y) = \left(\int pdf(\hat{\mathbb{S}}_i) ln \frac{pdf(\hat{\mathbb{S}}_i)}{pdf(\hat{\mathbb{S}}_j)} dy \right) \tag{4.10}$$

$$J_D^{KLDs} = \frac{1}{c} \sum_{k=1}^{K} (y(\hat{\mathbb{S}}_j, t_k, u) - y(\hat{\mathbb{S}}_i, t_k, u))^2 \tag{4.11}$$

Based on the mathematical models, an optimal stimulus, u_{opt}, is chosen which is expected to maximise J_D^{KLDs}. Subsequently, after solving the optimisation problem posed in Eq. (4.12) a new experimental run has to be conducted utilising u_{opt}.

$$u_{opt} = \arg \max_u J_D^{KLDs}(\hat{\mathbb{S}}_i, \hat{\mathbb{S}}_j, u) \tag{4.12}$$

The resultant data, $y_{opt}^{data}(t_k)$, are incorporated for a further model assessment cycle, e.g., to rerun an hypotheses test or to evaluate Eq. (4.8). The steps of model assessment, determination of informative operating conditions, and conducting new experiments are reiterated until the best model can be selected properly. Usually, the total number of reiterations is considerably high. One major reason of this high number of reiterations is imperfections [GBB09, BAGW10], i.e., the optimally designed experiments are based on imprecise assumptions which are addressed in more detail in what follows.

Up to this point, uncertainties about estimated model parameters, $\hat{\theta}$, and operating conditions are neglected. In general, these main sources of uncertainty are likely to influence the outcome of optimal design and model selection strongly. It is well known that measurement noise, $\epsilon(t_k)$, leads to some scatter about identified parameters, $\hat{\theta}$, see Sec. 3. Consequently, the evaluation of the cost function, J_D^{KLDs}, at a single point

in the parameter space, $\theta \in \mathbb{R}^l$, might be misleading. In the best case, this effect is compensated for by the previously mentioned reiteration of experimental design and experimental (re)run, i.e., newly generated measurement data, $y_{opt}^{data}(t_k)$, are used to refine parameter estimates before any assessment of the model candidates is done. Whenever this cycle is performed several times, the negative effect of parameter uncertainties can be reduced iteratively. The actual rerun of optimally designed experiments, however, comprises inherent uncertainties, too. That means, the implementation of an experiment at previously determined operating conditions is usually not free of errors, e.g., certain operating conditions may not be precisely controllable. Taking into account the fact that most of the treated models are non-linear in their operating conditions, minor deviations of u_{opt} may influence the outcome of an experiment critically. Further on, optimally model-based designed experiments are non-standard experiments, i.e., for the lab assistants those designed experiments are non-routine jobs and prone to error. In conclusion, an optimally designed experiment is likely to become suboptimal in practice.

Consequently, an obvious remedy might be the *robustification* of OED. That means, to make the process of model selection and OED robust against non-avoidable uncertainties. Although several concepts for this purpose exist in literature, they usually fail to treat uncertainties about parameters, as well as about operating conditions in equal measure. Furthermore, a rerun of experiments is still necessary for these standard approaches. An alternative approach which copes with uncertainties about model parameters and about operating conditions is presented in the next subsection. The essential idea is to perform OED in parallel to the experimental run in a fully automated fashion.

4.2 Online Model Selection Framework

As mentioned previously, the assessment of model candidates by probability values, $\Pi(\hat{\mathbb{S}}_i)$, is of practical interest. Assuming some statistical information about the model candidates as well as measurement data these probabilities can be determined directly from Bayes' theorem. That means, immediately after a new measurement data sample, $y^{data}(t_k)$, becomes available the conditioned model probabilities are determined

according to

$$\Pi(\hat{\mathbb{S}}_j|y_k^{data}) = \frac{pdf(y_k^{data}|\hat{\mathbb{S}}_j)\Pi(\hat{\mathbb{S}}_j|y_{k-1}^{data})}{\sum_{i=1}^{M}\left[pdf(y_k^{data}|\hat{\mathbb{S}}_i)\Pi(\hat{\mathbb{S}}_i|y_{k-1}^{data})\right]}; \quad \forall j = 1,\ldots,M \tag{4.13}$$

The link of $\Pi(\hat{\mathbb{S}}_j|y_k^{data})$ to Akaike weights is given by Eq. (4.9), details can be found in [BA02]. Obviously, for the purpose of model selection the probability of a single model candidate should converge iteratively to one by incorporating an increased number of data samples. To steer the process in the desired direction the stimulus $u(k)$ of the system is optimally chosen by evaluating a feasible cost function that takes the latest model probabilities into account. A suitable scalar measure of the probability distribution is given by Shannon's Entropy (Eq. (4.14)), which was introduced by G. E. P. Box and W. J. Hill [BH67] in the field of model selection. Generally, the Shannon's Entropy, SE, has a maximum value if all candidates have the same probability and a minimum value if one candidate is assigned by a probability value of one. Consequently, an optimal stimuli has to increase the difference of Shannon's entropy, ΔSE, at the latest time point t_k and the expected time point in future t_{k+1} as shown below.

$$SE = -\sum_{i=1}^{M}\Pi(\hat{\mathbb{S}}_i)ln\Pi(\hat{\mathbb{S}}_i) \tag{4.14}$$

$$\Delta SE = SE(t_{k+1}) - SE(t_k) \tag{4.15}$$

Moreover, the maximum change in entropy, D, which can be expected from new measurement data is expressed by Eq. (4.16)-(4.17), details can be found in [BH67].

$$K_I^{ij} = \int pdf(\hat{\mathbb{S}}_i)ln\frac{pdf(\hat{\mathbb{S}}_i)}{pdf(\hat{\mathbb{S}}_j)}dy + \int pdf(\hat{\mathbb{S}}_j)ln\frac{pdf(\hat{\mathbb{S}}_j)}{pdf(\hat{\mathbb{S}}_i)}dy \tag{4.16}$$

$$D = \sum_{i=1}^{M}\sum_{j=i+1}^{M}\Pi_k(\hat{\mathbb{S}}_i)\Pi_k(\hat{\mathbb{S}}_j)\cdot K_I^{ij} \tag{4.17}$$

By assuming Gaussian probability density functions the Kullback's total measure of

Information is given below, see [BH67] for explanation.

$$K_I^{N;ij} = \frac{1}{2}\left[\frac{(\sigma^2_{\hat{\mathbb{S}}_i} - \sigma^2_{\hat{\mathbb{S}}_j})^2}{(\sigma^2_{y^{data}_{k+1}} + \sigma^2_{\hat{\mathbb{S}}_i})(\sigma^2_{y^{data}_{k+1}} + \sigma^2_{\hat{\mathbb{S}}_j})} + (\bar{y}_{\hat{\mathbb{S}}_i,k+1} - \bar{y}_{\hat{\mathbb{S}}_j,k+1})^2 \left(\frac{1}{\sigma^2_{y^{data}_{k+1}} + \sigma^2_{\hat{\mathbb{S}}_i}} + \frac{1}{\sigma^2_{y^{data}_{k+1}} + \sigma^2_{\hat{\mathbb{S}}_j}}\right)\right] \tag{4.18}$$

Thus, during the experimental run an optimal stimulus, $u_{opt}(\Delta t)$, acting on a finite time-interval, $\Delta t = t_{k+1} - t_k$, can be calculated by solving the optimisation problem given in Eq. (4.19) in parallel. For the sake of simplicity, but with no loss of generality, a piecewise constant input profile is assumed in ongoing optimisation steps, $u_{opt}(\Delta t) = u_{const}$.

$$\arg \max_{u(\Delta t)} D(K_I^N(t_{k+1})) \tag{4.19}$$

For implementation purposes necessary statistics about model candidates, $\hat{\mathbb{S}}_i$, and about measurement data, y^{data}, are needed. Here, the Kalman Filter is tailor-made in terms of applicability and, therefore, shortly described in the following subsection. In contraction to the original work of [BH67] the Kalman Filter framework enables an online design. Moreover, the Kalman Filter copes well even with non-linear systems, under the prerequisite that a proper algorithm of Kalman Filter is put in operation.

4.2.1 Kalman Filter

At this point, the basics of the Kalman Filter (KF) approach are presented. For this purpose, only discrete-time systems are considered as

$$x_{k+1} = f(x_k, u_k) \tag{4.20}$$

$$y_{k+1} = h(x_{k+1}) \tag{4.21}$$

Further details about KF and its application for non-linear systems can be found in vast number of literature, e.g., [Kal60, Gel74, Ste94, Sim06]. In short, the KF operates in two steps. Firstly, it makes some inferences about the process states, $\hat{x}^-_{k+1}$, and its covariances, P^-_{k+1}, at some future time point, t_{k+1} (Eq. (4.22)-(4.23)). This part is

known as the prediction step.

$$\hat{x}^-_{k+1} = f(\hat{x}^+_k, u_k) \tag{4.22}$$

$$P^-_{k+1} = E\left[(\hat{x}^-_{k+1} - x_{k+1})(\hat{x}^-_{k+1} - x_{k+1})^T\right] + Q_{k+1} \tag{4.23}$$

Here, Q_{k+1} is a positive definite matrix taking into account model imperfections. In the second step, the correction step, the feedback in terms of measurement data, y^{data}_{k+1}, is used to improve the previously done inferences according to

$$SR_{k+1} = E\left[(h(\hat{x}^-_{k+1}) - h(x_{k+1}))(h(\hat{x}^-_{k+1}) - h(x_{k+1}))^T\right] + R_{k+1} \tag{4.24}$$

$$K_{k+1} = E\left[(\hat{x}^-_{k+1} - x_{k+1})(h(\hat{x}^-_{k+1}) - h(x_{k+1}))^T\right] SR^{-1}_{k+1} \tag{4.25}$$

$$\hat{x}^+_{k+1} = \hat{x}^-_{k+1} + K_{k+1}(y^{data}_{k+1} - h(\hat{x}^-_{k+1})) \tag{4.26}$$

$$P^+_{k+1} = P^-_{k+1} - K_{k+1} SR_{k+1} K^T_{k+1} \tag{4.27}$$

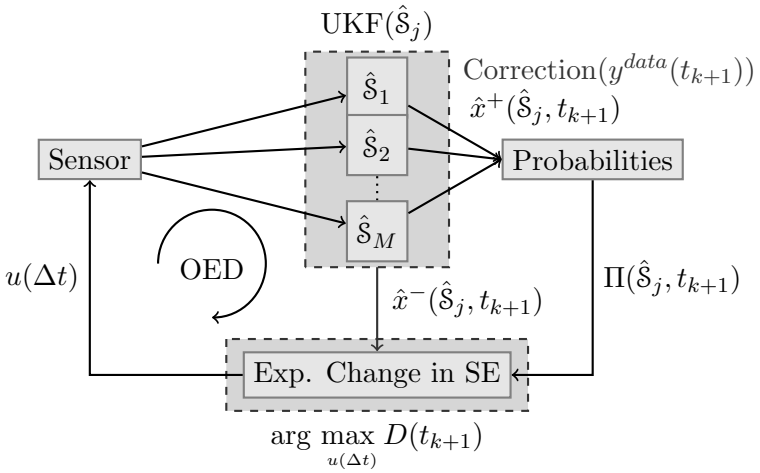


Figure 4.1: Workflow of the Online Optimal Design: The inherent input optimisation, $u(\Delta t)$, is based on the prediction step of the Unscented Kalman Filter (UKF), i.e., just simulation results are utilised for the online optimisation circle (OED). Once an optimal input has been derived, the system is steered in this way in real life. Thus, new measurement data, $y^{data}(t_{k+1})$, are generated which are immediately incorporated by the correction step of UKF. That is, a potential model mis-specification (e.g., biased model parameters) is compensated for to a certain extent by every new available data sample.

Here, R_{k+1} is a positive definite matrix taking into account measurement noise. The Kalman Filter, however, can only be implemented successfully if the system, Eq. (4.20)-(4.21), is observable [Gel74, Ste94, Sim06]. Thus, for all proposed models of the subse-

Algorithm 1: Pseudo-code of the overall online model selection framework

input : $\Pi(\hat{\mathbb{S}}_1|y_1^{data}) = \Pi(\hat{\mathbb{S}}_2|y_1^{data}), \ldots, \Pi(\hat{\mathbb{S}}_m|y_1^{data}) = \frac{1}{M}$
output: Optimal input profile $u^{opt}(t_k),\ \forall k = 1, \ldots, K$

for $k = 1$ **to** K **do**
  $\Delta t = t_{k+1} - t_k$;
  begin Prediction & Optimisation
    while $D(K_I^N(t_{k+1}), u^*(\Delta t)) < D(K_I^N(t_{k+1}), u^{opt}(\Delta t))$ **do**
      $u^*(\Delta t) \leftarrow$ `Optimiser`$(\{u^{min}\ u^{max}\})$;
      for $i = 1$ **to** M **do**
        $\left(\hat{y}^-_{\hat{\mathbb{S}}_i}(t_{k+1}), \hat{\sigma}^-_{\hat{\mathbb{S}}_i}(t_{k+1})\right) \leftarrow$ `Prediction`$(\hat{\mathbb{S}}_i, u^*(\Delta t))$;
        `// Prediction step based on Eq.`(4.22)`,`(4.23)
        $\Omega_i^- \leftarrow [\hat{y}^-_{\hat{\mathbb{S}}_i}(t_{k+1})\hat{\sigma}^-_{\hat{\mathbb{S}}_i}(t_{k+1}), \Pi(\hat{\mathbb{S}}_i|y_{t_k}^{data})]$;
        `// Uncorrected information about model candidate` $\hat{\mathbb{S}}_i$
      end
      $D(K_I^N(t_{k+1}), u^*) \leftarrow \Xi\left(\Omega_1^-, \Omega_2^-, \ldots, \Omega_M^-\right)$;
      `// Determination (Eq.`(4.17)`,`(4.18)`) of the maximum change in entropy` D `which is expected applying` $u^*(\Delta t)$
    end
  end
  $y^{data}(t_{k+1}) \leftarrow$ `ExperimentalRun`$(u^{opt}(\Delta t))$;
  `// Optimised stimulus` $u^{opt}(\Delta t)$ `is applied to the physical system providing an informative data sample` $y^{data}(t_{k+1})$
  begin Correction & Model Assessment
    for $i = 1$ **to** M **do**
      $\left(\hat{y}^+_{\hat{\mathbb{S}}_i}, \hat{\sigma}^+_{\hat{\mathbb{S}}_i}\right) \leftarrow$ `Correction`$(\hat{y}^-_{\hat{\mathbb{S}}_i}, \hat{\sigma}^-_{\hat{\mathbb{S}}_i}, y^{data}(t_{k+1}))$;
      `// Correction step based on Eq.`(4.26)`,`(4.27) `utilising the latest data sample` $y^{data}(t_{k+1})$
      $\Omega_i^+ \leftarrow [\hat{y}^+_{\hat{\mathbb{S}}_i}(t_{k+1})\hat{\sigma}^+_{\hat{\mathbb{S}}_i}(t_{k+1}), \Pi(\hat{\mathbb{S}}_i|y_{t_k}^{data})]$;
      `// Corrected information about model candidate` $\hat{\mathbb{S}}_i$
    end
    for $i = 1$ **to** M **do**
      $\Pi(\hat{\mathbb{S}}_i|y_{t_{k+1}}^{data}) \leftarrow$ `ModelAssessment` $\left(\Omega_1^+, \Omega_2^+, \ldots, \Omega_M^+\right)$;
      `// Applying Bayes' theorem Eq.`(4.29)
    end
  end
end

quent case studies, the observability has been verified successfully via a method which is based on differential algebra [Sed02].

4.2.2 Online Optimal Design by Kalman Filtering

The overall strategy of online optimal design can be split into two essential parts, i.e., model assessment and model discrimination. As described previously, the objective of model assessment is to assign every model candidate with a conditional probability value (Eq. (4.13)) according to current measurement data sample, $y^{data}(t_k)$. For this purpose one has to quantify the conditional probability density functions, $pdf(y_k^{data}|\hat{\mathbb{S}}_j);\ \forall j = 1 \dots M$. In [Ste94] it is shown how these probability density functions can be expressed by the estimated process state, $\hat{x}_k(\hat{\mathbb{S}}_j)$, of the model candidates at time point t_k according to

$$pdf(y_k^{data}|\hat{x}_k(\hat{\mathbb{S}}_j)) = \frac{1}{(2\pi)^{\frac{1}{2}}\sqrt{S_k}} e^{-\frac{1}{2} r_k^T S_k^{-1} r_k} \tag{4.28}$$

Here, the measurement residual is given by $r_k = y^{data}(t_k) - h(\hat{x}_k^+)$, where the corresponding residual covariance matrix, S_k, is defined by Eq. (4.24). Now the model probability can be calculated approximately via

$$\Pi(\hat{\mathbb{S}}_j|y_k^{data}) \approx \frac{pdf(y_k^{data}|\hat{x}_k(\hat{\mathbb{S}}_j))\Pi(\hat{\mathbb{S}}_j|y_{k-1}^{data})}{\sum_{i=1}^{M}\left[pdf(y_k^{data}|\hat{x}_k(\hat{\mathbb{S}}_i))\Pi(\hat{\mathbb{S}}_i|y_{k-1}^{data})\right]} \tag{4.29}$$

Clearly, for the purpose of model discrimination one is interested to increase the differences in the model outputs in the time-interval $\Delta t = t_{k+1} - t_k$ solving the optimisation problem in parallel, see Eq. (4.19). On that account, predictions of the expected model outputs, $\hat{y}(\hat{\mathbb{S}}_j, t_{k+1})$, and their covariance matrix, $\sigma^2(\hat{\mathbb{S}}_j, t_{k+1}) = SR_{k+1}$ (Eq. (4.24)), have to be determined by the prediction step of the Kalman Filter as indicated by Eq. (4.22)-(4.23). The overall scheme of the Online Optimal Design framework is illustrated in Fig. 4.1.

4.3 Case Study

The ability of living cells to react on external stimuli by a proper reply is essential in altering environments. Hence, signalling pathways capturing external stimuli, converting them into an intracellular signal which generates an associated response are of high interest in systems biology. A malfunction of a signalling pathway can cause a number

of serious diseases. Naturally, a better understanding of the underlying processes may lead to novel treatment strategies. Hence, mathematical modelling and model analysis can play a crucial part in contributing improvements in this field of biology/medicine. For this purpose, there is a strong need for highly predictive models. That means, an applied model has to describe the real process adequately even under conditions that had not been part of a former parameter identification step.

Obviously, a proper model structure, $\hat{\mathbb{S}}_i$, is a prerequisite to ensure predictive simulation results. Here, the proposed method of online OED comes into play to figure out the most plausible model structure for a finite number of different model candidates. In detail, the method of online OED is demonstrated for a well known signalling motif, the mitogene activated protein kinase (MAP kinase) [BHDE07]. Generally, the MAP kinase pathways mediates various processes ranging from gene transcription right up to programmed cell death. The cascade consists at least of three enzymes that are activated sequentially which allows a wide range of different response patterns.

Here, too, the very detailed mechanisms of the signalling pathway are unknown. As a consequence several plausible mechanisms coexist in literature. In the following, three different hypotheses of the MAP cascade are considered. As illustrated in Fig. 4.3, these model candidates are different in their topology, i.e., it is assumed that the total number of species, as well as their interactions are not known precisely. In addition to the sequential activation of the three enzymes, KKK, KK, and K, two feedback paths are postulated in the first model candidate, $\hat{\mathbb{S}}_1$. As shown, the feedback from KKK^* to K^* includes a forth activation step of an intermediate enzyme IP. The second model candidate $\hat{\mathbb{S}}_2$ has a similar topology to $\hat{\mathbb{S}}_1$. Here, the feedback from KKK^* to K^* is simplified, i.e., KKK^* and K^* are directly linked via a Michaelis-Menten kinetic. Finally, the most simplification is made in the third model candidate, $\hat{\mathbb{S}}_3$. Here, only the feedback loop starting from KKK^* to K^* is preserved, whereas the feedback loop from KK^* to K^* is neglected totally. The actual ODE systems of the individual model candidates are presented in the Appendix, see Tab. A.5.

Now, the proposed method of online model selection is applied to figure out the most plausible model candidate at three different test case scenarios presented below.

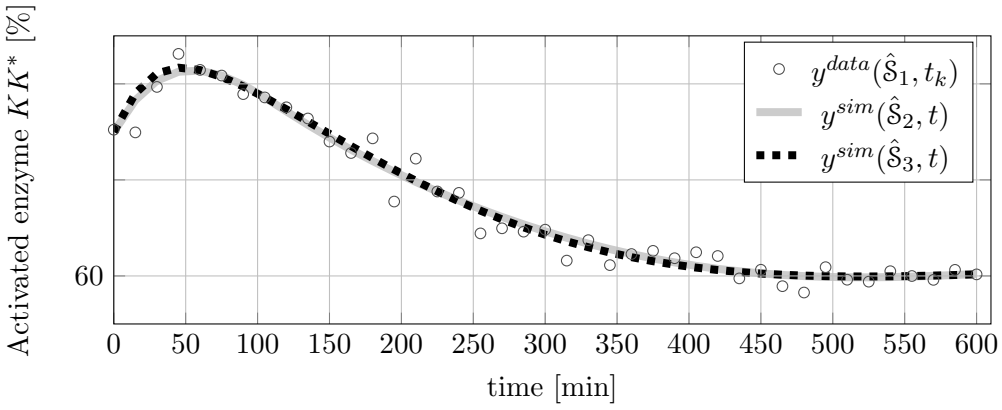
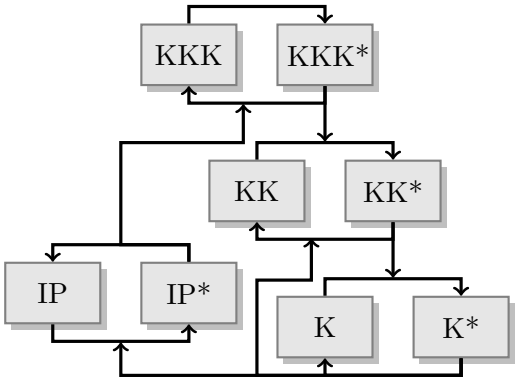


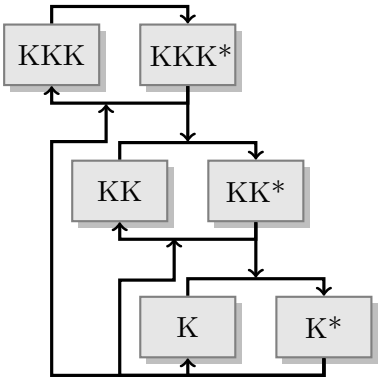
Figure 4.2: Simulation results after parameter identification, $\hat{\theta}(\hat{\mathbb{S}}_2)$ and $\hat{\theta}(\hat{\mathbb{S}}_3)$, via in-silico data, $y^{data}(\hat{\mathbb{S}}_1, t_k)$.

4.3.1 Ideal Case

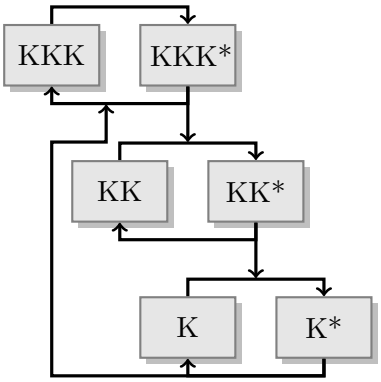
In the absence of real measurement data, in-silico measurement data, $y^{data}(\hat{\mathbb{S}}_j, t_k)$, are gathered by one of the three model candidates, $\hat{\mathbb{S}}_j,\ \forall j = 1, 2, 3$. In spite of this very idealised assumption this strategy is helpful to assess the proposed method as the correct result is known in advance. In the first step, model $\hat{\mathbb{S}}_1$ is chosen to provide the measurement data, $y^{data}(\hat{\mathbb{S}}_1, t_k)$. As stated previously, the data, $y^{data}(\hat{\mathbb{S}}_1, t_k)$, are corrupted by measurement noise, $\epsilon(t_k) \sim \mathcal{N}(0, \sigma^2_{y_k^{data}})$. The associated variance is set equal to $\sigma^2_{y_k^{data}} = 0.025^2$. The data are limited to the activity of the enzyme KK, whereas K^* and KKK^* are unmeasured. Data sampling is done at discrete time points for every 15 minutes, $\Delta t = t_{k+1} - t_k = 15$ min. Now, using a standard optimisation routine, fminsearch of Matlab® optimisation toolbox, the differences in the simulation results, $y^{sim}(\hat{\mathbb{S}}_2, t_k)$ and $y^{sim}(\hat{\mathbb{S}}_3, t_k)$, to the data, $y^{data}(\hat{\mathbb{S}}_1, t_k)$, are minimised at fixed operating conditions, see Tab. A.7 first row. Remember, the theoretical identifiability of the associated model parameters was checked in advance by a method based on differential algebra [Sed02]. Practically, estimates of model parameters, $\hat{\theta}(\hat{\mathbb{S}}_2)$ and $\hat{\theta}(\hat{\mathbb{S}}_3)$, can be reconstructed properly providing a suitable accordance of simulation results and data, see Fig. 4.2 and Tab. A.6 for details. Consequently, after the parametrisation step, all model candidates, at least at the previously applied operating condition, provide a similar input-output behaviour. In this case, methods of model selection are usually put into operation. Thus, for the initial conditions given Tab. A.7 the proposed online model selection method is applied. To demonstrate the significance of the optimisation step, i.e., maximising the change in Shannon's entropy, the stimulus is fixed to u = 0.2.



(a) Wiring diagram of candidate $\hat{\mathcal{S}}_1$



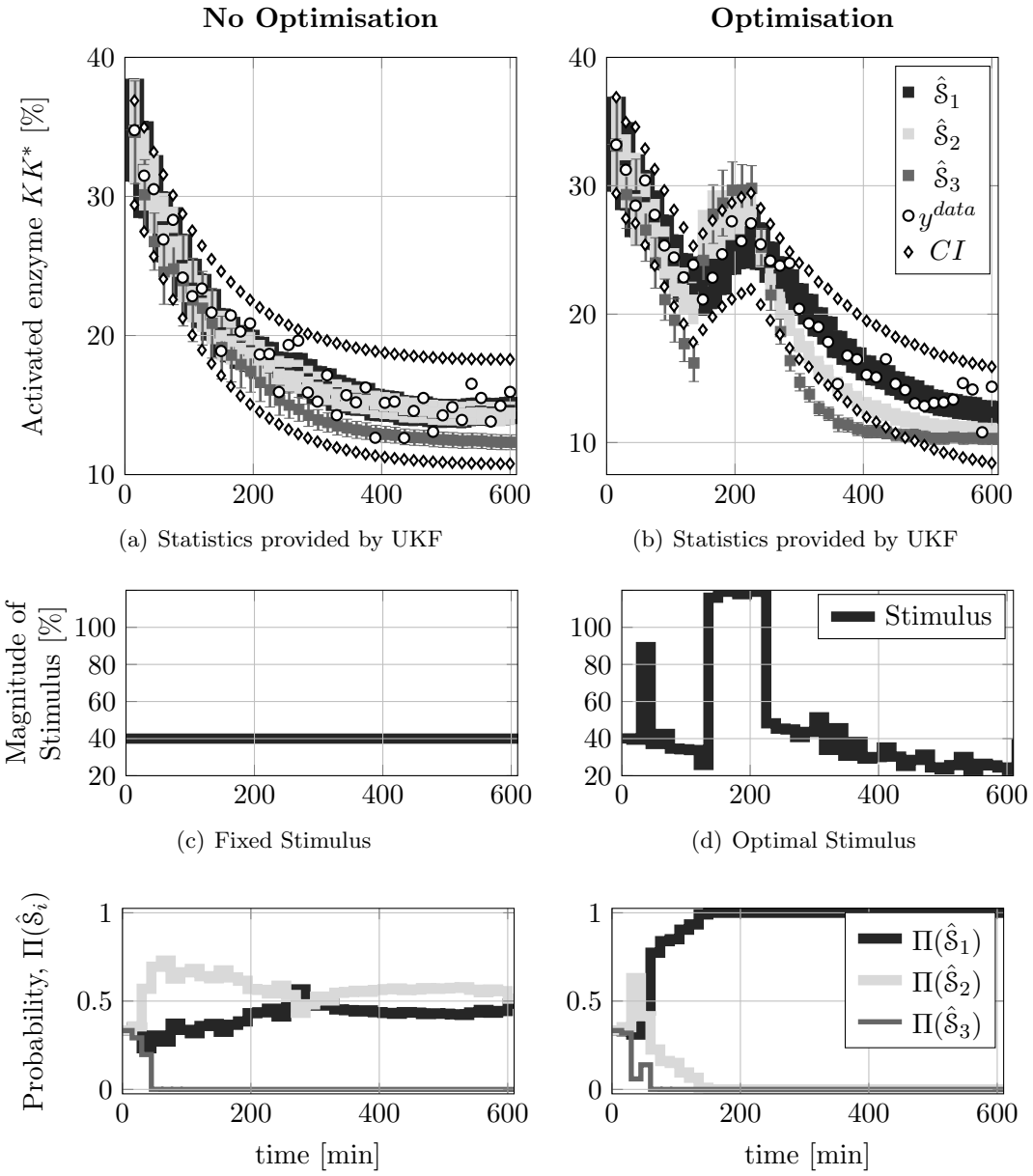
(b) Wiring diagram of candidate $\hat{\mathcal{S}}_2$



(c) Wiring diagram of candidate $\hat{\mathcal{S}}_3$

Figure 4.3: The topology of the three competing model candidates, $\hat{\mathcal{S}}_1$,$\hat{\mathcal{S}}_2$, and $\hat{\mathcal{S}}_3$, are illustrated.

