**Ghent University**
Faculty of Bioscience Engineering

# OPTIMAL EXPERIMENTAL DESIGN TO DISCRIMINATE AMONG RIVAL DYNAMIC MATHEMATICAL MODELS

ir. Brecht Donckels

Thesis submitted in fulfillment of the requirements for the degree of
Doctor (Ph.D) in Applied Biological Sciences

Academic year 2008-2009

*"Life comes down to a few moments. This is one of them."*

Oliver Stone, film director

*"It is good to have an end to journey toward;*
*but it is the journey that matters, in the end."*

Ernest Hemmingway, writer and journalist

*"I love it when a plan comes together."*

Colonel John "Hannibal" Smith, leader of the A-Team

Supervisors: Prof. dr. Bernard De Baets
Department of Applied Mathematics, Biometrics and Process Control
Ghent University, Belgium

Prof. dr. ir. Peter Vanrolleghem
Département de génie civil
Université de Laval, Canada

Dr. ir. Dirk De Pauw
Phyto-IT, Belgium

Dean: Prof. dr. ir. Guido Van Huylenbroeck

Rector: Prof. dr. Paul Van Cauwenberge

ir. Brecht Donckels

# OPTIMAL EXPERIMENTAL DESIGN TO DISCRIMINATE AMONG RIVAL DYNAMIC MATHEMATICAL MODELS

Thesis submitted in fulfillment of the requirements for the degree of

Doctor (Ph.D) in Applied Biological Sciences

Academic year 2008-2009

*Dutch translation of the title:*
Ontwerp van optimale experimenten voor het discrimineren van rivaliserende dynamische wiskundige modellen

*Please refer to this work as follows:*
Brecht M.R. Donckels (**2009**). *Optimal experimental design to discriminate among rival dynamic mathematical models*, PhD Thesis, Department of Applied Mathematics, Biometrics and Process Control, Ghent University, Ghent, Belgium.

ISBN 978-90-5989-291-0

# Dankwoord

Jawel, na vier jaar hard werken is mijn doctoraat eindelijk af. Ooit verklaarde iemand me zot dat ik na vijf jaar bio-ingenieur nog eens de moed vond om voor dokter te studeren. Dat is het dus niet. Ik ben de voorbije vier jaar ook niet echt student geweest, maar toch bedankt voor de vele succeswensen tijdens de examenperiodes. Neen, als doctoraatstudent doe je aan wetenschappelijk onderzoek. Maar een doctoraat is meer dan dat. In het begin valt alles best mee. Je leest wat, werkt je in in de materie en dan begint het pas echt. Als je je doctoraat wil behalen, word je namelijk geacht een bijdrage te leveren aan de wetenschap. Makkelijker gezegd dan gedaan. Voortdurend spoken diezelfde vragen door je hoofd, ook wanneer je er liever eens niet aan denkt. Stilaan wordt het doctoraat een stukje van jezelf.

Op de kaft van dit doctoraat staat dan wel mijn naam, maar verschillende mensen hebben er rechtstreeks of onrechtstreeks mee voor gezorgd dat dit doctoraat er *überhaupt* is gekomen. Een dankwoordje is dan ook zeker op zijn plaats.

Eerst en vooral zou ik mijn promotoren willen bedanken. Peter, je gaf me vier jaar geleden de kans om aan dit doctoraat te beginnen. Jammer dat je kort daarna naar Canada vertrok, maar gelukkig staat de technologie niet stil en kon ik alsnog op je inbreng rekenen. Bernard, in de eerste plaats om me te *adopteren*, maar uiteraard ook om tijd vrij te maken om resultaten te bespreken, teksten na te lezen en me op jouw specifieke manier te motiveren. Mijn onderzoek stond ver af van het jouwe, maar ik hoop dat je achteraf geen spijt hebt dat je je over mij ontfermd hebt. En dan Dirk. Ik denk dat ik wel mag zeggen dat dit doctoraat er niet ging gekomen zijn moest jij er niet geweest zijn. Ik heb enorm veel van jou geleerd, zowel op wetenschappelijk als op persoonlijk vlak. Bedankt!

Uiteraard ook een woordje van dank voor de mensen waarmee ik de voorbije vier jaar een bureau heb gedeeld. Ik denk dat de sfeer in onze bureau uniek was. Als persoon zijn

we allemaal verschillend, maar als groep gingen we toch verdomd goed samen. Aditya, je frisse en open kijk op de wereld zal me altijd bij blijven en ik ben ervan overtuigd dat je een schitterend doctoraat zult afleveren. Succes ermee. Jo, bedankt om zo dikwijls je deskundige mening te geven over vanalles en nog wat en voor al het duw- en trekwerk achter de schermen. Gaspard, om me met een nieuw dilemma op te zadelen: Linux of toch maar Windows. Bedankt ook om me nu en dan te depaneren als ik weer eens te veel geupgrade had. Gino, je enthousiasme en inzet is uniek, en ik zou je willen bedanken voor hetgeen je voor het project hebt gedaan en hebt betekend. Hetzelfde kan gezegd worden van Marjan, Joeri, Ellen en Hendrik. Jullie zijn stuk voor stuk bijzonder sympathieke mensen en jullie hulp en aanmoedigingen werden enorm geapprecieerd.

Ik zou ook alle collega's willen bedanken die zorg(d)en voor de plezante werksfeer op de vakgroep, en Tinne en Michael in het bijzonder. Tinne. Luidruchtig en direct, dat is meteen duidelijk, maar bovenal iemand met een groot hart. Bedankt voor de vele cakes en andere zoetigheden, de babbels, het tolereren van mijn (meestal) grappig bedoelde opmerkingen en de vriendschap *tout court*. Michael, bedankt dat ik altijd bij jou terecht kon met mijn verhaal en voor *alweer* een nederlaag in het fameuze kaartspel...

Jeanke, mijnen beste maat. Nu dit doctoraat eindelijk achter de rug is, zal ik terug wat meer tijd hebben om eens een pint te gaan drinken, een bordspel te spelen en om je eraf te rijden. Jochen, Wim, Tim M., Korneel, Pim, Sandy, Ans, Katrien, Ilse, Mattias, Wesley, Lander, Tim V., Joke, Walter en iedereen waar ik nu niet direct aan denk, bedankt om contact te blijven houden en ik probeer in de toekomst wat meer uit mijn kot te komen.

Ma en pa, jullie gaven me altijd kansen en ik ben jullie daarvoor eeuwig dankbaar. Jullie gaven me ook een warm nest en een toffe broer en zus. Bedankt voor alles! Timmelootje, 't zusje, maar ook Joline'tje en Bram maakten de weekenduitstapjes naar Zuid-West-Vlaanderen steevast de moeite waard. Bedankt ook aan de schoonfamilie in spe voor de tweede thuis die jullie me gaven.

Isabelle, mijn lief kleintje. Je bent een mooi mens(je) in alle opzichten. Ik wil je bedanken voor de steun die je bent geweest tijdens dit avontuur. Bedankt om me op te vangen als het weer eens tegen zat. Bedankt om mij te overtuigen om door te doen als ik het weer eens niet meer zag zitten. Bedankt om te luisteren naar mijn gezaag over grafiekjes en simulaties. Bedankt om zo goed voor mij te zorgen. Bedankt om mij graag te zien. Ik hou van jou.

Gent, April 2009 - Brecht Donckels

# Contents

# List of abbreviations and symbols

**Abbreviations**

| | |
|---|---|
| **FIM** | $n_p$-dimensional Fisher information matrix (Eqs. (2.20) and (2.21)) |
| AIC | Akaike information criterion (Eq. (2.33)) |
| BIC | Bayesian information criterion (Eq. (2.34)) |
| MEF | modelling efficiency (Eq. (2.30)) |
| WSSE | weighted sum of squared errors (Eq. (2.4)) |

**Symbols used in the glucokinase models**

| | |
|---|---|
| ADP | concentration of adenosine diphosphate [mM] |
| $ADP_p$ | concentration of adenosine diphosphate in the pulse [mM] |
| ATP | concentration of adenosine triphosphate [mM] |
| $ATP_p$ | concentration of adenosine triphosphate in the pulse [mM] |
| G6P | concentration of glucose-6-phosphate [mM] |
| $G6P_p$ | concentration of glucose-6-phosphate in the pulse [mM] |
| GLK | concentration of glucokinase [mg/L] |
| GLU | concentration of glucose [mM] |
| $GLU_p$ | concentration of glucose in the pulse [mM] |

| | |
|---|---|
| PEP | concentration of phosphoenolpyruvate [mM] |
| $\text{PEP}_p$ | concentration of phosphoenolpyruvate in the pulse [mM] |
| V | volume of the reaction vessel [L] |
| $F_p$ | flowrate of the pulse [L/s] |
| $k$ | catalytic rate constant for the reaction catalyzed by glucokinase [U/mg] |
| $K_{\text{ATP}}$ | equilibrium constant associated with ATP [mM] |
| $K_{\text{GLU}}$ | equilibrium constant associated with glucose [mM] |
| $K_{\text{PEP}}$ | equilibrium constant associated with PEP [mM] |
| $v_{glk}$ | velocity equation describing the kinetic behavior of glucokinase [mM/s] |

**Symbols**

| | |
|---|---|
| $\boldsymbol{\mathcal{P}}$ | Pareto-optimal front |
| $\ell_1$ | taxicab or Manhattan distance function |
| $\ell_2$ | Euclidean distance function |
| $\ell_\infty$ | Chebyshev distance function |
| $\hat{p}\,(x)$ | estimation of the probability density function of random variable $x$ (Eq. (7.1)) |
| $\boldsymbol{\Omega}$ | $n_m$-dimensional model prediction error covariance matrix (Eq. (2.24)) |
| $\boldsymbol{\Phi}$ | $n_p$-dimensional parameter estimation error covariance matrix (Eq. (4.26)) |
| $\boldsymbol{\Psi}_{ij}$ | uncertainty on the difference between the predicted outcomes of an experiment by models $i$ and $j$ (Eq. (4.21)) |
| $\boldsymbol{\Sigma}$ | $n_m$-dimensional measurement error covariance matrix |
| $\boldsymbol{\Theta}$ | $n_p$-dimensional space of possible model parameters |
| $\boldsymbol{\Xi}$ | feasible values of the experimental degrees of freedom |
| $\hat{\boldsymbol{\theta}}$ | $n_p$-dimensional vector of model parameters estimates |

| | |
|---|---|
| $\hat{\boldsymbol{y}}$ | $n_m$-dimensional vector of predicted response variables |
| $\boldsymbol{\theta}$ | $n_p$-dimensional vector of model parameters |
| $\boldsymbol{\xi}$ | an experiment, determined by its experimental degrees of freedom |
| $\boldsymbol{\xi}^\star$ | optimally designed experiment |
| $\boldsymbol{\xi}_c$ | compromise experiment |
| $\boldsymbol{f}^\circ$ | ideal point (Eq. (7.9)) |
| $\boldsymbol{x}$ | $n_s$-dimensional vector of state variables |
| $\boldsymbol{y}$ | $n_m$-dimensional vector of measured response variables |
| $\kappa$ | kernel density function |
| $\mathrm{D}(\boldsymbol{\xi})$ | D-optimality design criterion value associated with experiment $\boldsymbol{\xi}$ |
| $\mathrm{modE}(\boldsymbol{\xi})$ | modE-optimality design criterion value associated with experiment $\boldsymbol{\xi}$ |
| $\pi_i$ | relative model probability associated with model $m_i$ (Eq. (4.24)) |
| $\rho_\Omega$ | $n_m$-dimensional model prediction error correlation matrix (Eq. (2.25)) |
| $\rho_\Phi$ | $n_p$-dimensional parameter estimation error correlation matrix (Eq. (2.22)) |
| $\rho_i$ | model probability associated with model $m_i$ (Eq. (4.23)) |
| $\varsigma_y$ | constant minimal relative error on the measurement of $y$ |
| $d$ | index of agreement (Eq. (2.32)) |
| $h$ | smoothing parameter |
| $lb_y$ | lower accuracy bound on the measurement of $y$ |
| $m$ | number of rival models |
| $m_i$ | model $i$ |
| $n_e$ | number of experiments |
| $n_m$ | number of measured state variables |

| | |
|---|---|
| $n_p$ | number of model parameters |
| $n_s$ | number of state variables |
| $n_{sp}$ | number of sampling times |
| $t$ | time |
| $T(\boldsymbol{\xi})$ | overall discriminatory potential of experiment $\boldsymbol{\xi}$ |
| $T_a$ | approach of Hunter and Reiner (1965) to design optimal discriminatory experiments |
| $T_b$ | modified approach of Hunter and Reiner (1965) to design optimal discriminatory experiments |
| $T_c$ | approach of Buzzi-Ferraris et al. (1984) to design optimal discriminatory experiments |
| $T_d$ | anticipatory approach to design optimal discriminatory experiments |
| $T_{ij}(\boldsymbol{\xi})$ | discriminatory potential of experiment $\boldsymbol{\xi}$ to discriminate between model $m_i$ and model $m_j$ |
| $t_{min}$ | minimum time interval between two subsequent sampling times |
| $\Gamma_{\mathrm{D}_{ij}}$ | ratio between the D-optimality criterion value associated with the compromise experiment and the D-optimal experiment |
| $\Gamma_{\mathrm{modE}_{ij}}$ | ratio between the modE-optimality criterion value associated with the compromise experiment and the modE-optimal experiment |

# CHAPTER 1

## Problem statement, research objectives and outline

## Abstract

The aim of a modelling exercise is to obtain a mathematical model that adequately describes and even predicts the process behavior. However, it is important to realize that the lack of insight in the modelled process may result in the proposal of several so-called rival models, each of which represents a certain hypothesis of how the process works. The problem of identifying the best model from a set of rival models, often referred to as the *problem of model discrimination*, is dealt with in this dissertation. In this chapter, the objectives of this research are described and a number of research questions are formulated in order to meet these research objectives. In addition, this chapter provides an overview of how this dissertation is organized.

## 1.1 Introduction

A (mechanistic) mathematical model can be defined as a mathematical representation of the mechanism that governs the behavior of a process being studied, and the aim of a modelling exercise is to obtain a mathematical model that adequately describes and even predicts the process behavior. Although the complexity of most processes generally makes it impossible to exactly describe all aspects of the process behavior, mathematical models that are able to describe the most important ones have already shown to be very useful tools for both scientists and engineers.

The main use of a mathematical model is probably to act as a surrogate for the actual process, making it possible to investigate the process behavior under various input conditions both rapidly and inexpensively, and without necessarily tampering with the actual process (Ogunnaike and Ray, 1994). Examples of such applications are found in the context of process design, optimization and control. For instance, the performance of different operating conditions for a given process can be evaluated based on model predictions and the one for which the expected operating costs are minimal can eventually be implemented. Or, different control strategies for a given process can be developed, tested and compared *in silico*, without risking process failure. Mathematical models can also be used for online process control. Indeed, information obtained from online process measurements can be analyzed in real-time using mathematical models, that represent knowledge about the process. In this way, process failures can be detected in an early stage and appropriate control actions can be proposed to restore the process.

However, if the model is intended to increase the understanding of the (often complex) process, it may be more important that the model is able to capture all the different aspects and mechanisms that govern the behavior of the process. In such a case, some deviation from reality can be accepted as long as the model structure permits an acceptable description of the overall process. The models used to predict global warming are examples of such models. Given the enormous complexity of the climate system, simplifications are unavoidable when trying to capture its behavior. But even then, the system is too complex for a human being to grasp, and therefore mathematical models are used to integrate the many different aspects. Model simulations are then used to get an idea of how our climate will evolve over the next decades. Although these predictions may not be perfectly accurate nor precise, these models are useful decision support tools for policy makers.

## 1.2 Problem statement

From the discussion above, it is clear that once a proper mathematical model is available, it becomes a powerful tool for both scientists and engineers (Wiechert, 2002). However, it is important to realize that the lack of insight in the modelled process may result in the proposal of several so-called rival models, each of which represents a certain hypothesis of how the process works. Obviously, one is especially interested in the model that describes the process behavior in the best way. The problem of identifying the best model from a set of rival models, often referred to as the *problem of model discrimination*, is dealt with in this dissertation. As additional experiments have to be performed to identify the most appropriate model, optimal experimental design will be of crucial importance in order to minimize the required experimental effort.

## 1.3 Objectives of this research

To address the problem of model discrimination described above in a systematic way, the following research objectives were formulated:

1. obtain insight in the different aspects of the problem of model discrimination and how they influence each other,
2. propose a general procedure to discriminate among a set of rival models,
3. provide methods to design optimal discriminatory experiments,
4. position the model discrimination procedure in a more general procedure for building mathematical models.

If these research objectives are met, a systematic framework should become available to deal with a situation where several models are proposed to describe a particular process. To meet these research objectives, the five research questions described in the following section will be answered throughout this dissertation.

## 1.4 Research questions

When confronted with a situation where several rival models are proposed for a given process, it is of crucial importance to have an idea of how the problem can be dealt with in a systematic way. The first research question can thus be stated as:

1. *What is the general procedure to identify the most appropriate model from a set of rival models?*

When a number of rival models is proposed for a given process, it is clear that some of these models may not be able to accurately describe the experimental data. Of course, it is important to be able to detect these inadequate models. The second research question is related to the adequacy of mathematical models and can be formulated as:

2. *How can the adequacy of a model be evaluated and (how) can this be translated into quantitative model evaluation criteria?*

Often, it is necessary to collect new information about the modelled system to allow (further) model discrimination, and thus new experiments have to be performed. The third research question deals with model-based experimental design methods, in which the rival models themselves are used to predict the outcome of an experiment and the experiment is evaluated based on these model predictions.

3. *If additional experimental data has to be collected to allow further model discrimination, how can experiments be designed such that model discrimination is achieved with a minimum of additional experimental effort?*

As the parameters of the models used in the experimental design have to be estimated from experimental data which may not contain the information required to get accurate parameter estimates, the resulting uncertainty on the model predictions has to be taken into account in the experimental design. Since the design of optimal discriminatory experiments may become problematic when the uncertainty on the model predictions is too large, a fourth research question is formalulated as follows:

4. *What is the importance of the uncertainty on the parameter estimates with regard to the design of optimal discriminatory experiments?*

Since experimental design methods have been developed to design experiments that allow a more accurate estimation of the model parameters, they may be applied in this context as well. It is therefore important to investigate how these experimental design methods for more precise parameter estimation can be integrated in or combined with the model discrimination procedure. So, a fifth research question can be formulated as follows:

5. *Can optimal experimental design for parameter estimation be integrated with the procedure for model discrimination, and is it beneficial to do so?*

The following section describes how this dissertation is organized.

## 1.5 Roadmap through this dissertation

This dissertation consists of three main parts. In the first part, some basic modelling concepts are explained and a general procedure to discriminate among a number of rival models is proposed. The second part of this dissertation is devoted to the design of optimal discriminatory experiments. The integration of model discrimination with optimal experimental design for parameter estimation is dealt with in the third part.

The first part consists of two chapters. In a first chapter, basic modelling concepts such as parameter estimation, uncertainty on the parameter estimates and on the model predictions are described (Chapter 2). These will reappear regularly in the subsequent chapters and a good understanding of these concepts is crucial. In addition, this chapter describes a general procedure to discriminate among rival models. Chapter 3 contains a description of the working example which will be used throughout this dissertation. In this chapter, nine rival models are proposed to describe the kinetics of an enzyme (glucokinase).

The second part also consists of two chapters and deals with the design of optimal discriminatory experiments. In Chapter 4 an overview is given of the most important design criteria that were described in literature to design such experiments. One of these design criteria will be further improved and will be called the anticipatory approach to design optimal discriminatory experiments. The performance of this approach will be studied and compared to the performance of three existing approaches by applying them to a case study. Several design strategies are possible when the number of rival models is larger than two. Some design strategies are formalized and evaluated in Chapter 5 after applying them to the working example discussed in Chapter 3.