That means, the online framework allows only an assessment of the model candidates at given operating conditions, see Fig. 4.4, left column. Obviously, all three model candidates are able to describe the major trend in the measurement data adequately. Due to the available statistics, model candidate $\hat{\mathcal{S}}_3$ might be excluded, i.e., it is assigned with a probability value close to zero. On the contrary, the left two candidates, $\hat{\mathcal{S}}_1$ and $\hat{\mathcal{S}}_2$, are indistinguishable and in a good agreement to the data. Consequently, more effort has to put in operation to figure out the most plausible candidate. By implementing the essential step of entropy maximisation the result given in Fig. 4.4 is derived. Obviously, the step-wise optimally determined stimulus, $u(\Delta t)$, renders all three model candidates distinguishable. After a short time of convergence, model $\hat{\mathcal{S}}_1$ is preferred as the best candidate, i.e., $\hat{\mathcal{S}}_1$ is assigned with a probability value close to one. As the data, $y^{data}(\hat{\mathcal{S}}_1, t_k)$, are provided by model candidate $\hat{\mathcal{S}}_1$, too, the online selection strategy has done a good job. For the purpose of validation, however, the



(a) Statistics provided by UKF

(b) Statistics provided by UKF

(c) Fixed Stimulus

(d) Optimal Stimulus

(e) Probabilities of model candidates $\Pi(\hat{S}_i)$

(f) Probabilities of model candidates $\Pi(\hat{S}_i)$

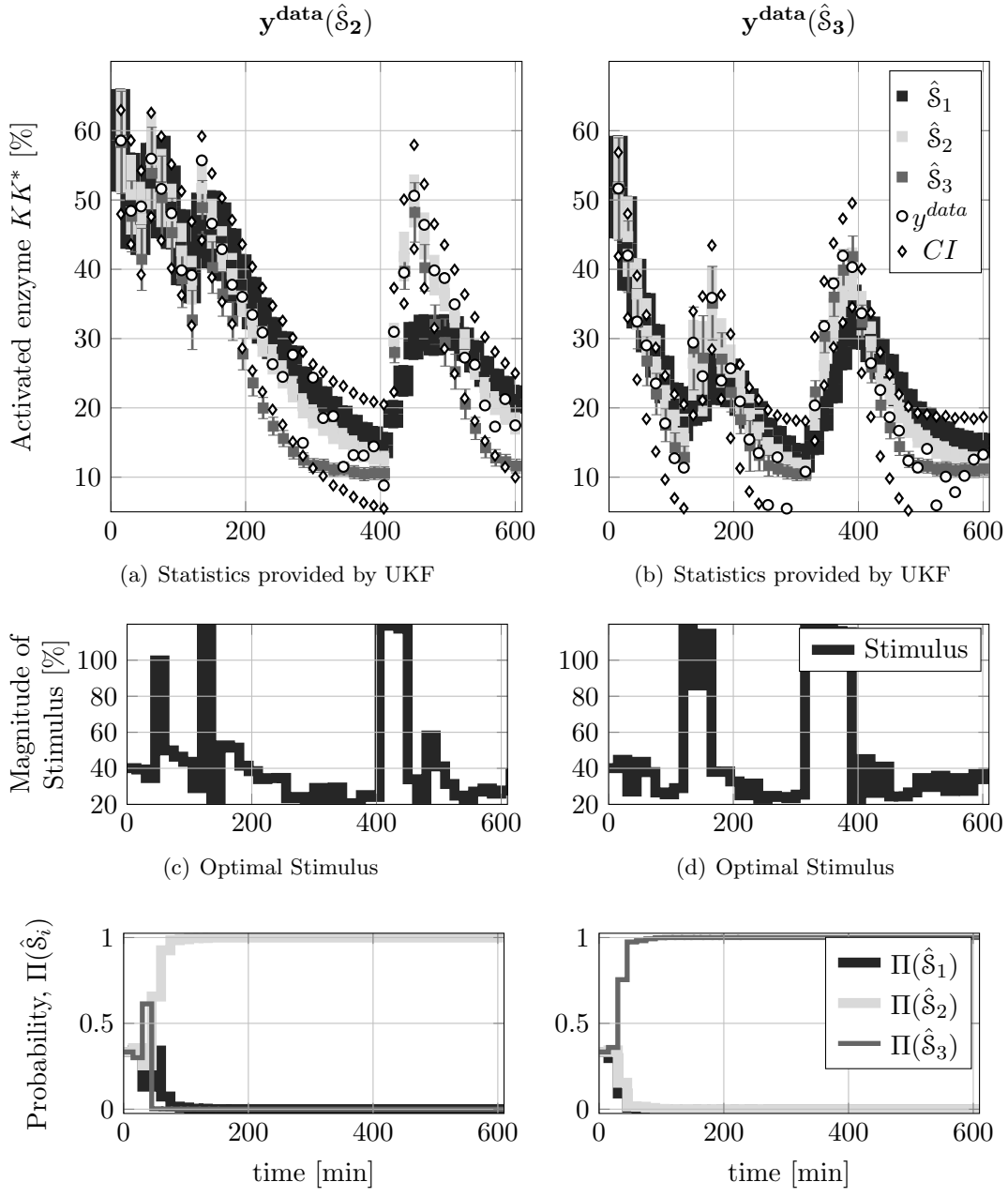
Figure 4.4: Online Optimal Experimental Design: Here, a non-optimal design vs. an optimal design is compared. In both cases, in-silico data $y^{data}(\hat{S}_1)$ are provided by model candidate $\hat{S}_1$ which acts as an surrogate of the real process. CI indicates the 99% confidence interval of measurements.

Data-generating model	$y^{data}(\hat{\mathbb{S}}_1)$	$y^{data}(\hat{\mathbb{S}}_2)$	$y^{data}(\hat{\mathbb{S}}_3)$
Selected model candidate	$\hat{\mathbb{S}}_1$	$\hat{\mathbb{S}}_2$	$\hat{\mathbb{S}}_3$

Table 4.1: Online model selection results of different data-generating models: For any test case scenario the correct model candidate is identified properly by the proposed online selection framework.

model selection process is repeated 100 times in the presence of measurement noise. In this case, the actual objective of the model selection strategy is to turn a preliminary incorrect model preference, $\Pi(\hat{\mathbb{S}}_1) \approx 0$, into the correct decision, $\Pi(\hat{\mathbb{S}}_1) \approx 1$. Moreover, the online optimally designed stimulus is compared with (i) a constant input profile, and (ii) with a pseudo binary amplitude modulated input profile which is known to provide informative data probably [NRPH03, Ise11]. In doing so, results shown in Fig. 4.10 are derived. Obviously, for the total number of 100 test runs the online OED approach determines the correct model candidate, $\hat{\mathbb{S}}_1$, as the most plausible candidate reliably. In comparison to the constant as well as pseudo random input profiles the most informative data samples are generated by the proposed online framework. For instance, in the majority of experimental test runs the constant input is insufficient to ensure a proper model selection. Indeed, the pseudo random input seems to provide more informative data in relation to the constant input profile, but, in comparison to the online approach an increased number of data samples has to be gathered to figure out the correct model candidate. Similar results can be derived for different conditions of the data-generating model, see Tab. 4.1 and Fig. 4.5 for details. In all cases, the proposed method of online selection is able to detect the correct model candidate properly.

Certainly, one has to comment on the optimised stimulus profiles. For all practical purposes, much simpler profiles might be desirable when taking lab equipment limitations into account. In this case, additional stimulus constraints can be easily implemented in the optimisation routine. For instance, the previous model selection task ($y^{data}(\hat{\mathbb{S}}_1)$) is repeated assuming a binary stimulus, $u_{opt} \in [0\ 0.5]$. Caused by the simpler shape of the stimulus profile the discrimination power is reduced, see Fig. 4.6. More measurement data have to be evaluated until the probability values converge and $\hat{\mathbb{S}}_1$ is selected selected correctly as the most plausible one.



(a) Statistics provided by UKF

(b) Statistics provided by UKF

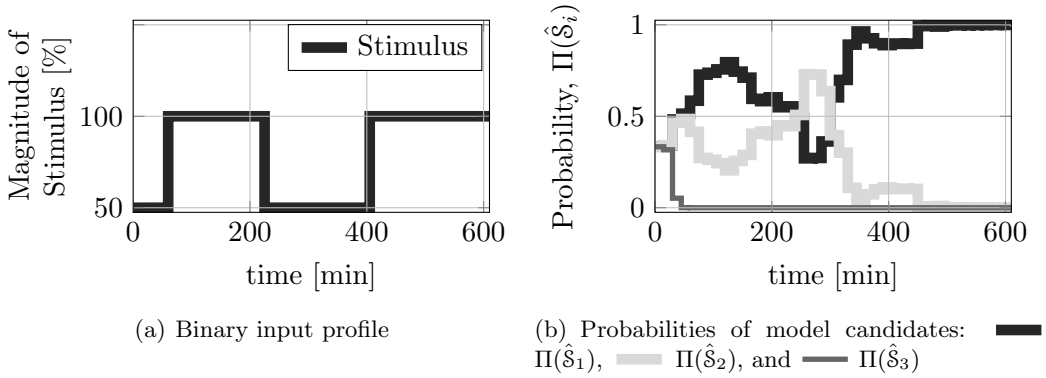
(c) Optimal Stimulus

(d) Optimal Stimulus

(e) Probabilities of model candidates $\Pi(\hat{\mathcal{S}}_i)$

(f) Probabilities of model candidates $\Pi(\hat{\mathcal{S}}_i)$

Figure 4.5: Online Optimal Experimental Design: In-silico data $y^{data}(\hat{\mathcal{S}}_i)$ are generated by model candidate $\hat{\mathcal{S}}_2$ (left column), as well as by model candidate $\hat{\mathcal{S}}_3$ (right column). In both cases, the correct model is selected via the proposed online method.



(a) Binary input profile

(b) Probabilities of model candidates: $\Pi(\hat{S}_1)$, $\Pi(\hat{S}_2)$, and $\Pi(\hat{S}_3)$

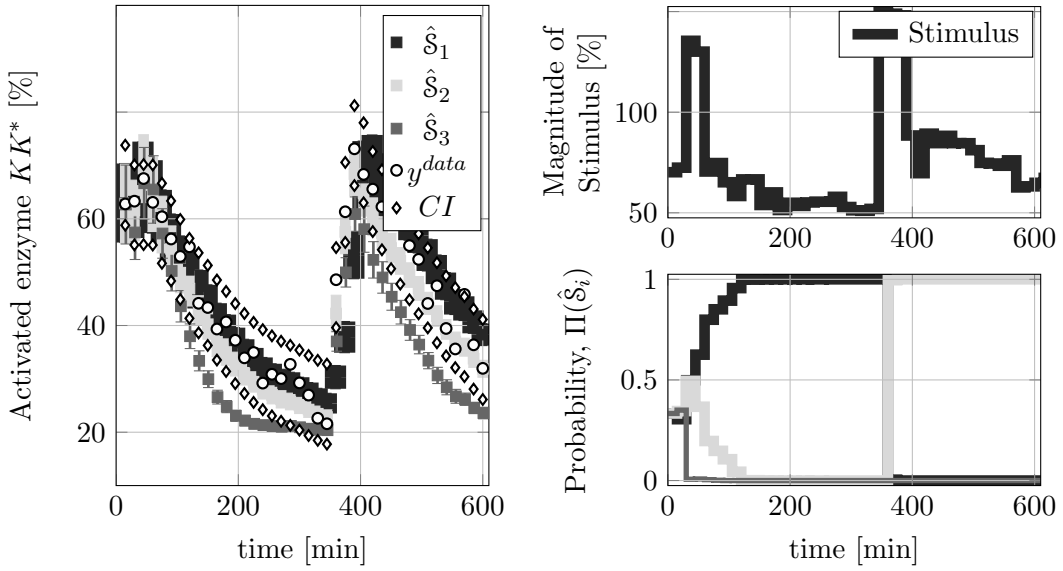
Figure 4.6: In case of a binary input profile the online model selection approach needs an increased number of data samples to figure out the correct model candidate due to the decreased flexibility of the input profile.

4.3.2 Switching Model Case

Complex systems like living organisms are regulated in a hierarchical manner. Thus, there are several regulatory layers that influence each other and usually act at different time scales. By way of example, the stimulus response of a cell is usually determined by the interaction of gene regulation [FPPM09] and post-translational protein modification (PTM) [DPD10]. Strictly speaking, the gene regulation specifies which genes are translated to proteins and which not, whereas PTM defines if synthesised proteins are active or inactive.

Frequently, models exist that describe a certain regulatory layer more precisely than other model candidates, whereas different regulatory layers are more appropriately represented by different model candidates, too. As these regulatory layers are active at different time scales, various model candidates are more suitable at different time intervals. Even such an effect can be detected by the proposed online model selection method. As shown in Eq.(4.29), the determined model probabilities, $\Pi(\hat{S}_i, t_k)$, are functions of time. Consequently, a temporal change in the very best model candidate can be traced. For the purpose of demonstration, an artificial switch in the data-generating process is implemented.

At the first time interval, $TI_1 = [0 \dots 300]$ min, the data are provided by model $\hat{S}_1$, whereas subsequently, $TI_2 = [300 \dots 600]$ min, the in-silico data are generated by model $\hat{S}_2$ instead. In this particular test case, the proposed online method is able to detect the



(a) Statistics provided by UKF

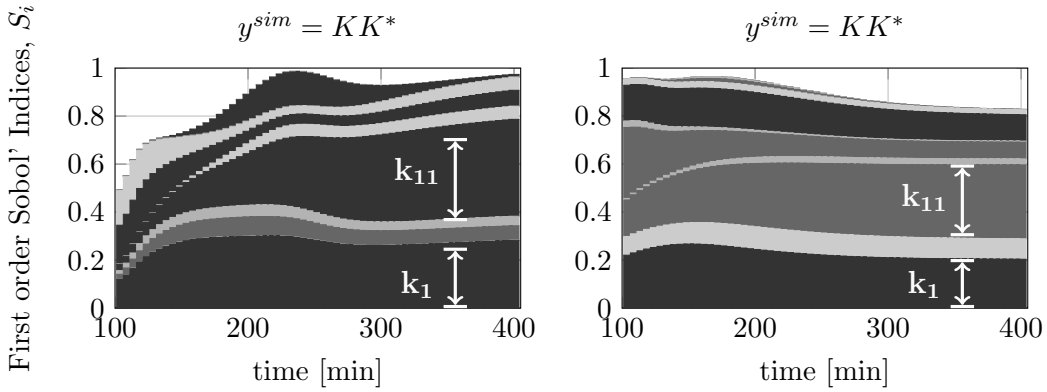
(b) Optimal Stimulus & Probabilities of model candidates: $\Pi(\hat{\mathbb{S}}_1)$, $\Pi(\hat{\mathbb{S}}_2)$, and $\Pi(\hat{\mathbb{S}}_3)$

Figure 4.7: Online Optimal Experimental Design: Here, a shift during the process of measurement data generation is introduced, $y^{data}(\hat{\mathbb{S}}_1) \rightarrow y^{data}(\hat{\mathbb{S}}_2)$. In detail, up to the time-point of 300 min the in-silico data are provided by $\hat{\mathbb{S}}_1$, whereas subsequent data are generated by $\hat{\mathbb{S}}_2$.

switch in the data, see Fig. 4.7 for details. Obviously, after a short time of convergence the model $\hat{\mathbb{S}}_1$ possesses a probability value close to one at TI_1. Subsequently, after the switch to $y^{data}(\hat{\mathbb{S}}_2, t_k)$, the candidate $\hat{\mathbb{S}}_2$ is assigned by a probability value close to one at TI_2 with a delay of a few data sample points. Here, too, simulations have been performed at operating conditions given in Tab. A.7.

4.3.3 True-to-Life Case

In the previous test case scenarios it has been assumed that the measurement data are generated by one of the model candidates. That means, the true model is part of the pool of candidates which are up for election. In practise, however, it would mean that one model candidate describes the physical process perfectly - a quite idealised assumption. A more true-to-life case represents the following scenario. A master model is used to generate in-silico measurement data, $y^{data}(\hat{\mathbb{S}}_{Master}, t_k)$, exclusively. Thus, this master model is not part of the set of potential model candidates, i.e., it is only used as a surrogate for the physical process. In detail, model $\hat{\mathbb{S}}_1$ is chosen as the master model. Hence, the left model candidates, $\hat{\mathbb{S}}_2$ and $\hat{\mathbb{S}}_3$, are applied as two



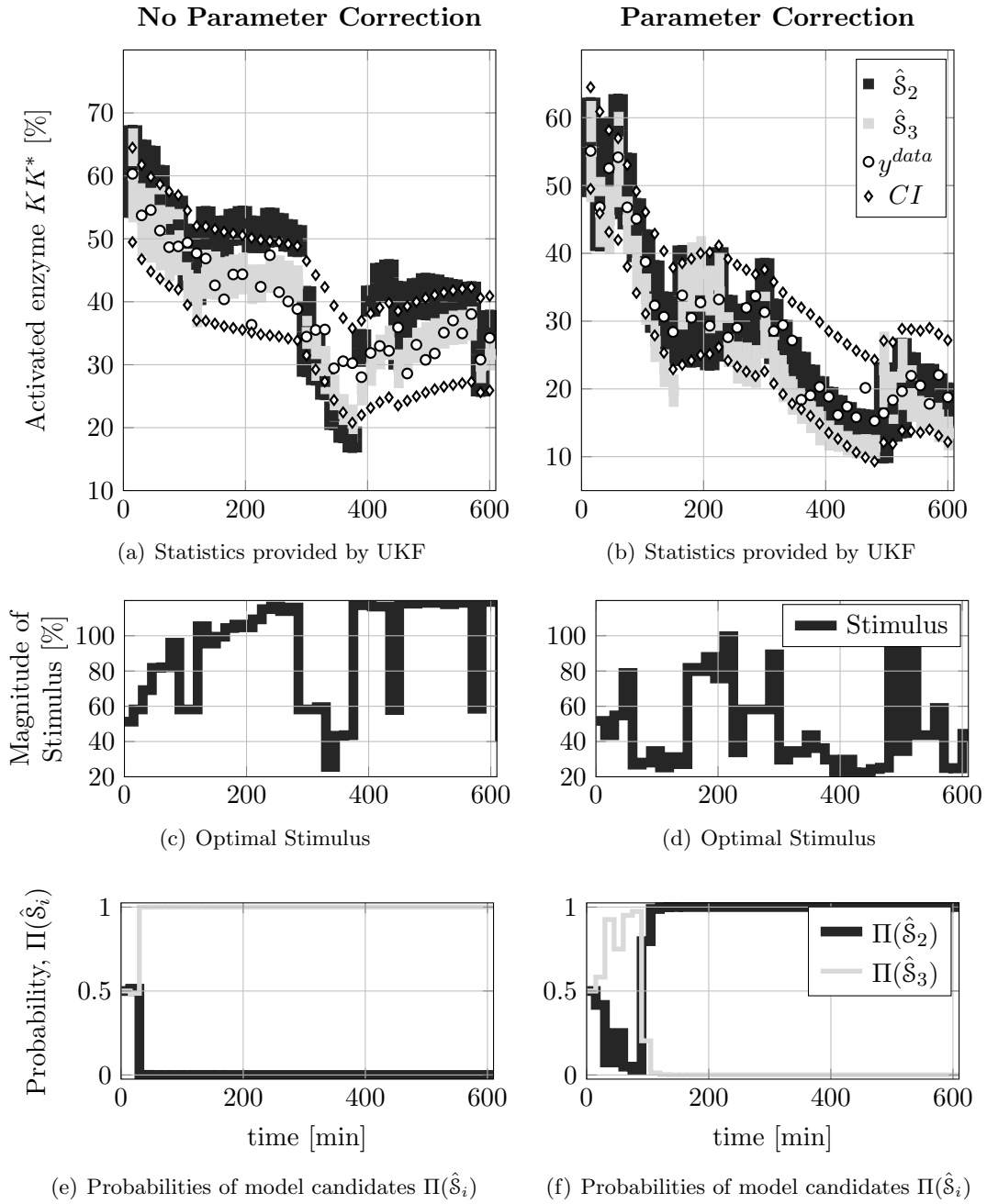
(a) Global Sensitivities associated to $\hat{\mathbb{S}}_2$

(b) Global Sensitivities associated to $\hat{\mathbb{S}}_3$

Figure 4.8: Global Sensitivities associated to $\hat{\mathbb{S}}_2$ & $\hat{\mathbb{S}}_3$: k_1 and k_{11} are the most sensitive parameters for both model candidates, $\hat{\mathbb{S}}_2$ and $\hat{\mathbb{S}}_3$.

different hypotheses about the vaguely known process of interest, $\hat{\mathbb{S}}_1$. Moreover, another crucial point in practise is the influence of parameter uncertainties. In principle, a correct model candidate can be excluded from the set of potential candidates due to a faulty model parametrisation. Naturally, to make the model selection process robust against these parameter uncertainties is a challenging issue. The proposed online selection method, however, compensates for minor uncertainties of model parameters by the Kalman Filter correction step inherently. In addition, to tackle even severe uncertainties of important model parameters, θ_h (Sec. 2.2.5.4), one has to slightly modify the implementation as described below.

In a first step, the most sensitive parameters, θ_h, are determined. Using formulas given in Sec. 2.2.5.4, the Sobol's Indices associated to $\hat{\mathbb{S}}_2$ and $\hat{\mathbb{S}}_3$ are calculated for a relative parameter perturbation of 20% at fixed operating conditions (Tab. A.7), see Fig. 4.8 for illustration. Obviously, two parameters, k_1 and k_{11}, are the most sensitive parameters for both model candidates. Therefore, these two parameters are added to the states, $x(t)$, of the related ODE systems (Tab. A.5). That means, the two most sensitive parameters are estimated directly by the Kalman Filtering process.
For demonstration purposes, k_1 and k_{11} are changed to k_1^c and k_{11}^c (Tab. A.7). Subsequently, an online selection run is conducted without any parameter correction (Fig. 4.9). Due to the wrong parametrisation, model $\hat{\mathbb{S}}_2$ is not able to follow the major trend defined by the measurement data, i.e., model $\hat{\mathbb{S}}_3$ is preferred and gets a probability value close to one. A close look to the topology (Fig. 4.3), however, shows that the



(a) Statistics provided by UKF

(b) Statistics provided by UKF

(c) Optimal Stimulus

(d) Optimal Stimulus

(e) Probabilities of model candidates $\Pi(\hat{S}_i)$

(f) Probabilities of model candidates $\Pi(\hat{S}_i)$

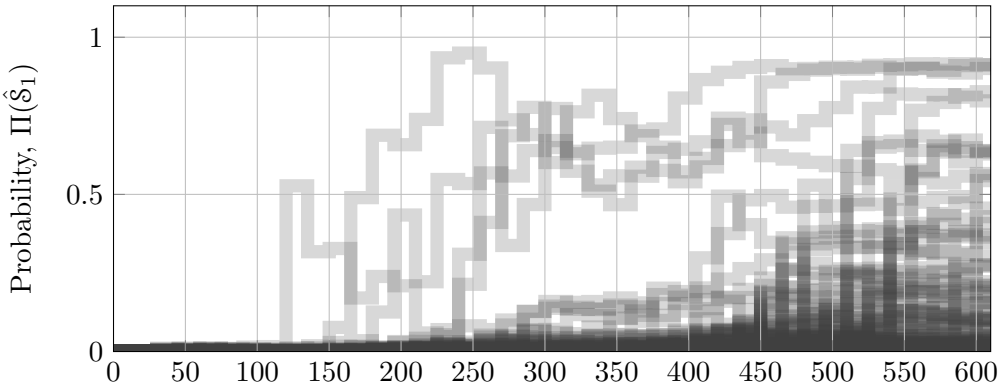
Figure 4.9: Online Optimal Experimental Design: Here, a master model ($\hat{S}_1$) generates in-silico data. Without model parameter correction (left column) a sub-optimal model candidate is determined. Only the joint estimation of states and sensitive parameters (right column) leads to a proper model selection.

feedback from K^* to KK^* is overlooked. If we allow now the parameter correction, the model candidate $\hat{\mathcal{S}}_2$ is assigned by a probability value close to one after short time of convergence. Obviously, the topologies of $\hat{\mathcal{S}}_2$ and of the "real process", $\hat{\mathcal{S}}_1$, are in good agreement, i.e., both comprise the two feedback mechanisms.

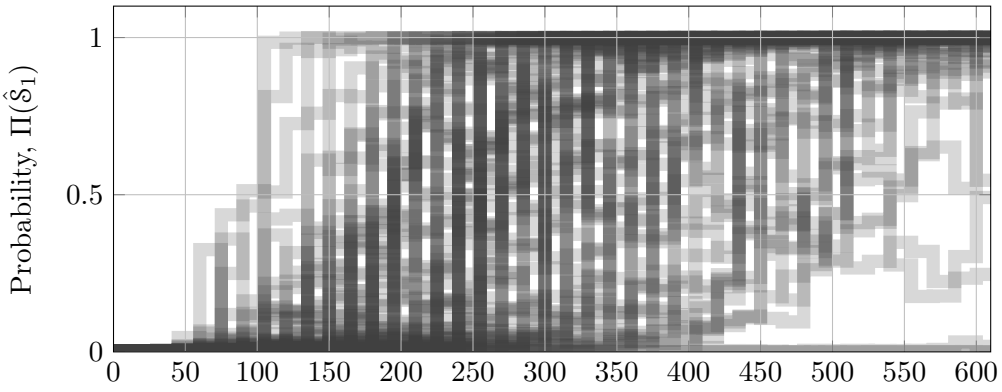
Here, too, the proposed online method is able to figure out the most plausible model candidate. Nevertheless, there is no guarantee of such a desired outcome, i.e., if a system cannot be excited sufficiently, or if the measurement imperfection is too dominant the online selection process is likely to fail like other approaches, too. In general, no method of model selection is free of flaw. For instance, the presented AIC's weights, $\mathcal{W}(\hat{\mathcal{S}}_i)$, are very sensitive to measurement noise, $\eta(t_k)$. Consequently, $\mathcal{W}(\hat{\mathcal{S}}_i)$ should be considered as random variables, too, instead of single scalar values. In doing so, the additional information of the variance of $\mathcal{W}(\hat{\mathcal{S}}_i)$ is helpful to take the measure of confidence about $\mathcal{W}(\hat{\mathcal{S}}_i)$. In Fig. 4.11 the expected value of $\mathcal{W}(\hat{\mathcal{S}}_i)$ as well as its confidence region are illustrated. Here, the stimulus is fixed to 0.6 and the statistics about $\mathcal{W}(\hat{\mathcal{S}}_i)$ are determined for an increasing measurement noise-levels, σ_y^2. Obviously, the expected value of $\mathcal{W}(\hat{\mathcal{S}}_i)$ is strongly affected by σ_y^2. Even under the assumption of almost noise-free data the confidence interval is serious prohibiting meaningful inferences.

4.4 Overlap Approach

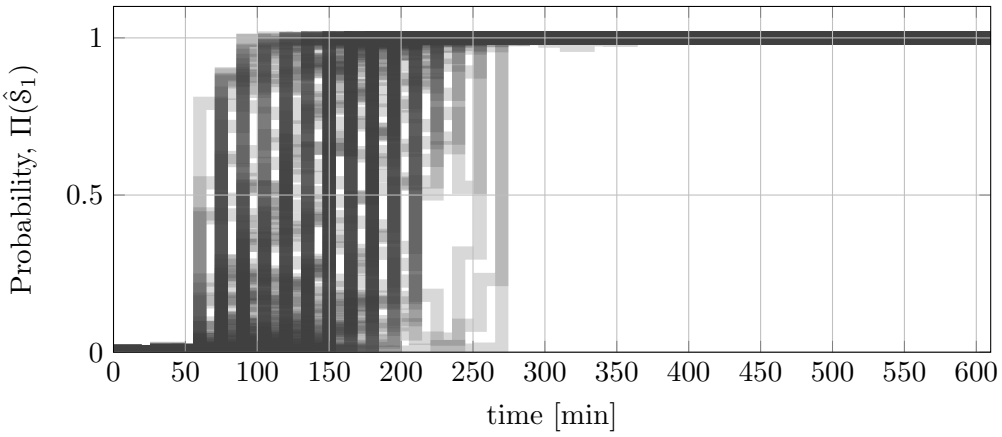
In cases at which no online measurement data can be provided the previously presented approach of online model selection fails. Here, the overlap approach [Lor05] might be an interesting alternative. In contrast to the online strategy the model selection process is performed in the traditional batch mode. That means, firstly the optimal model-based design is determined before the associated experiment is put in operation. In consequence, during the actual experimental run the previously optimal operating conditions are unchanged. Thus, available measurement samples are not incorporate immediately for the purpose of an experimental redesign, but, the measurements are collected over the entire running time. After the experiment is finished the whole set of data is used for the purpose of a new model assessment. Subsequently, a refined OED is determined in case of need. In doing so, a fully automated OED strategy cannot be performed, however, the overlap approach provides an efficient and elegant way to take at least the imperfection of parameter estimates in the field of model selection into account. Here, too, the inherent propagation of uncertainties is performed by crude linearisation principles and by computational cumbersome Monte Carlos simulations, respectively,



(a) Constant input



(b) Pseudo binary amplitude modulated input



(c) Online optimised input

Figure 4.10: Benchmark study comparing a constant input (a), a pseudo binary amplitude modulated input (b), and an online optimised input (c). For any type of input profile 100 test runs are evaluated to take account of measurement imperfections. The most informative data - at least for the purpose of model selection - are provide by the online model selection algorithm. By starting from a false probability, $\Pi(\hat{\mathcal{S}}_1) \approx 0$, the correct result, $\Pi(\hat{\mathcal{S}}_1) = 1$, is determined with the lowest number of evaluated data samples.