The third part consists of three chapters and deals with the integration of optimal experimental design for parameter estimation and model discrimination. In Chapter 6, the theory on optimal experimental design for parameter estimation is explained and its relation with model discrimination is briefly discussed. Chapter 7 is devoted to the design of compromise experiments. For this purpose, two methods are proposed, the kernel-based method and the ideal point method, and the performance to design such compromise experiments is evaluated after applying them to three case studies. The benefit of using such a compromise experiment is investigated in Chapter 8, by comparing its performance to the classical sequential procedure in which no compromise experiment is used. In addition, a simultaneous procedure is presented in this chapter, where model discrimination and optimal experimental design for parameter estimation are combined in a joint design criterion.

Finally, the conclusions drawn from the results and some suggestions for future research are formulated in Chapter 9. To structure the conclusions section, the research questions formulated above will be used and an answer will be provided for each of them.

# PART I

## MATHEMATICAL MODELLING AND MODEL DISCRIMINATION

# CHAPTER 2

## Introduction to mathematical modelling and model discrimination

*"Essentially, all models are wrong, but some are useful."*

George Box, statistician

## Abstract

In this chapter, some basic modelling concepts such as parameter estimation, uncertainty on the parameter estimates and on the model predictions are described. These concepts will reappear regularly in the following chapters and a good understanding of these concepts is of crucial importance to understand what follows. In addition, a general procedure for model discrimination is described which consists of four steps that are performed in an iterative manner until the best model is identified, all models appear to be inadequate and new models thus have to be proposed, or when discrimination among the remaining model candidates is no longer possible. Further, some approaches to evaluate the adequacy of a model are discussed and the difference between model discrimination, model evaluation, model selection and model validation is explained. Finally, the optimization algorithm used in this work (for parameter estimation and experimental design) is briefly described.

## 2.1 Introduction

In this introductory chapter, some basic modelling concepts are described, which will reappear regularly in the following chapters and a good understanding of these concepts is crucial to understand what follows. In addition, this chapter will introduce some notations that will be used in the subsequent chapters. Further, a general procedure to discriminate among rival models is presented.

## 2.2 Mathematical model representation

In what follows, general deterministic models in the form of a set of (possibly mixed) differential and algebraic equations are considered, using the following notations:

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{f}\left(\boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{\theta}, t\right) ; \quad \boldsymbol{x}\left(t_0\right) = \boldsymbol{x}_0 , \tag{2.1}$$

$$\hat{\boldsymbol{y}} = \boldsymbol{g}\left(\boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{\theta}, t\right) , \tag{2.2}$$

where $\boldsymbol{x}$ represents an $n_s$-dimensional vector of time-dependent state variables, $\boldsymbol{\theta}$ represents an $n_p$-dimensional vector of model parameters taken from a continuous, realizable set $\boldsymbol{\Theta}$, and $\hat{\boldsymbol{y}}$ represents an $n_m$-dimensional vector of measured response variables that are function of the state variables, $\boldsymbol{x}$. An experiment is denoted as $\boldsymbol{\xi}$ and is determined by the experimental degrees of freedom, such as sampling times, initial conditions and time-varying or constant process inputs. Quite often, $\boldsymbol{g}$ simply acts as a selector, selecting those state variables that are actually measured. Note that this will also be the case for the models used in this dissertation.

As explained in the introductory chapter, mathematical models are used as a surrogate for the actual process and are used to predict the behavior of the studied process under specified conditions. For instance, the outcome of an experiment can be predicted using a mathematical model, and the results of this *virtual experiment* can then be used to evaluate the information content of the experiment. Generally, the term *model simulation* is used to refer to the act of solving a model (defined through Eqs. (2.1) and (2.2)) for given values of the model parameters, initial values for the different state variables and well-defined process inputs. Note that in many cases, it is not possible to solve the mathematical model analytically and numerical techniques (or solvers) are required (Vanrolleghem and Dochain, 1998).

## 2.3 Parameter estimation

The values of the model parameters, which by definition do not change during the course of the simulation, have to be determined from experimental data. This process is called parameter estimation, and typically consists of minimizing the weighted sum of squared errors (WSSE) functional by optimal choice of the parameters $\boldsymbol{\theta}$. The optimal parameter estimates, denoted as $\hat{\boldsymbol{\theta}}$, are thus given by

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \text{WSSE}(\boldsymbol{\theta}) \ . \tag{2.3}$$

Here, $\text{WSSE}(\boldsymbol{\theta})$ is calculated as

$$\text{WSSE}(\boldsymbol{\theta}) = \sum_{k=1}^{n_e} \sum_{l=1}^{n_{sp_k}} \Delta\hat{\boldsymbol{y}}(\boldsymbol{\xi}_k, \boldsymbol{\theta}, t_l)' \cdot \mathbf{Q} \cdot \Delta\hat{\boldsymbol{y}}(\boldsymbol{\xi}_k, \boldsymbol{\theta}, t_l) \tag{2.4}$$

and

$$\Delta\hat{\boldsymbol{y}}(\boldsymbol{\xi}_k, \boldsymbol{\theta}, t_l) = \boldsymbol{y}(\boldsymbol{\xi}_k, t_l) - \hat{\boldsymbol{y}}(\boldsymbol{\xi}_k, \boldsymbol{\theta}, t_l) \tag{2.5}$$

represents the difference between the vector of the $n_m$ measured response variables and the model predictions at time $t_l$ $(l = 1, \ldots, n_{sp_k})$ of experiment $\boldsymbol{\xi}_k$ $(k = 1, \ldots, n_e)$. Further, $n_e$ represents the number of experiments from which data are used for estimating the parameters, $n_{sp_k}$ represents the number of samples in experiment $\boldsymbol{\xi}_k$, and $\mathbf{Q}$ is an $n_m$-dimensional matrix of user-supplied weighting coefficients. Typically, $\mathbf{Q}$ is chosen as the inverse of the measurement error covariance matrix $\boldsymbol{\Sigma}$ (Marsili–Libelli et al., 2003; Omlin and Reichert, 1999; Vanrolleghem and Dochain, 1998). In this way, the measurement uncertainty is incorporated in the WSSE, and the resulting equation is given as

$$\text{WSSE}(\boldsymbol{\theta}) = \sum_{k=1}^{n_e} \sum_{l=1}^{n_{sp_k}} \Delta\hat{\boldsymbol{y}}(\boldsymbol{\xi}_k, \boldsymbol{\theta}, t_l)' \cdot \boldsymbol{\Sigma}(\boldsymbol{\xi}_k, t_l)^{-1} \cdot \Delta\hat{\boldsymbol{y}}(\boldsymbol{\xi}_k, \boldsymbol{\theta}, t_l) \ , \tag{2.6}$$

where $\boldsymbol{\Sigma}(\boldsymbol{\xi}_k, t_l)$ represents the measurement error covariance matrix at sampling time $t_l$ of experiment $\boldsymbol{\xi}_k$.

## 2.4 Sensitivities of the state variables to the values of the model parameters

The model predictions are directly determined by the values of the model parameters. In this respect, it is often interesting to calculate the so-called parameter sensitivities, which indicate how sensitive the model predictions are to a change in the values of the parameters. A response variable is called sensitive to a certain parameter when a small change in the value of that parameter results in a significant change of the response variable, and vice versa.

For a model with $n_m$ measured response variables ($\hat{\boldsymbol{y}}$) and $n_p$ parameters ($\boldsymbol{\theta}$), one has to calculate $n_m \times n_p$ parameter sensitivities. Usually, these parameter sensitivities are collected in an $(n_m \times n_p)$-dimensional matrix, given by

$$\left.\frac{\partial \hat{\boldsymbol{y}}\left(\boldsymbol{\xi},\boldsymbol{\theta},t_l\right)}{\partial \boldsymbol{\theta}}\right|_{\hat{\boldsymbol{\theta}}} = \begin{bmatrix} \left.\frac{\partial \hat{y}_1(\boldsymbol{x},\boldsymbol{\xi},\boldsymbol{\theta},t_l)}{\partial \theta_1}\right|_{\hat{\boldsymbol{\theta}}} & \left.\frac{\partial \hat{y}_1(\boldsymbol{x},\boldsymbol{\xi},\boldsymbol{\theta},t_l)}{\partial \theta_2}\right|_{\hat{\boldsymbol{\theta}}} & \cdots & \left.\frac{\partial \hat{y}_1(\boldsymbol{x},\boldsymbol{\xi},\boldsymbol{\theta},t_l)}{\partial \theta_{n_p}}\right|_{\hat{\boldsymbol{\theta}}} \\ \cdots & \cdots & \cdots & \cdots \\ \left.\frac{\partial \hat{y}_{n_m s}(\boldsymbol{x},\boldsymbol{\xi},\boldsymbol{\theta},t_l)}{\partial \theta_1}\right|_{\hat{\boldsymbol{\theta}}} & \left.\frac{\partial \hat{y}_{n_m}(\boldsymbol{x},\boldsymbol{\xi},\boldsymbol{\theta},t_l)}{\partial \theta_2}\right|_{\hat{\boldsymbol{\theta}}} & \cdots & \left.\frac{\partial \hat{y}_{n_m}(\boldsymbol{x},\boldsymbol{\xi},\boldsymbol{\theta},t_l)}{\partial \theta_{n_p}}\right|_{\hat{\boldsymbol{\theta}}} \end{bmatrix}. \tag{2.7}$$

Here, the sensitivity of the predicted response variable $\hat{y}_i$ ($i \in \{1,\ldots,n_m\}$) to a change in parameter $\theta_j$ ($j \in \{1,\ldots,n_p\}$) is denoted as

$$\left.\frac{\partial \hat{y}_i\left(\boldsymbol{\xi},\boldsymbol{\theta},t_l\right)}{\partial \theta_j}\right|_{\hat{\boldsymbol{\theta}}}, \tag{2.8}$$

where $\hat{\boldsymbol{\theta}}$ represents the $n_p$-dimensional vector containing the available estimates of the model parameters and $t_l$ represents the time at which the parameter sensitivity is calculated. From Eq. (2.2) one can see that the predicted response variables $\hat{\boldsymbol{y}}$ are a function of the state variables $\boldsymbol{x}$. Therefore, the effect of a change in parameter $\theta_j$ on the value of response variable $\hat{y}_i$ at time $t_l$ can be calculated after applying the chain rule to determine the total differential of $\boldsymbol{g}\left(\boldsymbol{x},\boldsymbol{\xi},\boldsymbol{\theta},t\right)$. This results in the following equation:

$$\left.\frac{\partial \hat{y}_i\left(\boldsymbol{\xi},\boldsymbol{\theta},t_l\right)}{\partial \theta_j}\right|_{\hat{\boldsymbol{\theta}}} = \left.\frac{\partial g_i\left(\boldsymbol{x},\boldsymbol{\xi},\boldsymbol{\theta},t\right)}{\partial \theta_j}\right|_{\hat{\boldsymbol{\theta}}} + \left.\frac{\partial g_i\left(\boldsymbol{x},\boldsymbol{\xi},\boldsymbol{\theta},t\right)}{\partial \boldsymbol{x}}\right|_{\hat{\boldsymbol{\theta}}} \cdot \left.\frac{\partial \boldsymbol{x}\left(\boldsymbol{x},\boldsymbol{\xi},\boldsymbol{\theta},t\right)}{\partial \theta_j}\right|_{\hat{\boldsymbol{\theta}}}. \tag{2.9}$$

Here, $\boldsymbol{x}$ represents the $n_s$-dimensional vector of state variables and Eq. (2.9) can also be written with the explicit summation over all $n_s$ state variables, which results in the following equation:

$$\left.\frac{\partial \hat{y}_i\left(\boldsymbol{\xi},\boldsymbol{\theta},t_l\right)}{\partial \theta_j}\right|_{\hat{\boldsymbol{\theta}}} = \left.\frac{\partial g_i\left(\boldsymbol{x},\boldsymbol{\xi},\boldsymbol{\theta},t\right)}{\partial \theta_j}\right|_{\hat{\boldsymbol{\theta}}} + \sum_{k=1}^{n_s} \left.\frac{\partial g_i\left(\boldsymbol{x},\boldsymbol{\xi},\boldsymbol{\theta},t\right)}{\partial x_k}\right|_{\hat{\boldsymbol{\theta}}} \cdot \left.\frac{\partial x_k\left(\boldsymbol{x},\boldsymbol{\xi},\boldsymbol{\theta},t\right)}{\partial \theta_j}\right|_{\hat{\boldsymbol{\theta}}} . \tag{2.10}$$

Equation (2.10) clearly shows that a change in the value of a particular parameter $\theta_j$ has a direct effect on the value the response variable (first term of Eq. (2.10)), and an indirect effect (second term of Eq. (2.10)). The latter represents the change in $\hat{y}_i$ caused by the fact that also the values of the state variables will be influenced by the change in the parameter value, which on their turn change the value of $\hat{y}_i$. The equations required to calculate $\left.\frac{\partial x_k(\boldsymbol{x},\boldsymbol{\xi},\boldsymbol{\theta},t)}{\partial \theta_j}\right|_{\hat{\boldsymbol{\theta}}}$ are derived in what follows.

As shown in Eq. (2.1), the change of the state variables over time $\left(\frac{d\boldsymbol{x}}{dt}\right)$ is a function of the state variables themselves, the experimental degrees of freedom $\boldsymbol{\xi}$, the model parameters $\boldsymbol{\theta}$ and time $t$. For dynamic models, the effect of a changed parameter value on the value of state variable $x_i$ will propagate over time, which results in the following equation (Ternbach et al., 2005):

$$\left.\frac{\partial x_i\left(\boldsymbol{x},\boldsymbol{\xi},\boldsymbol{\theta},t_l\right)}{\partial \theta_j}\right|_{\hat{\boldsymbol{\theta}}} = \left.\frac{\partial x_i\left(\boldsymbol{x},\boldsymbol{\xi},\boldsymbol{\theta},0\right)}{\partial \theta_j}\right|_{\hat{\boldsymbol{\theta}}} + \int_{t=0}^{t_l} \frac{d}{dt}\left(\left.\frac{\partial x_i\left(\boldsymbol{x},\boldsymbol{\xi},\boldsymbol{\theta},t\right)}{\partial \theta_j}\right|_{\hat{\boldsymbol{\theta}}}\right) \cdot dt\,. \tag{2.11}$$

Here, $x_i\left(\boldsymbol{x},\boldsymbol{\xi},\boldsymbol{\theta},0\right)$ represents the initial value of state variable $x_i$. When parameter $\theta_j$ is different from $x_i\left(\boldsymbol{x},\boldsymbol{\xi},\boldsymbol{\theta},0\right)$, which is generally the case, the first term of Eq. (2.11) is equal to zero and can thus be omitted. Sometimes, however, the initial value(s) of one or some of the state variables of the model has to be estimated from experimental data. If this is the case, $x_i\left(\boldsymbol{x},\boldsymbol{\xi},\boldsymbol{\theta},0\right)$ can be considered as a model parameter (say $\theta_j$), and then the first term of Eq. (2.11) is equal to one (Leis and Kramer, 1988). In the following, however, it will be assumed that the first term of Eq. (2.11) can be omitted because it would unnecessarily complicate the derivation.

To solve Eq. (2.11), one can apply Scharwz' theorem (Ellwein et al., 2008; Turányi and Rabitz, 2000), which states that

$$\frac{d}{dt}\left(\left.\frac{\partial x_i\left(\boldsymbol{x},\boldsymbol{\xi},\boldsymbol{\theta},t\right)}{\partial \theta_j}\right|_{\hat{\boldsymbol{\theta}}}\right) = \left.\frac{\partial}{\partial \theta_j}\left(\frac{dx_i\left(\boldsymbol{x},\boldsymbol{\xi},\boldsymbol{\theta},t\right)}{dt}\right)\right|_{\hat{\boldsymbol{\theta}}} . \tag{2.12}$$

As seen from Eq. (2.1), the value of $x_i$ is not only directly dependent on the parameters, but also on the values of the (other) state variables $\boldsymbol{x}$, which on their turn depend on the parameters. So, a change in parameter $\theta_j$ will directly affect the value of state variable $x_i$ through its corresponding differential equation, but it will also have an indirect effect because the change in the parameter will also affect the other state variables. Therefore, the chain rule has to be applied to calculate the total differential of $f_i(\boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{\theta}, t)$. After substituting Eq. (2.12) into Eq. (2.11) and applying the chain rule, the following equation is found

$$\left.\frac{\partial x_i(\boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{\theta}, t_l)}{\partial \theta_j}\right|_{\hat{\boldsymbol{\theta}}} = \int_{t=0}^{t_l} \left( \left.\frac{\partial f_i(\boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{\theta}, t)}{\partial \theta_j}\right|_{\hat{\boldsymbol{\theta}}} + \left.\frac{\partial f_i(\boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{\theta}, t)}{\partial \boldsymbol{x}}\right|_{\hat{\boldsymbol{\theta}}} \cdot \left.\frac{\partial \boldsymbol{x}(\boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{\theta}, t)}{\partial \theta_j}\right|_{\hat{\boldsymbol{\theta}}} \right) \cdot dt\,. \tag{2.13}$$

The first term of Eq. (2.13) can be seen as the direct effect of a change in parameter $\theta_j$, whereas the second term reflects the indirect effect caused by the changes of the other state variables. As in Eq. (2.9), $\boldsymbol{x}$ represents the $n_s$-dimensional vector that contains the state variables, and Eq. (2.13) can therefore be written with the explicit summation over all $n_s$ state variables. This results in the following equation:

$$\left.\frac{\partial x_i(\boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{\theta}, t_l)}{\partial \theta_j}\right|_{\hat{\boldsymbol{\theta}}} = \int_{t=0}^{t_l} \left( \left.\frac{\partial f_i(\boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{\theta}, t)}{\partial \theta_j}\right|_{\hat{\boldsymbol{\theta}}} + \sum_{k=1}^{n_s} \left.\frac{\partial f_i(\boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{\theta}, t)}{\partial x_k}\right|_{\hat{\boldsymbol{\theta}}} \cdot \left.\frac{\partial x_k(\boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{\theta}, t)}{\partial \theta_j}\right|_{\hat{\boldsymbol{\theta}}} \right) \cdot dt\,, \tag{2.14}$$

which can be rewritten as

$$\frac{d}{dt}\left( \left.\frac{\partial x_i(\boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{\theta}, t)}{\partial \theta_j}\right|_{\hat{\boldsymbol{\theta}}} \right) = \left.\frac{\partial f_i(\boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{\theta}, t)}{\partial \theta_j}\right|_{\hat{\boldsymbol{\theta}}} + \sum_{k=1}^{n_s} \left.\frac{\partial f_i(\boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{\theta}, t)}{\partial x_k}\right|_{\hat{\boldsymbol{\theta}}} \cdot \left.\frac{\partial x_k(\boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{\theta}, t)}{\partial \theta_j}\right|_{\hat{\boldsymbol{\theta}}}\,. \tag{2.15}$$

From Eq. (2.15), one can see that the differential equations that describe the parameter sensitivities are coupled. So, to calculate the parameter sensitivities of the response variables to the parameters (Eq. (2.7)), $n_s \times n_p$ additional ordinary differential equations have to be defined and solved together with the actual model, which is given by Eqs. (2.1) and (2.2) (Atherton et al., 1975; Leis and Kramer, 1988; Li et al., 2000; Munack, 1989). This system of coupled differential equations can also be written in matrix notation, as follows:

$$\frac{d}{dt}\left( \left.\frac{\partial \boldsymbol{x}(\boldsymbol{\xi}, \boldsymbol{\theta}, t_l)}{\partial \boldsymbol{\theta}}\right|_{\hat{\boldsymbol{\theta}}} \right) = \left.\frac{\partial \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{\theta}, t)}{\partial \boldsymbol{\theta}}\right|_{\hat{\boldsymbol{\theta}}} + \left.\frac{\partial \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{\theta}, t)}{\partial \boldsymbol{x}}\right|_{\hat{\boldsymbol{\theta}}} \cdot \left.\frac{\partial \boldsymbol{x}(\boldsymbol{\xi}, \boldsymbol{\theta}, t_l)}{\partial \boldsymbol{\theta}}\right|_{\hat{\boldsymbol{\theta}}} \tag{2.16}$$

and

$$\left.\frac{\partial \hat{\boldsymbol{y}}\left(\boldsymbol{\xi},\boldsymbol{\theta},t_l\right)}{\partial \boldsymbol{\theta}}\right|_{\hat{\boldsymbol{\theta}}} = \left.\frac{\partial \boldsymbol{g}\left(\boldsymbol{x},\boldsymbol{\xi},\boldsymbol{\theta},t\right)}{\partial \boldsymbol{\theta}}\right|_{\hat{\boldsymbol{\theta}}} + \left.\frac{\partial \boldsymbol{g}\left(\boldsymbol{x},\boldsymbol{\xi},\boldsymbol{\theta},t\right)}{\partial \boldsymbol{x}}\right|_{\hat{\boldsymbol{\theta}}} \cdot \left.\frac{\partial \boldsymbol{x}\left(\boldsymbol{\xi},\boldsymbol{\theta},t_l\right)}{\partial \boldsymbol{\theta}}\right|_{\hat{\boldsymbol{\theta}}} . \tag{2.17}$$

Note that, as already stated in Section 2.2, it is often the case that some of the measured response variables $\hat{\boldsymbol{y}}$ are equal to one of the state variables $\boldsymbol{x}$. It is obvious that the parameter sensitivities of these response variables equal those of the corresponding state variables.