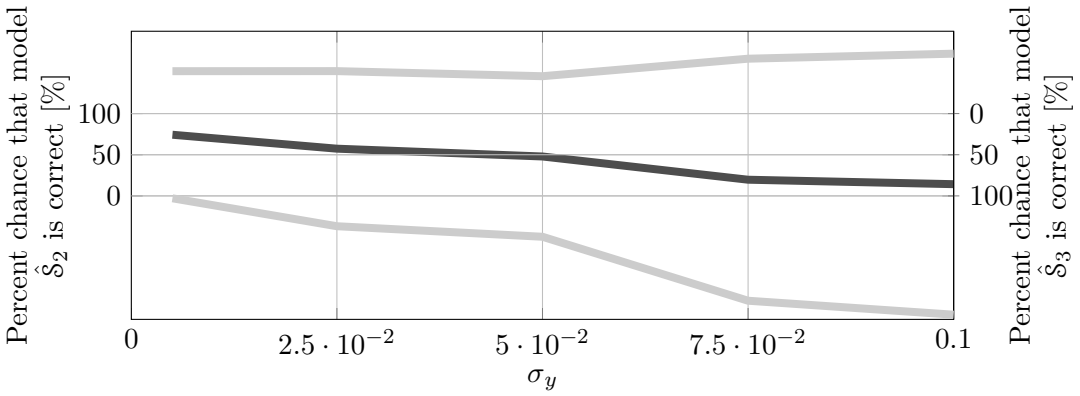


Figure 4.11: Uncertainty of AIC weights: Expected value of selection probability (black line) and the corresponding 99% confidence interval (grey line) are shown as a functions of the measurement error.

see [Lor05, LDTS07] for details.

Alternatively, the universal concept of the Unscented Transformation enables a proper exploration of the influence of parameter uncertainties on simulation results, $y^{sim}(t)$, see Sec. 2.2.3. For example, approximations of confidence intervals of simulation results can be determined in this way. In agreement to the previous results of model selection, the task of OED is to find appropriate operating conditions, e.g., a system stimulus which renders competing model candidates distinguishable. As indicated in Sec. 4.1.2, the general problem in OED is to find a suitable measure of model differences which can be used as a proper cost function for the purpose of model selection. Hence, the additional statistical information about simulation results is incorporate in the cost function definition. Thus, OED aims to provide operating conditions which turns simulation results of competing model candidates into distinguishable quantities, whereby the related confidence intervals indicate in which extend these differences have to be.

Obviously, the overlap of the system states of competing model hypothesis is a suitable measure of these scaled model differences [Lor05]. By assuming a Gaussian density distribution of $y^{sim}(\hat{S}_i)$ the definition of the associated overlap becomes

$$O_j = \sum_{t_k} \frac{2 \cdot \sqrt{C^{jj}_{x_{\hat{\mathcal{S}}_1}(t_k)} \cdot C^{jj}_{x_{\hat{\mathcal{S}}_2}(t_k)}}}{C^{jj}_{x_{\hat{\mathcal{S}}_1}(t_k)} + C^{jj}_{x_{\hat{\mathcal{S}}_2}(t_k)}} exp\left[\frac{-0.5 \cdot (E[x^j_{\hat{\mathcal{S}}_1}(t_k)] - E[x^j_{\hat{\mathcal{S}}_2}(t_k)])^2}{C^{jj}_{x_{\hat{\mathcal{S}}_1}(t_k)} + C^{jj}_{x_{\hat{\mathcal{S}}_2}(t_k)}}\right] \tag{4.30}$$

Here, $C_{(\cdot)}^{jj}$ the j'th diagonal element of the related (co)variance matrix and $E[(\cdot)^j]$ the j'th element of the expectation vector. To incorporate overlap regions of different system states their superposition is used according to

$$O_{\sum} = \sum_j O_j. \tag{4.31}$$

As already mentioned, all needed information to calculate the overlap, Eq. (4.30), is given by the Unscented Transformation. Uncertainties about estimated parameters which might be determined in advance by UT are transferred to the simulation results. Subsequently, the derived statistics of the outputs is used to make competing model hypothesis more distinguishable in a credible fashion. Furthermore, if no measurement data can be explained by the confidence regions of the system states, the model structure has to be corrected, as it remains the only source of disagreement.

For the purpose of illustration, the UT based overlap approach is implemented for a unstructured growth model:

$$\dot{c}_B = \mu \cdot c_B - D \cdot c_B \tag{4.32}$$

$$\dot{c}_S = -\frac{1}{Y_{B|S}} \cdot \mu \cdot c_B + D \cdot (c_{s,in} - c_s) \tag{4.33}$$

Here, too, c_B means the concentration of biomass, c_S the concentration of the substrate, D the dilution rate, and $Y_{B|S}$ the yield factor which is considered as given by literature. Moreover, it is assumed that there is some uncertainty about the growth rate, μ, resulting in two different hypothesis given below

$$\mu(\hat{\mathbb{S}}_1) = \mu_{max} \cdot \frac{c_S}{c_S + K_S} \tag{4.34}$$

$$\mu(\hat{\mathbb{S}}_2) = \mu_{max} \cdot \frac{c_S}{c_S + K_S + \frac{c_S^2}{K_I}} \tag{4.35}$$

Obviously, the rate $\mu(\hat{\mathbb{S}}_1)$ describes the classical Monod kinetic, whereas the competing rate $\mu(\hat{\mathbb{S}}_2)$ is known as the Andrew's equation [ON05] and takes account of the substrate inhibition effect. The two model parameters, μ_{max} (the maximum growth rate) and K_S (the substrate affinity constant) of the kinetics, are assumed to be unknown and have to be identified (before the actual model selection) by measurement data of the biomass concentration.

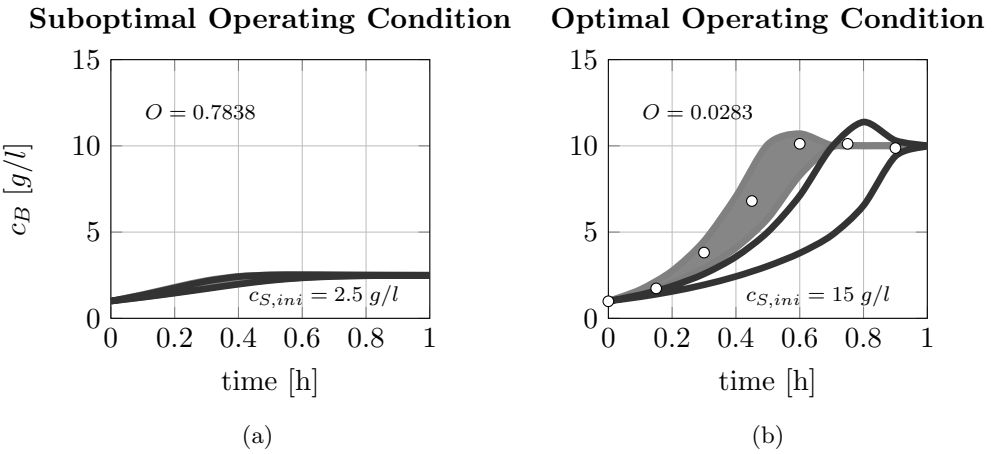


Figure 4.12: Here, the resulting confidence intervals are illustrate according to (▬) $\mu(\hat{\mathrm{S}}_1)$ and (▬) $\mu(\hat{\mathrm{S}}_2)$, respectively. At low initial substrate concentration (a) both candidates cannot be distinguished uniquely. Only in case of an increased initial substrate concentration (b) the two competing candidates can be distinguished properly. Moreover, measurement data (○) clearly support the correct model associated to $\mu(\hat{\mathrm{S}}_1)$.

For the case of simplicity, a batch process is assumed, D=0, and the artificial measurement data of biomass concentration at seven time-points, $t_k = [0, 0.15, 0.3, \ldots, 0.9]$ h, are generated by the original Monod model. The artificial data are corrupted by measurement noise, $v \sim N(0, 0.1)$, additionally. In case of suboptimal initial conditions, i.e., low initial concentration of biomass and substrate, the two models describe a similar dynamic behaviour. That means, both candidates provide almost equal state confidence intervals which are caused by parameter uncertainties, see Fig. 4.12(a). Thus, no advice of a proper model candidate selection can be given at this stage. Subsequently, by applying the overlap approach [LDTS07], the discrimination function, Eq. (4.31), is put in operation. Hence, the primary objective of OED is to find initial conditions which reduce the overlap of the states confidence intervals as low as in any way possible. As might be expected by (bio)chemical principles, an increased initial substrate concentration reduces the overlap significantly, see Fig. 4.12(b). Naturally, the influence of the substrate inhibition becomes substantially more important in case of $\mu(\hat{\mathrm{S}}_2)$. Moreover, measurement data, which are associated to the optimised experiment, clearly support model candidate $\hat{\mathrm{S}}_1$. Thus, the correct model has been select properly, see Fig. 4.12(b) for details.

4.5 Chapter Summary

In this chapter an unified framework of an online model selection algorithm has been presented. As shown by various test case scenarios the proposed approach copes well with serious uncertainties. The most essential merits are summarised below:

i) Designed as an online approach it may render succeeding experiments redundant. The online approach aims to reduce the total number of additional experiments by an immediate incorporation of measurement data followed by an immediate adaptation of operating conditions. Thus, one major source of uncertainty, i.e., to put optimally designed experiments into practice, is addressed appropriately.

ii) Minor uncertainties, e.g., uncertainties of initial conditions and of less sensitive model parameters, that would have an undesired influence on the optimal experimental design are compensated for by default, due to the correction step of the Kalman Filter.

iii) Sensitive model parameters, which may be known vaguely, are the principle reason that even an optimally designed experiment is frequently suboptimal in practice. Hence, the joint/parallel determination of an optimal design, $u(\Delta t)$, and the estimation of sensitive model parameters make this approach particularly well suited for the purpose of robust experimental design.

iv) The proposed method is flexible enough to detect fundamental changes in the measurement data. That means, different plausible model candidates can be detected for different time intervals.

Moreover, in cases at which no online measurement data are at hand the overlap approach has been presented as a proper alternative. By operating in a batch mode, the Unscented Transformation method is directly applied to determine some statistics about simulation results which are caused by parameter uncertainties and measurement imperfections, respectively. In doing so, the overall framework of OED becomes more robust and takes intrinsic uncertainties into account adequately.

5

Flatness Approach for Parameter Identification

5.1 Introduction[1]

As shown so far, mathematical models described by ODE systems are capable to represent various types of processes in biology, e.g., metabolism (Sec. 3.3, 3.4) as well as cell signaling processes (Sec. 4.3). The underlying principles of these biochemical reaction networks, however, are neither known in detail nor directly accessible by measurement data. As a consequence, different model hypotheses about basic processes may exist, which have to be validated by methods presented in Sec. 4. Alternatively, in case of insufficient physical insight into certain subprocesses these vague known parts might be approximated more appropriately by time-delay systems. Here, the sole focus is about the timespan which a certain mechanism consumes, but not about the detailed mechanism itself. Instead of ODEs, those systems are described by delay differential equations (DDEs). In what follows, input affine DDEs are of current interest:

$$\dot{x}(t) = f(x(t), x(t-\tau), \theta) + g(u(t)) \quad ; x \in \mathbb{R}^n \, , u \in \mathbb{R}^s \, , \theta \in \mathbb{R}^l \tag{5.1}$$

$$y^{sim}(t) = h(x(t)) \quad ; y^{sim} \in \mathbb{R}^m \tag{5.2}$$

$$x(t) = \Xi(t) \quad ; t \in [-\tau, 0], \tag{5.3}$$

Here, $\Xi(t)$ represents the initial function in the range of $t \in [-\tau, 0]$. Obviously, DDEs are the more general case of ODEs, i.e., for zero time delay values, $\tau = 0$, a DDE system simplifies to an ODE system (Eq. (4.1)). In case of DDEs, however, the additional

[1] A large majority of results contained in this chapter have already been published in the peer-reviewed literature, [SRM12, SM14].

time delay parameters, τ, influence the input/output behaviour of the mathematical model and have to be identified in addition to the model parameters, θ. Therefore, the parameter identification problem of DDEs (θ & τ) is generally more challenging in comparison to ODEs. Novel identification strategies that are beneficial for DDEs, as well as for ODEs are highly desired in systems biology.

In the previous sections, the parameter identification process has been applied in the traditional way, see Fig. 5.1(a). That means, an optimisation routine is used to minimise differences between simulation results, $y^{sim}(t) \in \mathbb{R}^1$, and measurement data, $y^{data}(t) \in \mathbb{R}^1$, by evaluating a suitable cost function, J_y, according to

$$\arg \min_{\theta,\tau} J_y(\theta,\tau) = \sum_{k=1}^{K} (y^{data}(t_k) - y^{sim}(t_k))^2 \tag{5.4}$$

Though being the most frequently applied approach in practice, this strategy has some serious shortcomings:

a) One limiting factor in evaluating the cost function, $J_y(\theta,\tau)$, is the repeated need of numerical integration for solving the underlying ODE/DDE system. Although very efficient ODE solvers are implemented in the most standard simulation tools, they slow down the actual parameter identification step dramatically. Depending on the problem at hand, the numerical integration subpart consumes up to 90% of the overall cpu-time in the framework of parameter identification [MMB03a]. In general, the situations becomes much worse for DDE systems. Here, numerical solvers are likely to suffer in efficiency and robustness, e.g., the history of the dynamic process before the experiment's starting time has to be known or alternatively to be estimated. Thus, with respect to the computational load it would be of great benefit to eliminate the numerical integration part in total.

b) Another disadvantage concerns the initialisation of the ODE/DDE solvers. Generally, before solving any ODE system one has to provide related initial values of the states, $x(t_0)$, at the initial time point, t_0. If these initial values are unknown they have to be identified by measurement data which might be challenging. The complexity of the identification process increases substantially. Moreover, to solve DDEs related initial functions (Eq. (5.3)) have to be provided. Similar to the initial conditions of ODEs, these initial functions have to be identified by measurement data in most applications. Consequently, by using the same amount of data, but reconstructing an increased number of unknowns, the identification

process leads to an increased level of uncertainties about the identified quantities in all. Therefore, an approach that is capable to identify model parameters without the explicit need for initial states/functions is preferable. In this case, the exclusive use of measurement data for the purpose of parameter identification ensures more precise parameter estimates or may even render the principle identification of model parameters possible in special cases.

c) The third major criticism concerns directly the cost function J_y, which has to be evaluated through the process of parameter identification. As previously mentioned, in the traditional approach one minimises the differences between simulation results, $y^{sim}(t)$, and measurement data, $y^{data}(t)$. Here, the functional dependency of the simulation results on model parameters is utilised. The relation of model parameters and outputs, however, becomes non-linear even in case of linear ODE systems. Hence, the resulting cost function for parameter identification is also non-linear in relation to the identified parameters. Thus, the non-linearity is one major reason why traditional cost functions for parameter identification are commonly non-convex. This basically means, an unambiguous set of parameter estimates is difficult to identify. In consequence, due to multiple local minima of the cost function strongly diverse parameter sets might be the result of the parameter identification process depending on the initial parameter values as well as on measurement data. To provide a remedy, global optimisation routines, e.g., simulated annealing, particle swarm, and genetic algorithms, could be applied for the purpose of parameter identification. Here, the computational demand increases significantly without guarantee of uniqueness in general. More efficient approaches are desired in practice. For instance, a reformulation of the cost function providing a global minima avoids the application of cpu-intensive global optimisation routines. Naturally, such a reformulation cannot be derived in general. Thus, in the case of non-convex cost functions, sophisticated global optimisation routines might be applied with a very low computational load if no ODE/DDE system has to be solved in parallel, see item a).

In summary it can be said, therefore, that the need of simulation results, $y^{sim}(t)$, turns the actual parameter identification step into a computational demanding and challengingly to evaluate process. In the following, the concept of flat inputs is introduced to overcome these previously mentioned limitations to a certain extent. On that account, a short recall about the basics of differential flatness is given which provides the link to the concept of flat inputs and their merits for parameter identification issues. Furthermore,

another point of interest is the model based design of new informative experiments providing additional measurement data which facilitate the parameter identification process, see Sec. 3. So far, new measurement data are of vital importance to turn insensitive model parameters into more sensitive one. In a break from this tradition, the pure evaluation of a cost function which is based on flat inputs leads to a change in parameter sensitivities. That means, without any new experimental data insensitive model parameters may become sensitive just by a change in the utilised cost function. Consequently, the flat input based parameter identification strategy contributes an interesting perspective to the framework of optimal experimental design. Finally, some application examples demonstrate potential merits of the proposed method. All applied models describe (bio)chemical processes whereas the general framework of the presented method is not limited to those.

5.2 Comparison with Existing Approaches in Literature

Over the last decades a tremendous effort has been made to improve the original parameter identification process. The focus has been on robustness, efficiency, and credibility of optimisation routines. The parameter identification or rather the implemented algorithm has to provide parameter estimates which are the very best candidates in minimising the mismatch between simulation results and measurement data. Hence, suboptimal results due to local minima have to be avoided or even better they have to be excluded totally by the applied optimisation routine. On that account, various global optimisation strategies have been developed and are subject of ongoing research. Usually, their capability to extricate themselves from local minima is dearly paid by an increased computational load. Therefore, approaches which speed up the evaluation of the underlying cost functions are of high interest for global as well as in local optimisation algorithms. Finally, parameter estimates have to be assessed in relation to their credibility, i.e., to derive a meaningful model the identified model parameters should possess low parameter uncertainties at all. In this field, parameter sensitivities are of vital interest as the most sensitive parameters are likely to be identified with the utmost precision.

Methods which try to tackle at least one of the previously mentioned shortcomings have been developed frequently, but in real world application an “one size fits all” algorithm does not exist down to the present day. For instance, the proposed approach of flat

input based parameter identification is limited to differentially flat systems and may suffer in credibility due to the need of higher-order derivatives of measurement data. Nevertheless, similar approaches exist and have been applied successfully in the field of (bio)chemical modelling. Details about these cognate methods and their relation to the proposed approach are discussed in more detail in what follows.

In the field of ODE systems, a sequential strategy of parameter identification was published in 1975 by J. Swartz and H. Bremermann [SB75]. Here, a parameter identification concept was implemented which is based on two essential steps:

a) By assuming that all states are measurable, model parameters are determined that minimise the differences between first order derivatives of the data, $y^{data}(t_k)$, and the right hand side of the ODE systems, see Eq. (5.6). Hence, first basic concepts of data smoothing are presented and applied in practice. Surrogate functions, $y^{surr}(t_k, c)$, are used for this purpose as shown in Eq. (5.5). In detail, polynomial functions of low degree are applied as surrogates and differentiated analytically. Being aware of potentially large parameter uncertainties due to a misfit of the surrogate functions a subsequent second step is put into operation.

$$\arg \min_{c} J_c(c) = \sum_{k}^{K} (y^{data}(t_k) - y^{surr}(t_k, c))^2 \quad \rightarrow \quad \hat{c} \tag{5.5}$$

$$\arg \min_{\theta} J_a(\theta) = \sum_{k}^{K} (\mathcal{D}y^{surr}(t_k, \hat{c}) - f(x(t_k), u(t_k), \theta))^2 \quad \rightarrow \quad \hat{\theta}^a \tag{5.6}$$

b) Parameter estimates which are the result of this first identification step, $\hat{\theta}^a$, are used to initialise the subsequent part of parameter refinement. Here, the standard cost function is utilised in the traditional way by evaluating Eq. (5.7) which results in the final estimates, $\hat{\theta}^b$. Thus, numerical integration of the ODE system is involved.

$$\arg \min_{\theta} J_b(\theta) = \sum_{k}^{K} (y^{data}(t_k) - y^{sim}(x(t_k), \theta))^2 \quad \rightarrow \quad \hat{\theta}^b \tag{5.7}$$

The characteristic of these two step concept can be summarised according to:

> “The first is fast (economical in computation time), requires no initial estimates, but is not so accurate. The second requires more computational time, and fairly accurate initial estimates, but achieves high accuracy.
> *J. Swartz & H. Bremermann [SB75]* ”

Since that time lots of effort has been been made to change the first step into a fast as well as accurate approach. Hence, different strategies have been derived which combine the approximation step via surrogate functions, $J_c(c)$, and the original model parameter estimation, $J_a(\theta)$, in a beneficial manner. In general, both objectives are joint into an unified optimisation problem as shown below

$$\arg \min_{c,\theta} J(c,\theta) = \upsilon J_c(c) + (1-\upsilon) J_a(\theta) \tag{5.8}$$

This expression might be solved iteratively or simultaneously [Var08]. In addition, the determination of a suitable weighting factor, υ, has been addressed by probabilistic assumptions [VMM08a]. In cases where all states are directly measured the flat input based approach is equivalent to the joint estimation problem given in Eq. (5.8). In practice, however, it is likely that only a subset of quantities is measurable. To cope also with unmeasured states model-based constraints have been incorporated into the above optimisation strategy [VMM08b] which are defined as

$$(\mathcal{D}y^{surr}_{unmeasured}(t_k,\hat{c}) - f_{unmeasured}(x(t_k),u(t_k),\theta))^2 \tag{5.9}$$

Here, additional nuisance parameters, $\hat{c}$, have to be identified in relation to the dimension of the unmeasured states, i.e., the overall number of unknown parameters may increase significantly. Moreover, the corresponding surrogate functions of unmeasured sub-states ensure model consistency but do not have any direct relation to measurement data. In doing so, the resulting optimisation problem is different in comparison to the flat input approach where higher-order derivatives of measurement data are used to recalculate the unmeasured states immediately.

In recent times, also methods based on differential algebra enrich the field of practical parameter identification routines. For instance, the differential elimination approach renders a given ODE system into an equivalent representation according to a given state-elimination-order [Bou07, NHO$^+$10, NSL$^+$12]. Hence, ...

> “...the most basic step in the theory of differential elimination algorithms is to replace differential equations by algebraic ones... .
> *A. Wittkopf [Wit04]* ”

In practice this means, unmeasured states are sequentially eliminated (according to the predefined ranking) by given outputs and derivatives thereof. An algebraic expression of the resulting equivalent system is indicated by C^{DE} [NSL$^+$12]. Here, $C^{(\cdot)}$ gives a hint to the intended application, a constraint/penalty term as part of the cost function, whereas (DE) reflects its origin, the differential elimination. The algebraic expressions are averaged over discrete time points, $C_i^{DE}(\theta) = \sum_{k=1}^{K} \left| C_i^{DE}(\theta, t_k) \right| ;\ \forall i = 1, \ldots, n$, and might be summarised according to

$$C^{DE}(\theta) = \sum_{i=1}^{n} C_i^{DE}(\theta) \tag{5.10}$$

Subsequently, $C^{DE}(\theta)$ is incorporated into the parameter optimisation cost function as a penalty term, see [NSL$^+$12] and references therein, as shown below

$$\arg \min_{\theta} J_{DE}(\theta) = \upsilon J_b(\theta) + (1 - \upsilon) C^{DE}(\theta) \tag{5.11}$$

By making use of simulation results, $y^{sim}(t)$, to evaluate the cost function, $J_b(\theta)$, the associated differential equation system has to be solved in parallel. Therefore, it is scarcely to be expected that there is any benefit in the meaning of computational load. At best, an improvement of the credibility of parameter estimates might be observed in opposition to the standard approach for parameter identification as highlighted by the following quotation:

> “The introduction of the constraints by using differential elimination has effectively improved the parameter accuracy . . . . This clearly indicates that the ability of our method for estimating the parameter values was far superior to that of various methods with the standard error function.
> *M. Nakatsui et al. [NHO$^+$10]* ”

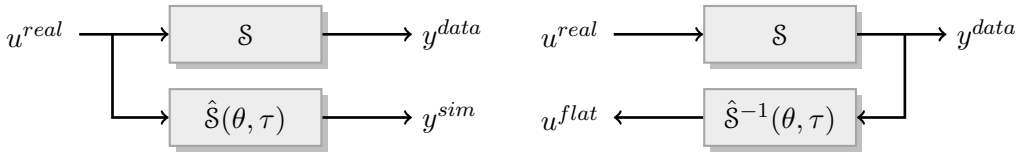
In contrast to the flat input approach only autonomous systems have been analysed by differential elimination so far, i.e., systems without any input, $u^{real}(t)$, have been subject of parameter identification [Bou07, NHO$^+$10, NSL$^+$12].

Up to now, in the flatness-based approach, an algebraic representation of the input/output behaviour of dynamic systems is derived for the purpose of flat input determination, $u^{flat}(t)$. In doing so, an analytical inverse model, $\hat{S}^{-1}$, is generated. As an alternative, the concept of inverse simulation [MS11] bypasses the analytical reformulation by applying control feedback methods. That means, the flat inputs are simulation results of the closed loop behaviour of the original system, $\hat{S}$, which has been extended by a proper feedback control strategy. In the field of non-linear dynamic systems the determination of suitable feedback concept as well as its parametrisation might be a tedious work if at all possible. Nevertheless, the concept of inverse simulation has been implemented successfully in various field of control problems, see [TB06, MS11] an references therein. Recently, also parameter sensitivities in conjunction with simulated inputs are taken into consideration [MS12b]. However, no application to the actual parameter identification problem can be found in literature as confirmed by the comment given below:

> “It appears that, despite the potential benefits, established parameter estimation methods such as the output error methods have not been applied to the inverse problem.
> *D. Thomson & R. Bradley [TB06]* ”

Recently, an approach which is also based on flat inputs has been introduced in the field of ODE systems [VGS+10]. In contradiction to the previously proposed concept the unknown model parameters, θ, are directly considered as flat inputs, $u^{flat}(t) = \theta(t)$. That means, the model parameters might be analytically recalculated by the output functions and derivatives thereof which obviates the need of any numerical optimisation routine. On the other hand, this also means that the number of identifiable model parameters is limited to the number of measurable quantities. Moreover, the analysed system has to be differentially flat for a given output configuration and for those model parameters which “pretend” to be flat inputs. Consequently, this approach is only applicable for a very limited number of practical parameter identification problems.

Here, it should be stressed that the previously reviewed methods are solely applied for parameter identification problems of pure ODE systems. Thus, non of these approaches finds use in the field of DDEs, i.e., to estimated model parameters, θ, as well as time-delay parameters, τ. In fact, for non-linear DDE systems the general concept of differential flatness is known as an appropriate tool to solve challenging problems in



(a) Standard Approach (b) Flat Input Approach

Figure 5.1: In the standard approach of parameter identification (left sub-figure) one attempts to minimise differences between simulation results, y^{sim}, and measurement data, y^{data}, by a proper selection of model parameters, θ, and time delay values, τ, respectively. In the right sub-figure the idea of flat inputs, u^{flat}, is illustrated. Here, one tries to minimise the mismatch between u^{flat} and physical inputs, u^{real}, by adjusting θ and τ jointly.

control theory [MR98]. For the task of parameter identification, however, additional research is required. Here, the proposed approach of flat input based parameter identification addresses this issue substantially. As demonstrated in Sec. 5.5.3, the proposed approach copes well with time delay parameters of non-linear DDEs.

5.3 Concept of Differential Flatness

Initially introduced by Fliess *et al.* [FLMR92, FLMR95] the concept of differential flatness has received much interest in control theory over the last two decades. The main field of application is devoted to trajectory tracking control problems [MR98, HLBA04, GWRG06, Ló9]. For instance, the flatness approach has been successfully implemented even in industrial practice [PRB$^+$02]. The crucial point of differential flatness is the determination of so-called flat outputs, $y^{flat}(t)$. In most cases, flat outputs do not agree with originally given output configurations, $y^{flat}(t) \neq y^{sim}(t)$, but have to be defined appropriately. In doing so, these outputs might be physical or fictitious, i.e., in the first case they correspond to the physical outputs of the analysed process, while in the latter case they have no physical counterpart in reality. For differentially flat systems, flat outputs and derivatives thereof parametrise the states, $x(t)$, as well as system inputs, $u(t)$, according to

$$x(t) = \Psi_x(y^{flat}(t), \mathcal{D}y^{flat}(t), \ldots, \mathcal{D}^{n-1}y^{flat}(t)) \tag{5.12}$$

$$u(t) = \Psi_u(y^{flat}(t), \mathcal{D}y^{flat}(t), \ldots, \mathcal{D}^{n}y^{flat}(t)), \tag{5.13}$$

where $\mathcal{D}^i$ represents the operator notation of the ith derivative, d^i/dt^i. That means, the original ODE system is transformed into an algebraic input/output representation.

Even though many control systems have been shown to be differentially flat in practice, necessary and sufficient conditions for the existence of flat outputs exist only for single input and single output systems (SISO systems). Here, the relative degree, ν, is of central relevance and is defined as follows.

Definition 1 (*Relative degree [WZ08]*)

The system (Eq. (5.1)-(5.3)) is said to have a relative degree ν locally at x_0 if

$$\begin{aligned} L_g L_f^i h(x) &= 0 \quad \forall x \in \mathcal{N}(x_0),\ i = 0, 1, \ldots, r-2 \\ L_g L_f^{r-1} h(x_0) &\neq 0, \end{aligned}$$

where $\mathcal{N}(x_0)$ is the neighbourhood of x_0 and $L_{(\cdot)}$ represents the Lie derivative of a function along the vector field $(\cdot)$.

In short, a SISO system is differentially flat if it has a relative degree of n ($x \in \mathbb{R}^n$), i.e., only after the nth derivative of $y^{flat}(t)$ the flat output becomes an explicit function of $u(t)$ in case of a differentially flat system.

Definition 2 (*Differential Flatness - Flat Outputs [WZ08]*)

The system (Eq. (5.1)-(5.3)) is said to be differentially flat if there exists an output $y^{flat}(t) = \bar{h}(x(t))$ such that the resulting SISO system

$$\dot{x}(t) = f(x(t), \theta, u(t)) \quad ; x \in \mathbb{R}^n\ , u \in \mathbb{R}^1 \tag{5.14}$$

$$y^{flat}(t) = \bar{h}(x(t)) \quad ; y^{flat} \in \mathbb{R}^1 \tag{5.15}$$

$$x(t_0) = x_0 \tag{5.16}$$

has a relative degree of n.

Obviously, the system output, $y^{sim}(t)$, is treated as a design variable that may transfer a non-flat ODE system into a differently flat system if at all possible, whereas the general model structure, $\hat{\mathbb{S}}$, model parameters, θ, initial conditions, $x(t_0)$, as well as system input, $u(t)$, are unchanged. As was intended, this strategy works satisfactorily for solving control issues in different areas of application. In the field of parameter estimation, however, a change in the output function causes serious problems. The need for the adjustment of simulation results to measurement data entails that flat outputs, $y^{flat}(t)$, are feasible in practice but not fictitious. Moreover, not all states are measurable in principle, which limits the amount of potential flat output candidates additionally. Hence, a given output configuration should remain unchanged in

the framework of parameter identification, too.

In comparison to the flat outputs, the concept of flat inputs seems to be more suited for solving parameter identification problems. In the first place, flat inputs have been introduced as a dual approach to the concept of flat outputs [WZ08]. Here, the output functions, $y^{sim}(t)$, are fixed but the so-called flat inputs, $u^{flat}(t)$, have to be determined in a proper way turning a non-flat ODE system into a flat system if at all possible. In doing so, potential generic system inputs are deleted by zeroing, $u(t) = 0$, in a first step. Subsequently, control affine flat inputs, $u^{flat}(t)$, are determined to derive a differentially flat system. The definition of flatness for the SISO case can be easily adapted in this context and is given below.

Definition 3 (*Differential Flatness - Flat Inputs [WZ08]*)

The system (Eq. (5.1)-(5.3)) is said to be differentially flat if there exists an input $u^{flat}(t)$ such that the resulting SISO system

$$\dot{x}(t) = f(x(t), \theta) + \gamma(x(t))u^{flat}(t) \quad ; x \in \mathbb{R}^n \, , u^{flat} \in \mathbb{R}^1 \tag{5.17}$$

$$y^{sim}(t) = h(x(t)) \quad ; y^{sim} \in \mathbb{R}^1 \tag{5.18}$$

$$x(t_0) = x_0 \tag{5.19}$$

has a relative degree of n. Here, $\gamma(x(t))$ represents the flat input vector field.

Now, for the purpose of parameter identification flat inputs, $u^{flat}(t)$, can be utilised to define suitable cost functions. If flat inputs have real physical counterparts, $u^{real}(t)$, a proper cost function has to measure the mismatch of $u^{real}(t)$ and $u^{flat}(t)$ as indicated in Fig. 5.1(b). Hence, a generic cost function might be defined as

$$\arg \min_{\theta,\tau} J_u(\theta, \tau) = \int_0^T (u^{flat}(t) - u^{real}(t))^2 dt \tag{5.20}$$

In cases of fictitious inputs, i.e., real physical counterparts do not exist in reality, the following cost function has to be implemented alternatively.

$$\arg \min_{\theta,\tau} J_u(\theta, \tau) = \int_0^T (u^{flat}(t))^2 dt \tag{5.21}$$

If needed time derivatives of output functions in Eq. (5.13) are unknown at certain

time intervals due to delays, τ, these parts have to be excluded in Eq. (5.20) and Eq. (5.21), respectively.

A much more serious problem in the framework of flat inputs for parameter identification are measurement imperfections. In practice, real measurement data, $y^{data}(t_k)$, are only available at discrete time points, t_k; $\forall k = 1, \ldots, K$, and are corrupted by measurement noise. Here, the concept of functional data analysis (FDA) [RS05, PVM$^+$06, VPM$^+$08] provides sophisticated methods for determining surrogate output functions (Eq. 5.22) that can be incorporated in Eq. (5.13) appropriately.

$$y^{surr}(t) = c^T \, \Phi(t) \tag{5.22}$$

In what follows, B-splines [RS05] are mainly applied as basis functions, $\Phi(t)$, which are fitted to the measurement data, $y^{data}(t_k)$, by optimally chosen coefficients, c. Moreover, different surrogate concepts are suitable as well. For instance, neural networks [ZP08] and wavelet concepts [DJ95] might be applied as feasible alternatives. Obviously, a proper surrogate function, $y^{surr}(t)$, and derivatives thereof, $\mathcal{D}^s y^{surr}(t)$, $\forall s = 1, \ldots, n$, are of high relevance for a meaningful evaluation of the cost functions Eq. (5.20) and Eq. (5.21), respectively. In practice, the surrogate function has to be made robust against measurement noise. For this purpose, a penalty term is frequently introduced in Eq. (5.22) [RS05] according to

$$\arg \min_c J_y^{surr}(c) = \upsilon \sum_{k=1}^{K} (y^{data}(t_k) - y^{surr}(t_k))^2 + (1-\upsilon) \int_{t_0}^{t_{end}} (\mathcal{D}^2 y^{surr}(t))^2 dt \tag{5.23}$$

By adding $\int_{t_0}^{t_{end}} (\mathcal{D}^2 y^{surr}(t))^2 dt$, the overall curvature of $y^{surr}(t)$ shall be kept low. That is, to fit the actual measurement signal but not the measurement noise. A weighting factor, $\upsilon \in [0, 1]$, has to be adapted in relation to the analysed data, $y^{data}(t_k)$, and is usually determined by intuition. As shown in Fig. 5.2, the parameter υ may strongly influence the outcome of the surrogate function parametrisation. For the purpose of illustration, measurement data are generated by a sine function which is corrupted by additive white noise, $y^{data}(t_k) = \sin(t_k) + v_k$. Only with a deliberately chosen weighting factor, $\upsilon = 0.8$, the original sine function can be approximated adequately.
To avoid the tedious work of manual tuning of υ this process is incorporated as a subroutine into the framework of parameter identification as well. In addition to the unknowns of the surrogate functions, c, and the model parameters, θ & τ, the weighting

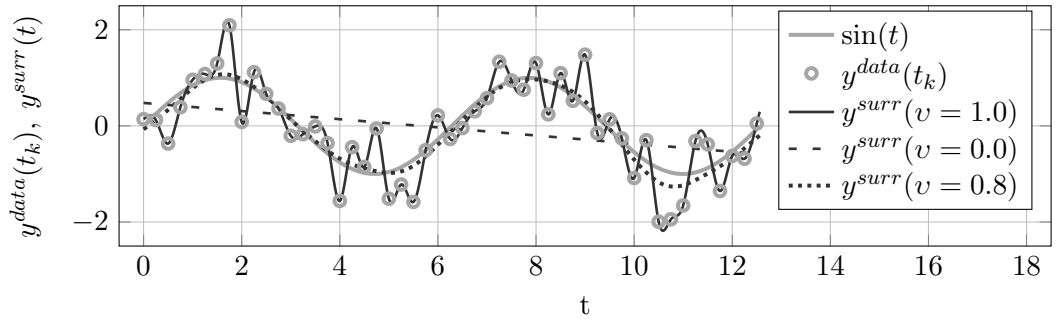


Figure 5.2: The influence of the additive penalty term, $(1-\upsilon)\int_{t_0}^{t_{end}}(\mathcal{D}^2 y^{surr}(t))^2 dt$, in determining a surrogate function, $y^{surr}(t)$, is illustrated. In the presence of discrete noisy data, $y^{data}(t_k) = \sin(t_k) + v_k;\ t_{k+1} - t_k = 0.25;\ v_k \sim \mathcal{N}(0, 0.2)$, an appropriate function $y^{surr}(t)$ can be derived using a proper υ value ($\upsilon = 0.8$).

factor, υ, is estimated by evaluating the cost functions Eq. (5.24) and Eq. (5.25), respectively. In the following test cases, one of the succeeding cost functions is applied.

$$\arg\min_{c,\upsilon,\theta,\tau} J_u(c,\upsilon,\theta,\tau) = \int_{t_0}^{t_{end}} (u^{flat}(\mathcal{D}^s y^{surr}(c,\upsilon,t),\theta,\tau) - u^{real}(t))^2 dt \tag{5.24}$$

$$\arg\min_{c,\upsilon,\theta,\tau} J_u(c,\upsilon,\theta,\tau) = \int_{t_0}^{t_{end}} (u^{flat}(\mathcal{D}^s y^{surr}(c,\upsilon,t),\theta,\tau))^2 dt \tag{5.25}$$

In cases of strongly fluctuating process dynamics, a penalty term minimising the curvature might be inappropriate. A possible way to bypass this issue consists in replacing the penalty term in Eq. (5.23) by an explicit incorporation of flat inputs, $u^{flat}(t)$, and their mismatch to $u^{real}(t)$ via

$$\arg\min_{c,\theta,\tau} J_u = \upsilon \sum_{k=1}^{K} (y^{surr}(c,t_k) - y^{data}(t_k))^2 + (1-\upsilon)\int_{t_0}^{t_{end}} (u^{flat}(t) - u^{real}(t))^2 dt \tag{5.26}$$

$$\arg\min_{c,\theta,\tau} J_u = \upsilon \sum_{k=1}^{K} (y^{c,surr}(t_k) - y^{data}(t_k))^2 + (1-\upsilon)\int_{t_0}^{t_{end}} (u^{flat}(t))^2 dt \tag{5.27}$$

The resulting cost functions are more flexible, but require a meaningful initial parameter set of θ and τ, too. The inverse model acts as a model based filter, $u^{flat}(\hat{\mathcal{S}}^{-1})$, minimising the influence of measurement noise on the determined $y^{surr}(t)$ and may be applied as an alternative to Eq. (5.24) and Eq. (5.25), respectively.

At this point it is worth noting that the presented approach can be easily implemented in existing tools for parameter identification. As shown in detail the only adaptation concerns the definition of the cost function, i.e., the standard cost function Eq. (5.20) is replaced by a suitable candidate of Eq. (5.24) -(5.27). The numerical optimisation routine on top, however, is not affected by this change. Thus, various optimisation strategies, e.g., genetic algorithms or multiple shooting methods [MMB03b, PT07], are still usable. In what concerns parameter statistics and parameter sensitivities a similar statement can be made. For instance, methods for the quantification of parameter uncertainties introduced in Sec. 3 are applicable as before. According to the presented global parameter sensitivities in Sec. 2.2.4 an equivalent counterpart of Sobol's first order indices can be defined as

$$S_i^u = \frac{\sigma^2(\underset{-i}{E}[u^{flat}|\theta[i]])}{\sigma^2(u^{flat})} \tag{5.28}$$

Finally, it should be pointed out that the concept of Optimal Experimental Design is beneficial for the flat input based parameter identification framework, too. One attempts to provide operating conditions that facilitate the overall parameter identification process. Here, the most desired outcome of OED are measurement data that are easily representable by surrogate output functions, $y^{surr}(t)$, and informative with respect to the estimates of θ and τ, respectively. Note moreover, due to the algebraic input/output representation of the ODE/DDE system there is no need of numerical integration routines in the OED framework. It is rather a question of whether an informative time course of the output function, $y^{sim}(t)$, exists at all. Therefore, the determination of an informative output function, $y^{opt}(t)$ (Eq. (5.29)), is of primary interest in this particular case. In detail, the coefficients, ζ, in Eq. (5.29) are the design parameter for OED and have to be determined by evaluating a suitable cost function, $J^{oed}(\zeta)$. For example, the basis functions, $\Phi^{oed}(t)$, might be B-splines and monomials, respectively.

$$y^{opt}(t) = \zeta^T \, \Phi^{oed}(t) \tag{5.29}$$

Once determined, the informative output functions are applied to the inverse model, $\hat{\mathrm{S}}^{-1}(\hat{\theta}, \hat{\tau})$, to calculate the corresponding input profiles. In a second step, these calculated input profiles are used to steer the physical system in the desired optimal way,

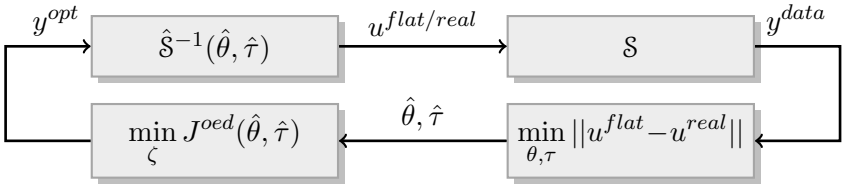


Figure 5.3: The overall scheme of OED in the field of flat input based parameter identification is shown. After a first experimental run, the resulting data, $y^{data}(t_k)$, are applied for the purpose of parameter identification, $\hat{\theta}\&\hat{\tau}$. In doing so, a new parametrised model is derived which is utilised in a subsequent part of the model-based experimental design. Subsequently, the resulting optimal output function, $y^{opt}(t)$, is applied to the inverse model, $\hat{\mathcal{S}}^{-1}$, providing an associated optimal input, $u^{flat/real}(t)$. Now, the real process, $\mathcal{S}$, is steered by $u^{flat/real}(t)$ generating new informative data. This procedure might be reiterated until sufficiently precise estimates are derived.

where optimal means to provide new informative data that are incorporated in the presented framework of flat input based parameter identification. The process of parameter identification ($\hat{\theta}$ & $\hat{\tau}$), the determination of new informative output functions, and the rerun of informative experiments might be reiterated until sufficiently precise parameter estimates are obtained. The overall work flow of OED associated to flat inputs is illustrated in Fig. 5.3.

Up to now, it has been assumed that flat inputs, $u^{flat}(t)$, have physical counterparts in reality, $u^{real}(t)$, for the purpose of OED. As previously mentioned, this is an very idealised assumption. At least a minor subset of $u^{flat}(t)$ is likely to have no corresponding physical inputs. By taking this into account, a constrained optimisation problem has to be solved. In detail, flat inputs that do not have a real equivalent ($u^{flat/fict}$) are set equal to zero by corresponding equality constraints. Whereas flat inputs which have a physical complement ($u^{flat/real}$) might be forced to stay in a certain range of physical bounds. This can be easily achieved by appropriate inequality constraints. In addition, constraints related to states might be incorporated as well. A resulting constrained optimisation problem is given below

$$\arg \min_{\zeta} J^{oed}\left(\min_{\theta,\tau} ||u^{flat} - u^{real}(\zeta)||\right) \tag{5.30}$$

$$\begin{aligned} \text{s.t. } c_1 \leq \quad & x_i & \leq c_2 \\ c_3 \leq \quad & u^{flat/real} & \leq c_4 \end{aligned}$$

In cases at which no real counterpart of $u^{flat}(t)$ exist at all, a potential workaround

might be - if determinable - the introduction of a so-called dynamic compensator (pre-filter) which turns fictitious inputs into physically applicable inputs on the condition that the resulting ODE system possesses stable internal dynamics [SSK09].

5.4 Determination of Flat Inputs

The aim of this subsection is to provide some helpful hints about the determination of flat inputs, $u^{flat}(t)$. Necessary and sufficient conditions for the existence of flat inputs are not known for the general multiple-input and multiple-output (MIMO) case [WZ10]. Nevertheless, a systematic procedure of determining flat inputs for observable systems exists [WZ10]. These results are a generalisation of conditions which have been provided for single-input and single-output (SISO) systems [WZ08]. The basic idea of flat input determination in the SISO case (Eq. (5.17)) is based on the determination of a proper input vector field, $\gamma(x(t))$. In case of observable systems, i.e., the observability matrix (Eq. (5.31))[1] is locally regular, $det(\mathcal{Q}(x)) \neq 0, \quad \forall x \in \mathcal{N}(x)$, the determination of $\gamma(x(t))$ can be done easily.

$$\mathcal{Q}(x) = \frac{\partial}{\partial x} \begin{pmatrix} h(x) \\ L_{f(x_0)}h(x) \\ \vdots \\ L^{n-1}_{f(x_0)}h(x) \end{pmatrix} \tag{5.31}$$

In particular, by assuming a non-zero scaling function, $\vartheta(x) \neq 0$, the flat input vector field can be expressed according to

$$\gamma(x) = \vartheta(x)\mathcal{Q}^{-1}(x)(\underbrace{0, \ldots, 0, 1}_{n \text{ entries}})^T \tag{5.32}$$

In doing so, the resulting system (Eq. (5.17)-(5.19)) has a relative degree that is equivalent to the dimension of the state space ($\nu = n$), i.e., steering the system with the determined flat input, $u^{flat}(t)$, the associated ODE system is rendered into a differentially flat system.

The same procedure can be applied for a broad class of MIMO systems [WZ10]. For special cases, however, the previous framework of flat input determination fails, i.e., for the general MIMO case no necessary conditions for existence of flat inputs can be given [WZ10]. The same is true for general DDE systems, i.e., there is no rigorous

[1] Here, L_f means the Lie derivative of a function along the vector field f as stated in [WZ08].

proof for the existence of flat inputs as well as no systematic procedure of flat input determination. Hence, a heuristic approach is shortly presented in the following. In flat system theory it is well known that the dimension of the input, $u \in \mathbb{R}^l$, has to be equivalent to the dimension of the output, $y \in \mathbb{R}^m$, $l \stackrel{!}{=} m$. Thus, for a given number of measurable outputs the number of needed flat inputs, $u^{flat}(t)$, is defined immediately. Now, the question at which states, $x(t)$, these flat inputs have to act on has to be addressed. In principle, this can be answered heuristically by the framework of structural analysis. First attempts in this direction can be found in the context of flat outputs [Wey02]. The structural analysis is based on directed graphs (digraphs), $D(v, e)$, with n different nodes, v_i, representing the states of the ODE/DDE system. That means, the existence of an edge, $e_{i,j}$, from node v_i to v_j is determined by non-zero elements, $a_{j,i}$, of the adjacency matrix, A^*. Here, the $a_{i,j}$ element of A^* is set to 1 if the associated derivative, $\frac{\partial f_i(x(t),x(t-\tau),\theta)}{\partial x_j(t)}$ or $\frac{\partial f_i(x(t),x(t-\tau),\theta)}{\partial x_j(t-\tau)}$, exists and equal to 0 if this is not the case. Equivalently, an adjacency matrix C^* of the output functions, $y^{sim}(t)$ (Eq. 5.2), can be derived. Now, flat inputs, $u^{flat}(t)$, should act on those nodes, v_i, that are the most distant nodes in combination to the outputs, $y^{sim}(t)$, i.e., there is a maximum number of edges to get from $u^{flat}(t)$ to $y^{sim}(t)$ travelling along the shortest path. Thus, a max-min optimisation problem has to be solved.

In order to illustrate the presented flat input selection strategy, the following parameter-free DDE system is analysed

$$\begin{aligned} \dot{x}_1(t) &= x_3(t-\tau_1) \\ \dot{x}_2(t) &= x_1(t) \\ \dot{x}_3(t) &= x_2(t) \\ y^{sim}(t) &= x_2(t) \end{aligned} \tag{5.33}$$

Due to the related adjacency matrices (Eq. 5.34) a digraph can be generated, see Fig.5.4. Obviously, a suitable flat input, $u^{flat}(t)$, should enter at node x_3 to provide the most distantly shortest path.

$$A^* = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}; \quad C^* = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \tag{5.34}$$

According to the above outcome, a potential input affine DDE system can be derived

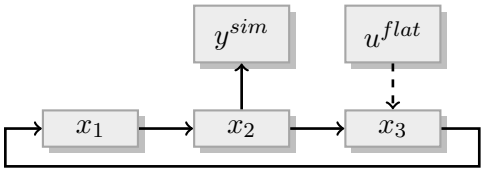


Figure 5.4: Digraph of the illustrative example: For a given output, y^{sim}, the flat input, u^{flat}, has to act on state x_3 to provide a maximum distance to get from u^{flat} to y^{sim} using a minimum number of edges.

as

$$\begin{aligned} \dot{x}_1(t) &= x_3(t-\tau_1) \\ \dot{x}_2(t) &= x_1(t) \\ \dot{x}_3(t) &= x_2(t) + u^{flat}(t) \\ y^{sim}(t) &= x_2(t) \end{aligned} \tag{5.35}$$

In fact, the resulting DDE system (Eq. 5.35) is differentially flat, i.e., all states, $x_i(t)$, as well as the flat input, $u^{flat}(t)$, can be recalculated by the output, $y^{sim}(t)$, and derivatives thereof as demonstrated below.

$$x_2(t) = y^{sim}(t) \tag{5.36}$$

$$x_1(t) = \mathcal{D}^1 y^{sim}(t) \tag{5.37}$$

$$x_3(t-\tau_1) = \mathcal{D}^2 y^{sim}(t) \tag{5.38}$$

$$u^{flat}(t) = \mathcal{D}^3 y^{sim}(t+\tau_1) - y^{sim}(t) \tag{5.39}$$

Remember, $\mathcal{D}^i$ represents the operator notation of the ith derivative, d^i/dt^i. In general, more complex systems can be checked in this way by highly efficient methods of graph analysis, e.g., the algorithm of Dijkstra shortest path [Dij59] might be applied for a systematic analysis of complex ODE/DDE systems.

5.5 Introduction to the Case Studies

Throughout this section the concept of flat inputs for parameter identification is applied for the following three case studies.

a) Parameter identification for the FitzHugh-Nagumo equations

b) Parameter identification for a MAP Cascade model

c) Parameter identification for an influenca A virus production model

In the first case study, in silico data, y^{data}, are generated and incorporated in the frame work of parameter identification. Here, a benchmark study demonstrates potential benefits of the presented method compared to the standard approach for parameter identification under almost perfect measurement data. That means, their is no measurement noise. Subsequently, measurement noise is introduced and its influence on the parameter identification is analysed. Therefore, the in silico data, $y^{data}(t_k)$, are perturbed by a Gaussian random variable, $\nu_k \sim \mathcal{N}(0, \sigma_y^2)$. Naturally, the presented concepts of Optimal Experimental Design are beneficial for flat input based parameter identification as well. Hence, a new OED is derived that provide new informative data, i.e., the influence of measurement imperfections to the identified model parameters is reduced significantly.

In what follows, the focus is not that much on the impact of measurement noise to identified parameters. In fact, shortcomings of the standard approach for parameter identification are illustrated that show up frequently even under conditions that are beneficial for the actual identification step, i.e., a large number of noise free measurement samples is assumed. For instance, in the second case study it is shown that the internal dynamics of a system might be insensitive to the measurement data, $y^{data}(t_k)$. In many cases, the measurements are an integral quantity of the overall process, i.e., internal highly dynamical fluctuations near the input of a model are attenuated through the underlying process itself. Consequently, only a minor subset of model parameters is sensitive to the output. The majority of unknown parameters, however, can be changed by order of magnitudes without a significant variation on $y^{sim}(t)$, i.e., these parameters are likely to be hard to identify practically. Here, it is shown that the reconstructed flat input is much more sensitive to the internal dynamics. Consequently, a different set of associated model parameters is sensitive to the flat input and might become more precisely identifiable.

A similar effect is shown in the third test case. By analysing a DDE system, in silico data are used to identify model parameters as well as a time delay parameter. Compared to the standard approach of parameter identification the flat input based identification framework is robust against initial model parameters and the time delay parameter, respectively.

5.5.1 Parameter Identification for FitzHugh-Nagumo Equations

The first considered model (Eq. (5.40)) describes the electro-physiology of a nerve axon. Here, the quantities V and R represent the voltage and the recovery of the membrane, respectively. This model is known as the so-called FitzHugh-Nagumo equations and was derived by FitzHugh [Fit61] and Nagumo [NAY62] independently. The aim of this model is not to explore the underlying biochemical principles in detail but to describe essential features qualitatively, i.e., the action potential and their spiking property are reproduced. In detail, the FitzHugh-Nagumo equations are a simplification of the well known Hodgkin and Huxley model [HH52] and have been intensively used in the field of non-linear analysis, see [Sey10] and references therein. Moreover, this ODE system was also used in [RHCC07] as a test case for their generalised smoothing approach for parameter identification.

$$\begin{aligned} \dot{V} &= c\left(V - \frac{V^3}{3} + R\right) \\ \dot{R} &= -\frac{1}{c}(V - a + bR) \end{aligned} \tag{5.40}$$

Throughout this case study, the parameters a and b have to be identified whereas c is assumed to be known from literature. For the purpose of parameter identification, in silico measurement data of the membrane voltage are provided by Eq. (5.41).

$$y^{data}(t_k) = V(t_k) \tag{5.41}$$

Obviously, the FitzHugh-Nagumo equations are an autonomous ODE system, i.e., there is no physical input, $u(t)$. As one quantity is measured, one fictitious flat input, $u^{flat}(t)$, has to be determined to render Eq. (5.40) into a differentially flat SISO representation. In the next step, the question has to be addressed at which state this flat input has to act on. For this purpose, concepts which are presented in Sec. 5.4 are applied. As a SISO ODE system is analysed a flat input vector field, $\gamma(x(t))$, can be derived systematically. The corresponding observability matrix has to be regular which is confirmed below.

$$\mathcal{Q}(x) = \begin{bmatrix} 1 & 0 \\ c(1-V)^2 & c \end{bmatrix} \tag{5.42}$$

$$det(\mathcal{Q}(x)) = c \tag{5.43}$$

By assuming $c \neq 0$ and $\alpha(x) = c$, a resulting flat input vector field can be derived according to

$$\gamma(x) = \alpha(x)\mathcal{Q}^{-1}(x)[0,1]^T = [0,1]^T \tag{5.44}$$

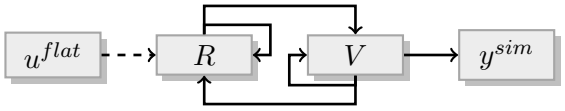


Figure 5.5: Digraph of the FitzHugh-Nagumo equations: For a given output y^{sim}, the flat input u^{flat} has to act on R to provide a maximum distance to get from u^{flat} to y^{sim} using a minimum number of edges.

Subsequently, a differentially flat SISO system can be determined immediately as

$$\begin{aligned} \dot{V}(t) &= c\left(V(t) - \frac{V^3(t)}{3} + R(t)\right) \\ \dot{R}(t) &= -\frac{1}{c}(V(t) - a + bR(t)) + u^{flat}(t) \\ y^{sim}(t) &= V(t) \end{aligned} \tag{5.45}$$

Alternatively, the structural analysis concept is put in operation. The associated adjacency matrices are given by

$$A^* = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}; \quad C^* = \begin{bmatrix} 0 & 1 \end{bmatrix} \tag{5.46}$$

The corresponding digraph in Fig. 5.5 agrees well with the above result of flat input detection. Accordingly, the flat input, u^{flat}, has to act on R to provide a maximum distance to get from u^{flat} to y^{sim} utilising a minimum number of edges.

After the determination of a flat input, the differentially flat ODE system (Eq. (5.45)) can be transformed into the algebraic input/output representation. As shown in Eq. (5.47) the states, $R(t)$ and $V(t)$, as well as the defined flat input, $u^{flat}(t)$, can be determined by $y^{sim}(t)$ and derivatives thereof. In the following, $u^{flat}(t)$ and its applicability for the purpose of parameter identification is analysed in detail.

$$\begin{aligned} V(t) &= y^{sim}(t) \\ R(t) &= \frac{1}{c}\mathcal{D}y^{sim}(t) - \frac{3y^{sim}(t) + {y^{sim}(t)}^3}{3} \\ u^{flat}(t) &= \frac{1}{c}\mathcal{D}^2 y^{sim}(t) - \mathcal{D}y^{sim}(t)\left(1 - {y^{sim}(t)}^2\right) - \frac{1}{c}\left(y^{sim}(t) - a + bR(t)\right) \end{aligned} \tag{5.47}$$

First of all, the cost functions of the standard approach, J_y (Eq. (5.4)), and of the flat input based method, J_u (Eq. (5.25)), are evaluated for different model parameter configurations. Therefore, noise free measurement data of the voltage are generated by Eq. (5.40) on conditions that are given in Tab. 5.1. Here, a sampling rate of $\Delta t = 0.05$ ms

is assumed and the overall experimental run is limited to $t^{end} = 20$ ms. Corresponding simulation results are shown in Fig. (5.6).

a	b	c	$V(t=0)$	$R(t=0)$	Δt	σ_y^2
0.2	0.2	3	-1	1	0.05	0.04

Table 5.1: Configuration parameters of the test case study at which the in silico data, $y^{data}(t_k)$, are generated.

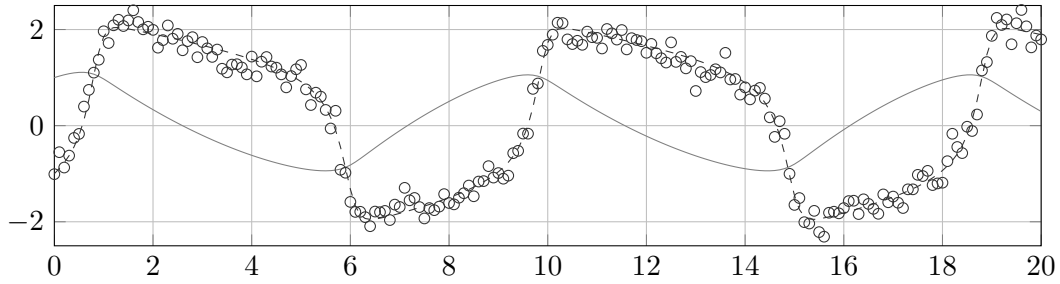
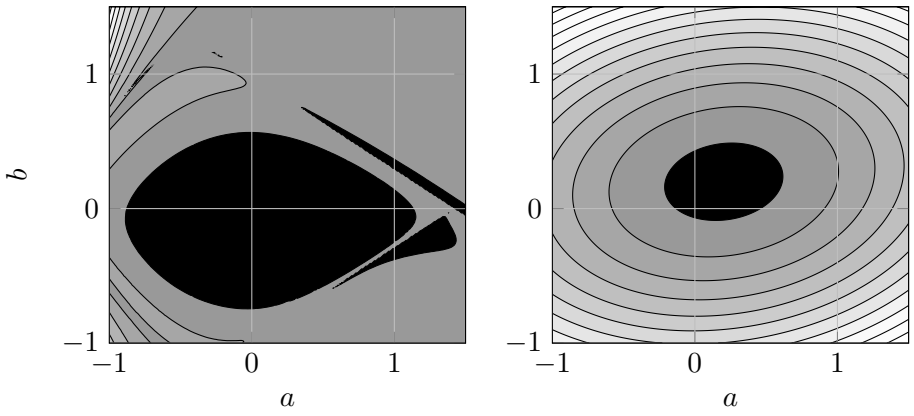


Figure 5.6: Simulation results of the FitzHugh-Nagumo equations. In detail, —— represents the membrane recovery (R), - - - the voltage (V), and ○ illustrates discrete noisy simulated data of V.