Note also that the equations used to calculate the parameter sensitivities (that is, Eqs. (2.16) and (2.17)) may become very complicated when one is working with large and complex models. In such cases, it may not be practically feasible to calculate the parameter sensitivities directly from these equations and they have to be approximated using numerical techniques. A nice overview of the existing methods is given in De Pauw and Vanrolleghem (2006b). In what follows, only the simplest approach, usually called finite difference approximation, is briefly discussed. In this approach, the sensitivity of response variable $\hat{y}_i$ to a change in parameter $\theta_j$ is approximated as

$$\left.\frac{\partial \hat{y}_i\left(\boldsymbol{\xi},\boldsymbol{\theta},t_l\right)}{\partial \theta_j}\right|_{\hat{\boldsymbol{\theta}}} = \lim_{\Delta\theta_j \to 0} \frac{\hat{y}_i\left(\boldsymbol{\xi},\boldsymbol{\theta}+\Delta\theta_j,t_l\right) - \hat{y}_i\left(\boldsymbol{\xi},\boldsymbol{\theta},t_l\right)}{\Delta\theta_j} , \tag{2.18}$$

where $\hat{y}_i\left(\boldsymbol{\xi},\boldsymbol{\theta}+\Delta\theta_j,t_l\right)$ represents the value of $\hat{y}_i$ obtained when $\Delta\theta_j$ is added to the value of parameter $\theta_j$. It is clear that Eq. (2.18) is only valid when $\Delta\theta_j$ is infinitesimally small (and thus approaches zero). However, in practice, it is not possible to choose such a small value for $\Delta\theta_j$ because this would result in numerical inaccuracies. On the other hand, $\Delta\theta_j$ should not become too large either, because then the non-linearity of the model will start to play an important role in the calculation of the parameter sensitivities. Despite the fact that the choice of a good value for $\Delta\theta_j$ is not straightforward, this approach is frequently applied because of its simplicity. In this dissertation, however, the parameter sensitivities will be calculated from Eqs. (2.16) and (2.17).

## 2.5 Uncertainty on the parameter estimates

Since the parameters are estimated from noise-corrupted experimental data, the resulting parameter estimates will be uncertain to some extent. The parameter estimates by themselves thus only become meaningful when they are accompanied by their so-called

confidence region. This is the region in which their true values are expected to lie with a given level of confidence. According to Seber and Wild (1989), it is justifiable to base the confidence region of the parameter estimates on the contours of the WSSE functional, since the latter gives an indication of how well the experimental data is described by the model. An exact confidence region can be determined from the following equation (Marsili–Libelli et al., 2003; Seber and Wild, 1989)

$$\left\{\boldsymbol{\theta} : \mathrm{WSSE}\left(\boldsymbol{\theta}\right) \leq c \cdot \mathrm{WSSE}\left(\hat{\boldsymbol{\theta}}\right)\right\} \tag{2.19}$$

for any $c > 1$. This region can be regarded as *exact* in the sense that it is not based on any approximation. However, when the model is nonlinear, it is difficult to select a value for $c$ with any statistical significance. Rather than trying to find this exact confidence region, one most often uses linear approximations to construct the confidence region. This approach was also adopted in this dissertation and is explained below.

The parameter estimation error covariance matrix, denoted as $\boldsymbol{\Phi}$, is used to represent the uncertainty on the parameter estimates. Obviously, the quality of these parameter estimates is determined by the information content of the experimental data from which they are determined. The latter can be quantified by means of the so-called Fisher information matrix (**FIM**). According to the Cramér–Rao inequality (Ljung, 1999), the inverse of the **FIM** is equal to the lower bound of the parameter estimation error covariance matrix if the measurement errors are independent samples from a normal distribution with zero mean (which is assumed in what follows).

The inverse of the **FIM** is therefore often used as an approximation of the parameter estimation error covariance matrix (Asprey and Macchietto, 2000; Atkinson and Donev, 1992; Goodwin and Payne, 1977; Ljung, 1999; Mehra, 1974; Munack, 1991; Shirt et al., 1994; Vanrolleghem and Dochain, 1998; Walter and Pronzato, 1997). The **FIM** is calculated as

$$\mathbf{FIM} = \sum_{k=1}^{n_e} \mathbf{FIM}\left(\boldsymbol{\xi}_k\right) , \tag{2.20}$$

where

$$\mathbf{FIM}\left(\boldsymbol{\xi}_k\right) = \sum_{l=1}^{n_{sp_k}} \left( \left.\frac{\partial \hat{\boldsymbol{y}}\left(\boldsymbol{\xi}_k, \boldsymbol{\theta}, t_l\right)}{\partial \boldsymbol{\theta}}\right|_{\hat{\boldsymbol{\theta}}} \right)' \cdot \boldsymbol{\Sigma}\left(\boldsymbol{\xi}_k, t_l\right)^{-1} \cdot \left( \left.\frac{\partial \hat{\boldsymbol{y}}\left(\boldsymbol{\xi}_k, \boldsymbol{\theta}, t_l\right)}{\partial \boldsymbol{\theta}}\right|_{\hat{\boldsymbol{\theta}}} \right) . \tag{2.21}$$

Here, $\partial\hat{\boldsymbol{y}}/\partial\boldsymbol{\theta}$ represents the $n_m \times n_p$-dimensional parameter sensitivity matrix (Eq. (2.7)) associated with measurement time $t_l$ of experiment $\boldsymbol{\xi}_k$, and is calculated as explained in Section 2.4. For a derivation of Eq. (2.21), the reader is referred to Chapter 6 where the calculation of the **FIM** is discussed in more detail.

A closer look at Eq. (2.21) shows that the **FIM** is composed of two components, the parameter sensitivities ($\partial\hat{\boldsymbol{y}}/\partial\boldsymbol{\theta}$) and the measurement error covariance matrix ($\boldsymbol{\Sigma}$). The sensitivity of a certain state variable with respect to a parameter expresses how much that state variable will change when this parameter is slightly perturbed. A state variable that is highly sensitive to a certain parameter will therefore contain a lot of information about this parameter, while a variable that is insensitive to the parameter does not contribute to the information content for that parameter. The role of the measurement error covariance matrix in the calculation of the **FIM** is rather straightforward, since it is obvious that a measurement associated with a large measurement error will contribute less to the information content than a measurement with a small measurement error.

The diagonal elements of the parameter estimation error covariance matrix are the variances of the errors of the parameter estimates and the off-diagonal elements are the covariances between the parameter estimation errors. Based on these variances and covariances the elements of the linear correlation matrix, denoted as $\rho_\Phi$, can be calculated as (De Pauw, 2005; Marsili–Libelli et al., 2003; Walter and Pronzato, 1997):

$$\rho_\Phi(i,j) = \frac{\boldsymbol{\Phi}(i,j)}{\sqrt{\boldsymbol{\Phi}(i,i)\cdot\boldsymbol{\Phi}(j,j)}}\,. \tag{2.22}$$

These elements give a measure for the linear correlation between two parameters and can range between -1 and 1. If the linear correlation coefficient is close to -1 or 1 the parameters are said to be highly correlated, while correlation coefficients close to zero imply low correlations.

From the parameter estimation error covariance matrix, the $100\cdot(1-\alpha)$ percent confidence interval associated with parameter estimate $i$ can be calculated as (Marsili–Libelli et al., 2003; Walter and Pronzato, 1997)

$$\sqrt{\boldsymbol{\Phi}(i,i)}\cdot t_{n-n_p}^{\alpha/2}\,, \tag{2.23}$$

where $\boldsymbol{\Phi}(i,i)$ represents the $(i,i)$-th element of $\boldsymbol{\Phi}$, $n$ represents the total number of data points, $n_p$ represents the number of parameters that were estimated from them, and $t_{n-n_p}^{\alpha/2}$

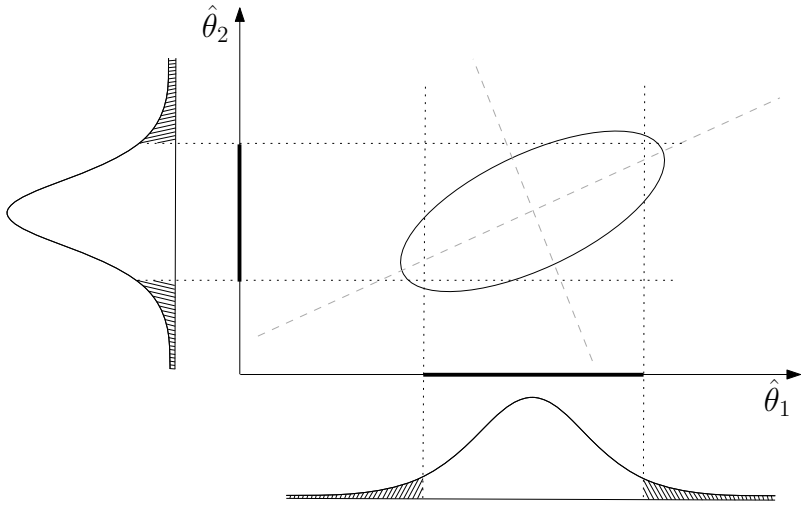



**Figure 2.1:** The shaded ellipse represents the joint 95% confidence region for parameter estimates $\hat{\theta}_1$ and $\hat{\theta}_2$, whereas the confidence intervals of the individual parameter estimates are indicated by the thick lines.

represents the upper $\alpha/2$ quantile of Student's $t$ distribution for the given confidence level $\alpha$ and $n - n_p$ degrees of freedom.

Note that only the diagonal elements of the parameter estimation error covariance matrix are considered in the calculation of the confidence intervals, and the covariance between the parameter estimates is thus ignored. This is illustrated in Fig. 2.1, where the joint confidence region is shown, as well as the confidence intervals calculated from Eq. (2.23). The latter only consider the uncertainty on the corresponding parameter estimate and do not take the correlation or covariance with other parameter estimates into account.

## 2.6 Model prediction uncertainty

The uncertainty on the parameter estimates will propagate when simulating the model, and the model predictions will consequently be uncertain as well. Also for model predictions, a covariance matrix is used to quantify the uncertainty. The model prediction error covariance matrix associated with time $t_l$ of experiment $\boldsymbol{\xi}_k$, denoted as $\boldsymbol{\Omega}(\boldsymbol{\xi}_k, t_l)$, is calculated by propagating the uncertainty on the parameter estimates, denoted as $\boldsymbol{\Phi}$, according to (Omlin and Reichert, 1999)

$$\boldsymbol{\Omega}(\boldsymbol{\xi}_k, t_l) = \left( \frac{\partial \hat{\boldsymbol{y}}(\boldsymbol{\xi}_k, \boldsymbol{\theta}, t_l)}{\partial \boldsymbol{\theta}} \bigg|_{\hat{\boldsymbol{\theta}}} \right) \cdot \boldsymbol{\Phi} \cdot \left( \frac{\partial \hat{\boldsymbol{y}}(\boldsymbol{\xi}_k, \boldsymbol{\theta}, t_l)}{\partial \boldsymbol{\theta}} \bigg|_{\hat{\boldsymbol{\theta}}} \right)' . \tag{2.24}$$

The correlation matrix and the confidence intervals associated with the model predictions are calculated in a similar way as those of the parameter estimates (Eqs. (2.22) and (2.23), respectively). The correlation matrix, denoted as $\rho_\Omega$, is calculated as (Omlin and Reichert, 1999; Walter and Pronzato, 1997; Seber and Wild, 1989)

$$\rho_\Omega(i,j) = \frac{\mathbf{\Omega}(i,j)}{\sqrt{\mathbf{\Omega}(i,i) \cdot \mathbf{\Omega}(j,j)}} \,. \tag{2.25}$$

whereas the $100 \cdot (1-\alpha)$ percent confidence interval associated with the model prediction of response variable $\hat{y}_i$ can be calculated from the corresponding element in the covariance matrix

$$\sqrt{\mathbf{\Omega}(i,i)} \cdot t_{n-n_p}^{\alpha/2} \,, \tag{2.26}$$

where $n$ represents the total number of data points used for parameter estimation, $n_p$ represents the number of parameters that were estimated, and $t_{n-n_p}^{\alpha/2}$ represents the upper $\alpha/2$ quantile of Student's $t$ distribution for the given confidence level $\alpha$ and $n-n_p$ degrees of freedom.

## 2.7 Model discrimination

As described earlier, this dissertation mainly deals with the problem of model discrimination. Here, the term model discrimination is defined as the procedure in which the most appropriate model has to be identified from a set of rival models. Model discrimination differs from model selection, which will be discussed in Section 2.9, by the fact that additional experiments are designed and performed during the model discrimination procedure. In model selection, on the other hand, the best performing model is selected from a set of adequate models based on the available data and the characteristics of the model structures (for instance, the number of parameters). Note that model discrimination has also been called model structure characterization to emphasize that the aim is to identify the most appropriate model structure (Spriet, 1985; Vanrolleghem and Dochain, 1998), but in the following the term model discrimination will be used because it is more common.

A general procedure to discriminate among $m$ rival models is depicted in Fig. 2.2. It basically consists of four steps that are performed in an iterative manner until a stopping criterion is met. In the first step, the parameters of the rival models are estimated from

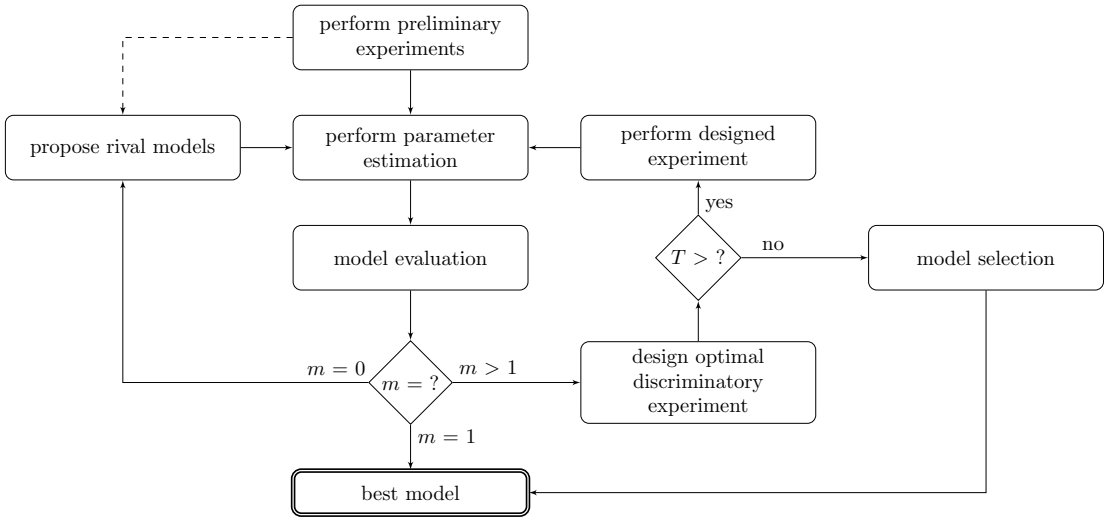



**Figure 2.2:** General procedure to discriminate among $m$ rival models (adapted from Chen and Asprey (2003) and Schwaab et al. (2006)).

all the data that is available by minimizing the WSSE-functional described in Section 2.3. A second step involves an evaluation of the model's adequacy to describe the available data and is thus performed in order to find out which models are able to describe the data in a reasonable manner and which ones do not. Models that pass this test are used in a third step, where an optimal discriminatory experiment is designed using one of the design criteria that will be described in Chapter 4. This experiment is then performed in a fourth and last step, after which the loop is closed by re-estimating the parameters of all rival models using all data available at that time. By reconsidering all models in the model evaluation step, one accounts for the possibility that one (or more) models were wrongly rejected by chance in a previous iteration (Buzzi-Ferraris et al., 1990).

The iterative procedure described above continues until the best model is identified ($m = 1$), all models appear to be inadequate and new models thus have to be proposed ($m = 0$), or when discrimination among the remaining model candidates is no longer possible. The latter is indicated by the fact that $T$ is smaller than a predefined cut-off value, where $T$ represents the discriminatory potential of the designed experiment. In the latter case, the best model is selected from the remaining rival models as a trade-off between model fit, model parsimony and identifiability (Spriet, 1985), as will be briefly discussed in Section 2.9.

## 2.8 Model evaluation

From the discussion above, it is clear that the model evaluation step is a crucial step in the model discrimination procedure (Dochain and Vanrolleghem, 2001). Model evaluation is the step in which the model predictions of the response variables are confronted with the measured data, and the adequacy of the rival models is questioned. According to Vanrolleghem and Dochain (1998), the methods that have been developed for this purpose can be classified in two categories: so-called *a priori* model evaluation methods and *a posteriori* model evaluation methods. The former comprises those methods that are capable of determining whether a model is adequate or not without the need to estimate the model parameters from the available data, while parameter estimation is required for the latter group of methods. In this dissertation, only the methods for *a posteriori* structure characterization are considered because successful applications of the *a priori* model evaluation methods are limited and parameter estimation is required anyhow when the model discrimination procedure described in the previous section is applied.

The model evaluation step deals with the model structure itself and one is interested in its ability to represent the available data or, put differently, in its inability to capture the dynamics of the studied process. Since the model parameters are estimated prior to the model evaluation step, the observed model fit is the best one possible given that particular model structure. If the model structure appears to be able to describe the available data, the model is considered adequate. If not, it is labeled inadequate and the model is not considered for the experimental design step. This rationale is illustrated in Fig. 2.3, which explains the difference between accuracy and precision using the analogy of a target practice. If the gray dots represent the measurements and if the center of the bullseye represents the model prediction, one could state that the adequate models are those that correspond to cases 1 and 2.

Note that the information content of the data used to estimate the model parameters will determine the quality of the parameter estimates, and thus the uncertainty on the model predictions. However, although the latter is important and is one of the required characteristics of what is considered a good model, it is not taken into account here.