The analysed cost functions are evaluated iteratively in a predefined parameter range of $-1 \leq a, b \leq 1.5$. Each individual parameter range is separated in 251 intervals which leads to an overall number of 63001 evaluations of J_y (Eq. (5.4)) and J_u (Eq. (5.25)), respectively. Subsequently, an assessment of the used cpu-time (t^{cpu}) is done. As previously mentioned, the need of an ODE solver in the subroutine of the standard cost function slows down the evaluation process significantly. The total computation of corresponding cost function values consumes $t^{cpu}(y^{sim}) = 5707.1$ sec. Alternatively, the overall computation of cost function values which are based on $u^{flat}(t)$ needs $t^{cpu}(u^{flat}) = 14.8$ sec in total. Obviously, the algebraic input/output representation of the original ODE system leads to a significant reduction in terms of the computational load. As any parameter optimisation routine has to evaluate the cost functions Eq. (5.4) or Eq. (5.25) inherently, the cpu-non-intensive concept of flat inputs is beneficial for a wide class of optimisation algorithms.

Analogous to the golden rule of mechanics which essentially states that whatever one saves in power one has to invest in displacement, the question of negative side effects of the flat input based approach comes up. As shown, the proposed cost function, J_u

(Eq. (5.25)), may reduce the need of cpu-power drastically in the framework of parameter identification. This effect might be compensated for by an increased number of the overall function evaluations due to an increased complexity of cost function surface, i.e., the actual parameter identification problem might become ill-conditioned and, therefore, challengingly to explore. For instance, if the resulting cost function possesses several local minima a global parameter optimisation algorithm has to be used which usually corresponds to an increased number of function evaluations. For this particular application, the opposite is true. The method of flat inputs saves not only cpu-power, but also changes the cost function complexity in a beneficial manner. The surfaces of the corresponding cost functions, J_y and J_u, are illustrated in Fig. 5.7(a) and 5.7(b). Obviously, the implementation of the standard cost function leads to a non-convex optimisation problem, i.e., the cost function has a number of local minima which turns the parameter identification into a hard to solve issue for common optimisation routines. On the contrary, the cost function J_u which is based on $u^{flat}(t)$ renders the optimisation process into a well-posed problem. That means, the resulting cost function surface (Fig. 5.7(b)) relates to a convex optimisation surface which can be efficiently evaluated by standard optimisation algorithms.

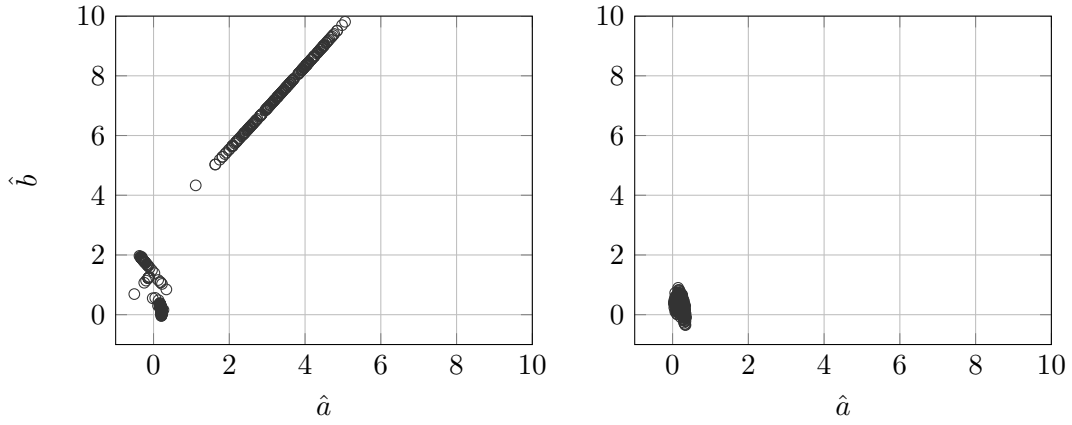


(a) Cost Function - Standard Approach (b) Cost Function - Flat Input Approach

Figure 5.7: The resulting cost function of the standard approach (left sub-figure) possesses several local minima, i.e., the optimisation problem is ill-posed and difficult to evaluate properly. Whereas using the same measurement data as done before, the cost function which is based on the flat input (Eq. (5.25)) ensures a global minima, i.e., the optimisation problem is well-posed. (Low values are shown from dark grey to black, whereas high values are shown from light grey to white.)

Up to now, only noise free measurement data are considered. In addition, the influence

of measurement imperfection is analysed in what follows. For this purpose, the simulated data are corrupted by a Gaussian random variable, $\epsilon \sim \mathcal{N}(0, 0.04)$, which corresponds to a standard deviation of at least 10% of the generated in silico data. Moreover, also the initial estimates values are changed randomly in the range of $0 \leq a, b \leq 10$. In doing so, 1000 optimisation problems are generated and solved by evaluating J_y (Eq. (5.4)) and J_u (Eq. (5.25)), respectively. In detail, the Matlab® built-in Levenberg-Marquardt algorithm (lsqnonlin) at its default configuration is applied for the actual parameter identification where the parameter range is limited to $-1 \leq a, b \leq 10$. The resulting parameter estimates, $\hat{\theta}$, are illustrated in Fig. 5.8(a) and 5.8(b), respectively. As might be expected, the standard approach (J_y) produces a strong variation in the identified parameters due to the non-convexity of the cost function. By contrast, the usage of the flat input based cost function, J_u, ensures a credible identification of the two parameters, $\hat{\theta} = (\hat{a}, \hat{b})$.



(a) Scatter Plot - Standard Approach

(b) Scatter Plot - Flat Input Approach

Figure 5.8: Running 1.000 parameter identifications the resulting estimates, $\hat{a}$ & $\hat{b}$, are visualised. In doing so, the initial parameters are changed randomly in the range of $0 \leq a, b \leq 10$, and measurement data are corrupted by white noise. In the left sub-figure the standard approach is utilised which leads to significant uncertainty about parameter estimates. Using the flat-input based approach instead the estimates are closely grouped in the near of the nominal parameter values, $a = b = 0.2$.

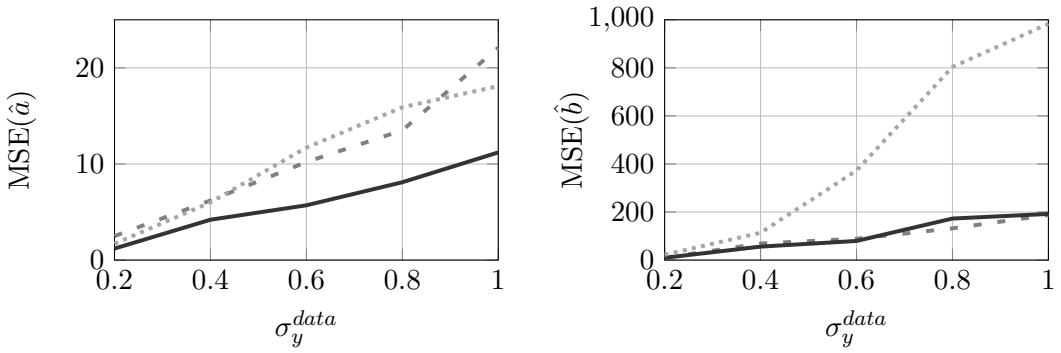
Additionally, the computational effort of both approaches is compared and summarised in Tab. 5.2. By assessing the consumed cpu-time and reliability of parameter estimates, the concept of flat inputs outperforms the standard approach even in the presence of measurement imperfections. For instance, the total cpu-time of the parameter identification process is reduced from 24394.1 seconds to 661.4 seconds, respectively. By

taking the overall number of function calls into account (Tab. 5.2) the average cpu-time for evaluating J_y is roughly about 0.34 seconds due to the inherent use of an ODE solver. In the case of J_u, a single evaluation takes 0.01 seconds, i.e., the input/output representation of the ODE system in the framework of parameter identification reduces the computational load approximately by a factor of 30. As mentioned previously, wavelets and neural networks may also be suitable candidates in providing surrogate output functions, $y^{surr}(t)$. Hence, a benchmark study is performed which compares the credibility of parameter estimates in relation to an increased level of measurement noise. The result is summarised in Fig. 5.9(a) and 5.9(b), respectively. Further details about the de facto applied algorithms are given in Appendix A.8. As might be expected, an increased level of measurement imperfection leads to an increased level of parameter uncertainties which can be seen by high mean square error values. In this particular case, surrogate functions based on wavelets and neural networks seem to be suitable candidates in the presence of measurement noise.

	$\hat{a}$	$\sigma_{\hat{a}}$	$\hat{b}$	$\sigma_{\hat{b}}$	t^{cpu}(PI)[sec]	$t^{cpu}(J_{y/u})$[sec]	calls($J_{y/u}$)
PI(J_y)	0.825	1.355	1.765	2.952	24394.1	24004.8	70503
PI(J_u)	0.190	0.058	0.327	0.168	661.4	551.4	47000
PI(J_u^{opt})	0.201	0.050	0.208	0.024	704.3	652.7	110943

Table 5.2: Result of parameter identification by minimising a cost function based on the traditional (J_y) as well as on flat input-based (J_u) approach, respectively. Here, the concept of flatness outperforms the traditional approach in the meaning of credibility and computational effort as well. An optimal experimental design (J_u^{opt}) ensures a precise estimate of the two parameters, a and b. t^{cpu}(PI)[sec] represents the total cpu-time of the overall parameter identification process and $t^{cpu}(J_{y/u})$[sec] the cpu-time dedicated to the actual cost function evaluation.

Obviously, the estimates of b are strongly affected by the measurement noise independently of the applied surrogate output function concept. Subsequently, more informative measurement data are needed to reduce the uncertainties about $\hat{b}$. Hence, the parameter sensitivities (Sec. 2.2.4) associated to $u^{flat}(t)$ are of high interest. By analysing Eq. (5.48) and Eq. (5.49) it is evident that u^{flat} depends linearly on a and b, respectively. Consequently, sensitivities based on linearisation are appropriate for the purpose of Optimal Experimental Design. When interested in low parameter uncertainties, i.e., the standard deviations of $\sigma_{\hat{\theta}}$ have to be small, one has to increase the parameter sensitivities as stated in Sec. 3. Here, only the parameter sensitivity about



(a) Benchmark Surrogate Methods

(b) Benchmark Surrogate Methods

Figure 5.9: Various approaches are applicable for the purpose of determining $y^{surr}(t)$. Here, wavelets (——), neural network (- -), and B-Splines (·····) concepts are compared. In doing so, the mean square error (MSE) of parameter estimates is shown at different levels of measurement noise, σ_y^{data}.

b (Eq. (5.49)) can be influenced by OED whereas the corresponding sensitivity about a depends solely on the constant c. In conclusion, the objective of OED is to maximise the parameter sensitivity about b (Eq. (5.49)) by the determination of an informative output function, $y^{opt}(t)$.

$$\frac{\partial u^{flat}(t)}{\partial a} = -\frac{1}{c} \tag{5.48}$$

$$\frac{\partial u^{flat}(t)}{\partial b} = \left(\frac{1}{c}\right)^2 \mathcal{D}y^{sim}(t) - \frac{1}{c}\left(\frac{3y^{sim}(t) - y^{sim}(t)^3}{3}\right) \tag{5.49}$$

A proper candidate of $y^{opt}(t)$ has to be easily representable by a surrogate output function, $y^{surr}(t)$. In what follows, the previous "worst case" approach in the presence of measurement noise, the cubic B-splines, is used. Now, a tailor-made OED strategy is performed, i.e., by using different approaches for the determination of $y^{surr}(t)$ also different OED strategies might be required. As shown previously, a penalty term ($\mathcal{D}^2 y^{surr}(t)$) is incorporated in Eq. (5.23) for the determination of $y^{surr}(t)$. That means, $y^{opt}(t)$ should have a minimum curvature inherently (Eq. (5.50)) to ensure a reliable discrimination between the measurement signal and the measurement noise.

$$\arg\min_{\zeta} \left\{ \int_{t_0}^{t_{end}} (\mathcal{D}^2 y^{opt}(t))^2 dt \right\} \tag{5.50}$$

Consequently, a potential candidate of $y^{opt}(t)$ is the equation of a straight line (Eq.

(5.51)) which fulfils the previous condition (Eq. (5.50)) by definition.

$$y^{opt}(t) = \zeta_0 \tag{5.51}$$

Now, the question has to be addressed how to determine the design parameter, ζ_0, for the purpose of sensitivity maximisation. Therefore, a constrained optimisation problem is formulated according to

$$\begin{aligned} \arg\ \max_{\zeta_0} J^{opt}(\zeta_0) = & \int_{t_0}^{t_{end}} \left(\frac{\partial u^{flat}(t)}{\partial b} \right)^2 dt \\ \zeta_0 \geq & \ 0; \\ u^{flat/real}(t) \leq & \ 1 \end{aligned} \tag{5.52}$$

In fact, an analytic solution can be derived as

$$y^{opt}(t) = 2.4269 \tag{5.53}$$

In the next step, the determined optimal output, $y^{opt}(t)$, is applied to the inverse model, $\hat{S}^{-1}$, providing the corresponding optimal input, $u^{opt}(t) = u^{flat/real}(t) = 1$. In practice, $u^{opt}(t)$ is used to steer the next experimental run optimally, i.e., to provide informative measurement data, $y^{data}(t_k)$. The new data are illustrated in Fig. 5.10. Obviously, the expected value of the data, $y^{data}(t)$, is close to the desired optimal output of $y^{opt}(t) = 2.4269$. The minor mismatch of both can be explained by the previous estimates of θ which are similar but not equal to the nominal parameter values. That means, the recalculated optimal flat input, $u^{flat/real}$, depends on the credibility of the estimates, $\hat{\theta} = (\hat{a}, \hat{b})$ and might be suboptimal in cases at which $\hat{\theta}$ differs strongly in comparison to the the nominal values of θ. To a certain extent this effect might be compensated for by a number of OED reiterations and data generation, respectively, see Fig. 5.3.

After the experimental run, the new data, which are expected to be informative for $\hat{b}$, are used in addition to the previous data to determine $\hat{\theta} = (\hat{a}, \hat{b})$ by minimising J_u (Eq. (5.25)). In fact, the credibility of $\hat{b}$ is improved significantly, see Tab. 5.2 as well as Fig. 5.11(a),5.11(b).

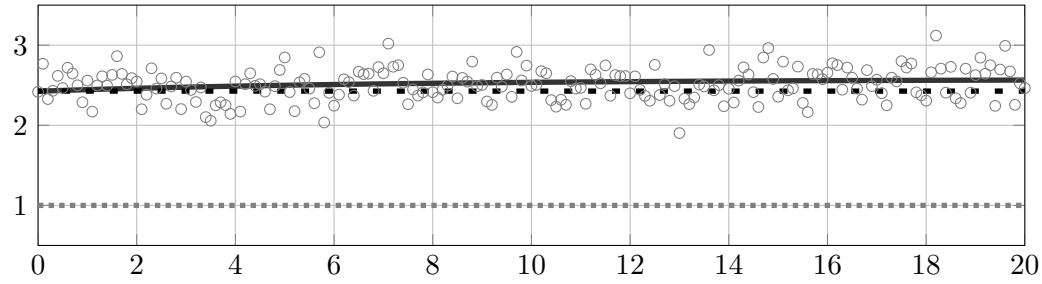
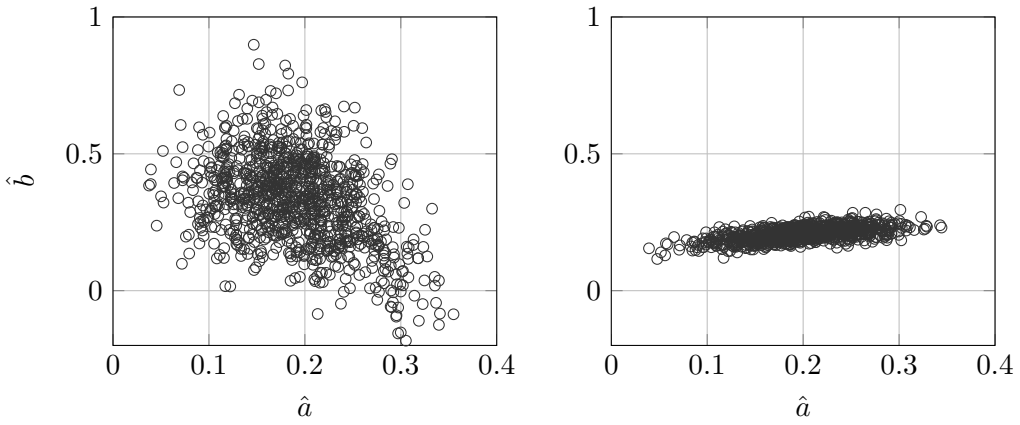


Figure 5.10: Based on the parametrised model an optimal stimulus, $u^{opt}(t)$ ·····, is calculated which is associated to the optimal output function, $y^{opt}(t) = 2.4269$ ▪ ▪ . By applying $u_{opt}(t)$ to the physical system, S, measurement data, $y^{data}(t_k)$ ○, are generated. The mean values of these data, ▬, are in good agreement with the previously determined $y^{opt}(t)$ function.

5.5.2 Parameter Identification for a MAP Kinase Model

The biological meaning of the MAP kinase was presented in Sec.4.3 in detail. Here, the focus is on parameter sensitivities in relation to $y^{sim}(t)$ and $u^{flat}(t)$, respectively. In this case study, the ODE system given in Eq. (5.54) is assumed to describe the signalling process approximately. First of all, the ODE system is analysed if the predefined input/output configuration is a differential flat system. After the adjacency matrices are derived in Eq. (5.55) the corresponding digraph illustrated in Fig. 5.12 is analysed. Obviously, the input u enters at the node which is the most distant node in relation to the output y^{sim}. As shown previously, this outcome is a basic prerequisite for u to be a flat input, $u^{flat} = u$.

$$
\begin{aligned}
\dot{KKK^*} &= \frac{k1 \cdot u \cdot (1 - KKK^*)}{k1m + (1 - KKK^*)} - \frac{v2 \cdot KKK^*}{k2m + KKK^*} - \frac{k5 \cdot K^* \cdot KKK^*}{k5m + KKK^*} \\
\dot{KK^*} &= \frac{k6 \cdot KKK^* \cdot (1 - KK^*)}{k6m + (1 - KK^*)} - \frac{v7 \cdot KK^*}{k7m + KK^*} \\
\dot{K^*} &= \frac{k3 \cdot KK^* \cdot (1 - K^*)}{k3m + (1 - K^*)} - \frac{v4 \cdot K^*}{k4m + K^*} \\
&\text{with} \\
1 &= KKK + KKK^* \\
1 &= KK + KK^* \\
1 &= K + K^* \\
y^{sim}(t) &= K^*
\end{aligned}
\tag{5.54}
$$



(a) Scatter Plot - Non-OED (b) Scatter Plot - OED

Figure 5.11: Here, 1.000 parameter estimates are performed using noisy data. In the left sub-figure the result of an initial experiment (Non-OED experiment) is shown. Using in addition measurement data which are generated by OED, the uncertainty about $\hat{b}$ is reduced significantly.

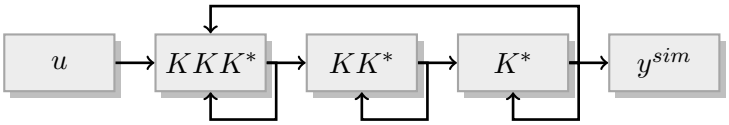


Figure 5.12: Digraph of the MAP kinase model: For a given input/output configuration the input u enters at the most distant node in relation to y^{sim}. This result is a basic prerequisite for a system to be differentially flat.

$$A^* = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}; \quad C^* = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \tag{5.55}$$

Actually, the MAP kinase model can be reformulated to an algebraic input/output representation by $y^{sim}(t)$ and derivatives thereof as shown by the condensed expressions below.

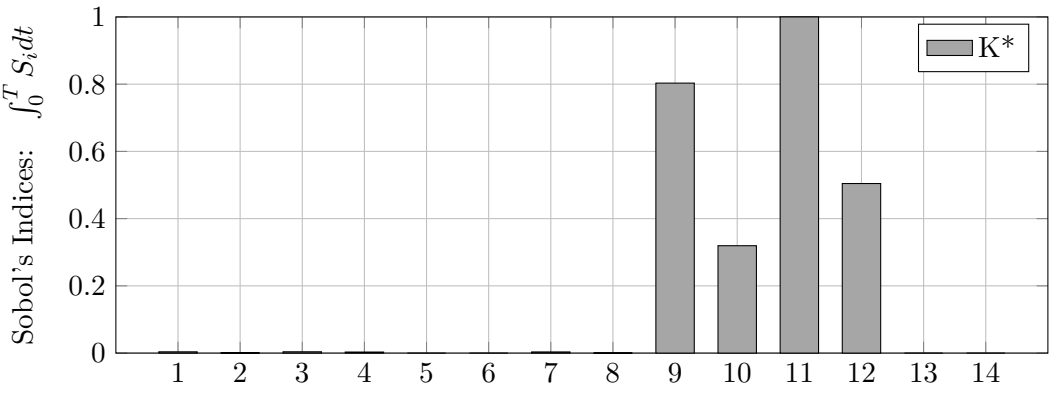
$$\begin{aligned} K^*(t) &= y^{sim}(t) \\ KK^*(t) &= \Psi_{KK^*}\left(\theta, y^{flat}(t), \mathcal{D}y^{sim}(t)\right) \\ KKK^*(t) &= \Psi_{KKK^*}\left(\theta, y^{flat}(t), \mathcal{D}y^{sim}(t), \mathcal{D}^2 y^{sim}(t)\right) \\ u^{flat}(t) &= \Psi_{u}\left(\theta, y^{flat}(t), \mathcal{D}y^{sim}(t), \mathcal{D}^2 y^{sim}(t), \mathcal{D}^3 y^{sim}(t)\right) \end{aligned} \tag{5.56}$$

The detailed functional expressions are given in App. A.10.

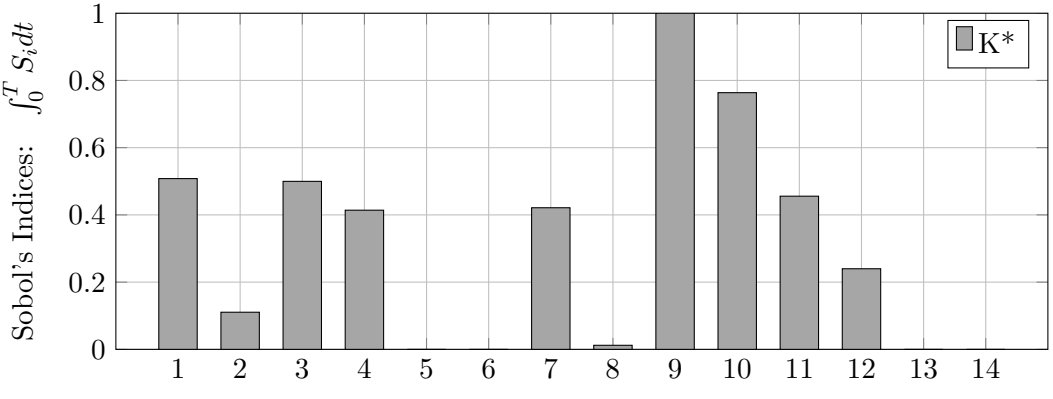
Moreover, Sobol' indices of first order are determined for $y^{sim}(t)$ as well as for $u^{flat}(t)$. Initial conditions, x_0, the input, $u^{real}(t)$, the model parameters, θ, and the measurement layout have to be specified as done in App. A.6. Subsequently, the Sobol' indices associated to $y^{sim}(t)$ are calculated by the presented approach given in Sec. 2.2.4. Here, the Sobol' indices are integrated with respect to simulation time and normalised by the largest entry of these integral quantities. The result given in Fig. 5.13(a) illustrates clearly that $y^{sim}(t)$ is only sensitive to a minor subpart of θ. Thus, most of the parameters are likely to be poorly identified, i.e., estimates of θ might be strongly biased and/or possess large confidence intervals. Consequently, the desired outcome of modelling, the development of a predictive model, is counteracted by an imprecise parametrisation. In this case, methods of OED presented in Sec. 3 are of inestimable value for the purpose of a reliable parameter identification. A serious drawback of OED, however, is the inevitable need of new measurement data, $y^{data}(t_k)$, at optimised operating conditions. In cases at which no rerun of an experiment is feasible at all, e.g., due to limitations in term of money and/or time, the concept of OED has to be fundamentally reconsidered. In this context, it is rather a question of whether existing measurement data can be recycled more informatively for the purpose of parameter identification. In detail, one has to address the questions how insensitive parameters become sensitive using the same "old" measurement data. This may sound counterintuitive at the first glance but becomes more clear in application.

Therefore, Sobol' indices, S_i^u, related to the flat input, $u^{flat}(t)$, are determined as defined previously in Eq. (5.28). Though, the same in-silico date are utilised as before the range of parameter sensitivities, presented in Fig. 5.13(b), looks quite different to those based on the standard approach (Fig. 5.13(a)). Using the same measurement data but applying the concept of flat inputs an extended set of model parameters, θ, is sensitive to u^{flat} and J_u, respectively. An increased number of parameters become practically identifiable by evaluating J_u instead of J_y. Such an outcome satisfies the overall objective of OED evidently, i.e., an increased number of parameters is likely to be identified more precisely.

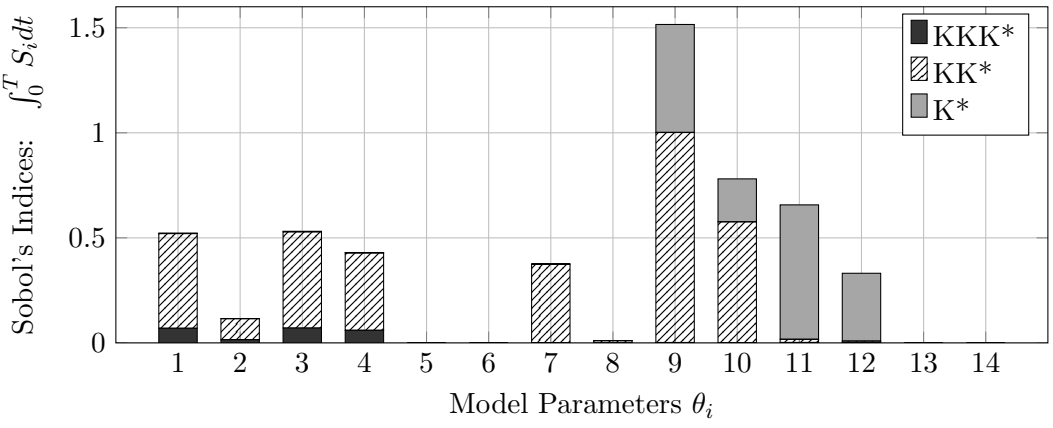
As shown in Fig. 5.13(c), by applying the standard approach a similar range of parameter sensitivities can only be derived by measuring at least the phosphorylated enzyme, $y_2^{sim}(t) = KK^*(t)$, additionally.



(a) Standard approach

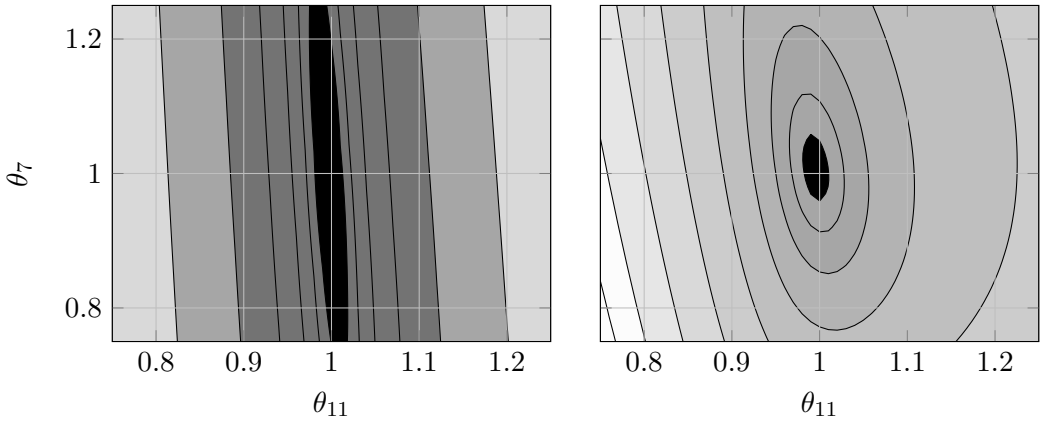


(b) Flat input based approach



(c) Standard approach - measuring all states

Figure 5.13: Integral measures of Sobol's indices, $\int_0^T S_i dt$, normalised to the most sensitive parameter are presented for the two different strategies of parameter identification, sub-figures (a) and (b), assuming $y^{data}(t_k) = K^*$. In (a), the spectrum of parameter sensitivities is shown for the standard approach, i.e., parameter sensitivities related to outputs, $y^{sim}(t)$. In (b), parameter sensitivities related to flat inputs, $u^{flat}(t)$, are given. In subfigure (c), Sobol' indices are given for the standard approach but assuming that all states are measurable.



(a) Standard approach

(b) Flat input based approach

Figure 5.14: The surface of the cost function associated to the parameters, θ_7 and θ_{11}, is compared. (Low values are shown dark grey to black, whereas high values are shown light grey to white.) In the left sub-figure the standard approach is used. Here, the insensitivity of θ_7 in relation to J_y leads to an ill-posed parameter identification problem, i.e., a proper identification of θ_7 is hard to derive. In contrast, using the flat input based approach (right sub-figure) the identification is turned into an easier to evaluate issue, i.e., the two parameters, θ_7 as well as θ_{11}, are sensitive to J_u.

5.5.3 Parameter Identification for a Virus Replication Model

The following DDE system (Eq. 5.57) describes the influenza A virus production in large-scale microcarrier culture by a model similar to [MFSR05]. Uninfected Madin-Darby canine kidney (MDCK) cells, $U_c(t)$, are infected by active viruses, $V_i(t)$. After a certain delay time of τ_1, infected MDCK cells, $I_c(t)$, release active and inactive virus particles, $V_i(t)$ and $V_d(t)$, respectively. The active virus particles either infect the remaining uninfected cells or are degraded to inactive virions.

$$\begin{aligned}
\frac{dU_c(t)}{dt} &= \theta_6 \frac{C_{max} - (U_c(t) + I_c(t))}{C_{max}} U_c(t) - \\
&\quad \theta_1 U_c(t) V_i(t) \\
\frac{dI_c(t)}{dt} &= \theta_1 U_c(t) V_i(t) - \theta_2 I_c(t) \\
\frac{dV_i(t)}{dt} &= \theta_3 I_c(t - \tau_1) - \theta_4 V_i(t) - \theta_1 U_c(t) V_i(t) \\
\frac{dV_d(t)}{dt} &= \theta_5 I_c(t - \tau_1) + \theta_4 V_i(t)
\end{aligned} \tag{5.57}$$

Here, it is assumed that the concentrations of active and inactive virus particles are

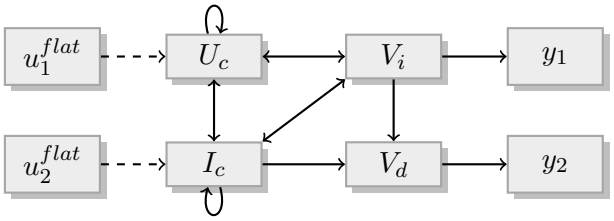


Figure 5.15: Digraph of the influenza A virus production model (Eq. (5.57)). Here, the two potential flat inputs, u_1^{flat} and u_2^{flat}, should act on the states U_c and I_c to fulfil the distance criteria which has been introduced in Sec. 5.4 as a rule of thumb.

measurable according to

$$y_1(t) = V_i(t) \tag{5.58}$$

$$y_2(t) = V_d(t) \tag{5.59}$$

With these measurement quantities and Eq. (5.57) an associated digraph (Fig. 5.15) can be derived. As two quantities are measured, two flat inputs, have to be determined, $u_1^{flat}(t)$ and $u_2^{flat}(t)$. Due to the distance criteria (Sec. 5.4), as rule of thumb, these inputs should act on $U_c(t)$ and $I_c(t)$ (Eq. (5.60)), respectively.

$$\begin{aligned}
\frac{dU_c(t)}{dt} &= \theta_6 \frac{C_{max} - (U_c(t) + I_c(t))}{C_{max}} U_c(t) - \\
&\quad \theta_1 U_c(t) V_i(t) + u_1^{flat}(t) \\
\frac{dI_c(t)}{dt} &= \theta_1 U_c(t) V_i(t) - \theta_2 I_c(t) + u_2^{flat}(t) \\
\frac{dV_i(t)}{dt} &= \theta_3 I_c(t - \tau_1) - \theta_4 V_i(t) - \theta_1 U_c(t) V_i(t) \\
\frac{dV_d(t)}{dt} &= \theta_5 I_c(t - \tau_1) + \theta_4 V_i(t)
\end{aligned} \tag{5.60}$$

As states and inputs (Eq. (5.61)-(5.65)) can be expressed explicitly by the outputs (Eq. (5.58)-(5.59)) and derivatives thereof, the associated DDE system (Eq. (5.60)) is differentially flat.

$$I_c(t-\tau_1) = \frac{1}{\theta_5}\left(\dot{y}_2(t) - \theta_4 y_1(t)\right) \tag{5.61}$$

$$U_c(t) = \frac{\theta_3 I_c(t-\tau) - \theta_4 y_1(t) - \dot{y}_1(t)}{\theta_1 y_1(t)} \tag{5.62}$$

$$\begin{aligned} u_2^{flat}(t) = \frac{1}{\theta_5}\left(\ddot{y}_2(t+\tau_1) - \theta_4 \dot{y}_1(t+\tau_1)\right) - \\ \theta_1 U_c(t) y_1(t) + \theta_2 I_c(t) \end{aligned} \tag{5.63}$$

$$\begin{aligned} u_1^{flat}(t) = \mu(t+\tau_1) U_c(t+\tau_1) + \theta_1 U_c(t+\tau_1) y_1(t+\tau_1) - \\ \frac{\theta_4 \dot{y}_1(t+\tau_1) - \ddot{y}_1(t+\tau_1)}{\theta_1 y_1(t+\tau_1)} - \\ \frac{\theta_3\left(\theta_1 U_c(t) y_1(t) - \theta_2 I_c(t) + u_{f_2}(t)\right)}{\theta_1 y_1(t+\tau_1)} - \\ \frac{U_c(t+\tau_1)\dot{y}_1(t+\tau_1)}{y_1(t+\tau_1)} \end{aligned} \tag{5.64}$$

with

$$\mu(t+\tau_1) = \theta_6 \frac{C_{max} - \left(U_c(t+\tau_1) + I_c(t+\tau_1)\right)}{C_{max}} \tag{5.65}$$

After the successful transformation to a flat system, the proposed method of parameter identification can be applied. The cultivation of MDCK cells is done in a batch mode, i.e., the generated two flat inputs (Eq. (5.60)) are just fictitious. Consequently, the cost function (Eq. (5.27)) has to be evaluated for the identification of model parameters and the delay parameter, respectively. (Technical Note: Here, the upper limit of the integral in Eq. (5.27) has to be replaced by $t_{end} - \hat{\tau}_1$).

Assuming almost perfect measurement data, i.e., high sample rate without measurement noise, the model parameters, θ, are estimated for 100 different delay parameter values equally spaced in the the range of 5 to 15 hours, $\tau_1 \in [5, 15]$ h. As the in-silico measurement data, $y^{data}(t_k)$, are noise-free, the υ value is set equal to 1 in Eq. (5.27). In figure Fig. 5.16, the cost function (Eq. (5.27)) has a global minimum at $\tau_1 = 7.5$ h which is the correct result. The estimated model parameters, $\hat{\theta}$, at this optimally determined delay parameter are given in Tab. 4.1. Although the initial parameter values, $\hat{\theta}_{Ini}$, deviates strongly from the true parameter values, θ_{True}, the proposed optimisation framework is able to provide reliable estimates, $\hat{\theta}_{Opt}$, in a feasible computational time. The overall cpu-time is less than 10 seconds in this case. In comparison to the

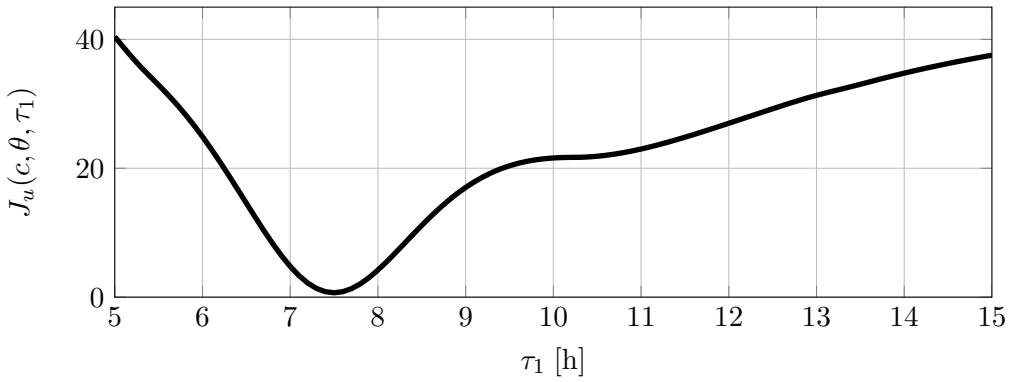


Figure 5.16: The novel cost function, $J_u(c, \theta, \tau_1)$ (Eq. (5.27)), evaluated at different time-delay parameter values τ_1. In detail, an optimiser is initialised iteratively at 100 different τ_1 values, $\tau_1 \in [5, 15]$ h. The overall cpu-time is less than 10 seconds in this case.

	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
$\hat{\theta}_{Ini}/\theta_{True}$	750	750	750	750	750	750
$\hat{\theta}_{Opt}/\theta_{True}$	1.0032	1.0014	0.9974	1.000	0.9974	0.9976

Table 5.3: Result of parameter identification by minimising a cost function based on flat inputs $u_{flat}(t)$ (Eq. (5.27)). Despite the initial parameter deviation, θ_{Ini}, the identified model parameters, $\hat{\theta}_{Opt}$, are close to the true values, θ_{True}.

evaluation of the standard cost function (Eq. (5.4)) it is significant speedup.

In the framework of parameter identification also robustness against initial parameter values, θ^{Ini} and τ_1^{Ini}, is an important issue. Here, the standard approach (Eq. (5.4)) is much more sensitive to initial parameter values than the flat input approach. For instance, only under the ideal condition of $\tau_1 = 7.5$ h the model parameters, θ, are identified properly. A slight deviation from τ_1, however, leads to a divergence of the optimisation routine as demonstrated in Fig. 5.17.

Finally, the change in parameter sensitivities is demonstrated for the differentially flat DDE system (Eq. (5.60)). In the following, a relative parameter perturbation of 25% and fixed operating conditions are assumed. The corresponding parameter sensitivity spectrum of the standard approach (Eq. (2.68)) indicates that model parameter θ_4 is the least sensitive one, see Fig. 5.18(a). Here, as previously suggested, the parameter sensitivities are investigated by the novel approach described above (Eq. (5.28)). As shown in Fig. 5.18(b), the spectrum of parameter sensitivities is changed significantly. The previously insensitive model parameter θ_4 is now the most sensitive one, i.e., parameter θ_4 is likely to be identified precisely. This result agrees well with the previous

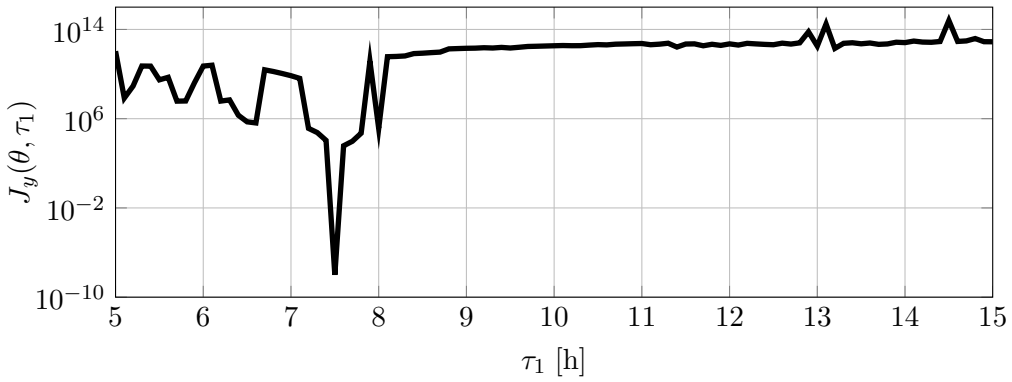


Figure 5.17: The standard cost function, $J_y(\theta, \tau_1)$ (Eq. (5.4)), evaluated at different time-delay parameter values, τ_1. In detail, an optimiser is initialised iteratively at 100 different τ_1 values, $\tau_1 \in [5, 15]$ h. A slight deviation of the true τ_1 value, $\tau_1 = 7.5$ h, leads to a divergence of the optimisation routine.

outcome of the actual parameter identification (Tab. 4.1), i.e., the model parameter θ_4 is reconstructed best. Remember, the existing measurement data set is used just in a different way by evaluating Eq. (5.27). Thus, there is no need of OED and any additional experiment to improve the sensitivity of θ_4.

5.6 Chapter Summary

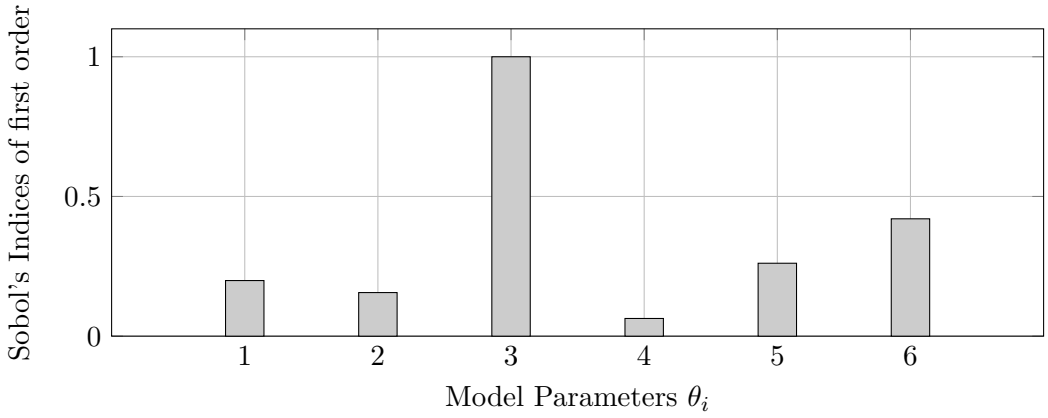
The essential step of parameter identification in model developing is known to be a cpu-intensive process. In an attempt to avoid a computational overload, the flatness property of dynamical systems for the purpose of parameter identification is reviewed. A property of practical interest is, that for flat systems the states and inputs are given as analytical expressions of the outputs and a finite number of derivatives thereof. Thus, there is no need to solve differential equations numerically as part of a cpu-intensive subroutine in the overall parameter identification framework. As a result, this strategy leads to a significant speed-up in the parameter identification process and it circumvents the problem of determining possibly unknown initial conditions of the states. These features are particularly useful for delay differential equations, whose numerical solution and the feasible choice of initial functions are challenging.

Moreover, it has also been demonstrated that the flat input approach might render a non-convex cost function into a more suitable expression, i.e., into a cost function, J_u, with less local minima in comparison to the traditional approach, J_y. This result can hardly be generalised to all classes of systems, in fact, even the opposite may holds. But

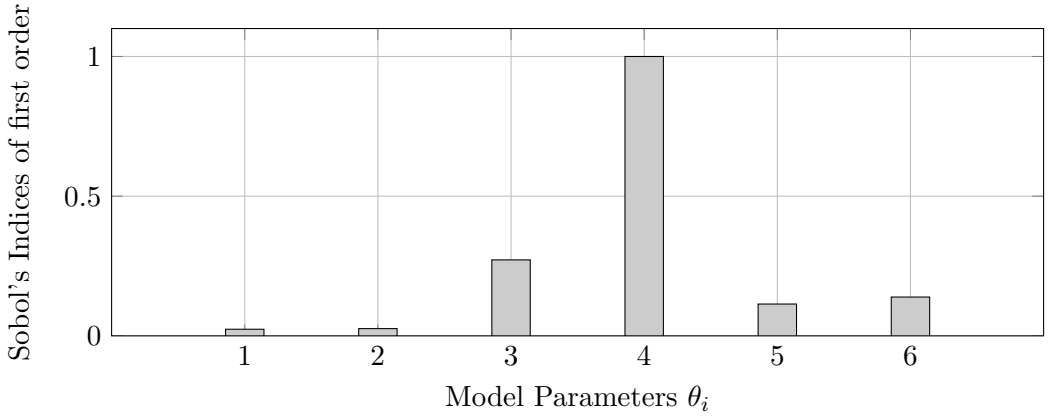
still, one can also conclude that expressions based on flat inputs are likely to depend on the model parameters in a less severe non-linear way than the outputs. For example, model parameters of linear systems appear as arguments of exponential functions in the expressions of the system outputs, but as polynomial or rational functions in the expressions of flat inputs. Thus, in cases of parameter identification problems with many local minima in the traditional approach it may be worthwhile looking at the flat inputs instead. It has to be stressed, too, that there is no need of additional data and new experimental runs, respectively. The same recorded data are evaluated just in a different way. The obstacle which may limits the applicability of the proposed approach is the determination of suitable flat input candidates. Fortunately, there is a method in literature that makes this issue surprisingly simple for the large class of observable systems. Additionally, the construction of flat inputs for given outputs is much easier than the opposite problem of finding flat outputs for the purpose of flatness based control.

The most crucial point of the flat input method, however, is need to form derivatives of measured outputs, $y^{data}(t_k)$. In this thesis, this problem is solved satisfactorily by introducing surrogate output functions based on B-splines and other approximation techniques. The extra effort of determining a larger set of unknowns could be over-compensated by the very efficient solution of the ODE/DDE model equations. Naturally, the accuracy of the proposed method is likely to suffer if higher order output derivatives are required and the measurements are noisy or the measurement sampling rates are low. Thus, systems where the number of states is not orders of magnitude larger than the number of outputs are most attractive for the flat input approach.

Finally, the close link to OED principles is demonstrated by parameter sensitivity analyses. As mentioned previously, OED aims to provide informative data which may render insensitive parameters to sensitive quantities. Traditionally, a novel experimental run and subsequently new measurement data are utilised to provide a desired change in the range of parameter sensitivities. Thus, before running a new experiment to transform insensitive parameters to sensitive expressions the evaluation of the flat input based cost function might be worthwhile.



(a) Standard approach (Eq. 2.68)



(b) Flat input approach (Eq. 5.28)

Figure 5.18: Integral measures of Sobol's indices, $\int_0^T S_i dt$, normalised to the most sensitive parameter are presented for two different strategies of parameter identification. In the upper case, the spectrum of parameter sensitivities is shown for the standard approach, i.e., parameter sensitivities related to the simulated outputs, $y^{sim}(t)$. Below, parameter sensitivities related to flat inputs, $u^{flat}(t)$, are given. Obviously, by using the same measurement data, $y^{data}(t_k)$, there is a significant change in the range of sensitivities.

6

Conclusions and future work

In this thesis, three different aspects of Optimal Experimental Design are addressed in order ...

1.) ... to provide operating conditions which are expected to ensure the most informative data in relation to parameter identification problems. Here, OED aims to provide the most precise parameter estimates and the most credible simulation results, respectively.

2.) ... to provide operating conditions which are expected to ensure a proper model selection in case of several model candidates. Here, OED aims to facilitate the model selection process by providing suitable measurement data.

3.) ... to provide a changed spectrum of parameter sensitivities without the need of additional experimental data. That means, to improve the parameter accuracy by utilising existing measurement data in a more sophisticated way.

In Chapter 2 the general problem of uncertainty quantification and propagation is presented. The concept of the Unscented Transformation is compared with more traditional approaches, e.g., the Fisher Information Matrix (FIM) and Monte Carlo simulation. It is shown, that the Unscented Transformation outperforms the FIM in precision and the Monte Carlos simulation in computational load. Moreover, the basics of the Polynomial Chaos Expansion (PCE) are addressed shortly. Here, too, in comparison to the Unscented Transformation the PCE approach might become too computational demanding for the intended application for OED. Thus, the Unscented Transformation is identified to be a more reliable as well as practicable concept for the purpose of uncertainty propagation. It is also shown how the UT approach can be applied to

non-Gaussian distributions by evaluating suitable transformation equations. The universal concept of UT provides an efficient calculation of global parameter sensitivities additionally. Therefore, the Unscented Transformation is a versatile approach which is applied in the following chapters intensively.

The principles of Optimal Experimental Design for parameter identification are introduced in Chapter 3. Here, it is demonstrated that the overall performance of OED depends on the credibility of the uncertainty propagation critically. Most applied mathematical models are associated to their model parameters in a non-linear fashion. Therefore, concepts which are based on linearisation principles are likely to fail. In the framework of OED this means, a sub-optimal experimental design is put in operation. The UT-based OED, however, has the ability to gather informative data which ensure more precise parameter estimates and simulation results, respectively.

OED for model selection, which is addressed in Chapter 4, aims to provide measurement data which facilitate the original model selection task. That is, to figure out the most plausible model. Here, too, a proper consideration of uncertainties is mandatory. The Unscented Transformation as part of the Unscented Kalman Filter is implemented to run an online model selection concept. The immediate processing of available measurement data contributes essentially to the robustness of the proposed algorithm. In more detail, even in case of a poor initial model parametrisation the most plausible model candidate can be quantified. Moreover, the online OED concept contributes to reduce the impact of "human-error", because optimally designed experiments which are usually non-standard experiments are error-prone in implementation. In all presented test case studies the correct model candidate is selected by the proposed algorithm.

By implementing the previously presented OED strategies it has been assumed inherently that new experiments are feasible in principle. Thus, new experiments are designed to provide the most informative data depending on the intend task. In case of non-repeatable / non-additional experiments or in the case that even optimally designed experiments are insufficient, new concepts have to be derived. A potential strategy is to utilise the flatness property of a certain class of dynamical systems. For this purpose, the basics of flatness and how flat inputs contribute to solve parameter identification problems are revealed in Chapter 5.

In future work following aspects might be worthwhile to be analysed in more detail:

- In the framework of OED for parameter identification scalar values are evaluated which represent the uncertainty of model parameters and simulation results, respectively. Multi-objective criteria, however, might be more suitable compared to scalar values, because the individual contribution of parameters or states becomes visible.

- In case of the online model selection framework it would be worthwhile taking state and parameter constraints into account as well as delay times in measurements. It is expected that the flexibility of the UKF approach enables an straightforward incorporation of these issues of practical relevance.

- For the purpose of flat input based parameter identification there is need for model inversion. Moreover, the concept of inverse simulation [Lu07] might be an interesting alternative/extension to the proposed approach.

A

Appendix

A.1 An n-Dimensional Input Problem Settings

In case of the non-linear mapping of

$$\eta = g(\xi) = \xi^T \xi, \quad \xi \sim \mathcal{N}(0_n, I_{n\times n}) \tag{A.1}$$

the UT3 approach performs in the following way. The sample points and associated weights are determined according to the Equations (2.30)-(2.31) and (2.36)-(2.37), respectively. By assuming $\vartheta = \sqrt{3}$, see Section 2.2.3 for explanation, the resulting η-samples are: $\eta^0 = 0$ and $\eta^i = 3$; $\forall i = 1, \ldots, 2n$. Thus, the expected value of η is determined correctly as

$$\begin{aligned} E[\eta] &= w_0 \cdot \eta^0 + 2n \cdot w_1 \cdot \eta^i && \text{(A.2)} \\ &= w_0 \cdot 0 + 2n \cdot w_1 \cdot 3 && \text{(A.3)} \\ &= (1 - \frac{n}{3}) \cdot 0 + 2n \cdot \frac{1}{6} \cdot 3 && \text{(A.4)} \\ &= n && \text{(A.5)} \end{aligned}$$

Remember, the analytical results are $E[\eta] = n$ and $\sigma_\eta^2 = 2n$ [HMGB03b]. A credible approximation of the associated variance depends critically on the correction factor, β. As stated previously, the corrected weighting factor is defined as $w_0^c = w_0 + \beta$. The

associated variance is approximated according to

$$\begin{aligned}
\sigma_\eta^2 &\approx w_0^c \cdot (\eta^0 - n)^2 + 2n \cdot w_1 \cdot (\eta^i - n)^2 && \text{(A.6)}\\
&\approx w_0^c \cdot (0 - n)^2 + 2n \cdot w_1 \cdot (3 - n)^2 && \text{(A.7)}\\
&\approx (1 + \frac{n}{3} + \beta) \cdot (0 - n)^2 + 2n \cdot \frac{1}{6} \cdot (3 - n)^2 && \text{(A.8)}\\
&\approx 3n - (\beta - 1)n^2 && \text{(A.9)}
\end{aligned}$$

Obviously, the most suitable choice of the correction factor is $\beta = 1$. For the sake of completeness, by a minor modification of the ϑ-value ($\vartheta = \sqrt{2},\ \beta = 1$) the UT3 approach determines the variance correctly as $\sigma_\eta^2 = 2n$.

A.2 Global Sensitivity Analysis Settings

Each term in

$$S_i^y = \frac{\underset{i}{\sigma^2}(\underset{-i}{E}[y^{sim}|\theta[i]])}{\sigma^2(y^{sim})} \tag{A.10}$$

can be determined immediately by applying UT3 and UT5, respectively. In doing so, a quite a number of redundant samples are generated. The associated computational effort is unnecessarily high. For example, by applying the UT3 approach to an n-dimensional input problem the total number of samples points is made up of

$$\begin{aligned}
\sigma^2(y^{sim}) &\rightarrow 2n + 1 && \text{(A.11)}\\
\underset{-i}{E}[y^{sim}|\theta[i]]) &\rightarrow 2(n-1) + 1 && \text{(A.12)}\\
\underset{i}{\sigma^2}(\cdot) &\rightarrow 2 \cdot 1 + 1 && \text{(A.13)}
\end{aligned}$$

The last two terms have to be evaluated for all n individual elements: $n \cdot (2 \cdot 1 + 1 \cdot (2(n-1)+1))$. Thus, the overall number becomes

$$\begin{aligned}
\#SP &= 2n + 1 + n \cdot ((2 \cdot 1 + 1) \cdot (2(n-1) + 1)) && \text{(A.14)}\\
&= 6n^2 - n + 1 && \text{(A.15)}
\end{aligned}$$

By eliminating all redundant samples, however, the number of sample evaluations can be reduced to

$$\#SP = 2n^2 + 1 \tag{A.16}$$

Obviously, the resulting set of samples is equivalent to the set of samples which is generated by applying the UT5 method to approximate $\sigma^2(y^{sim})$. In the meaning of

precision, the Sobol' indices should be determined in the following way: (i) to determine $\sigma^2(y^{sim})$ by UT5, and (ii) to reuse these evaluated samples to get the n expressions of $\underset{i}{\sigma^2}(\underset{-i}{E}[y^{sim}|\theta[i]])$. In this way, the most difficultly to approximate term, $\sigma^2(y^{sim})$, is calculated by the precision associated to UT5.

The numerical values associated to the parameter vectors, a_i, as well as to the matrix M of the O'Hagan & Oakley function read:

$$a_1 = \begin{bmatrix} 0.0118 \\ 0.0456 \\ 0.2297 \\ 0.0393 \\ 0.1177 \\ 0.3865 \\ 0.3897 \\ 0.6061 \\ 0.6159 \\ 0.4005 \\ 1.0741 \\ 1.1474 \\ 0.788 \\ 1.1242 \\ 1.1982 \end{bmatrix} ; \quad a_2 = \begin{bmatrix} 0.4341 \\ 0.0887 \\ 0.0512 \\ 0.3233 \\ 0.1489 \\ 1.036 \\ 0.9892 \\ 0.9672 \\ 0.8977 \\ 0.8083 \\ 1.8426 \\ 2.4712 \\ 2.3946 \\ 2.0045 \\ 2.2621 \end{bmatrix} ; \quad a_3 = \begin{bmatrix} 0.1044 \\ 0.2057 \\ 0.0774 \\ 0.273 \\ 0.1253 \\ 0.7526 \\ 0.857 \\ 1.0331 \\ 0.8388 \\ 0.797 \\ 2.2145 \\ 2.0382 \\ 2.4004 \\ 2.0541 \\ 1.9845 \end{bmatrix}$$

A.3 Kriging Interpolation

The evaluation of cost functions of complex systems can be very cpu-time intensive. To reduce the computational effort one is interested in an easily to evaluate surrogate cost function, which is used to solve the original optimisation problem in a proper manner. One possibility is to apply Kriging interpolation. The roots of the Kriging interpolation go back to geostatistics (spatial statistics) [Mat63]. The original application was to get maps of underground sedimentations based on samples that are taken from the area of interest, e.g., regularly or irregularly spaced borehole sites for the search of noble metals or oil [DJ07]. Kriging has been adapted quite soon in the field of global optimisation [Rat01, LP06, EVBM09]. The key idea is to evaluate only a few points of the design-parameter space of the original cost function. These points provide an input sample set for the Kriging interpolation. In contrast to the standard polynomial regression approaches, these input samples are assumed to be correlated in the following way: the closer the input samples (design-parameter) are, the more positively correlated are their outputs (evaluations of cost function (Eq.3.10)). Finally, Kriging provides estimates, $\hat{y}(x)$, with a minimum error variance at unexplored points, x, of the region of interest. A detailed mathematical background of this method can be found in [Kle07]. In the following, the Matlab Toolbox DACE$^{\text{TM}}$ is used for running Kriging interpolation. A

-0.022483	-0.18502	0.13418	0.36867	0.17173	0.13651	-0.44034	-0.081423	0.71321	-0.44361	0.50383	-0.024101	-0.04594	0.21666	0.055887
0.2566	0.053792	0.258	0.23796	-0.59126	-0.081627	-0.28749	0.41582	0.49752	0.083893	-0.11057	0.033222	-0.13979	-0.031021	-0.22319
-0.056	0.19542	0.095529	-0.28627	-0.14441	0.22369	0.14527	0.28998	0.23105	-0.3193	-0.29039	-0.20957	0.43139	0.024429	0.044904
0.66448	0.4307	0.29925	-0.16202	-0.3148	-0.39027	0.1768	0.057953	0.1723	0.13466	-0.35275	0.25147	-0.018811	0.36482	-0.32505
-0.12128	0.12463	0.10657	0.046562	-0.21679	0.19492	-0.065521	0.024405	-0.096829	0.19366	0.33355	0.31296	-0.083615	-0.25342	0.37326
-0.28376	-0.3282	-0.10496	-0.22073	-0.13708	-0.14426	-0.11503	0.22424	-0.030395	-0.51506	0.017255	0.038957	0.36069	0.30902	0.05003
-0.077876	0.0037457	0.86686	-0.2659	-0.079325	-0.042735	-0.18654	-0.35605	-0.17497	0.0887	0.40026	-0.05598	0.13724	0.21486	-0.011266
-0.092295	0.5921	0.031338	-0.033081	-0.24309	-0.099799	0.03446	0.09512	-0.33802	0.006386	-0.61207	0.081325	0.88683	0.14255	0.14776
-0.13189	0.52878	0.12652	0.045114	0.58374	0.37292	0.11395	-0.29479	-0.57014	0.46292	-0.09405	0.13959	-0.38607	-0.44897	-0.14602
0.058108	-0.32289	0.093139	0.072427	-0.56919	0.52554	0.23657	-0.011782	0.071821	0.078277	-0.13356	0.22723	0.14369	-0.45199	-0.55575
0.66146	0.34633	0.14098	0.51883	-0.2802	-0.16032	-0.068413	-0.20428	0.069672	0.23113	-0.044369	-0.16455	0.21621	0.0042702	-0.087399
0.316	-0.027552	0.13434	0.13497	0.054006	-0.17375	0.17525	0.060359	-0.17914	-0.31057	-0.25359	0.025848	-0.43006	-0.62266	-0.033997
-0.29038	0.034101	0.034903	-0.12122	0.026031	-0.33546	-0.41424	0.053248	-0.27099	-0.026251	0.41024	0.26636	0.15583	-0.186666	0.019896
-0.24389	-0.44099	0.012619	0.24945	0.071102	0.24624	0.17485	0.0085287	0.25147	-0.1466	-0.084625	0.36931	-0.29955	0.11044	-0.7569
0.041494	-0.25981	0.46402	-0.36112	-0.94981	-0.16504	0.0030943	0.052793	0.22524	0.3839	0.45562	-0.18632	0.0082334	0.16671	0.16046

Table A.1: The matrix M of the O'Hagan & Oakley function.

lot of work has been done in this field pointing out that Kriging is superior over the most commonly used interpolation approaches [GSB+06]. For illustration, interpolation results of a benchmark function, the original six-hump camel back function (Eq.A.17) [DS75], are visualised, see Fig. A.1. Over the region of interest an input set of 50 randomly spaced samples is used for interpolation of

$$f(x_1, x_2) = (4 - 2.1x_1^2 + x_1^{4/3})x_1^2 + x_1x_2 + (-4 + 4x_2^2)x_2^2 \tag{A.17}$$

The contour-plot based on Kriging agrees quite well with the original function while the Matlab built-in triangle based cubic interpolation algorithm provides just a very crude approximation.