To evaluate the adequacy of a model, several approaches are possible. They can roughly be categorized in two groups: model adequacy tests (statistical tests developed with the purpose of detecting lack-of-fit) and so-called model evaluation criteria (measures developed

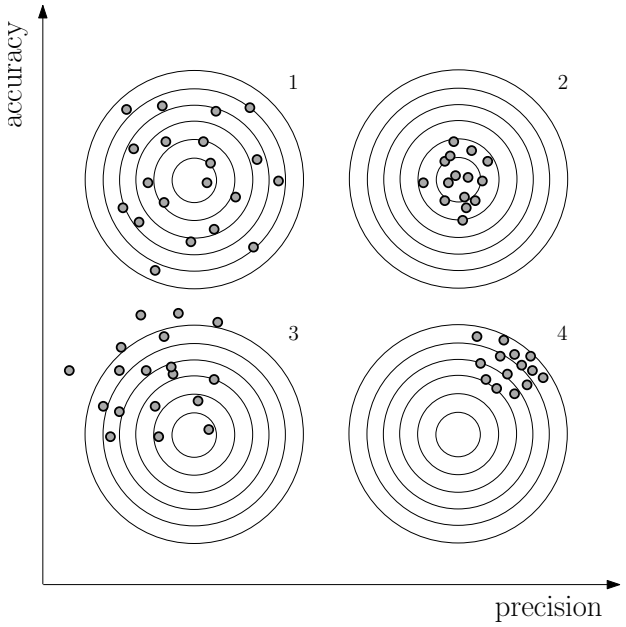



**Figure 2.3:** Illustration of the notions of accuracy versus precision using the analogy of a target practice. Case 1 is accurate and imprecise, case 2 is accurate and precise, case 3 is inaccurate and imprecise, and case 4 is inaccurate and precise. Note that the figure was adopted from Tedeschi (2006).

to evaluate the quality of the model fit). These approaches are respectively discussed in Section 2.8.1 and Section 2.8.2.

### 2.8.1 Model adequacy testing

After estimating the model parameters, the residual error between the measurements and the model predictions (WSSE) has two components: the model variance error and the model bias error. The model variance error is caused by the measurement errors that are associated with the experimental data from which the parameters were estimated, while the model bias error reflects the unmodeled dynamics of the process. This can be formalized by rewriting the WSSE as follows

$$\text{WSSE} = \text{WSSE}_{\boldsymbol{\xi}} + \text{WSSE}_{\text{LOF}} \,. \tag{2.27}$$

Here, $\text{WSSE}_{\boldsymbol{\xi}}$ represents the contribution due to measurement error (model variance error) and $\text{WSSE}_{\text{LOF}}$ represents the contribution due to lack-of-fit (model bias error). Since $\text{WSSE}_{\boldsymbol{\xi}}$ is known or estimated from data, the statistical tests described below evaluate

whether the contribution due to lack-of-fit is significant. If so, the model is rejected and thus considered inadequate.

The lack-of-fit test used in this dissertation is based on the property of the WSSE functional being a sample from a $\chi^2$-distribution with $n - n_p$ degrees of freedom (Buzzi-Ferraris et al., 1990; Chen and Asprey, 2003; de Brauwere et al., 2005; Franceschini and Macchietto, 2008). However, this property only holds under two assumptions (de Brauwere et al., 2005): (i) the measurements are disturbed with random zero mean normally distributed noise with known (or a priori estimated) variance, and are not subject to systematic errors; and (ii) no model errors are present.

In this work, data is generated by adding noise to the simulation results of which the characteristics are known, so the first assumption is always valid. Consequently, when the WSSE is significantly larger than the expected value of the appropriate $\chi^2_{n-n_p}$-distribution, one can conclude that the model is not able to describe the experimental data in a reasonable manner and the model can thus be rejected. Chen and Asprey (2003) recommend the use of this $\chi^2$-test for simulation studies where noise is added to simulated data.

The $\chi^2$-test described above is valid when the characteristics of the measurement errors are known. If this is not the case, a different statistical test has to be used because the variance of the measurements has to be estimated from the available data (Chen and Asprey, 2003; Dochain and Vanrolleghem, 2001; Franceschini and Macchietto, 2008). For this purpose, repeated measurements are required at each sampling time. When $n_r$ represents the number of repetitions, the variance of the measured response variable $y_i$ at sampling time $t_l$, denoted as $s_i^2(t_l)$, can be estimated as follows:

$$s_i^2(t_l) = \frac{\sum_{r=1}^{n_r} \left(y_i(t_{l,r}) - \overline{y}_i(t_l)\right)^2}{n_r - 1}, \tag{2.28}$$

where $\overline{y}_i(t_l)$ represents the mean of the repeated measurements at sampling time $t_l$. Using the estimated variances of each measured response variable, possible lack-of-fit can be detected using an F-test. Therefore, the following test statistic has to be compared with the reference value of an F-distribution with $(n - n_p - n_r)$ and $(n_r - 1)$ degrees of freedom:

$$F = \frac{\mathrm{WSSE}_{\mathrm{LOF}}}{\mathrm{WSSE}_{\boldsymbol{\xi}}} \sim F_{n-n_p-n_r,n_r-1}\,. \tag{2.29}$$

When the value of $F$ is larger than the reference value, the contribution of $\mathrm{WSSE}_{\mathrm{LOF}}$ is considered significant and the model is considered as inadequate.

### 2.8.2 Model evaluation criteria

The statistical tests described above in a way assume that a *true* model exists. However, in reality, the processes being studied are often very complex, and it is not realistic to assume that such a model can be found. Indeed, a mathematical model is an abstract representation of reality, and it can never mimic reality under all conditions. Nevertheless, mathematical models are often very useful tools, as discussed in the introduction (Section 1.1). Therefore, the use of a statistical test may not be a good option for practical applications, because it may reject these useful models.

As an alternative to the statistical tests described above, literature provides many criteria or measures that are developed to evaluate the model fit or the adequacy of a particular model. The WSSE is one of them and is often used for this purpose, but also the modelling efficiency and the index of agreement are promising ones (Janssen and Heuberger, 1995; Krause et al., 2005; Moriasi et al., 2007; Willmott et al., 1985). These two criteria are discussed in the following.

The modelling efficiency (MEF), which is also known as the Nash-Sutcliffe efficiency (Nash and Sutcliffe, 1970), is calculated as

$$\mathrm{MEF} = 1 - \frac{\sum_{k=1}^{n_e}\sum_{i=1}^{n_m}\sum_{l=1}^{n_{sp_k}}\left(y_i\left(\boldsymbol{\xi}_k, t_l\right) - \hat{y}_i\left(\boldsymbol{\xi}_k, \hat{\boldsymbol{\theta}}, t_l\right)\right)^2}{\sum_{k=1}^{n_e}\sum_{i=1}^{n_m}\sum_{l=1}^{n_{sp_k}}\left(y_i\left(\boldsymbol{\xi}_k, t_l\right) - \overline{y}_i\left(\boldsymbol{\xi}_k\right)\right)^2} \,. \tag{2.30}$$

Here, $y_i\left(\boldsymbol{\xi}_k, t_l\right)$ represents the measured response variable at sampling time $t_l$ of experiment $\boldsymbol{\xi}_k$, $\hat{y}_i\left(\boldsymbol{\xi}_k, \hat{\boldsymbol{\theta}}, t_l\right)$ represents the model prediction of that response variable, and $\overline{y}_i\left(\boldsymbol{\xi}_k, \hat{\boldsymbol{\theta}}\right)$ represents the mean of the response variables that is calculated as

$$\overline{y}_i\left(\boldsymbol{\xi}_k\right) = \frac{\sum_{l=1}^{n_{sp_k}} y_i\left(\boldsymbol{\xi}_k, t_l\right)}{n_{sp_k}} \,. \tag{2.31}$$

The denominator of Eq. (2.30), used to calculate the modelling efficiency, represents the variance of the measured response variables. The sum of the absolute differences between the measured and the predicted response variables is normalized by dividing it by this variance. The modelling efficiency ranges between $-\infty$ and 1, with MEF = 1 being the optimal value. A value of zero indicates that the model performs as good as the mean of the measured response variables, whereas values below zero indicate that the mean of the measured response variables is a better predictor than the model prediction. Therefore,

values between zero and one are generally viewed as acceptable levels of performance, and negative values indicate unacceptable performance (Krause et al., 2005; Legates and McCabe Jr., 1999; Moriasi et al., 2007; Nash and Sutcliffe, 1970).

The index of agreement, denoted as $d$, was developed by Willmott (1981) as a standardized measure of the degree of model prediction error and varies between 0 and 1. A value of 1 indicates a perfect agreement between the measured and predicted response variables, and 0 indicates no agreement at all (Willmott, 1981; Legates and McCabe Jr., 1999). The index of agreement is calculated as:

$$d = 1 - \frac{\sum_{k=1}^{n_e} \sum_{i=1}^{n_m} \sum_{l=1}^{n_{sp_k}} \left( y_i\left(\boldsymbol{\xi}_k, t_l\right) - \hat{y}_i\left(\boldsymbol{\xi}_k, \hat{\boldsymbol{\theta}}, t_l\right)\right)^2}{\sum_{k=1}^{n_e} \sum_{i=1}^{n_m} \sum_{l=1}^{n_{sp_k}} \left( \left| \hat{y}_i\left(\boldsymbol{\xi}_k, \hat{\boldsymbol{\theta}}, t_l\right) - \overline{y}_i\left(\boldsymbol{\xi}_k\right)\right| + \left| y_i\left(\boldsymbol{\xi}_k, t_l\right) - \overline{y}_i\left(\boldsymbol{\xi}_k\right)\right|\right)^2}. \tag{2.32}$$

The denominator of Eq. (2.32) represents the so-called potential error (Willmott, 1981). This potential error represents the largest value that the squared difference of the pairs of measured and predicted values can attain and it is used to normalize the sum of the absolute differences between the observed and the predicted values.

Note that in both criteria, MEF and $d$, the differences between the measured response variables and the predicted ones are squared. Because of this, both criteria (as well as the WSSE) are sensitive to extreme values, whereas lower values are neglected (Legates and McCabe Jr., 1999; Krause et al., 2005). To overcome this issue, Legates and McCabe Jr. (1999) proposed modified versions of these criteria by taking the absolute values of the differences. In a comparative study, Krause et al. (2005) showed that these modified versions outperformed the original criteria for the examples they considered. Also note that the differences between the measured response variables and the predicted ones are not weighed using the uncertainty associated with the measurements, as for instance in the calculation of the WSSE (Eq. (2.4)). It may however be a good idea to include this weighing in the calculation of both the modeling efficiency and the index of agreement. If the information on the measurement errors is not included, these criteria may reject an adequate model in situations where the measurements are imprecise (as depicted in case 1 of Fig. 2.3).

## 2.9 Model selection

As suggested in Section 2.7, a situation may occur in which it is very difficult or even impossible to further discriminate among a number of remaining rival models with the available experimental setup. In such a case, one of the remaining rival models can be chosen or selected. Model selection can be defined as the task of selecting the best performing model from a set of adequate models based on the available data and the characteristics of the model structures. Spriet (1985) argued that the best model is selected as a trade-off between model fit, model parsimony and balanced accuracy. Each of these aspects is briefly discussed below.

Model fit deals with the ability of the different models to describe the available data. From the discussion in the previous section, it is clear that this is an important aspect that also has to be reflected in the model selection criterion. Early methods for model selection only considered this aspect, and the model with the smallest WSSE was selected (Spriet, 1985).

Next to the ability of the model to fit the data, also the model parsimony is important and has to be considered. Model parsimony is related with the complexity of the model. A particular model structure is believed to be most effective and plausible if it is simple. Indeed, the model fit can be made as good as wanted by increasing its complexity, that is, by adding additional parameters. When the number of parameters becomes too high compared to the information content of the data, the ability of the model to generalize beyond the available experiments is reduced. This phenomenon is called overfitting, and can be detected by adding an additional term to the model selection criterion that penalizes for the number of parameters. Therefore, a more complicated model will only be accepted if its improvement in model fit is large enough, and thus balances the penalty term. Two popular criteria that take into account the model complexity are the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). These are respectively given by Eqs. (2.33) and (2.34) (de Brauwere et al., 2005; Spriet, 1985; Vanrolleghem and Dochain, 1998):

$$\mathrm{AIC} = n \cdot \log\left(\frac{\mathrm{WSSE}}{n}\right) + 2 \cdot n_p \,, \tag{2.33}$$

$$\mathrm{BIC} = n \cdot \log\left(\frac{\mathrm{WSSE}}{n}\right) + n_p \cdot \log\left(n\right) \,, \tag{2.34}$$

where $n$ represent the number of data points and $n_p$ represents the number of model parameters.

A third and final aspect that has to be taken into account is balanced accuracy, which we prefer to call structural identifiability. The notion of identifiability is related to the possibility to give a unique value to each of the model parameters (Bellman and Aström, 1970; Vanrolleghem and Dochain, 1998), and the outcome of a structure identifiability analysis is a *yes* or *no* answer. A lack of identifiability can be related to the model structure itself (structural identifiability) or to the quantity and quality of the experimental data (practical identifiability). The structural identifiability of a particular model structure is examined under the assumption that perfect or *error-free* measurements are available for the response variables. From the structural identifiability analysis, one may conclude that only certain combinations of the model parameters can be identified. For instance, in the model $y = a \cdot x_1 + b \cdot x_2 + c \cdot (x_1 + x_2)$, it is impossible to give a unique value to $a$, $b$ or $c$ and only two parameters will be identifiable if the other one is known *a priori*. This becomes clear after rewriting the model as $y = (a + c) \cdot x_1 + (b + c) \cdot x_2$. Practical identifiability, on the other hand, determines whether the available data is sufficiently informative to identify the model parameters. When selecting a model, there is little sense in retaining a possibly more valid model structure if its parameters can only be poorly estimated. An approximate model structure of which the parameters can be estimated more accurately may be more adequate, and thus a better model. To our knowledge, there are currently no model selection criteria that explicitly take this identifiability into account.

## 2.10 Model validation

Once a model is selected, its validity has to be examined before it can actually be used. Therefore, the model predictions are confronted with new experimental data, that is, data that have not been used for estimating the model parameters. Model validation may seem a simple and straightforward task, but it is not. One reason why this is a difficult task is because model validation is strongly linked with the purpose of the model, and thus its intended use (Mitchell, 1997; Moriasi et al., 2007; Power, 1993; Rykiel, 1996; Tedeschi, 2006). Indeed, the term *valid model* has different meanings to different users, and one must maintain some flexibility as to what is considered as a valid model.

Mathematical models can be classified according to many different criteria, but an interesting classification is the one where a distinction is made between *models for prediction* and *models for understanding* (Vanrolleghem and Dochain, 1998). Models for prediction are used in applications where it is required that the model is able to predict the behavior of the studied process very accurately. In such applications, the model will primarily be used

to make predictions that may eventually replace real observations of the process. Often, this may require extrapolation, for which the mechanisms that govern the behavior of the process must be included adequately in the model. However, good predictions can also be achieved without understanding the basic mechanisms of the studied process. This is, for instance, the case with empirical, so-called black box models. The second type of models are intended to increase the understanding in the (often complex) process. Although a full understanding of the process is too ambitious, some applications may require that the model captures most or at least the most important mechanisms that determine the behavior of the process. In such applications, the model holds until significant deviations from reality are observed.

From the above discussion, one can conclude that a mathematical model can thus not be proved valid as such. But one can only evaluate whether it is appropriate for its intended purpose for given conditions or not (Tedeschi, 2006). Based on an extensive review of the vast amount of literature on this subject, Rykiel (1996) define model validation as *"the process of demonstrating that a model possesses a satisfactory range of accuracy consistent with the intended application of the model within its domain of applicability"*. This indicates that the model is acceptable for use, not that it embodies any absolute truth, nor that it is the best model available.

## 2.11 Optimization algorithms

Both parameter estimation and optimal experimental design are optimization problems. To find the optimum, the use of an optimization algorithm is required. In this work, the SIMPSA optimization algorithm proposed by Cardoso et al. (1996) was used. This algorithm combines the nonlinear simplex (Nelder and Mead, 1965) and the simulated annealing algorithm (Kirkpatrick et al., 1983), and it showed good performance for both the parameter estimation and the experimental design problems encountered in this work. Below, these optimization algorithms will be briefly described. For more detailed information on these optimization algorithms, the reader is referred to the cited literature.

### 2.11.1 Local and global optima

Before describing the optimization algorithms, this section briefly discusses the difference between local and global optima and its importance with regard to the (choice of the) optimization algorithm. Suppose one is dealing with a minimization problem, then a

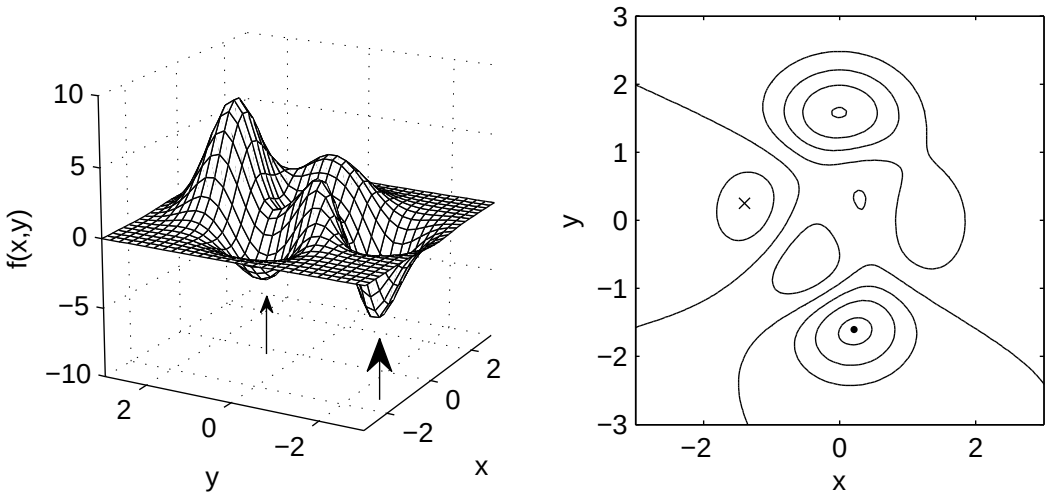



**Figure 2.4:** Illustrating the difference between local and global optima for a hypothetical optimization problem in which the minimum of function $f(x, y)$ is sought. The global minimum is indicated by the arrow with the large arrow head in the left figure and the • symbol in the right figure, while the local minimum is indicated by the small arrow and the × symbol.

global minimum is defined as the lowest objective function value in the whole parameter space. For the hypothetical optimization problem shown in Fig. 2.4, the global minimum is indicated by the arrow with the largest arrow head. As indicated by the other arrow, the objective function also has a so-called local minimum, which is defined as the lowest objective function value in a bounded neighborhood around it.

Despite the fact that a significant amount of research has been done to develop efficient optimization algorithms that guarantee that the global optimum for a given optimization problem is found, no perfect optimization algorithm exists (so far) (Dochain and Vanrolleghem, 2001). It is thus important to be aware of this problem, and care must be taken to maximize the confidence in the result of the optimization. Nevertheless, some optimization algorithms are more sensitive to this problem than others and the choice of the optimization algorithm is thus important.