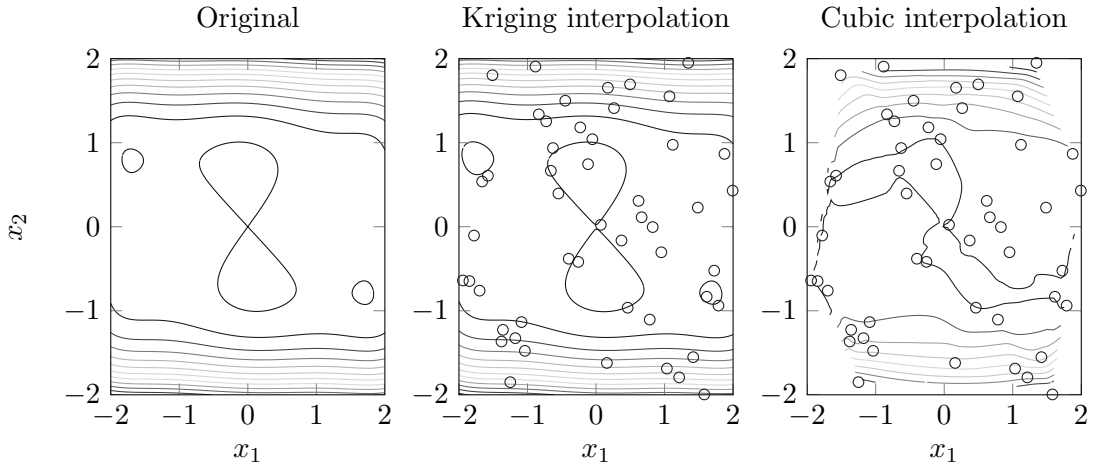


Figure A.1: Benchmark-contour plot of the Kriging method related to the Matlab built-in triangle-based cubic interpolation. Small black circles represent the initial input set of 50 samples of the original six-hump camel back function.

A.4 ODEs of the Two-Substrate Model

The following ODE system is implemented in Matlab™:

```
function yp = sysdyn(t,y,p)
% simulates Glc/Glc_6P
parameter_global
% --------  modified parameters --------
K1        =   p(1);
K2        =   p(2);
Kg6p      =   p(3);
```

```
Keiiap   =   p(4);
% --------  states ---------------------
XX = y(1,:);  % biomass
S1 = y(2,:);  % Glc_6P
S2 = y(3,:);  % Glc
E1 = y(4,:);  % transporter Glc_6P
E2 = y(5,:);  % transporter Glc
M1 = y(6,:);  % G6p
M2 = y(7,:);  % Pep
M3 = y(8,:);  % Pyruvate
XP = y(9,:);  % EIIAP (Pts Protein)
% ---------  rates   ------------------
%  uptake carbo 1/2
rup1 = kg6p*(E1.*S1)./(Kg6p+S1);
rup2 = kptsup*XP.*(E2.*S2)./...
       (Kglc*Keiiap*x0+S2*Keiiap*x0+XP*Kglc+XP.*S2);

%  enzyme syn 1/2
rsyn1= k1 * (kb+ksyn*(XP).^6./(XP.^6+K^6))...
       .*rup1./(K1+rup1);
rsyn2= k2 * (KI./(E1+KI)).*(kb+ksyn*(XP).^6./(XP.^6+K^6))...
       .*rup2./(K2+rup2);

% rates
rgly = kgly * M1;                              % Glycolyse
f =  (M1.^n).*M2.^m;
rpyk = kpyk*M2.*f;                             % Pyk
rpts = kpts * M2.*(x0-XP) - km_pts*M3.*XP;     % PTS rate
rabb =  kpdh * M3;                             % Pdh
% ------------------------------
mu = Y1_sim * rup1 + Y2_sim * rup2;            % growth rate

yp =[ (mu).* XX                                % biomass
-mw1 * rup1.*XX                                % carbo 1
-mw2 * rup2.*XX                                % carbo 2
```

```
rsyn1 - (kd + mu).*E1;                    % trans 1
rsyn2 - (kd + mu).*E2;                    % trans 2
rup1 + rup2 - rgly                        % G6p
2*rgly - rpyk - rpts                      % Pep
rpyk + rpts  - rabb                       % Pyruvate
rpts - rup2];                             % EIIAP
```

A.5 Box-Bias Approach

As a more traditional approach of uncertainty analysis for non-linear regression models the so called Box Bias has been introduced [Box71] to overcome the assumption of an ideal unbiased estimator. Traditionally, the Box Bias is calculated according to

$$Bi_{Box} = -\frac{1}{2} FIM^{-1} \sum_{t_k=1}^{K} S_{t_k}^T C_y^{-1} \begin{bmatrix} tr\left[FIM^{-1} H_{1,t_k}\right] \\ \vdots \\ tr\left[FIM^{-1} H_{m,t_k}\right] \end{bmatrix}, \tag{A.18}$$

where

$$FIM = \sum_{t_k=1}^{K} S_{t_k}^T C_y^{-1} S_{t_k}, \tag{A.19}$$

$$S_{t_k} = \begin{bmatrix} \frac{\partial y_1}{\partial \theta_1}\Big|_{t_k} & \frac{\partial y_1}{\partial \theta_2}\Big|_{t_k} & \cdots & \frac{\partial y_1}{\partial \theta_l}\Big|_{t_k} \\ \frac{\partial y_2}{\partial \theta_1}\Big|_{t_k} & \frac{\partial y_2}{\partial \theta_2}\Big|_{t_k} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial \theta_1}\Big|_{t_k} & \cdots & \frac{\partial y_m}{\partial \theta_{l-1}}\Big|_{t_k} & \frac{\partial y_m}{\partial \theta_l}\Big|_{t_k} \end{bmatrix} \tag{A.20}$$

and

$$H_{i,t_k} = \begin{bmatrix} \frac{\partial^2 y_i}{\partial \theta_1^2}\Big|_{t_k} & \frac{\partial^2 y_i}{\partial \theta_1 \partial \theta_2}\Big|_{t_k} & \cdots & \frac{\partial^2 y_i}{\partial \theta_1 \partial \theta_l}\Big|_{t_k} \\ \frac{\partial^2 y_i}{\partial \theta_2 \partial \theta_1}\Big|_{t_k} & \frac{\partial^2 y_i}{\partial \theta_2^2}\Big|_{t_k} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 y_i}{\partial \theta_l \partial \theta_1}\Big|_{t_k} & \cdots & \frac{\partial^2 y_i}{\partial \theta_p \partial \theta_{l-1}}\Big|_{t_k} & \frac{\partial^2 y_i}{\partial \theta_l^2}\Big|_{t_k} \end{bmatrix}. \tag{A.21}$$

Obviously, there is a need for first and second-order derivatives of the model output with respect to the parameters (Eq.A.20,Eq.A.21). Especially, the determination of second derivatives is numerically quite challenging. Furthermore, the inverse of the Fisher Information matrix (Eq.A.19) is part of Eq.A.18, so the Box Bias is likely to be inaccurate for ill-conditioned matrices. Again the UT method seems to be in favour

as neither inversion nor derivatives are part of this approach. Commonly, a high bias indicates a severe nonlinearity. In [Rat83] a model is quantified to be non-linear with respect to a parameter θ_j if the percentage bias $\%Bi_j$ (Eq.A.22) exceeds 1%.

$$\%Bi_j = \frac{|Bi_j|}{\theta_j} \cdot 100, \tag{A.22}$$

where Bi_j is the jth element of the bias vector.

A.6 MAP Kinase Model Settings

a	b	c	$V(t=0)$	$R(t=0)$	Δt	σ_{y^2}
0.2	0.2	3	-1	1	0.05	0.04

Table A.2: Configuration parameters of the test case study at which the in-silico data, $y^{data}(t_k)$, are generated.

θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}	θ_{11}	θ_{12}
3.26	0.87	2.38	0.14	9.32e-1	23.63	0.03	0.20	0.02	0.37	0.02	1.07

θ_{13}	θ_{14}	a	w	t_s	t_{end}	Δt	$KKK^*(t_0)$	$KK^*(t_0)$	$K^*(t_0)$
8.57e-6	1.40e3	1	0.1e-2	1.57	600	5.00	0.00	0.50	0.50

Table A.3: Model parameters of model candidates $\hat{\mathbb{S}}_1$, $\hat{\mathbb{S}}_2$, and $\hat{\mathbb{S}}_3$

θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}	θ_{11}	θ_{12}
3.26	0.87	2.38	0.14	9.32e-1	23.63	0.03	0.20	0.02	0.37	0.02	1.07

θ_{13}	θ_{14}	a	w	t_s	t_{end}	Δt	$KKK^*(t_0)$	$KK^*(t_0)$	$K^*(t_0)$
8.57e-6	1.40e3	1	0.1e-2	1.57	600	5.00	0.00	0.50	0.50

Table A.4: Model parameters of model candidates $\hat{\mathbb{S}}_1$, $\hat{\mathbb{S}}_2$, and $\hat{\mathbb{S}}_3$

A.7 Model Selection Equations and Settings

$\hat{\mathcal{S}}_1$	$K\dot{K}K^*$	$=$	$\frac{k_1 \cdot U(1-KKK^*)}{k_{1m}+(1-KKK^*)} - \frac{k_2 \cdot KKK^*}{k_{2m}+KKK^*} - k_5 \cdot IP^* \cdot KKK^*$
	$K\dot{K}^*$	$=$	$\frac{k_3 \cdot KKK^*(1-KK^*)}{k_{3m}+(1-KK^*)} - \frac{k_4 \cdot KK^*}{k_{4m}+KK^*} - k_{10} \cdot K^* \cdot KK^*$
	$\dot{K}^*$	$=$	$\frac{k_6 \cdot KK^*(1-K^*)}{k_{6m}+(1-K^*)} - \frac{k_7 \cdot K^*}{k_{7m}+K^*}$
	$I\dot{P}^*$	$=$	$\frac{k_8 \cdot K^*(1-IP^*)}{K_{8m}+(1-IP^*)} - \frac{k_9 \cdot IP^*}{k_{9m}+IP^*}$
$\hat{\mathcal{S}}_2$	$K\dot{K}K^*$	$=$	$\frac{k_1 \cdot U(1-KKK^*)}{k_{1m}+(1-KKK^*)} - \frac{k_2 \cdot KKK^*}{k_{2m}+KKK^*} - \frac{k_{11} \cdot K^* \cdot KKK^*}{k_{11m}+KKK^*}$
	$K\dot{K}^*$	$=$	$\frac{k_3 \cdot KK^*(1-KK^*)}{k_{3m}+(1-KK^*)} - \frac{k_4 \cdot KK^*}{k_{4m}+KK^*} - \frac{k_{10} \cdot K^* \cdot KK^*}{k_{10m}+KK^*}$
	$\dot{K}^*$	$=$	$\frac{k_6 \cdot KK^*(1-K^*)}{k_{6m}+(1-K^*)} - \frac{k_7 \cdot K^*}{k_{7m}+K^*}$
$\hat{\mathcal{S}}_3$	$K\dot{K}K^*$	$=$	$\frac{k_1 \cdot U(1-KKK^*)}{k_{1m}+(1-KKK^*)} - \frac{k_2 \cdot KKK^*}{k_{2m}+KKK^*} - \frac{k_{11} \cdot K^* \cdot KKK^*}{k_{11m}+KKK^*}$
	$K\dot{K}^*$	$=$	$\frac{k_3 \cdot KKK^*(1-KK^*)}{k_{3m}+(1-KK^*)} - \frac{k_4 \cdot KK^*}{k_{4m}+KK^*}$
	$\dot{K}^*$	$=$	$\frac{k_6 \cdot KK^*(1-K^*)}{k_{6m}(1-K^*)} - \frac{k_7 \cdot K^*}{k_{7m}+K^*}$

Table A.5: Related ODE system of model candidates $\hat{\mathcal{S}}_1$, $\hat{\mathcal{S}}_2$, and $\hat{\mathcal{S}}_3$

A.8 Concepts Used to Derive Surrogate Functions

By utilising the Matlab Spline Toolbox™ following setting is applied:

- B-Splines up to order 8 are tested
- 350 equally spaced knots are used

By utilising the Matlab Wavelet Toolbox™ following setting is applied:

- a biorthogonal wavelet family 'bior3.3' is selected
- a decomposition level of 5 is used

By utilising the Matlab Neural Network Toolbox™ following setting is applied:

- a generalised regression neural network is selected
- the spread of radial basis functions is set equal to 0.4

	k_1	k_{1m}	k_2	k_{2m}	k_3	k_{3m}	k_4	k_{4m}	k_5	k_6
$\hat{\mathbb{S}}_1$	0.30	0.46	0.1e-2	0.18	0.25e-1	0.66	0.71e-2	1.96	3.86	0.12e-1
$\hat{\mathbb{S}}_2$	1.13	0.15e-1	0.17e-2	28.95	0.95e-1	3.62	0.17e-2	0.68	-	0.29e-2
$\hat{\mathbb{S}}_3$	3.26	0.87	0.93-7	23.63	0.31e-1	0.2	0.15e-1	0.37	-	0.21e-1

k_{6m}	k_7	k_{7m}	k_8	k_{8m}	k_9	k_{9m}	k_{10}	k_{10m}	k_{11}	k_{11m}
0.56	0.64e-2	2.10	0.23e-2	0.68	0.25e-2	1.11	0.1e-1	-	-	-
0.12e-1	0.53e-2	21.82	-	-	-	-	0.34e-1	2.1	1.98	0.24
1.07	0.86e-5	1.4e3	-	-	-	-	-	-	2.38	0.14

Table A.6: Model parameters of model candidates $\hat{\mathbb{S}}_1$, $\hat{\mathbb{S}}_2$, and $\hat{\mathbb{S}}_3$

	$KKK^*(t_0)$	$KK^*(t_0)$	$K^*(t_0)$	$IP^*(t_0)$	u	$P_0^+ = I. \cdot P_{0,i}^+$	$Q = I. \cdot Q_i$
Fig. 4.2	0	0.5	0.5	0	0.6	-	-
Fig. 4.4	0.25	0.5	0.5	0.1	0.2	0.03	1e-6
Fig. 4.7	0.25	0.5	0.5	0.1	-	0.03	1e-6
Fig. 4.8	0.25	0.5	0.5	-	1	-	-
Fig. 4.9	0.25	0.5	0.5	0.1	-	0.03	1e-6
Fig. 4.11	0	0.5	0.5	0	0.6	-	-
Fig. 4.5	0.25	0.5	0.5	0.1	-	0.03	1e-6

	$Q_\theta = I^{2\times 2}. \cdot Q_{\theta,i}$	R	$k_1^c(\hat{\mathbb{S}}_2)$	$k_1^c(\hat{\mathbb{S}}_3)$	$k_{11}^c(\hat{\mathbb{S}}_2)$	$k_{11}^c(\hat{\mathbb{S}}_3)$	α	β	κ
Fig. 4.2	-	-	-	-	-	-	0.1	2	0
Fig. 4.4	-	0.025^2	-	-	-	-	0.1	2	0
Fig. 4.7	-	0.025^2	-	-	-	-	0.1	2	0
Fig. 4.8	-	-	-	-	-	-	0.1	2	0
Fig. 4.9	1e-3	0.025^2	$1.5 \cdot k_1$	$1.5 \cdot k_1$	$0.5 \cdot k_{11}$	$1.5 \cdot k_{11}$	0.1	2	0
Fig. 4.11	-	-	-	-	-	-	0.1	2	0
Fig. 4.5	-	0.025^2	-	-	-	-	0.1	2	0

Table A.7: Operating conditions for different scenarios. Here, I is the $n \times n$ identity matrix.

A.9 Analytical Solution of the Optimal Flat Input

According to the Equations ((5.49)-(5.51)) the sensitivity function becomes

$$\frac{\partial u^{flat}(t)}{\partial b} = -\frac{1}{c}\left(\frac{3\zeta_0 - \zeta_0^3}{3}\right) \tag{A.23}$$

Thus, the actual optimisation problem reads as follows

$$\begin{aligned} \arg\ \max_{\zeta_0} J^{opt}(\zeta_0) &= \int_{t_0}^{t_{end}} \left(\frac{\partial u^{flat}(t)}{\partial b}\right)^2 dt \\ &= t_{end}\left(-\zeta_0\frac{1}{c} + \frac{1}{c}\zeta_0^3\right)^2 - t_0\left(-\zeta_0\frac{1}{c} + \frac{1}{c}\zeta_0^3\right)^2 \end{aligned} \tag{A.24}$$

By assuming $t_0 = 0$, $t_{end} = 20$, and $c = 3$, the following derivatives of the cost function

with respect to the design parameter, ζ_0, exist

$$\begin{aligned}\frac{dJ^{opt}(\zeta_0)}{d\zeta_0} &= -40\left(\zeta_0/3-\zeta_0^3/9\right)\left(\zeta_0^2/3-1/3\right)\\ \frac{d^2J^{opt}(\zeta_0)}{d\zeta_0^2} &= 40\left(\zeta_0^2/3-1/3\right)-\left(80\zeta_0\left(\zeta_0/3-\zeta_0^3/3\right)\right)/3\end{aligned} \tag{A.25}$$

The resulting singularities, $\frac{dJ^{opt}(\zeta_0)}{d\zeta_0}=0$, are

$$\zeta_0^* = \left[-1, 1, 0, \sqrt{3}, -\sqrt{3}\right] \tag{A.26}$$

The associated second-order derivatives are determined as

$$\left.\frac{d^2J^{opt}(\zeta_0)}{d\zeta_0^2}\right|_{\zeta_0^*} = [-160/27, -160/27, 40/9, 160/9, 160/9] \tag{A.27}$$

Obviously, the optimisation problem has 2 local maxima and 3 local minima [NW99], see Fig. A.2 for illustration. By a closer inspection of Eq.(A.24), however, it can be concluded that a design parameter, ζ_0, tending to plus/minus infinity ensures the highest sensitivity of the cost function with respect to parameter b. By introducing some plausible physical constraints of the system input, $u(t)$, the following constrained optimisation problem has to be solved

$$\begin{aligned}&\arg\ \max_{\zeta_0} J^{opt}(\zeta_0)\\ &\text{subject to } u(t)\leq 1\\ &\qquad\qquad u(t)\geq 0\end{aligned} \tag{A.28}$$

By utilising the Karush-Kuhn-Tucker conditions [NW99] the optimisation problem is recast according to

$$\mathcal{L}(\zeta_0,\mu) = J^{opt}(\zeta_0) + \mu_1\left(u(t)-1\right) + \mu_2\left(-u(t)\right) \tag{A.29}$$

Here, in addition to the local maxima, $\zeta_0^*=1$, an optimum at $\zeta_0^*=2.4268$ becomes the most feasible solution ($J^{opt}(\zeta_0=1) < J^{opt}(\zeta_0=2.4268)$), i.e., the constraints are fulfilled ($\mu_1=29.34,\ \mu_2=0$) and the second derivation is negative, $\left.\frac{d^2\mathcal{L}(\zeta_0,\mu)}{d\zeta_0^2}\right|_{\zeta_0^*} = -81.7$.

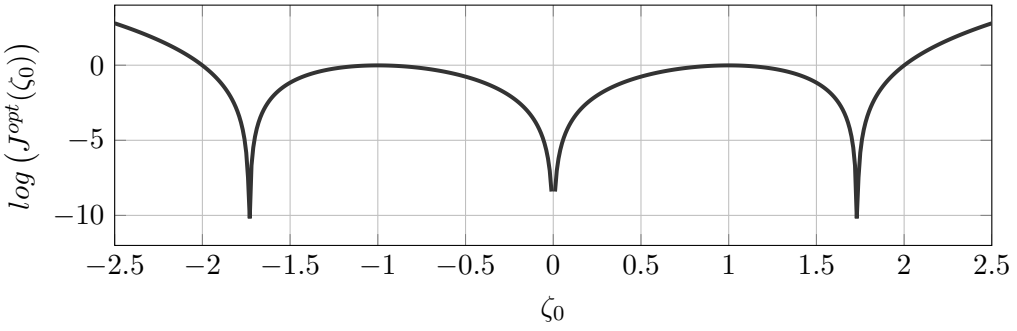


Figure A.2: OED for flat inputs: Here, the goal is to figure out an optimum output function by evaluating the unconstrained version of the optimisation problem given in Eq.(A.24).

A.10 Equations for the Flat Input of MAP

```
%-----------------------------------------------------------
% Measurements and time derivatives thereof
y  = Ks;     % measuring the acitve K - Enzyme, K*
yd;          % first-order derivative of y
ydd;         % second-order dervivative
yddd;        % third-order derivative
%-----------------------------------------------------------
% Reconstruction of KK*
KKs =    (yd +(v7.*y)./(k7m+y)).*(k6m+1-y)./(k6.*(1-y));
%-----------------------------------------------------------
% Reconstruction of KKK*
M   =    (-yd.*(k6-k6.*y)-(k6m+1-y).*(-k6.*yd))./...
         (k6.*(1-y)).^2;
L   =    ydd+(v7.*yd.*(k7m))./(k7m+y).^2;
KKsd=    L.*(k6m+1-y)./(k6.*(1-y))+(yd+(v7.*y)./...
         (k7m+y)).*M;
KKKs=    (KKsd+(v4.*KKs)./(k4m+KKs)).*((k3m+1-KKs)./...
         (k3.*(1-KKs)));
%-----------------------------------------------------------
% Reconstruction of the flat input U
H4  =    2.*(k6.*(1-Ks)).*(-k6.*Ksd);
H3  =    -Ksdd.*(k6-k6.*Ks)+(-Ksd.*(-k6.*Ksd))-...
         (-Ksd.*(-k6.*Ksd))-(k3m+1-Ks).*(-k6.*Ksdd);
```

```
Md   =   (H3.*(k6.*(1-Ks)).^2-(-Ksd.*(k6-k6.*Ks)-...
         (k3m+1-Ks).*(-k6.*Ksd)).*H4)./(k6.*(1-Ks)).^4;
Ld   =   Ksddd+(v7.*k7m.*Ksdd.*(k7m+Ks).^2-v7.*k7m.*...
         Ksd.*2.*Ksd.*(k7m+Ks))./(k7m+Ks).^4;
KKsdd=   Ld.*(k6m+1-Ks)./(k6.*(1-Ks)) + L .*...
         (-Ksd.*(k6.*(1-Ks))-(k6m+1-Ks).*(-k6.*Ksd))./...
         (k6.*(1-Ks)).^2+(Ksdd + (v7.*Ksd.*(k7m + Ks)...
         -(v7.*Ks).*Ksd)./(k7m+Ks).^2).*M+...
         (Ksd+(v7.*Ks)./(k7m+Ks)).*Md;
H2   =   (-KKsd.*(k3.*(1-KKs))-((k3m+1-KKs).*...
         (-k3.*KKsd))).*(k3.*(1-KKs)).^2;
H1   =   KKsdd+((v4.*KKsd.*(k4m+KKs)-v4.*KKsd.*KKs))./...
         (k4m+KKs).^2;
KKKsd=   H1.*((k3m+1-KKs)./(k3.*(1-KKs)))+...
         (KKsd + v4.*KKs./(k4m+KKs)).*H2;
U    =   (KKKsd+v2.*KKKs./(k2m+KKKs)+...
         k5.*Ks.*KKKs./(k5m + KKKs)).*...
         (k1m+1-KKKs)./(k1.*(1-KKKs));
```

References

[AKM05] A. Andreev, A. Kanto, and P. Malo. Simple approach for distribution selection in the pearson system. In *Helsinki School of Economics Working Papers*, 2005. 34

[And11] T. V. Anderson. *Efficient, Accurate, and Non-Gaussian Statistical Error Propagation Through Nonlinear, Closed-Form, Analytical System Models.* PhD thesis, Brigham Young University, 2011. 18

[ATE+08] J. F. Apgar, J. E. Toettcher, D. Endy, F. M. White, and B. Tidor. Stimulus design for model selection and validation in cell signaling. *PLoS Computational Biology*, 4:1940–1952, 2008. 70

[AWWT10] J. F. Apgar, D. K. Witmer, F. M. White, and B. Tidor. Sloppy models, parameter uncertainty, and the role of experimental design. *Molecular BioSystems*, 6:1890–1900, 2010. 4

[BA02] K. P. Burnham and D. R. Anderson. *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach.* Springer, 2002. 7, 70, 72, 74, 76

[BAGW10] T. Barz, H. Arellano-Garcia, and G. Wozny. Handling uncertainty in model-based optimal experimental design. *Ind. Eng. Chem. Res.*, 49:5702–5713, 2010. 74

[BCRFB07] E. Balsa-Canto, M. Rodriguez-Fernandez, and J. R. Banga. Optimal design of dynamic experiments for improved estimation of kinetic parameters of thermal degradation. *Journal of Food Engineering*, 82:178–188, 2007. 49

[BD87] G. E. P. Box and N. R. Draper. *Empirical Model-Building And Response Surfaces.* WILEY, 1987. 32, 49

[BD06] C. Baudritt and D. Dubois. Practical representations of incomplete probabilistic knowledge. *Computational Statistics & Data Analysis*, 51:86–108, 2006. 13

[Bel66] R. Bellman. *Adaptive Control Processes - A Guided Tour.* Princeton University Press, 1966. 19

[BH67] G. E. P. Box and W. J. Hill. Discrimination among mechanistic models. *Technometrics*, 9:57–71, 1967. 74, 76, 77

[BHC+04] K. S. Brown, C. C. Hill, G. A. Calero, K. H. Lee, J. P. Sethna, and R. A. Cerione. The statistical mechanics of complex signaling networks: nerve growth factor signaling. *Quantitative Biology*, 1:184 195, 2004. 53

[BHDE07] M. Behar, N. Hao, H. G. Dohlman, and T. C. Elston. Mathematical and computational analysis of adaptation via feedback inhibition in signal transduction pathways. *Biophysical Journal*, 93:806–821, August 2007. 81

[Bou07] F. Boulier. Differential elimination and biological modelling. *Radon Series Comp. Appl. Math.*, 2:111–139, 2007. 102, 103

[Box71] M. J. Box. Bias in nonlinear estimation. *J. R. Stat. Soc. Series B*, 33:171–201, 1971. 145

[Box76] G. E. P. Box. Science and statistics. *Journal of the American Statistical Association*, 71:791–799, 1976. 2

[BPZ+11] P. Baraldi, N. Pedroni, E. Zio, E. Ferrario, A. Pasanisi, and M. Couplet. *Advances in Safety, Reliability and Risk Management: Esrel 2011.* Crc Pr Inc, 2011. 13

[Bre70] A. M. Breipohl. *Probabilistic System Analysis.* JOHN WILEY & SONS, INC., 1970. 15, 16, 17

[BSSR94] M. Baltes, R. Scheider, C. Sturm, and M. Reuss. Optimal experimental design for parameter estimation in unstructured growth models. *Biotechnology Progress*, 10, 1994. 49, 58

[BW04] B. Bogacka and F. Wright. Comparison of two design optimality criteria applied to a nonlinear model. *Journal of Biopharmaceutical Statistics*, 14:4:909–930, 2004. 51

[CAR+98] S. Carter, K. L. Auer, D. B. Reardon, M. Birrer, P. B. Fisher, K. Valierie, R. Schmidt-Ullrich, R. Mikkelsen, and P. Dent. Inhibition of the mitogen activated protein (map) kinase cascade potentiates cell killing by low doses ionizing radiation in a431 human squamous carcinoma cells. *Oncogene*, 16:2787–2796, 1998. 71

[CCFCO12] S.-W. Chung, C. R. Cooper, M. C. Farach-Carson, and B. A. Ogunnaike. A control engineering approach to understanding the tgf-β parardox in cancer. *J. R. Soc. Interface*, 9:1389–1397, 2012. 2

[CF05] G. Cumming and S. Finch. Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist*, 60:170–180, 2005. 12

[CL12] A. Cheng and T. K .Lu. Synthetic biology: An emerging engineering discipline. *Annu. Rev. Biomed. Eng.*, 14:155–178, 2012. 3

[Coc73] W. G. Cochran. Experiments for nonlinear functions. *J. Am. Stat. Assoc.*, 68:771–781, 1973. 65

[CS13] R. Chen and M. Snyder. Promise of personalized omics to precision medicine. *WIREs Syst Biol Med*, 5:73–82, 2013. 3