### 2.11.2 Steepest descent algorithm

A wide range of optimization algorithms has been developed for locating the minimum of a given objective function, and many among them make use of information about the gra-

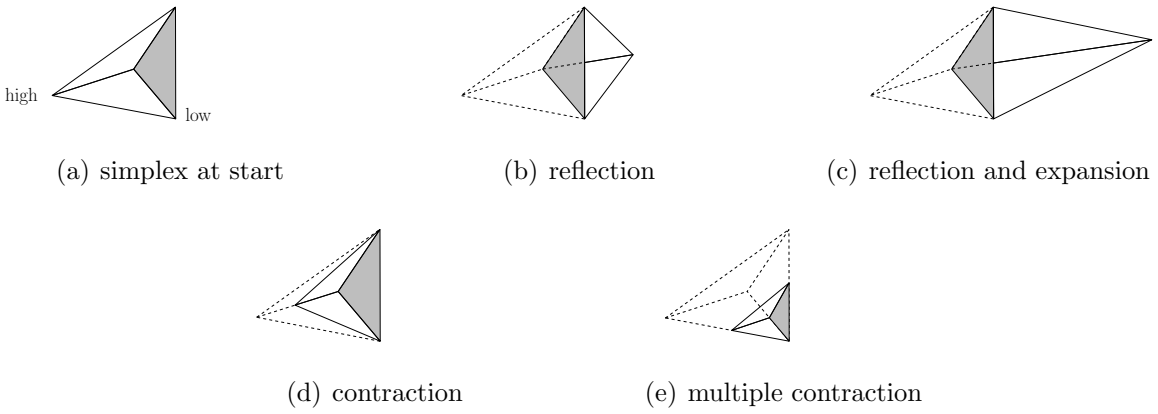
dient of the objective function with respect to the variables to be optimized (Dochain and Vanrolleghem, 2001). One of these algorithms is the so-called steepest descent algorithm.

The basic idea of this algorithm is that the minimum is sought in that direction in the variable space where the decrease in the objective function is largest. So, first, the direction in the variable space along which the objective function decreases fastest is determined (based on the gradient), and once the direction of steepest descent is identified, this path is followed until the minimum along this direction is reached. If so, the optimization variables are changed accordingly, and a new direction of steepest descent is identified.

It is clear that this approach does not guarantee that the global optimum is found. Indeed, this algorithm will converge to the optimum that is closest to the initial guess, regardless of its nature (local or global). In addition, this algorithm requires information on the gradient of the objective function. The latter is not always available, and its computation by numerical approximation is often computationally demanding.

### 2.11.3 Nonlinear simplex algorithm

The nonlinear simplex optimization algorithm was proposed by Nelder and Mead (1965). This algorithm uses the geometrical concept of a simplex, which is the $n$-dimensional analogue of a triangle in two dimensions. In $n$ dimensions, the simplex is the geometrical figure that arises when $n + 1$ points (or vertices) are connected. This is illustrated in Fig. 2.5a, where a simplex is shown for the case of three dimensions.




(a) simplex at start (b) reflection (c) reflection and expansion

(d) contraction (e) multiple contraction

**Figure 2.5:** Elementary operations in the simplex optimization algorithm, illustrated on a hypothetical case in three dimensions.

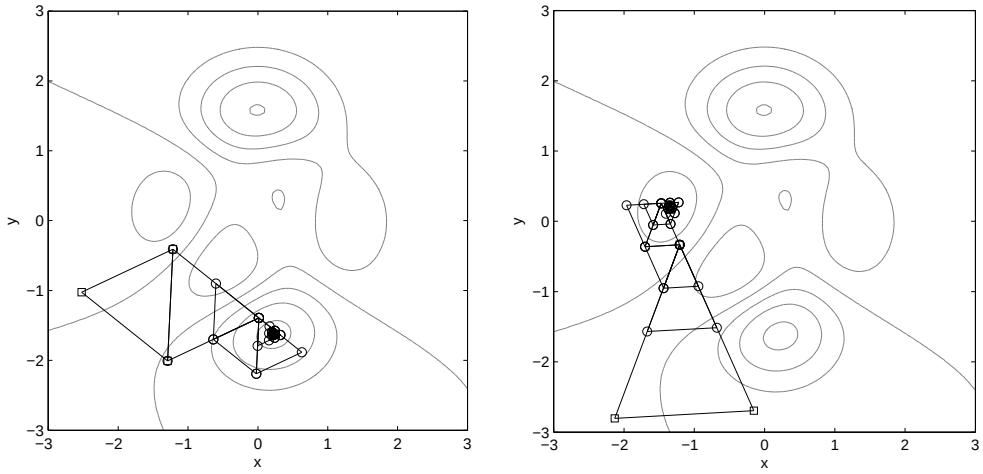



**Figure 2.6:** Illustration of the nonlinear simplex algorithm on the example shown in Fig. 2.4. The simplex algorithm converges to the global minimum in (a) and to the local minimum in (b).

Starting from an initial simplex (Fig. 2.5a), the optimum (here, minimum) is found by evaluating the objective function value at the vertices of the simplex and replacing the vertex with the highest value by a new point in $n$-dimensional space for which the corresponding objective value is lower. To determine this new point, a number of operations can be performed on the working simplex. For instance, the objective function value of the point obtained by reflecting the worst vertex is calculated, and the worst performing vertex is replaced by this new point if its corresponding objective value is smaller. If not, the points obtained from the operations shown in Fig. 2.5c and Fig. 2.5d are evaluated in a similar way. The optimization process is terminated when the working simplex becomes sufficiently small and the vertices thus almost coincide, or when the objective function values of the different vertices are nearly identical.

The nonlinear simplex algorithm is less sensitive to local optima than the steepest descent algorithm described in the previous section, although better alternatives are available (for instance, the simulated annealing algorithm described in the next section). However, the simplex optimization algorithm is appreciated for its ease of implementation, the fact that it does not require information on the gradient and because of its reasonable convergence rate (Dochain and Vanrolleghem, 2001).

### 2.11.4 Simulated annealing algorithm

Algorithms based on simulated annealing employ a stochastic generation of solution vectors and employ similarities between the physical process of annealing in metallurgy (Kirkpatrick et al., 1983). The latter is a technique involving heating and controlled cooling of a material to increase the size of its crystals and reduce their defects. The heat causes the atoms to become unstuck from their initial positions (a local minimum of the internal energy) and wander randomly through states of higher energy; the slow cooling gives them more chances of finding configurations with lower internal energy than the initial one. In other words, during the cooling process, transitions are accepted to occur from a low to a high energy level through a Boltzmann probability distribution.

By analogy with this physical process, each step of the simulated annealing algorithm replaces the current solution by a random *nearby* solution, chosen with a probability that depends on the difference between the corresponding objective function values and on a global parameter called the temperature, which is gradually decreased during the process. The dependency is such that the current solution changes almost randomly when the temperature is large, but increasingly *downhill* as the temperature goes to zero. The allowance for *uphill* moves prevent the method from becoming stuck at local minima, which is for instance the case with the steepest descent algorithm described in Section 2.11.2.

### 2.11.5 SIMPSA algorithm

The SIMPSA algorithm is based on the combination of the nonlinear simplex and simulated annealing algorithms (Cardoso et al., 1996). In principle, the SIMPSA algorithm is not much different from the simulated annealing algorithm described above. The difference between both algorithms lies in the way in which the new solutions are chosen. In the original simulated annealing algorithm, these new solutions are chosen randomly in the neighborhood of the current solution. The SIMPSA algorithm, on the other hand, works with simplexes (as discussed in Section 2.11.3) and the new solutions are determined by the simplex operations shown in Fig. 2.5. So, the role of the nonlinear simplex is to generate potential solutions for the optimization problem.

The SIMPSA algorithm was selected among the many different optimization algorithms described in literature (for instance, in Banga et al. (2003, 2005); Mendes and Kell (1998); Moles et al. (2003)) for two reasons. First, because it is a global optimization algorithm. This is important because both parameter estimation and experimental design problems

can suffer from local minima/maxima, and the use of a global optimization algorithm is therefore required. Several examples can be found in literature where such algorithms were used, both for parameter estimation (Checchi and Marsili-Libelli, 2005; Moles et al., 2003; Rodriguez-Fernandez et al., 2006) and experimental design exercises (Banga et al., 2002; De Pauw and Vanrolleghem, 2006a; Moles et al., 2003; Ternbach et al., 2005). Second, because it requires little to no configuration. Typically, several parameters are available to tune an algorithm, and the values at which they are set determine the efficiency of the optimization algorithm. However, the optimal settings are mostly problem specific, and manual tuning is typically required. With the SIMPSA algorithm, some of the parameters are automatically tuned and only two parameters remain to be tuned manually.

### 2.11.6 Handling constraints

Often, the optimization variables, which can be parameters and experimental degrees of freedom in the context of this dissertation, are only allowed to vary between well-defined boundaries. Nevertheless, the unconstrained optimization algorithm presented above may propose a value for these variables that lies beyond these boundaries. To handle constraints on the optimization variables, a penalty function is used (Dochain and Vanrolleghem, 2001). When the optimization algorithm proposes a value lying beyond the upper or lower bound, the corresponding value of the objective function is decreased (maximization) or increased (minimization) with a large penalty term that increases as the proposed value is further away from the bound (as illustrated in Fig. 2.7). By adding a gradient to the penalty function, the optimization algorithm gets an idea about the direction in which it should progress.

## 2.12 Summary and conclusions

In this introductory chapter, a number of important aspects of mathematical modelling were explained and the notations used throughout this dissertation were introduced. One should remember from this chapter that the mathematical models considered in the following are given by a set of coupled differential and algebraic equations, which constitute the model structure. These equations describe the evolution of the state variables over time and the models are therefore called dynamic models. Some of these state variables correspond to actually measured response variables, while other ones can be calculated from one or more state variables. To be able to simulate a well-defined experiment with a given model, values for the model parameters have to be supplied. These parameter

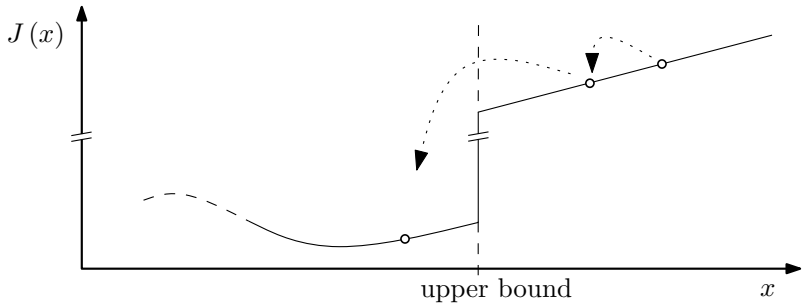



**Figure 2.7:** Illustrating of the approach used in this work to deal with constraint violation. When the optimization algorithm proposes a value of a particular variable which lies beyond the upper boundary (dotted vertical line), a penalty is added to the objective function value ($J(x)$), which decreases in the direction of the allowed variable range. In this way, the values proposed in the following iterations are more likely to be within the allowed variable range.

values have to be determined from experimental data, which are inevitably corrupted by measurements errors/noise. As a consequence, some uncertainty will be associated with the parameter estimates, which will propagate to the model predictions when the model is used for simulation. These uncertainties are very important and the methods described in this chapter can be used to quantify them.

In addition, this chapter presented a general procedure to discriminate among a number of rival models, which consisted of four steps that are performed in an iterative manner. After estimating the parameters of the rival models from the available experimental data sets, the adequacy of the models are evaluated. The models that are found to be adequate are considered in the design of optimal discriminatory experiments that are eventually performed and the procedure is repeated. Ideally, this procedure is repeated until the most appropriate model is identified. As the case studies described in this dissertation are simulation studies, statistical tests can be described to evaluate the adequacy of the rival models. However, in practical applications, these test may reject useful models because it is unrealistic to expect that the true model can be found. Therefore, some alternative methods to evaluate the adequacy of a model were presented as well. The design of optimal discriminatory experiments is dealt with in the following chapters.

# CHAPTER 3

## Rival kinetic models for glucokinase: working example

*"The purpose of computing is insight, not numbers."*

Richard Hamming, mathematician

## Abstract

The experimental design methods described in this work will be evaluated after applying them (*in silico*) to the model discrimination problem described in this chapter. In this working example, nine rival models are proposed to describe the kinetics of the enzymatic reaction catalyzed by glucokinase. In addition, the approach used throughout this work to generate experimental data is discussed.

## 3.1 Introduction

Throughout this dissertation, the mathematical modelling and experimental design concepts will be illustrated on a working example in which nine rival models are proposed to describe the enzymatic conversion of glucose (GLU) and adenosine triphosphate (ATP) to glucose-6-phosphate (G6P) and adenosine diphosphate (ADP). The enzyme that catalyzes this biochemical conversion is called glucokinase (*glk*, EC: 2.7.1.2), and is reported to be inhibited by phosphoenolpyruvate (PEP) (Ogawa et al., 2007). For illustratory purposes, a three-dimensional representation of this enzyme is shown in Fig. 3.1.

Glucokinase is found in several micro-organisms, for instance in *Escherichia coli*, where it catalyzes the first reaction of the so-called glycolysis. The latter is a pathway present in most organisms and consists of a sequence of reactions that converts glucose into pyruvate with the concomitant production of ATP. This ATP plays a crucial role in the cell's metabolism (Madigan et al., 2000; Mathews et al., 2000). It is generated in the cell by energy-releasing processes and is broken down by energy-consuming processes. In this way, ATP transfers cellular energy between spatially-separate metabolic reactions.

To determine the kinetics of an enzyme, so-called *in vitro* enzyme assays can be conducted. In such an enzyme assay, the isolated and purified enzyme is brought together with its substrates in a reaction vessel, and the consumption of the substrate(s) or the production of the product(s) are measured over time. The concentration profiles obtained in this way can then be used to determine the enzyme kinetics.

In the following, nine rival models will be proposed to describe the kinetics of glucokinase, each of which can be seen as a mathematical representation of a plausible reaction mechanism. Based on information found in literature on glucokinase, one of these models was assumed to represent the true reaction mechanism. This *presumed* reaction mechanism is discussed in the next section, and the corresponding kinetic equation is derived in Section 3.3. The corresponding model will be used to generate experimental data in the *in silico* examples described in the remainder of this dissertation. The kinetic equations corresponding to the other reaction mechanisms are described in Section 3.5.

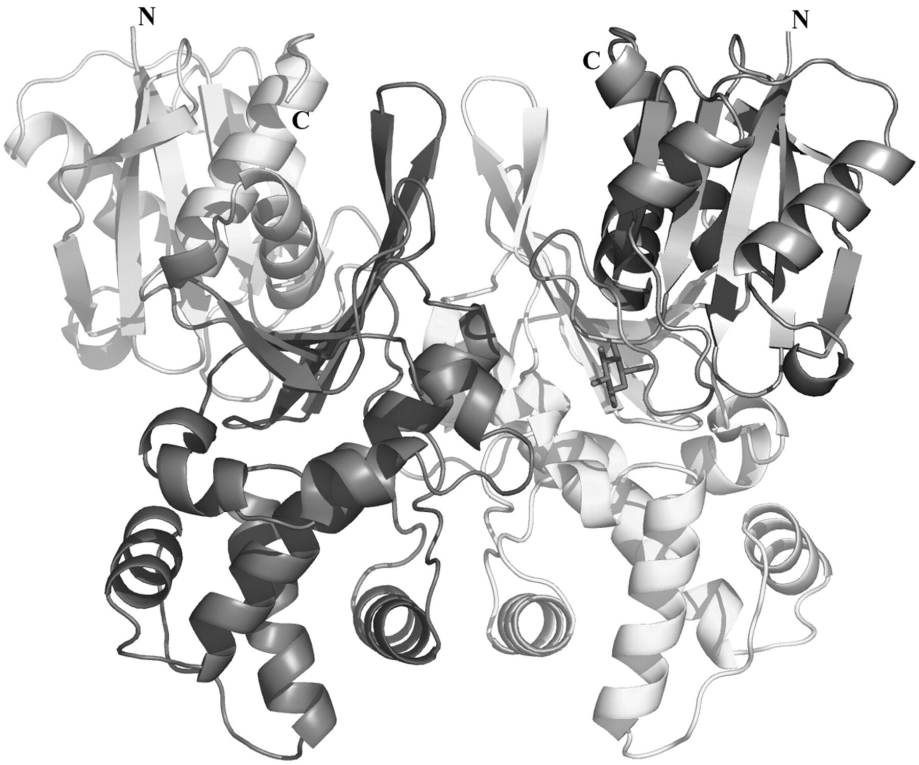



**Figure 3.1:** Three-dimensional represenatation of glucokinase found in *Escherichia coli*, adopted from Lunin et al. (2004)

## 3.2 Reaction mechanism of glucokinase

The reaction catalyzed by glucokinase is a reaction in which two substrates are converted into two products, and it is therefore called a bi-reactant system (Segel, 1975). For such enzymes, two reaction mechanisms are possible, random and ordered. If both substrates can bind to the enzyme independently, the reaction is called a random bi-reactant system. The conversion takes place as soon as the two substrates are bound to the enzyme, and the products are released subsequently. However, such a mechanism may not always be possible. Often, one of the substrate has to bind to the enzyme first, before the other substrate can bind. Therefore, the first binding substrate is often called the activator. This activator might induce a conformational change in the enzyme so that substrate binding groups become available, or the activator itself is modified by the enzyme to a form that then participates in positioning the substrate properly in relation to the catalytic groups of the enzyme (Segel, 1975).

As stated in the introduction, Ogawa et al. (2007) reported that glucokinase is inhibited by PEP. An inhibitor is defined as any substance that reduces the velocity of an enzyme-catalyzed reaction, and inhibition can occur in many different forms (Fersht, 1999; Segel, 1975). According to Ogawa et al. (2007), PEP inhibits glucokinase because it acts as a competitive inhibitor for ATP. Competitive inhibition occurs when a substance (here, PEP) combines with free enzyme in a manner that prevents subsequent substrate binding (Segel, 1975). In other words, PEP can bind with glucokinase and this binding prevents the binding of ATP. This of course slows down the overall conversion rate. Monasterio and Cárdenas (2003) state that it has been postulated that the reaction mechanism is ordered with glucose as the first binding substrate, although this subject is not closed and it remains unclear whether the reaction mechanism is totally ordered.

The presumed reaction mechanism of glucokinase is shown in Fig. 3.2, where A and B stand for glucose and ATP, and $A'$ and $B'$ stand for G6P and ADP, respectively. The inhibitor is represented by I, and corresponds with PEP in the case of glucokinase. Note that these notations (A, B, $A'$, $A'$ and I) were introduced for simplicity. In the following section, a kinetic equation will be derived that allows a mathematical description of the enzymatic conversion process that works according to this reaction mechanism.

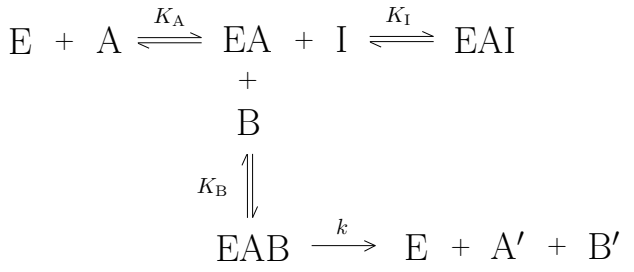



**Figure 3.2:** Ordered reaction mechanism for an enzyme (E) that catalyzes the conversion of A and B into A′ and B′, where A is the first binding substrate and I inhibits the binding of the second substrate to the enzyme (competitive inhibition). For the case of glucokinase, A and B stand for glucose and ATP, respectively, A′ and B′ stand for G6P and ADP, and the inhibitor I is PEP.

## 3.3 Kinetic equation for glucokinase

To derive an equation that describes the kinetics of an enzyme, the so-called rapid equilibrium approach can be followed (Segel, 1975). The kinetics derived in this way are often called steady state kinetics (Fersht, 1999). In the rapid equilibrium approach, it is assumed that all binding and dissociation steps are very rapid compared to the catalytic step, in which the products are formed and released from the enzyme. For the presumed reaction mechanism of glucokinase, this catalytic step (as depicted in Fig. 3.2) is represented by

$$\mathrm{EAB} \xrightarrow{k} \mathrm{E} + \mathrm{A}' + \mathrm{B}' \,. \tag{3.1}$$

This step is thus assumed to be the rate-limiting step, and the velocity-dependence equation states that the velocity at which the conversion proceeds is equal to the product of the concentrations of all product-forming species, each multiplied by their catalytic rate constant. In the case of glucokinase, there is only one product-forming species (that is, EAB) and the velocity equation can be written as

$$v = k \cdot [\mathrm{EAB}] \,. \tag{3.2}$$

Here, $k$ represents the so-called catalytic rate constant and [EAB] stands for the concentration of the EAB complex. To continue the derivation, both terms of Eq. (3.2) are divided by $[E]_t$. The latter expresses the concentration of the enzyme in all its forms, that is, the

free enzyme [E] as well as the different enzyme complexes. With the notations introduced in Fig. 3.2, $[E]_t$ can be calculated as

$$[\mathrm{E}]_t = [\mathrm{E}] + [\mathrm{EA}] + [\mathrm{EAB}] + [\mathrm{EAI}] \ . \tag{3.3}$$

This results in the following equation for $v$:

$$\frac{v}{[\mathrm{E}]_t} = k \cdot \frac{[\mathrm{EAB}]}{[\mathrm{E}] + [\mathrm{EA}] + [\mathrm{EAB}] + [\mathrm{EAI}]} \ . \tag{3.4}$$

To express the concentration of each enzyme species in terms of the concentration of free enzyme, the dissociation or equilibrium constants ($K$) for the different reactions preceding the catalytic step (see Fig. 3.2) are integrated in Eq. (3.4). These equilibrium constants are defined as follows

$$K_\mathrm{A} = \frac{[\mathrm{E}] \cdot [\mathrm{A}]}{[\mathrm{EA}]} \ , \quad K_\mathrm{B} = \frac{[\mathrm{EA}] \cdot [\mathrm{B}]}{[\mathrm{EAB}]} \quad \text{and} \quad K_\mathrm{I} = \frac{[\mathrm{EA}] \cdot [\mathrm{I}]}{[\mathrm{EAI}]} \ .$$

Substitution of these equilibrium constants into Eq. (3.4) results in the following equation for $v$:

$$v = k \cdot [\mathrm{E}]_t \cdot \frac{\frac{[\mathrm{A}] \cdot [\mathrm{B}]}{K_\mathrm{A} \cdot K_\mathrm{B}}}{1 + \frac{[\mathrm{A}]}{K_\mathrm{A}} + \frac{[\mathrm{A}] \cdot [\mathrm{B}]}{K_\mathrm{A} \cdot K_\mathrm{B}} + \frac{[\mathrm{A}] \cdot [\mathrm{I}]}{K_\mathrm{A} \cdot K_\mathrm{I}}} \ . \tag{3.5}$$

In the case of glucokinase, this equation can be rewritten as:

$$v_{glk} = k \cdot \mathrm{GLK} \cdot \frac{\frac{\mathrm{GLU} \cdot \mathrm{ATP}}{K_\mathrm{GLU} \cdot K_\mathrm{ATP}}}{1 + \frac{\mathrm{GLU}}{K_\mathrm{GLU}} + \frac{\mathrm{GLU} \cdot \mathrm{ATP}}{K_\mathrm{GLU} \cdot K_\mathrm{ATP}} + \frac{\mathrm{GLU} \cdot \mathrm{PEP}}{K_\mathrm{GLU} \cdot K_\mathrm{PEP}}} \ , \tag{3.6}$$

where the GLK stands for the concentration of glucokinase, and GLU, ATP and PEP stand for the concentrations of glucose, ATP and PEP, respectively. The catalytic rate constant or maximum specific reaction rate is represented by $k$, and the equilibrium constants are represented by $K_\mathrm{GLU}$, $K_\mathrm{ATP}$ and $K_\mathrm{PEP}$. With this equation, it is thus possible to calculate the rate of conversion for given concentrations of GLK, GLU, ATP and PEP. However, to describe the outcome of an *in vitro* enzyme assay, this kinetic equation has to be plugged into a more general model, the structure of which is discussed in the next section.

## 3.4 General model

Before describing the (nine) different kinetics for glucokinase, a general mass balance model for the enzymatic conversion process is formulated. For this, it is assumed that the experimental setup used for the *in vitro* enzyme assay allows the experimenter to give a pulse of glucose, ATP and PEP, or a mixture thereof.

The volume of the reaction vessel [L], denoted as V, is determined by the pulse and the sampling (frequency and volume). The flowrate of the pulse [L/s], denoted as $F_p$, is calculated as the ratio between the volume of the pulse [L] and the time in which the pulse is given [s]. In this example, the sampling volume will be neglected and the volume can thus be described by

$$\frac{d\mathrm{V}}{dt} = F_p \,. \tag{3.7}$$

For the concentration of glucokinase [mg/L], denoted as GLK, only a dilution effect is considered. Inactivation of the enzyme is neglected, which is a reasonable assumption since a typical experiment ends after 20 minutes. The resulting equation for describing the enzyme concentration is given as

$$\frac{d\mathrm{GLK}}{dt} = -\frac{F_p}{\mathrm{V}} \cdot \mathrm{GLK} \,. \tag{3.8}$$

The equations used to describe the other state variables (all of which are expressed in [mM]) are given as:

$$\begin{aligned}
\frac{d\mathrm{GLU}}{dt} &= \frac{F_p}{\mathrm{V}} \cdot (\mathrm{GLU}_p - \mathrm{GLU}) - v_{glk} \,, && (3.9)\\
\frac{d\mathrm{ATP}}{dt} &= \frac{F_p}{\mathrm{V}} \cdot (\mathrm{ATP}_p - \mathrm{ATP}) - v_{glk} \,, && (3.10)\\
\frac{d\mathrm{G6P}}{dt} &= -\frac{F_p}{\mathrm{V}} \cdot \mathrm{G6P} + v_{glk} \,, && (3.11)\\
\frac{d\mathrm{ADP}}{dt} &= -\frac{F_p}{\mathrm{V}} \cdot \mathrm{ADP} + v_{glk} \,, && (3.12)\\
\frac{d\mathrm{PEP}}{dt} &= \frac{F_p}{\mathrm{V}} \cdot (\mathrm{PEP}_p - \mathrm{PEP}) \,. && (3.13)
\end{aligned}$$

Here, $\mathrm{GLU}_p$, $\mathrm{ATP}_p$ and $\mathrm{PEP}_p$ represent the concentrations [mM] of glucose, ATP and PEP in the pulse, respectively, and $v_{glk}$ represents the velocity equation describing the kinetic behavior of glucokinase [mM/s].

## 3.5 Rival models

As stated above, the conversion catalyzed by glucokinase is a bi-reactant system (Segel, 1975) for which two reaction mechanisms are possible: random and ordered. Recently, it was suggested that glucokinase may be inhibited by phosphoenolpyruvate (PEP) (Ogawa et al., 2007). Based on these considerations, nine different models were defined to describe the enzyme kinetics (Segel, 1975), each of which is based on a particular hypothesis of how the enzyme works. The derivation of their respective kinetic equations is not explicitly described here for brevity, but these derivations are very similar to the one described in Section 3.3. Although the resulting kinetic equations are different for each of the rival models. Each one is of the following form:

$$v_{glk} = k \cdot \mathrm{GLK} \cdot \frac{\frac{\mathrm{GLU}}{K_{\mathrm{GLU}}} \cdot \frac{\mathrm{ATP}}{K_{\mathrm{ATP}}}}{\varphi\left(\mathrm{GLU}, \mathrm{ATP}, \mathrm{PEP}\right)}\,, \tag{3.14}$$

where the parameter $k$ expresses the maximum specific reaction rate [U/mg], where one unit [U] is defined as that amount of enzyme that catalyzes one $\mu$mol of substrate in one minute. The denominator $\varphi\left(\mathrm{GLU}, \mathrm{ATP}, \mathrm{PEP}\right)$ is different for each rival model and the equations for each of the models are derived below.

For models $m_1$, $m_2$ and $m_3$, it is assumed that the reaction mechanism is random. With regard to the inhibition by PEP, three scenarios are possible (also for the other models described further on): there is no inhibition by PEP (Eq. (3.15)), PEP inhibits the binding of ATP (Eq. (3.16)) and PEP inhibits the binding of glucose (Eq. (3.17)). This results in the following equations for $\varphi\left(\mathrm{GLU}, \mathrm{ATP}, \mathrm{PEP}\right)$:

$$1 + \frac{\mathrm{GLU}}{K_{\mathrm{GLU}}} + \frac{\mathrm{ATP}}{K_{\mathrm{ATP}}} + \frac{\mathrm{GLU}}{K_{\mathrm{GLU}}} \cdot \frac{\mathrm{ATP}}{K_{\mathrm{ATP}}}\,, \tag{3.15}$$

$$1 + \frac{\mathrm{GLU}}{K_{\mathrm{GLU}}} + \frac{\mathrm{ATP}}{K_{\mathrm{ATP}}} + \frac{\mathrm{PEP}}{K_{\mathrm{PEP}}} + \frac{\mathrm{GLU}}{K_{\mathrm{GLU}}} \cdot \frac{\mathrm{PEP}}{K_{\mathrm{PEP}}} + \frac{\mathrm{GLU}}{K_{\mathrm{GLU}}} \cdot \frac{\mathrm{ATP}}{K_{\mathrm{ATP}}}\,, \tag{3.16}$$

$$1 + \frac{\mathrm{GLU}}{K_{\mathrm{GLU}}} + \frac{\mathrm{ATP}}{K_{\mathrm{ATP}}} + \frac{\mathrm{PEP}}{K_{\mathrm{PEP}}} + \frac{\mathrm{ATP}}{K_{\mathrm{ATP}}} \cdot \frac{\mathrm{PEP}}{K_{\mathrm{PEP}}} + \frac{\mathrm{GLU}}{K_{\mathrm{GLU}}} \cdot \frac{\mathrm{ATP}}{K_{\mathrm{ATP}}}\,. \tag{3.17}$$

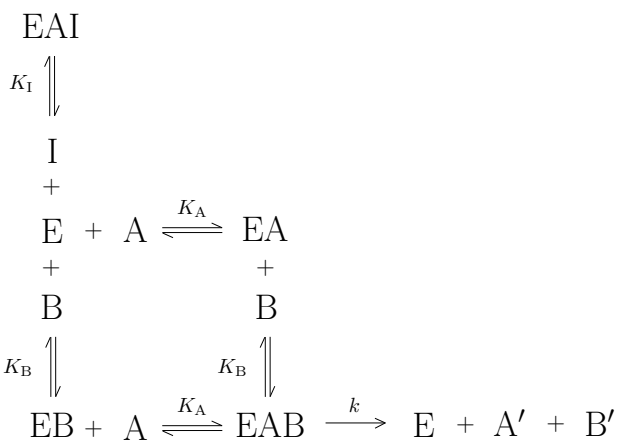



**Figure 3.3:** Random reaction mechanism for an enzyme (E) that catalyzes the conversion of A and B into A′ and B′, where A is the first binding substrate and I inhibits the binding of the second substrate to the enzyme (competitive inhibition). For the case of glucokinase, A and B stand for glucose and ATP, respectively, A′ and B′ stand for G6P and ADP, and the inhibitor I is PEP.

A schematic representation of the random reaction mechanism with a competitive inhibitor is given in Fig. 3.3. This figure is similar to Fig. 3.2, where an ordered reaction mechanism was presented and where the binding of the second binding substrate was competitively inhibited.

For the other six models, an ordered reaction mechanism is assumed. For models $m_4$, $m_5$ and $m_6$, it is assumed that glucose is the first binding substrate, which results in the following equations for $\varphi\,(\mathrm{GLU}, \mathrm{ATP}, \mathrm{PEP})$:

$$1 + \frac{\mathrm{GLU}}{K_{\mathrm{GLU}}} + \frac{\mathrm{GLU}}{K_{\mathrm{GLU}}} \cdot \frac{\mathrm{ATP}}{K_{\mathrm{ATP}}}\,, \tag{3.18}$$

$$1 + \frac{\mathrm{GLU}}{K_{\mathrm{GLU}}} + \frac{\mathrm{GLU}}{K_{\mathrm{GLU}}} \cdot \frac{\mathrm{PEP}}{K_{\mathrm{PEP}}} + \frac{\mathrm{GLU}}{K_{\mathrm{GLU}}} \cdot \frac{\mathrm{ATP}}{K_{\mathrm{ATP}}}\,, \tag{3.19}$$

$$1 + \frac{\mathrm{GLU}}{K_{\mathrm{GLU}}} + \frac{\mathrm{ATP}}{K_{\mathrm{ATP}}} \cdot \frac{\mathrm{PEP}}{K_{\mathrm{PEP}}} + \frac{\mathrm{GLU}}{K_{\mathrm{GLU}}} \cdot \frac{\mathrm{ATP}}{K_{\mathrm{ATP}}}\,. \tag{3.20}$$

The equations associated with models $m_7$, $m_8$ and $m_9$ are similar, but ATP is assumed to be the first binding substrate. The equations for $\varphi\,(\mathrm{GLU}, \mathrm{ATP}, \mathrm{PEP})$ are given by:

$$1 + \frac{\mathrm{ATP}}{K_{\mathrm{ATP}}} + \frac{\mathrm{GLU}}{K_{\mathrm{GLU}}} \cdot \frac{\mathrm{ATP}}{K_{\mathrm{ATP}}}\,, \tag{3.21}$$

**Table 3.1:** Overview of the kinetic equations of the nine rival models described in Section 3 and used as a working example throughout this dissertation.

| rival model | random | ordered | | inhibition by PEP | | kinetic equation |
|---|---|---|---|---|---|---|
| | | glucose | ATP | glucose | ATP | |
| $m_1$ | × | - | - | - | - | Eq. (3.15) |
| $m_2$ | × | - | - | - | × | Eq. (3.16) |
| $m_3$ | × | - | - | × | - | Eq. (3.17) |
| $m_4$ | - | × | - | - | - | Eq. (3.18) |
| $m_5$ | - | × | - | - | × | Eq. (3.19) |
| $m_6$ | - | × | - | × | - | Eq. (3.20) |
| $m_7$ | - | - | × | - | - | Eq. (3.21) |
| $m_8$ | - | - | × | - | × | Eq. (3.22) |
| $m_9$ | - | - | × | × | - | Eq. (3.23) |

$$1 + \frac{\mathrm{ATP}}{K_{\mathrm{ATP}}} + \frac{\mathrm{GLU}}{K_{\mathrm{GLU}}} \cdot \frac{\mathrm{PEP}}{K_{\mathrm{PEP}}} + \frac{\mathrm{GLU}}{K_{\mathrm{GLU}}} \cdot \frac{\mathrm{ATP}}{K_{\mathrm{ATP}}}\,, \tag{3.22}$$

$$1 + \frac{\mathrm{ATP}}{K_{\mathrm{ATP}}} + \frac{\mathrm{ATP}}{K_{\mathrm{ATP}}} \cdot \frac{\mathrm{PEP}}{K_{\mathrm{PEP}}} + \frac{\mathrm{GLU}}{K_{\mathrm{GLU}}} \cdot \frac{\mathrm{ATP}}{K_{\mathrm{ATP}}}\,. \tag{3.23}$$

An overview of the kinetics of the rival models is given in Table 3.1. Note that the kinetic equation for model $m_5$ corresponds to the one that was described in Section 3.3, and thus corresponds to the presumed reaction mechanism of glucokinase.

## 3.6 Real model and data generation

According to literature (Monasterio and Cárdenas, 2003; Ogawa et al., 2007), the reaction mechanism of glucokinase is ordered, with glucose as the first binding substrate, and PEP inhibits the binding of ATP to the enzyme (as discussed in Section 3.2). Based on these considerations, the fifth model was chosen as the real model ($m_5^\star$). This model was used to generate experimental data by simulating the experiment and adding random noise to mimic the measurement error. The parameters used to generate experimental data are given in Table 3.2. The standard deviations of the measurements were calculated in the same way as suggested by Ternbach et al. (2005):

$$\sigma_y = \hat{y} \cdot \varsigma_y \cdot \left( 1 + \frac{1}{\left(\frac{\hat{y}}{lb_y}\right)^2 + \frac{\hat{y}}{lb_y}} \right) . \tag{3.24}$$

Here, $\varsigma_y$ and $lb_y$ respectively represent a constant minimal relative error and a lower accuracy bound on the measurement of $y$. In this way, the standard deviations of the measurements are proportional to the value of $\hat{y}$, but increase when the latter approaches the detection limit or the lower accuracy bound of the measured state variable.

**Table 3.2:** Parameters of the real model ($m_5^\star$) that were used to generate experimental data.

| $k$ [U/mg] | $K_{\mathrm{GLU}}$ [mM] | $K_{\mathrm{ATP}}$ [mM] | $K_{\mathrm{PEP}}$ [mM] |
|---|---|---|---|
| 312.00 | 0.1500 | 0.1300 | 0.1000 |

## 3.7 Some typical simulation results

In this section, some simulation results will be briefly described in order to increase the understanding of the kinetic equations presented above.

### Example one

In this example, two experiments are simulated using model $m_5$. At the start of both experiments, the concentrations of glucose and ATP are 1.5 and 0.5 mM, respectively. In the first experiment, PEP is not present in the reaction medium, while the concentration of PEP is set at 0.5 mM in the second experiment. So, except for the PEP concentration, both experiments are the same. The experiment (and the enzymatic reaction) is initiated by adding glucokinase to the reaction medium (upto 0.016 mg/mL).

The concentration profiles of glucose, ATP, ADP and G6P obtained when simulating the experiment without PEP using the parameter values shown in Table 3.2 are shown in the left graphs of Fig. 3.4. The simulation results obtained for the experiment with PEP are shown in the middle graphs of Fig. 3.4. The graphs on the left represent the simulation results (for the experiment with PEP) obtained using a larger value for $K_{\mathrm{PEP}}$, that is used to characterize the inhibitory effect of PEP.

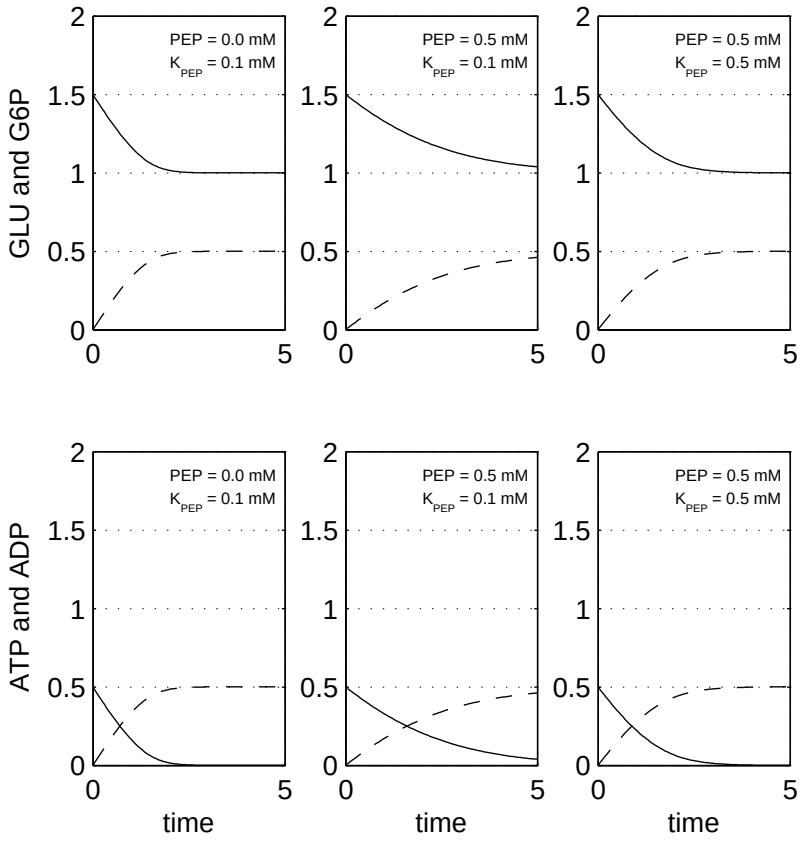



**Figure 3.4:** Simulation results obtained with model $m_5$. The PEP concentration is indicated in the figures as well as the value for parameter $K_{\text{PEP}}$.