[DBI10] E. Van Derlinden, K. Bernaerts, and J. F. Van Impe. Simultaneous versus sequential optimal experiment design for the identification of multi-parameter microbial growth kinetics as a function of temperature. *Journal of Theoretical Biology*, 264:347–55, 2010. 49

REFERENCES

[DBMM+11] A. Donoso-Bravo, J. Mailier, C. Martin, J. Rodriguez, C. A. Aceves-Lara, and A. V. Wouwer. Model selection, identification and validation in anaerobic digestion: A review. *WATER RESEARCH*, 45:5347–5364, 2011. 7, 70

[DBR10a] S. Dobre, T. Bastogne, and A. Richard. Global sensitivity and identifiability implications in systems biology. In *Proceedings of teh 11th International Symposium on Computer Applications in Biotechnology (CAB 2010)*, 2010. 46

[DBR10b] M. M. Donahue, G. T. Buzzard, and A. E. Rundell. Experiment design through dynamical characterisation of non-linear systems biology models utilising sparse grids. *IET Systems Biology*, 4:249–262, 2010. 20

[Dij59] E. W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, pages 269–271, 1959. 114

[DJ95] D. L. Donoho and I. M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90:1200–1225, 1995. 108

[DJ07] P. J. Diggle and P. J. Riberio Jr. *Model-based Geostatistics*. Springer, 2007. 141

[DPD10] Y. L. Deribe, T. Pawson, and I. Dikic. Post-translational modifications in signal integration. *Nature Structural & Molecular Biology*, 17:666–672, 2010. 87

[DR07] P. J. Davis and P. Rabinowitz. *Methods of Numerical Integration*. Dover Publications, 2007. 19

[DS75] L. C. W. Dixon and G. P. Szego. *Towards Global Optimization*. North Holland Publishing Company, 1975. 143

[Efr82] B. Efron. *The Jacknife, the Bootstrap and Other Resampling Plans*. siam - Society for Industrial and Applied Mathematics, 1982. 14

[ET94] B. Efron and R. J. Tibshirani. *An Introduction to the Bootrap*. Chapman & Hall, 1994. 14

[Eva67] D. H. Evans. An application of numerical integration techniques to statistical tolerancing. *Technometrics*, 9:441–456, 1967. 25, 26

[Eva74] D. H. Evans. Statistical tolerancing: The state of the art. *Journal of Quality Technology*, 6:188–195, 1974. 25

[EVBM09] J. A. Egea, E. Vazquez, J. R. Banga, and R. Marti. Improved scatter search for the global optimization of computationally expensive dynamic models. *Journal of Global Optimization*, 43:175–190, 2009. 141

[FdVN+10] D. Falck, J. S. B. de Vlieger, W. M. A. Nissen, J. Kool, M. Honing, M. Giere, and H. Irth. Development of an online p38α mitogen-activated protein kinase binding assay and integration of lc-hr-ms. *Anal Bioanal Chem*, 398:1771–1780, 2010. 71

[Fey13] R. Feynman. www.feynman.com, 2013. 3

[Fis35] R. A. Fisher. *The Design of Experiments*. Oliver & Boyd, 1935. 49

[Fis71] R. A. Fisher. *The Design of Experiments*. Macmillan Pub Co, 1971. 72

[Fit61] R. FitzHugh. Impulses and physiological states in theoretical models of nerve membrane. *Biophysical Journal*, 1:445–466, 1961. 116

[FLMR92] M. Fliess, J. Lévine, P. Martin, and P. Rouchon. On differentially flat non-linear systems. *C.R. de l'Académie des Sciences*, 315:619–624, 1992. 105

[FLMR95] M. Fliess, J. Lévine, P. Martin, and P. Rouchon. Flatness and defect of nonlinear systems: Introductory theory and examples. *Int. J. Control*, 61:1327–1361, 1995. 105

[FM08] G. Franceschini and S. Machietto. Model-based design of experiments for parameter precision: State of the art. *Chemical Engineering Science*, 63:4846–4872, 2008. 4, 49

[FPPM09] B. K. Fiil, K. Petersen, M. Peterson, and J. Mundy. Gene regulation by map kinase cascades. *Current Opinion in Plant Biology*, 12:615–621, 2009. 87

[FTM12] S. Franceschini, C. Tsai, and M. Marani. Point estimate methods based on taylor series expansion - the perturbance moments method - a more coherent derivation of the second order statistical moment. *Applied Mathematical Modelling*, 36:5445–5454, 2012. 25

[GBB09] F. Galvanin, M. Barolo, and F. Bezzo. Online model-based redesign of experiments for parameter estimation in dynamic systems. *Ind*, 48:4415–4427, 2009. 74

[Gel74] A. Gelb. *Applied Optimal Estimation*. The M.I.T. Press, 1974. 77, 78

[GH12] F. Gustafsson and G. Hendeby. Some relations between extended and unscented kalman filters. *IEEE Transactions on Signal Processing*, 60:545–555, 2012. 43, 48

[GSB+06] A. A. Giunta, L. P. Swiler, S. L. Browns, M. S Eldred, M. D. Richards, and C. Cyr. The surfpack software library for surrogate modeling of sparse irregularly spaced multidemnsional data. In *AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, 2006. 143

[GvO09] J. Gore and A. van Oudenaarden. The yin and yang of nature. *NATURE*, 457:271–272, 2009. 3

[GWC+07] R. N. Gutenkunst, J. J. Waterfall, F. P. Casey, K. S. Brown, C. R. Myers, and J. P. Sethna. Universally sloppy parameter sensititivities in systems biology models. *PLoS Computational Biology*, 3:1871–1878, 2007. 46, 53

[GWRG06] A. Gensior, O. Woywode, J. Rudolph, and H. Güldner. On differential flatness, trajectory planning, observers, and stabilization for dc-dc converters. *IEEE Transaction on Circuits and Systems*, 53:2000–10, 2006. 105

[HF96] C. Y. Huang and J. E. Ferrell. Ultrasensitivity in the mitogene-activated protein kinase cascade. *Proceeding of the National Academy of Sciences*, 93:10078–10083, 1996. 39

[HH52] A. L. Hodgkin and A. F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.*, 117:500–544, 1952. 116

[HKK08] T. Heine, M. Kawohl, and R. King. Derivative-free optimal experimental design. *Chemical Engineering Science*, 63:4873–4880, 2008. 65

[HLBA04] Y. Hao, B. Laxton, E. R. Benson, and S. K. Agrawal. Differential flatness-based formation following of a simulated autonomous small grain harvesting system. *Transaction of the ASAE*, 47:1–9, 2004. 105

[HMGB03a] W. W. Hines, D. C. Montgomery, D. M. Goldsman, and C. M. Borror. *Probability and Statistics in Engineering*. John Wiley & Sons, Inc., 2003. 72

[HMGB03b] W. W. Hines, Douglas C. Montgomery, D. M. Goldsman, and C. M. Borror. *Probability and Statistics in Engineering*. JOHN WILEY & SONS, INC., 2003. 15, 16, 32, 43, 139

[Ise11] R. Isermann. *Identification of Dynamic Systems: An Introduction with Applications*. Springer, 2011. 85

[Isu99] S. S. Isukapalli. *Uncertainty Analysis of Transport-Transformation Models*. PhD thesis, Rutgers, The State University of New Jersey, 1999. 33, 34

[Jam80] F. James. Monte carlo theory and practice. *Rep. Prog. Phys.*, 43:1145–1189, 1980. 20

[JSMK06] M. Joshi, A. Seidel-Morgenstern, and A. Kremling. Exploiting the bootstrap method for quantifying parameter confidence intervals in dynamical systems. *Metabolic Engineering*, 8:447–455, 2006. 4, 21, 51

[JU94] S. J. Julier and J. K. Uhlmann. A general method for approximating nonlinear transformations of probability distributions. Technical report, Dept. of Engineering Science, University of Oxford, 1994. 24, 31, 35

[JU04] S. J. Julier and J. K. Uhlmann. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92:401–422, 2004. 18, 24, 26, 27, 31, 35, 36, 43

[KAG09] S. Körkel and H. Arellano-Garcia. Online experimental design for model validation. In *Proceedings of the 10th International Symposium on Process Systems Engineering - PSE2009*, 2009. 7, 70

[KAGSW08] S. Körkel, H. Arellano-Garcia, J. Schöneberger, and Günter Wozny. Optimum experimental design for key performance indicators. In *18th European Symposium on Computer Aided Process Engineering - ESCAPE 18*, 2008. 12

[Kal60] R. E. Kalman. A new approach to linear filtering and prediction problems. *ASME-Journal of Basic Engineering*, 82:35–45, 1960. 77

[Kay93] S. M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall PTR, 1993. 4, 15, 18, 50, 52

[KBG07] A. Kremling, K. Bettenbrock, and E. D. Gilles. Analysis of global control of escherichia coli carbohydrate uptake. *BMC Systems Biology*, 1:42, 2007. 60

[KFG+04] A. Kremling, S. Fischer, K. Gadkar, F. J. Doyle, T. Sauter, E. Bullinger, F. Allgöwer, and E. D. Gilles. A benchmark for methods in reverse engineering and model discrimination: Problem formulation and solutions. *Genome Res.*, 14:1773–1785, 2004. 7, 70

[Kho00] B. N. Kholodenko. Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades. *Eur J Biochem*, 267:1583–1588, 2000. 39

[KK00] G. A. Korn and Theresa M. Korn. *Mathematical Handbook For Scientists And Engineers*. Dover Publications, 2000. 32

[KKBS04] S. Körkel, E. Kostina, H. G. Bock, and J. P. Schlöder. Numerical methods for optimal control problems in design of robust optimal experiments for nonlinear dynamic processes. *Optimization Methods and Software*, 19:327–338, 2004. 51

[Kle07] J. P. C. Kleijnen. Kriging metamodeling in simulation: A review. discussion paper, 2007. 141

[KT09] C. Kreutz and J. Timmer. Systems biology: experimental design. *FEBS Journal*, 276:923–942, 2009. 4, 49, 70

[KW60] J. Kiefer and J. Wolfowitz. The equivalence of two extremum problems. *Canad. J. Math.*, 12:363–366, 1960. 53

[LÓ9] J. Lévine. *Analysis and Control of Nonlinear Systems - A Flatness-based Approach*. Springer, 2009. 105

[Lan05] J. R. Lanzante. A cautionary note in the use of error bars. *journal of Climate*, 18:3699–3703, 2005. 12

[LBS94] T. M. Ludden, S. L. Beal, and L. B. Sheiner. Comparison of the akaike information criterion, the schwarz criterion and the f test as guides to model selection. *Journal of Pharmacokinetics and Biopharmaceutics*, 22:431–445, 1994. 7, 70

[LC07] S. H. Lee and W. Chen. A comparative study of uncertainty propagation methods for black-box type functions. In *ASME 2007 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference*, 2007. 33

[LDTS07] S. Lorenz, E. Diederichs, R. Telgmann, and C. Schütte. Discrimination of dynamical system models for biological and chemical processes. *J Comput Chem*, 28:1384–1399, 2007. 93, 95

[Ler02] U. N. Lerner. *Hybrid Bayesian Networks For Reasoning About Complex Systems*. PhD thesis, Stanford University, 2002. 18, 25, 26, 27, 29, 32

[Lev66] N. Levinson. Wiener's life. *Bull. Amer. Math. Soc.*, 72:1–32, 1966. 1

[Li92] K. S. Li. Point-estimate method for calculating statistical moments. *J. Eng. Mech.*, 118:1506–1511, 1992. 25, 33

REFERENCES

[LL13] Z. Lin and W. Li. Restrictions of point estimate methods and remedy. *Reliability Engineering & System Safety*, 111:106–111, 2013. 25

[LML+09] M. Lamboni, D. Makowksi, S. Lehuger, B. Gabrielle, and H. Monod. Multivariate global sensitivity analysis for dynamic crop models. *Field Crop Research*, 113:312–320, 2009. 53

[LMM10] M. Lamboni, D. Makowksi, and H. Monod. Empirical relationship between sensitivity index and mean square error of prediction: a case study for greenhouse gas emission. *Procedia Social and Behavioral Sciences*, 2:7698–7699, 2010. 53

[Lor05] S. Lorenz. *The model-data-overlap*. PhD thesis, Freie Universität Berlin, 2005. 91, 93

[LP06] K. H. Lee and G. J. Park. A global robust optimization using kriging based approximation model. *JSME International Journal*, 49, 2006. 141

[Lu07] L. Lu. *Inverse Modelling and Inverse Simulation for System Engineering and Control Applications*. PhD thesis, University of Glasgow, 2007. 137

[MALF12] C. A. Mattson, T. V. Anderson, B. J. Larson, and D. T. Fullwood. Efficient propagation of error through system models for functions common in engineering. *Journal of Mechanical Design*, 134, 2012. 18

[Mat63] G. Matheron. Principles of geostatistics. *Economic Geology*, 58:1246–1266, 1963. 141

[MB12] J. Mandur and H. Budman. A polynomial-chaos based algorithm for robust optimization in the presence of bayesian uncertainty. In *8th IFAC Symposium on Advanced Control of Chemical Processes*, 2012. 33

[MC04] H. Motulsky and A. Christopoulos. *Fitting Models to Biological Data Using Linear and Nonlinear Regression: A Practical Guide to Curve Fitting*. Oxford Univ Pr, 2004. 7, 70, 72, 73

[MFSR05] L. Möhler, D. Flockerzi, H. Sann, and U. Reichl. Mathematical model of influenza a virus production in large-scale microcarrier culture. *Biotechnology and Bioengineering*, 90:46–58, 2005. 128

[MK10] O. P. Le Maitre and O. M. Knio. *Spectral Methods for Uncertainty Quantification*. Springer, 2010. 16, 21, 22, 23, 32

[MMB03a] C. G. Moles, P. Mendes, and J. R. Banga. Parameter estimation in biochemical pathways: A comparison of global optimization methods. *Genome Res.*, 13:2467–2474, 2003. 98

[MMB03b] C. G. Moles, P. Mendes, and J. R. Banga. Parameter estimation in biochemical pathways: A comparison of global optimization methods. *Genome Res.*, 13:2467–2474, 2003. 110

[MPR07] J. M. Morales and J. Pérez-Ruiz. Point estimate schemes to solve the probabilistic power flow. *IEEE Transactions on Power Systems*, 22:1594–1601, 2007. 25

[MR98] H. Mounier and J. Rudolph. Flatness-based control of nonlinear delay systems: a chemical reactor example. *Int. J. Control*, 71:871–890, 1998. 105

[MS11] D. J. Murray-Smith. Feedback methods for inverse simulation of dynamic models for engineering systems applications. *Mathematical and Computer Modelling of Dynamical Systems*, 17:515–541, 2011. 104

[MS12a] M. Mangold and R. Schenkendorf. Methods for optimal experimental design of nonlinear systems using sigma points. In *Dechema-Regionalkolloquium: "Design of Optimal Experiments", Magdeburg*, 2012.

[MS12b] D. J. Murray-Smith. The application of parameter sensitivity analysis methods to inverse simulation models. *Mathematical and Computer Modelling of Dynamical Systems*, 19:67–90, 2012. 104

[MSM10] C. Michalik, M. Stuckert, and Wolfgang Marquardt. Optimal experimental design for discriminating numerous model candidates: The awdc criterion. *Ind. Eng. Chem. Res.*, 49:913–919, 2010. 70

[MV08] S. Mekid and D. Vaja. Propagation of uncertainty: Expressions of second and third order uncertainty with third and fourth moments. *Measurement*, 41:600–609, 2008. 18

[MWL02] N. McIntyre, H. Wheather, and M. Lees. Estimation and propagation of parametric uncertainty in environmental models. *Journal of HydroInformatics*, 04.3:177–198, 2002. 25

[NAY62] J. Nagumo, S. Arimoto, and S. Yoshizawa. An active pulse transmission line simulating nerve axon. *Proceedings of the IRE*, 50:2061–2070, 1962. 116

[NHO+10] M. Nakatsui, K. Horimoto, M. Okamoto, Y. Tokumoto, and J. Miyake. Parameter optimization by using differential elimination: a general approach for introducing constraints into objective functions. *BMC Systems Biology*, 4(Suppl 2):59, 2010. 102, 103

[NK04] H. Naranf and M. Krishna. Mitogen-activated protein kinases: Specificity of response to dose of ionizing radiation in liver. *J. Radiat. Res.*, 45:213–220, 2004. 71

[NRPH03] M. Nørgaard, O. Ravn, N. K. Poulsen, and L. K: Hansen. *Neural Networks for Modelling and Control of Dynamic System*. Springer, 2003. 85

[NSL+12] M. Nakatsui, A. Sedoglavic, F. Lemaire, F. Boulier, A Ürgüplü, and K. Horimoto. A general procedure for accurate parameter estimation in dynamic systems using new estimation errors. In K. Horimoto, M. Nakatsui, and N. Popov, editors, *Algebraic and Numeric Biology*, volume 6479 of *Lecture Notes in Computer Science*, pages 149–166. Springer Berlin Heidelberg, 2012. 102, 103

[NW99] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, 1999. 149

[ON05] G. C. Okpokwasili and C. O. Nweke. Microbial growth and substrate utilization kinetics. *African Journal of Biotechnology*, 5:305–317, 2005. 94

[OO04a] J. E. Oakley and A. O'Hagan. Probabilistic sensitivity analysis of complex models: a bayesian approach. *J. R. Statist. Soc. B*, 66:751–769, 2004. 46

[OO04b] A. O'Hagan and J. E. Oakley. Probability is perfect, but we can't elicit it perfectly. *Reliability Engineering & System Safety*, 85:239–248, 2004. 12

[PK12] G. Paramasivan and A. Kienle. Decentralized control system design under uncertainty using mixed-integer optimization. *Chemical Engineering & Technology*, 35:261–271, 2012. 25

[PP07] A. Pzman and Luc Pronzato. *mODa 8 - Advances in Model-Oriented Design and Analysis*. Physica-Verlag HD, 2007. 4, 51

[PRB+02] N. Petit, P. Rouchon, J.-M. Boueilh, F. Guérin, and P. Pinvidic. Control of an industrial polymerization reactor using flatness. *Journal of Process Control*, 12:659–665, 2002. 105

[Pro08] L. Pronzato. Optimal experimental design and some related control problems. *Automatica*, 44:303–325, 2008. 4

[PT07] M. Peifer and J. Timmer. Parameter estimation in ordinary differential equations for biochemical processes using the method of multiple shooting. *IET Systems Biology*, 1:78–88, 2007. 110

[PVM+06] A. A. Poyton, M. S. Varziri, K. B. McAuley, P. J. McLellan, and J. O. Ramsay. Parameter estimation in continuous-time dynamic models using principle differential analysis. *Computers and Chemical Engineering*, 30:698–708, 2006. 108

[QDSM11] T. Quaiser, A. Dittrich, F. Schaper, and M. Mönnigmann. A simple work flow for biologically inspired model reduction - application to early jak-stat signaling. *BMC Systems Biology*, 5:30, 2011. 39

[Rat83] D. A. Ratkowsky. *Nonlinear Regression Modelling: A unified Practical Approach*. Marcel Decker, Inc., 1983. 62, 146

[Rat01] A. Ratle. Kriging as a surrogate fitness landscape in evolutionary optimization. *Artificial Intelligence for Engineering Design , Analyses and Manufacturing*, 15:37–49, 2001. 141

[RHCC07] J. O. Ramsay, G. Hooker, D. Campbell, and J. Cao. Parameter estimation for differential equations: a generalized smoothing approach. *J. R. Statist. Soc. B*, 69:741–796, 2007. 116

[RHK10] N. Rossner, T. Heine, and R. King. Quality-by-design using a gaussian mixture density approximation of biological uncertainties. In *Proceedings of the 11th International Symposium in Computer Applications in Biotechnologyy (CAB 2010), Leuven, Belgium, July 7-9*, 2010. 35

[RMS12] S. Rollié, M. Mangold, and K. Sundmacher. Designing biological systems: Systems engineering meets synthetic biology. *Chemical Engineering Science*, 69:1–29, 2012. 3

[RP02] S. J. Roberts and W. D. Penny. Variational bayes for generalized autoregressive models. Technical report, Robotics Research Group, Oxford University, 2002. 74

[RS05] J. O. Ramsay and B. W. Silverman. *Functional Data Analysis*. Springer New York, 2005. 108

[RvdSHV07] J. C. Refsgaard, J. P. van der Sluijs, A. L. Hojberg, and P. A. Vanrolleghem. Uncertainty in the environmental modelling process - a framework and guidance. *Environmental Modelling & Software*, 22:1543–1556, 2007. 13, 14

[RWL+12] K. Rateitschak, F. Winter, F. Lange, R. Jaster, and O. Wolkenhauer. Parameter identifiablility and sensitivity analysis predict targets for enhancement of stat1 activity in pancreatic cancer and stellate cells. *PLoS Computational Biology*, 8(12), 2012. 39

[SA11] M. Sjöstrand and Özlem Aktas. *Cornish-Fisher Expansion and Value-at-Risk in Application to Risk Management of Large Portfolios*. LAP LAMBERT Academic Publishing, 2011. 35, 40

[SB75] J. Swartz and H. Bremermann. Discussion of parameter estimation in biological modelling: Algorithms for estimation and evaluation of the estimates. *Journal of Mathematical Biology*, 1:241–257, 1975. 101, 102

[Sed02] A. Sedoglavic. A probabilistic algorithm to test local algebraic observability in polynomial time. *Journal of Symbolic Computation*, 33(5):735–755, May 2002. 80, 82

[Sey10] R. U. Seydel. *Practical Bifurcation and Stability Analysis*. Springer, 2010. 116

[SG01] N. Schenker and J. F. Gentleman. On judging th significance of differences by examining the overlap between confidence intervals. *The American Statistican*, 55:182–186, 2001. 12

[Sim06] D. Simon. *Optimal state estimation*. Wiley-Interscience, 2006. 32, 77, 78

[SKM08] R. Schenkendorf, A. Kremling, and M. Mangold. Application of sigma points to the optimal experimental design of a biological system. In *The 20th International Symposium on Chemical Reaction Engineering- Green Chemical Reaction Engineering for a Sustainable Future*, 2008.

[SKM09a] R. Schenkendorf, A. Kremling, and M. Mangold. Optimal experimental design and model selection by a sigma point approach. In *Mathmod 2009 - 6th Vienna International Conference on Mathematical Modelling, Vienna, Austria*, 2009. 69

[SKM09b] R. Schenkendorf, A. Kremling, and M. Mangold. Optimal experimental design with the sigma point method. *IET Systems Biology*, 3:10–23, 2009. 56

[SKM12] R. Schenkendorf, A. Kremling, and M. Mangold. Influence of non-linearity to the optimal experimental design demonstrated by a biological system. *Mathematical and Computer Modelling of Dynamical Systems*, 18:413–426, 2012. 60

REFERENCES

[SL10] D. Skanda and D. Lebiedz. An optimal experimental design approach to model discrimination in dynamic biochemical systems. *Bioinformatics*, 26:939–945, 2010. 70

[SM11a] R. Schenkendorf and M. Mangold. Challenges of parameter identification for nonlinear biological and chemical systems. In *SIAM Conference on Optimization, Darmstadt*, 2011.

[SM11b] R. Schenkendorf and M. Mangold. Qualitative and quantitative optimal experimental design for parameter identification. In *18th IFAC World Congress Milano (Italy)*, 2011.

[SM13] R. Schenkendorf and M. Mangold. Online model selection approach based on unscented kalman filtering. *Journal of Process Control*, 23:44–47, 2013. 69

[SM14] R. Schenkendorf and M. Mangold. Parameter identification for ordinary and delay differential equations by using flat inputs. *Theoretical Foundations of Chemical Engineering*, accepted, 2014. 97

[Sob93] I. M. Sobol'. Sensitivity analysis for nonlinear mathematical models. *Mathematical Modeling and Computational Experiment*, 1:407–414, 1993. 38

[Sob01] I. M. Sobol'. Global sensitivity indices for nonlinear mathematical models and the monte carlo estimates. *Ma*, 55:271–280, 2001. 38

[SRM12] R. Schenkendorf, U. Reichl, and M. Mangold. Parameter identification of time-delay systems: A flatness based approach. In *MATHMOD 2012 - Full Paper Preprint Volume*, 2012. 97

[SRTC05] A. Saltelli, M. Ratto, S. Tarantola, and F. Campolongo. Sensitivity analysis for chemical models. *Chemical Reviews*, 105:28112828, 2005. 38

[SSK09] J.-F. Stumper, F. Svaricek, and R. Kennel. Trajectory tracking control with flat inputs and a dynamic compensator. In *Proceedings of the European Control Conference*, 2009. 112

[Ste94] R. F. Stengel. *Optimal Control and Estimation*. Dover Publications, 1994. 13, 77, 78, 80

[TB06] D. Thomson and R. Bradley. Inverse simulation as a tool for fight dynamics research - principles and applications. *Progress in Aerospace Sciences*, 42:174–210, 2006. 104

[Tem09] Brian A. Templeton. A polynomial chaos approach to control. Master's thesis, Virginia Polytechnic Institute ans State University, 2009. 22

[TF05] C. W. Tsai and S. Franceschini. Evaluation of probabilistic point estimate methods in uncertainty analysis for environmental engineering applications. *J. Environ. Eng.*, 131:387–395, 2005. 25

[TKA00] Y. Tanaka, K. Kimata, and H. Aiba. A novel regulatory role of glucose transporter of escherichia coli: membrane sequestration of a global repressor mlc. *The EMBO Journal*, 19:5344–5352, 2000. 60

[TLDI12a] D. Telen, F. Logist, E. Van Derlinden, and J. F. Van Impe. Robust optimal experiment design: A multi-objective approach. In *7th Vienna Conference on Mathematical Modelling*, 2012. 51

[TLDI12b] D. Telen, F. Logist, E. Van Derlinden, and J. Van Impe. Approximate robust optimal experiment design in dynamic bioprocess models. In *20th Mediterrnean Conference on Control & Automation (MED), Barcelona, Span*, 2012. 51

[Tyl53] G. W. Tyler. Numerical integration of functions of several variables. *Canadian Jn. Math.*, 5:393–412, 1953. 24, 25

[Var08] M. S. Varziti. *Parameter Estimation in Nonlinear Continuous-Time Dynamic Models with Modelling Errors and Process Disturbances*. PhD thesis, Queen's University, Kingston, Ontaria, Canada, 2008. 102

[vdM04] R. v. d. Merwe. *Sigma-Point Kalman Filters for Probabilistic Inference in Dynamic State-Space Models*. PhD thesis, OGI School of Science & Engineering at Oregon Health & Science University, 2004. 31, 32

[VG07] J.-P. Vila and J.-P. Gauchi. Optimal designs based on exact confidence regions for parameter estimation of a nonlinear regression model. *Journal of Statistical Planning and Inference*, 137:2935–2953, 2007. 51

[VGS+10] V. Vassilev, M. Gröschel, H.J. Schmid, W. Peukert, and G. Leugering. Interfacial energy estimation in a precipitation reaction using flatness based control of the moment trajectories. *Chemical Engineering Science*, 65:2183–2189, 2010. 104

[VMM08a] M. S. Varziri, K. B. McAuley, and P. J. McLellan. Approximate maximum likelihood parameter estimation for nonlinear dynamic models: Application to a laboratory-scale nylon reactor model. *Ind. Eng. Chem. Res.*, 47:7274–7283, 2008. 102

[VMM08b] M. S. Varziri, K. B. McAuley, and P. J. McLellan. Parameter estimation in continuous-time dynamic models in the presence of unmeasured states and nonstationary disturbances. *Ind. Eng. Chem. Res.*, 47:380–393, 2008. 102

[VPM+08] M. S. Varziri, A. A. Pyoton, K. B. McAuley, P. J. McLellan, and J. O. Ramsay. Selecting optimal weighting factors in ipda for parameter estimation in continuous-time dynamic models. *Computers and Chemical Engineering*, 32:3011–3022, 2008. 108

[WAJ+13] O. Wolkenhauer, C. Auffray, R. Jaster, Gustav Steinhoff, and O. Dammann. The road from systems biology to systems medicine. *Pediatric RESEARCH*, 73:502–507, 2013. 3

[Wey02] T. Wey. *Nichtlineare Regelungssysteme - Ein differentialalgebraischer Ansatz*. Teubner, 2002. 113

[Wit04] A. Wittkopf. *Algorithms and Implementations for Differential Elimination*. PhD thesis, Simon Fraser University, Canada, 2004. 103

[Won94] W. K. Wong. Comparing robust properties of a, d, e and g-optimal designs. *Computational Statistics & Data Analysis*, 18:441–448, 1994. 53

[Won95] W. K. Wong. On the equivalence of d and g-optimal designs in heteroscedastic models. *Statistics & Probability Letters*, 25:317–321, 1995. 53

[WP97] E. Walter and L. Pronzato. *Identification of Parametric Models*. Springer, 1997. 1, 4, 52, 57

[WZ08] S. Waldherr and M. Zeitz. Conditions for the existence of a flat input. *International Journal of Control*, 81:439–443, 2008. 106, 107, 112

[WZ10] S. Waldherr and M. Zeitz. Flat inputs in the mimo case. In *IFAC Symposium on Nonlinear Control Systems*, 2010. 112

[XK02] D. Xiu and G. E. Karniadakis. The wiener-askey polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing*, 24:619–644, 2002. 32

[XM12] J. Xue and J. Ma. A comparative study of several taylor expansion methods on error propagation. In *GEOINFORMATICS, 2012 20th International Conference on Geoinformatics*, 2012. 18

[Yen05] Ö. Yeniay. Penalty function methods for constrained optimization with genetic algorithms. *Mathemtical and Computational Applications*, 10, 2005. 39

[Zha06] J. Zhang. The calculating formulae, and experimental methods in error propagation analysis. *IEEE Transactions on Reliability*, 55:169–181, 2006. 18

[ZP08] Z. Zainuddin and O. Pauline. Function approximation using artificial neural networks. *WSEAS Transactions on Mathematics*, 7:333–338, 2008. 108