The results shown in left graphs of Fig. 3.4, for instance, clearly illustrate that the reaction can only take place when both substrates (glucose and ATP) are present. One can also see that the concentration of the substrates decrease almost linearly when both substrates are abundant. Comparing the results obtained for the experiments with and without PEP (left versus middle graphs) clearly show that the presence of PEP slows down the enzymatic reaction. The results in the graphs in the right side of Fig. 3.4 show that the inhibition effect of PEP reduces for larger values of parameter $K_{\text{PEP}}$. Or, put differently, when the inhibitory effect of a given molecule (here, PEP) is small, the value obtained for parameter $K_{\text{PEP}}$ after estimating the model parameter is expected to be large.

### Example two

In this example, the experiment with PEP described above is simulated with models $m_4$, $m_5$ and $m_6$. The parameter values shown in Table 3.2 were used for all three models.

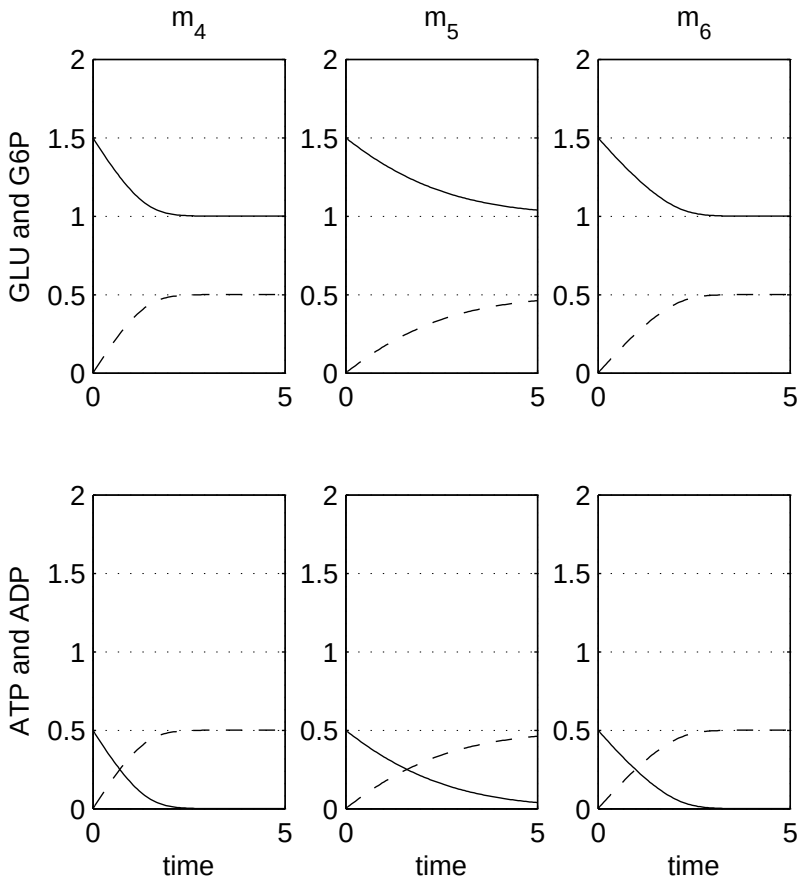



**Figure 3.5:** Simulation results obtained after simulating the experiment with PEP using model $m_4$, $m_5$ and $m_6$.

Note, however, that model $m_4$ does not consider inhibition by PEP and the corresponding kinetic equation does not contain parameter $K_{\mathrm{PEP}}$. Model $m_6$ differs from model $m_5$ by the fact that PEP inhibits the binding of ATP instead of the binding of glucose.

The simulation results are shown in Fig. 3.5. As expected, one can see that the results obtained with model $m_4$ are the same as the ones obtained with $m_5$ for the experiment without PEP (left graphs in Fig. 3.4). One can also see that the reaction simulated with model $m_6$ proceeds faster than with model $m_5$. This is because model $m_6$ assumes that PEP competitively inhibits the binding of glucose. As glucose is more abundant than ATP, the reaction proceeds relatively faster.

# PART II

## OPTIMAL EXPERIMENTAL DESIGN FOR MODEL DISCRIMINATION

# CHAPTER 4

## Criteria for the design of optimal discriminatory experiments

*"What is important cannot always be measured,*
*and what can be measured is not always important."*

Albert Einstein, theoretical physicist

## Abstract

The problem of model discrimination arises when several models are proposed to describe one and the same process. To identify the best model from the set of rival models, it may be necessary to collect new information about the process, and thus additional experiments have to be performed. This chapter deals with the experimental design methodologies that are used to find the experimental conditions that allow to discriminate among rival models with the least experimental effort. For this, the expected experimental results should be predicted differently by the rival models, and the uncertainty on the measurements and on the model predictions should not be too large. These aspects were included in the approach developed by Buzzi-Ferraris and co-workers (1984), but in their approach the uncertainties are estimated from the information content of the already performed experiments. This work presents a modification of the Buzzi-Ferraris approach in which the expected information content of the newly designed experiment is considered, even before the experiment is performed (anticipatory design). In this way, a better estimate of the uncertainties is achieved, and an experiment with an increased discriminatory potential is obtained. The approaches were illustrated and compared by applying them to a case study in which two rival models are proposed to describe the in vitro kinetics of an enzyme.

## 4.1 Introduction

As advocated before, mathematical models are useful tools for scientists and engineers. Next to increasing insight in often complex processes, mathematical models are used in process design, optimization and control. How one obtains such models will not be discussed here. However, it is important to realize that the lack of insight in the modelled process may result in the proposal of several rival models. Obviously, one is especially interested in the model that describes the process under study in the best way. To identify this model from a set of rival models, it may be necessary to collect new information about the process, and thus additional experiments have to be performed.

The methods to design experiments that allow discriminating among rival models, often referred to as optimal experimental design for model discrimination (OED/MD) or optimal experimental design for (model) structure characterization (Vanrolleghem and Van Daele, 1994), will be the main focus of this chapter. As explained in Section 2.7, these experimental design methods are part of a more general procedure for model discrimination in which four steps are performed in an iterative manner until a stopping criterion is met (Fig. 2.2). The rival models and the preliminary experimental data are used in a first step, in which the parameters of the rival models are estimated. A second step involves an evaluation of the adequacy of the models, and the adequate models are used in a third step, where an optimal discriminatory experiment is designed using one of the approaches developed and discussed in this chapter. The optimal discriminatory experiment is then performed in a fourth and last step, after which the loop is closed by re-estimating the parameters of all rival models using all data available at that time. This iterative procedure continues until the best model is identified. Of course, when all models appear to be inadequate, new models have to be proposed.

This chapter is organized as follows. After describing the most important design criteria for OED/MD in Section 4.4, a new approach will be proposed in Section 4.5 that is based on the design criterion originally proposed by Buzzi-Ferraris et al. (1984). In this approach, the expected information content of the newly designed experiment is considered, even before the experiment is performed. Therefore, this approach is called the anticipatory approach. Further, the most promising design criteria found in literature and the anticipatory approach are applied to the working example described in Chapter 3, where nine rival models are proposed to describe the kinetics of glucokinase.

## 4.2 Characteristics of a discriminatory experiment and the importance of uncertainty for its design

Before explaining how one can design an experiment that allows to discriminate among a number of rival models, the characteristics of such an experiment are discussed, as well as the importance of uncertainty for the experimental design. The illustrations depicted in Fig. 4.1 will be used to facilitate this discussion. They represent hypothetical situations one might encounter when trying to discriminate between two rival models $m_1$ and $m_2$.

Figure 4.1a depicts a situation in which both models are able to describe the experimental data in an adequate manner. It is clear that it will not be possible to discriminate between model $m_1$ and $m_2$ in such a situation, as both models describe the data (or the process) equally well. However, if an experiment is performed that results in the data shown in Fig. 4.1b, it it is clear that model $m_2$ is able to adequately describe the experimental data, while the other model does not. This example clearly illustrates the basic idea behind optimal experimental design for model discrimination. When designing an optimal discriminatory experiment, the rival models themselves are used to evaluate an experiment for its discriminatory potential, which is basically determined by the difference between the model predictions. Indeed, when designing the experiment, one assumes that one of the rival models is the *true* model and that the outcome of the designed experiment can be predicted by this model. Under this assumption, one expects that it will be possible to identify the most appropriate model when the other model predicts this experiment totally different (for instance as depicted in Fig. 4.1b). In other words, the design of an optimal discriminatory experiment basically comes down to finding that experiment that maximizes the difference between the model predictions.

However, when designing optimal discriminatory experiments, one should be aware of the fact that the experimental data that will be collected from the designed experiment will inevitably be corrupted with experimental error. Indeed, no measurement is free from experimental error and the uncertainty on the measurements, which can be seen as a measure of the reproducibility of the experiment(al) data, has to be taken into account when designing a discriminatory experiment. When this uncertainty is not considered, a situation may occur as depicted by the black dots in Fig. 4.1c. Here, it is not possible to appoint which of the two models is the most appropriate one, although the difference in the model predictions is significantly large. When the measurement error is not considered, the discriminatory potential of the experiment may thus be wrongly evaluated. In such a case,

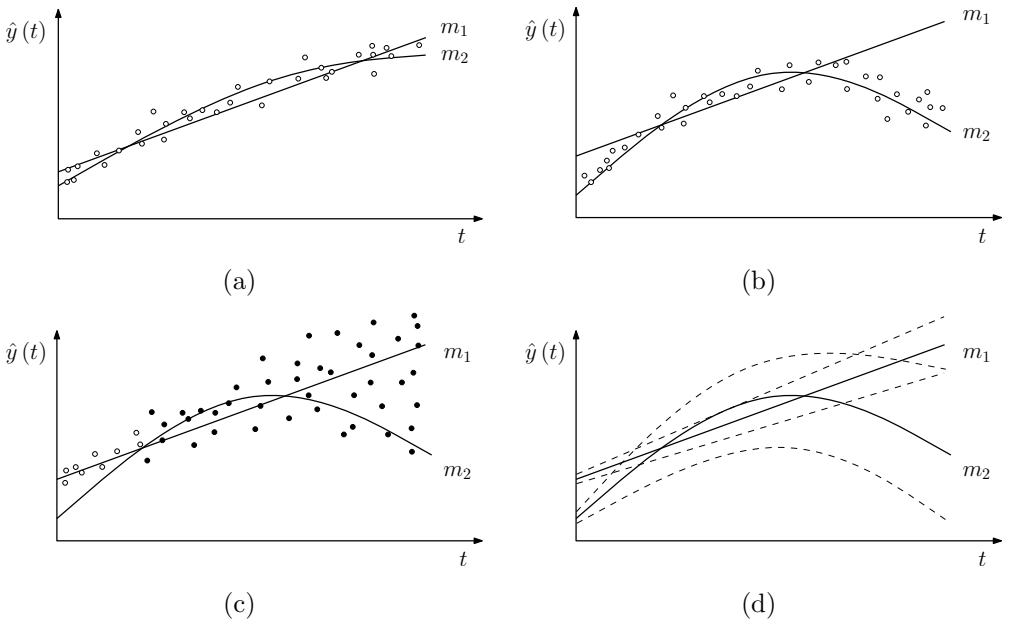



**Figure 4.1:** Hypothetical situations illustrating the required characteristics of a discriminatory experiment. In a situation as depicted in (a), model discrimination is not possible, whereas this is possible in situation (b). This illustrates that a discriminatory experiment should be predicted differently by the rival models. When the uncertainty on the measurements is considered (c), model discrimination will be possible in the beginning of the experiment, and not at the end. This is also the case when the uncertainty on the model predictions is taken into account (d).

more additional experiments may be required to achieve model discrimination compared to approaches that do take the measurement uncertainty into account. In this hypothetical example, it would have been better to sample in the beginning of the experiment (as indicated by the white dots), where the observed difference in the model predictions is small(er), but significant when the measurement error is considered.

The design of optimal discriminatory experiments is also hampered when the uncertainty on the model predictions is too large (Box and Hill, 1967; Burke et al., 1997; Buzzi-Ferraris et al., 1984). As stated above, the design of experiments is model-based and the evaluation of an experiment is based on how it is predicted by the rival models. In this respect, it is obvious that the uncertainty on these model predictions is important and has to be considered when designing the optimal discriminatory experiment. This is illustrated in Fig. 4.1d where the model predictions (the black lines) represent an estimate of the behavior of the real process, and the uncertainties on these estimates are represented by the confidence limits (dashed lines). The latter represent the limits within which the real values of the process variables are expected to lie according to the corresponding model, for a specified level of significance (Kennedy and Neville, 1985). If the limits associated with the predictions of the rival models (almost) overlap, there is a possibility that the expected difference in the model predictions will disappear once the newly collected information is considered and model discrimination may not be possible afterall (Box and Hill, 1967; Buzzi-Ferraris and Forzatti, 1983; Buzzi-Ferraris et al., 1984). In other words, if the uncertainty on the model predictions is not considered during the experimental design, the discriminatory potential of the experiment can be misjudged and the experiment may not result in model discrimination as was expected before performing it.

## 4.3 OED/MD as an optimization problem

In general, optimal experimental design is an optimization problem, where the optimum of a well-defined objective function, denoted as $T$, is sought by varying the experimental degrees of freedom. This can be formalized as follows

$$\boldsymbol{\xi}^{\star} = \arg\max_{\boldsymbol{\xi}\in\boldsymbol{\Xi}} \; T\left(\boldsymbol{\xi}\right) \; . \tag{4.1}$$

The experimental degrees of freedom, $\boldsymbol{\xi}$, are restricted by a number of constraints that define a set of possible experiments, denoted as $\boldsymbol{\Xi}$. These constraints are determined by the experimental setup and are specified before the start of the experimental design

exercise. Note that in this context, the objective functions are also called design criteria, and these terms will be used as synonyms in the following. This chapter is devoted to the design criteria one can use to design optimal discriminatory experiments.

## 4.4 Design criteria for OED/MD

Suppose, for simplicity, that one has to design an experiment to discriminate between two rival models ($m = 2$). To discriminate between more than two rival models ($m > 2$), several strategies can be followed. For instance, one can design an optimal discriminatory experiment for each model pair, and eventually perform the one associated with the highest value of the design criterion (Buzzi-Ferraris et al., 1990; Schwaab et al., 2006). In this way, it should be possible to eliminate the worst models faster. Some other strategies are further discussed in Chapter 5, but it is important to be aware of the fact that the design criteria described in the following are also applicable to experimental design problems with more than two models.

### 4.4.1 Design criterion proposed by Hunter and Reiner (1965)

From the hypothetical examples discussed in Section 4.2, it is clear that the data expected from the designed experiment should be predicted differently by the two models to allow for model discrimination. Hunter and Reiner (1965) translated this heuristic into an objective function denoted as $T_{ij}(\boldsymbol{\xi})$ and given by

$$T_{ij}(\boldsymbol{\xi}) = \sum_{l=1}^{n_{sp}} \Delta\hat{\boldsymbol{y}}_{ij}\left(\boldsymbol{\xi}, \hat{\boldsymbol{\theta}}_i, \hat{\boldsymbol{\theta}}_j, t_l\right)' \cdot \Delta\hat{\boldsymbol{y}}_{ij}\left(\boldsymbol{\xi}, \hat{\boldsymbol{\theta}}_i, \hat{\boldsymbol{\theta}}_j, t_l\right) , \tag{4.2}$$

where

$$\Delta\hat{\boldsymbol{y}}_{ij}\left(\boldsymbol{\xi}, \hat{\boldsymbol{\theta}}_i, \hat{\boldsymbol{\theta}}_j, t_l\right) = \hat{\boldsymbol{y}}_i\left(\boldsymbol{\xi}, \hat{\boldsymbol{\theta}}_i, t_l\right) - \hat{\boldsymbol{y}}_j\left(\boldsymbol{\xi}, \hat{\boldsymbol{\theta}}_j, t_l\right) \tag{4.3}$$

represents the difference between the $n_m$-dimensional vectors of the predicted outcomes of experiment $\boldsymbol{\xi}$ by model $i$ and model $j$ at time $t_l$, and $n_{sp}$ represents the number of samples taken. Note that this notation will be simplified to $\Delta\hat{\boldsymbol{y}}_{ij}(\boldsymbol{\xi}, t_l)$ in the following.

The design criterion of Hunter and Reiner (1965) represents the basic idea behind OED/MD. However, it is important to point out that this design criterion does not take into account

the uncertainty on the measurements, nor on the model predictions. Why it is important to do so was discussed in Section 4.2, and some design criteria that do take these sources of uncertainty into account are described in the next section.

### 4.4.2 Design criterion proposed by Atkinson and Fedorov (1975)

In the original approach of Atkinson and Fedorov (1975), which was later modified by Munack (1992), it is assumed that one of the two rival models is the *true* one and the experiment proposed by the optimization algorithm ($\boldsymbol{\xi}$) is simulated with this model, giving rise to an additional data set. If $m_i$ is assumed to be the *true* model, the parameters of the other model ($m_j$) are re-estimated using the already available data and the *new* data set generated by model $m_i$. Under the assumption that model $m_i$ is the *true* model, it is clear that the optimal discriminatory experiment is that experiment for which the residual sum of squared errors is largest. Indeed, under these conditions, one expects that model $m_j$ will not be able to describe the newly collected data properly and that the failure of model $m_j$ to adequately describe the experimental data will become most apparent when performing this experiment.

Since Atkinson and Fedorov (1975) did not consider the uncertainty on the measurements to estimate the model parameters, this residual sum of squared errors (denoted as SSE) can be calculated as follows

$$\text{SSE}\left(\hat{\boldsymbol{\theta}}_j\right) = \sum_{l=1}^{n_{sp}} \Delta\hat{\boldsymbol{y}}_{ij}\left(\boldsymbol{\xi}, \hat{\boldsymbol{\theta}}_i, \boldsymbol{\theta}_j, t_l\right)' \cdot \Delta\hat{\boldsymbol{y}}_{ij}\left(\boldsymbol{\xi}, \hat{\boldsymbol{\theta}}_i, \boldsymbol{\theta}_j, t_l\right) , \tag{4.4}$$

where $\Delta\hat{\boldsymbol{y}}_{ij}\left(\boldsymbol{\xi}, \hat{\boldsymbol{\theta}}_i, \boldsymbol{\theta}_j, t_l\right)$ is calculated as

$$\Delta\hat{\boldsymbol{y}}_{ij}\left(\boldsymbol{\xi}, \hat{\boldsymbol{\theta}}_i, \boldsymbol{\theta}_j, t_l\right) = \hat{\boldsymbol{y}}_i\left(\boldsymbol{\xi}, \hat{\boldsymbol{\theta}}_i, t_l\right) - \hat{\boldsymbol{y}}_j\left(\boldsymbol{\xi}, \boldsymbol{\theta}_j, t_l\right) \tag{4.5}$$

and represents the difference between the new data obtained by simulating the experiment with model $m_i$, denoted as $\hat{\boldsymbol{y}}_i(\boldsymbol{\xi}, \hat{\boldsymbol{\theta}}_i, t_l)$, and the predictions obtained with model $m_j$ for parameter values $\boldsymbol{\theta}_j$, denoted as $\hat{\boldsymbol{y}}_j(\boldsymbol{\xi}, \boldsymbol{\theta}_j, t_l)$.

The optimal discriminatory experiment is then found by maximizing the residual sum of squared errors, and the design criterion can thus be formalized as

$$T_{ij}\left(\boldsymbol{\xi}\right) = \min_{\boldsymbol{\theta}_j \in \boldsymbol{\Theta}} \quad \sum_{l=1}^{n_{sp}} \Delta\hat{\boldsymbol{y}}_{ij}\left(\boldsymbol{\xi}, \hat{\boldsymbol{\theta}}_i, \boldsymbol{\theta}_j, t_l\right)' \cdot \Delta\hat{\boldsymbol{y}}_{ij}\left(\boldsymbol{\xi}, \hat{\boldsymbol{\theta}}_i, \boldsymbol{\theta}_j, t_l\right) . \tag{4.6}$$

Note that, if the uncertainty on the measurements is considered when the model parameters are estimated, the design criterion will be similar to the modified design criterion of Hunter and Reiner (1965), which will be discussed in Section 4.4.6.

### 4.4.3 Modified design criterion of Atkinson and Fedorov (1975)

As already stated above, this design criterion represented by Eq (4.6) was adapted by Munack (1992). In his approach, the experiment proposed by the optimization algorithm ($\boldsymbol{\xi}$) is simulated with both rival models, giving rise to two additional data sets. Each data set corresponds to a scenario in which one of the rival models is assumed to be the *true* model and for each scenario the design criterion formalized in Eq. (4.6) is evaluated. The smallest design criterion value is eventually used as a measure of the discriminatory potential of the proposed experiment and is maximized. This design criterion can thus be formalized as follows:

$$T_{ij}\left(\boldsymbol{\xi}\right) = \min\left(T_{ij}^{j}\left(\boldsymbol{\xi}\right), T_{ij}^{i}\left(\boldsymbol{\xi}\right)\right) , \tag{4.7}$$

where

$$T_{ij}^{j}\left(\boldsymbol{\xi}\right) = \min_{\boldsymbol{\theta}_j \in \boldsymbol{\Theta}} \quad \sum_{l=1}^{n_{sp}} \Delta\hat{\boldsymbol{y}}_{ij}\left(\boldsymbol{\xi}, \hat{\boldsymbol{\theta}}_i, \boldsymbol{\theta}_j, t_l\right)' \cdot \Delta\hat{\boldsymbol{y}}_{ij}\left(\boldsymbol{\xi}, \hat{\boldsymbol{\theta}}_i, \boldsymbol{\theta}_j, t_l\right) , \tag{4.8}$$

and

$$T_{ij}^{i}\left(\boldsymbol{\xi}\right) = \min_{\boldsymbol{\theta}_i \in \boldsymbol{\Theta}} \quad \sum_{l=1}^{n_{sp}} \Delta\hat{\boldsymbol{y}}_{ij}\left(\boldsymbol{\xi}, \boldsymbol{\theta}_i, \hat{\boldsymbol{\theta}}_j, t_l\right)' \cdot \Delta\hat{\boldsymbol{y}}_{ij}\left(\boldsymbol{\xi}, \boldsymbol{\theta}_i, \hat{\boldsymbol{\theta}}_j, t_l\right) . \tag{4.9}$$

The discriminatory potential of the proposed experiment in the scenario where $m_i$ represents the *true* model and the parameters of model $m_j$ are re-estimated, is thus represented by $T_{ij}^{j}\left(\boldsymbol{\xi}\right)$.

### 4.4.4 Design criterion proposed by Munack (1992)

In another approach, also proposed by Munack (1992), the experimental design is based on the level of dependence of the parameter estimates on the experimental conditions. The latter in fact corresponds to a kind of adequacy evaluation of the rival models, where adequate models are characterized by the fact that the parameter values obtained after estimating them from experimental data are independent of the experimental conditions (Box and Hunter, 1967; Munack, 1992). In other words, if the parameters have to change considerably in order to obtain a significantly good fit to the newly collected experimental data, the model structure is probably inadequate.

To quantify the change in the values of the parameter estimates, Munack (1992) proposed to use the so-called Mahalanobis distance function. One can use this distance function to calculate the distance of a certain point to a distribution and thus differs from the Euclidian distance function by the fact that it allows to take into account the variance-covariance matrix of the distribution. This is illustrated in Fig. 4.2a where a contour plot of a two-dimensional multivariate normal distribution with mean $\overline{\boldsymbol{x}}$ and an unspecified variance-covariance matrix is shown. As indicated by the contour lines, the points $\boldsymbol{x}_a$ and $\boldsymbol{x}_b$ are both located on the same Mahalanobian distance from the mean.

The fact that uncertainty is considered is the reason why Munack (1992) proposed to use the Mahalanobis distance function. Indeed, in this context, the parameter estimates are uncertain to some extent (as discussed in Section 2.5) and this has to be taken into account to quantify the change in their values. Suppose that $\hat{\boldsymbol{\theta}}$ represents an $n_p$-dimensional vector of parameter estimates and $\boldsymbol{\Phi}$ represents the corresponding parameter estimation error covariance matrix, then one can calculate the Mahalanobis distance, denoted as $d$, from an arbitrarily chosen vector of parameters $\boldsymbol{\theta}$ as follows

$$d^2\left(\hat{\boldsymbol{\theta}}, \boldsymbol{\Phi}, \boldsymbol{\theta}\right) = \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)' \cdot \boldsymbol{\Phi}^{-1} \cdot \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) . \tag{4.10}$$

Now, the procedure proposed by Munack (1992) to design an optimal discriminatory experiment is very similar to the one described in Section 4.4.3. However, instead of evaluating the discriminatory potential of the experiment based on the differences in the model predictions (through Eqs. (4.7)–(4.9)), the change in the model parameter values is assessed. Indeed, if the values of the model parameters are expected to change significantly after performing a particular experiment, the latter may succeed in appointing an inappropriate

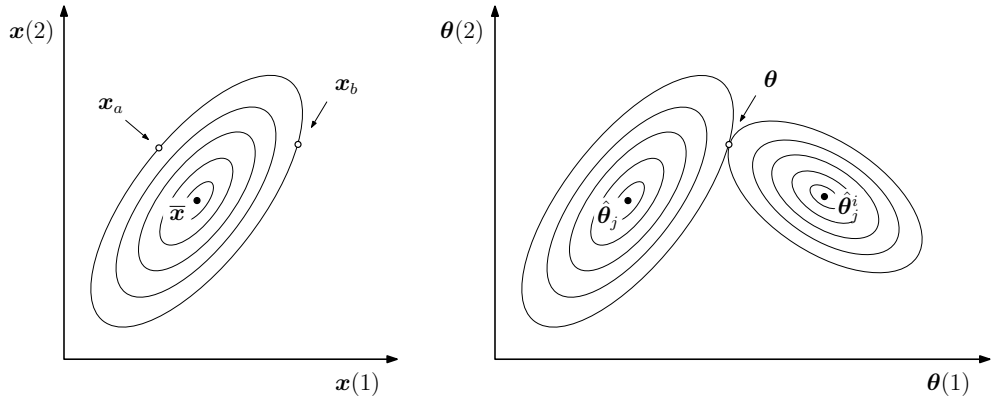



**Figure 4.2:** The principle of the Mahalanobis distance is illustrated in subfigure (a) where the contour plot of a two-dimensional multivariate normal distribution with mean $\overline{\boldsymbol{x}}$ is shown. The vectors $\boldsymbol{x}_a$ and $\boldsymbol{x}_b$ are located on the same Mahalanobian distance from the mean. How the Mahalanobis distance function is used to quantify the expected change in parameter estimates while designing an optimal discriminatory experiment according to Munack (1992) is illustrated in subfigure (b) for the scenario in which model $m_i$ is assumed to be the *true* model.

model and the experiment can thus be considered as a discriminatory experiment. Therefore, the optimal discriminatory experiment is found by maximizing the change in the parameter values observed when re-estimating the parameters of one of the rival model from the experimental data generated by simulating the proposed experiment with the other rival model.

However, the expected difference between the parameter estimates before ($\hat{\boldsymbol{\theta}}_j$) and after ($\hat{\boldsymbol{\theta}}_j^i$) performing the proposed experiment cannot simply be calculated from Eq. (4.10), because both sets of parameter estimates are uncertain. Therefore, the expected change in the parameter estimates is defined as the minimal Mahalanobis distance that any vector $\boldsymbol{\theta}$ may have to both the original parameter estimates ($\hat{\boldsymbol{\theta}}_j$) and the re-estimated parameter estimates ($\hat{\boldsymbol{\theta}}_j^i$), taking into account their corresponding parameter estimation error covariance matrices. This is illustrated in Fig. 4.2b for the scenario in which model $m_i$ is assumed to be the *true* model and the parameters of model $m_j$ have to be re-estimated. The change in the parameters of model $m_j$ can thus be formalized as

$$d_{ij}^{j} = \arg \min_{\boldsymbol{\theta}_j \in \boldsymbol{\Theta}} d\left(\hat{\boldsymbol{\theta}}_j, \boldsymbol{\Phi}_j, \boldsymbol{\theta}_j\right) , \tag{4.11}$$

subject to

$$d_{ij}^{j}\left(\hat{\boldsymbol{\theta}}_j, \boldsymbol{\Phi}_j, \boldsymbol{\theta}_j\right) = d_{ij}^{j}\left(\hat{\boldsymbol{\theta}}_j^{i}\left(\boldsymbol{\xi}\right), \boldsymbol{\Phi}_j^{i}\left(\boldsymbol{\xi}\right), \boldsymbol{\theta}_j\right) . \tag{4.12}$$

Here, $\hat{\boldsymbol{\theta}}_j^{i}\left(\boldsymbol{\xi}\right)$ represents the vector containing the parameter estimates for model $m_j$ after re-estimating them from the experiment data generated by simulating the proposed experiment $(\boldsymbol{\xi})$ with model $m_i$, and $\boldsymbol{\Phi}_j^{i}\left(\boldsymbol{\xi}\right)$ represents the corresponding parameter estimation error covariance matrix. Similar to Eq. (4.9), the design criterion is then given by

$$T_{ij}^{i}\left(\boldsymbol{\xi}\right) = \min\left(d_{ij}^{j}\left(\boldsymbol{\xi}\right), d_{ij}^{i}\left(\boldsymbol{\xi}\right)\right) , \tag{4.13}$$

and has to be maximized.

### 4.4.5 Design criterion proposed by Box and Hill (1967)

The importance of incorporating the uncertainties in the design criterion as discussed in Section 4.2, was also noted by Box and Hill (1967). They proposed an alternative design criterion that takes into account both the uncertainty on the measurements and the uncertainty on the model predictions. Their design criterion is based on Shannon's concept of entropy (Shannon, 1948; Kullback, 1959), which has been developed in the context of information theory. In the context of model discrimination, entropy (denoted as $S$) is used as a measure of the uncertainty as to which of the $m$ rival models is the most appropriate one (Box and Hill, 1967; Hunter et al., 1968). It is defined as

$$S = -\sum_{i=1}^{m} \pi_i \cdot \ln \pi_i \, , \tag{4.14}$$

where $m$ represents the number of rival models, and $\pi_i$ represents the probability that the $i$-th model is the true or most appropriate one. The least possible information as to which of the rival models is the appropriate model occurs when $\pi_1 = \pi_2 = \cdots = \pi_m = \frac{1}{m}$, and corresponds to maximum entropy. On the other hand, the entropy is minimal when one particular model, say model $m_k$, is undoubtedly the most appropriate one, and

$\pi_k \approx 1.0$. Since $\sum_{i=1}^{m} \pi_i = 1$, the model probabilities associated with the other models are (approximately) zero when this occurs.

At the start of a model discrimination exercise, the model probabilities are chosen such that they reflect the belief in the correctness of the particular rival models, or defined as $\frac{1}{m}$ when such information is absent. When a new experiment is performed, the results are analyzed by updating the model probabilities according to Bayes' theorem (Box and Hill, 1967; Burke et al., 1995; Hill, 1978). The posterior model probability for model $m_i$, denoted as $\pi_i\left(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{n_e}, \boldsymbol{\xi}_{n_e+1}\right)$, is then calculated as

$$\pi_i\left(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{n_e}, \boldsymbol{\xi}_{n_e+1}\right) = \frac{\pi_i\left(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{n_e}\right) \cdot p_i\left(\boldsymbol{\xi}_{n_e+1}\right)}{\sum_{k=1}^{m} \pi_k\left(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{n_e}\right) \cdot p_k\left(\boldsymbol{\xi}_{n_e+1}\right)}, \tag{4.15}$$

where $\pi_i\left(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{n_e}\right)$ represents the prior model probability, and $p_i\left(\boldsymbol{\xi}_{n_e+1}\right)$ represents the probability density function for the outcome of experiment $\boldsymbol{\xi}_{n_e+1}$ under the assumption that model $m_i$ is adequate.

The original design criterion of Box and Hill (1967) was formulated for steady state experiments, where only one sample was taken in each experiment. However, the design criterion can be extended for dynamic experiments by assuming that the different samples are independent, as suggested by Takors et al. (1997). Indeed, under this assumption, the outcome of a dynamic experiment can be considered as a set of samples taken at the different sampling times, and an experiment $\boldsymbol{\xi}$ can thus be defined as $\{\hat{\boldsymbol{y}}\left(\boldsymbol{\xi}, t_j\right) \mid \forall j = 1, \ldots, n_{sp}\}$, where $n_{sp}$ represents the number of samples taken in experiment $\boldsymbol{\xi}$. Using this notation, Eq. (4.15) can be rewritten as

$$\pi_i\left(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{n_e}, \boldsymbol{\xi}_{n_e+1}\right) = \frac{\pi_i\left(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{n_e}\right) \cdot \prod_{j=1}^{n_{sp}} p_i\left(\boldsymbol{y}\left(\boldsymbol{\xi}_{n_e+1}, t_j\right)\right)}{\sum_{k=1}^{m} \pi_k\left(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{n_e}\right) \cdot \prod_{j=1}^{n_{sp}} p_k\left(\boldsymbol{y}\left(\boldsymbol{\xi}_{n_e+1}, t_j\right)\right)}, \tag{4.16}$$

where $p_k\left(\boldsymbol{y}\left(\boldsymbol{\xi}_{n_e+1}, t_j\right)\right)$ represents the probability density function for $\boldsymbol{y}\left(\boldsymbol{\xi}_{n_e+1}, t_j\right)$ under the assumption that model $m_k$ is adequate. Assuming that the residuals (calculated as $\boldsymbol{y}\left(\boldsymbol{\xi}, t_j\right) - \hat{\boldsymbol{y}}_i\left(\boldsymbol{\xi}, t_j\right)$) are normally distributed, $p_i\left(\boldsymbol{y}\left(\boldsymbol{\xi}, t_j\right)\right)$ can be calculated as

$$p_i\left(\boldsymbol{y}\left(\boldsymbol{\xi}, t_j\right)\right) = \frac{1}{(2\pi)^{n_p/2} \cdot \left|\boldsymbol{\Sigma} + \boldsymbol{\Omega}_i\right|^{1/2}} \cdot \exp\left(-\frac{1}{2} \cdot \left(\boldsymbol{y}\left(\boldsymbol{\xi}, t_j\right) - \hat{\boldsymbol{y}}_i\left(\boldsymbol{\xi}, t_j\right)\right) \cdot \left(\boldsymbol{\Sigma} + \boldsymbol{\Omega}_i\right)^{-1} \cdot \left(\boldsymbol{y}\left(\boldsymbol{\xi}, t_j\right) - \hat{\boldsymbol{y}}_i\left(\boldsymbol{\xi}, t_j\right)\right)'\right), \tag{4.17}$$

where $\boldsymbol{\Sigma}$ represents the measurement error covariance matrix, $\boldsymbol{\Omega}_i$ represents the model prediction covariance matrix associated with model $m_i$, and $n_p$ represents the number of parameters of model $m_i$. Note that the notations $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}_i$ were used instead of $\boldsymbol{\Sigma}(\boldsymbol{\xi}, t_j)$ and $\boldsymbol{\Omega}_i(\boldsymbol{\xi}, t_j)$ to avoid complexity in the notations.

With Eqs. (4.15), (4.16) and (4.17) it is possible to update the model probabilities when new experimental data become available. To design an optimal discriminatory experiment, a design criterion is derived from the idea that the designed experiment should maximize the information gained when proceeding from a given state of uncertainty (on which rival model is the best one) to a state of lower uncertainty. Or, in other words, the experiment should be designed with the aim to maximize the rate at which the probability of the best model approaches one. A measure of this can be obtained from the expected change in entropy that will occur after performing the designed experiment. Box and Hill (1967) presented an upper bound of this expected change in entropy, which was extended for dynamic experiments by Ternbach (2005). The latter is denoted as $T(\boldsymbol{\xi}_{n_e+1})$ and can be calculated as

$$
\begin{aligned}
T(\boldsymbol{\xi}_{n_e+1}) = \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \sum_{k=1}^{n_{sp}} & \pi_i(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{n_e}) \cdot \pi_j(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{n_e}) \cdot \\
& \left( \int_{\boldsymbol{y}=-\infty}^{\boldsymbol{y}=+\infty} p_i(\boldsymbol{y}(\boldsymbol{\xi}_{n_e+1}, t_k)) \cdot \ln \frac{p_i(\boldsymbol{y}(\boldsymbol{\xi}_{n_e+1}, t_k))}{p_j(\boldsymbol{y}(\boldsymbol{\xi}_{n_e+1}, t_k))} \cdot d\boldsymbol{y} \; + \right. \\
& \left. \int_{\boldsymbol{y}=-\infty}^{\boldsymbol{y}=+\infty} p_j(\boldsymbol{y}(\boldsymbol{\xi}_{n_e+1}, t_k)) \cdot \ln \frac{p_j(\boldsymbol{y}(\boldsymbol{\xi}_{n_e+1}, t_k))}{p_i(\boldsymbol{y}(\boldsymbol{\xi}_{n_e+1}, t_k))} \cdot d\boldsymbol{y} \right) .
\end{aligned} \tag{4.18}
$$

After substitution of Eq. (4.17) into Eq. (4.18) and integration of the resulting equation, one finds that the optimal discriminatory experiment is found through the maximization of

$$
\begin{aligned}
T(\boldsymbol{\xi}_{n_e+1}) = & \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \sum_{k=1}^{n_{sp}} \pi_i(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{n_e}) \cdot \pi_j(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{n_e}) \cdot \\
& \left( \frac{(\boldsymbol{\Omega}_i - \boldsymbol{\Omega}_j)^2}{(\boldsymbol{\Sigma} + \boldsymbol{\Omega}_i) \cdot (\boldsymbol{\Sigma} + \boldsymbol{\Omega}_j)} + \left( \frac{1}{\boldsymbol{\Sigma} + \boldsymbol{\Omega}_i} + \frac{1}{\boldsymbol{\Sigma} + \boldsymbol{\Omega}_j} \right) \cdot \left( \hat{\boldsymbol{y}}_i(\boldsymbol{\xi}_{n_e+1}, t_k) - \hat{\boldsymbol{y}}_j(\boldsymbol{\xi}_{n_e+1}, t_k) \right)^2 \right) .
\end{aligned} \tag{4.19}
$$

An important and interesting characteristic of this design criterion is the fact that it focuses on the discrimination between the most probable models. Indeed, from Eq. (4.19) one can see that the factor which represents the discriminatory potential of the proposed experiment is multipled by $\pi_i(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{n_e}) \cdot \pi_j(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{n_e})$, which represents the product of the model probabilities of the model pair $m_i$ and $m_j$. In this way, the most probable models will be favored compared to models with low model probabilities.

Note that, in this approach, the experimental design and the evaluation of the adequacy of the rival models are carried out together (Hill, 1978; Buzzi-Ferraris and Forzatti, 1983). Indeed, the posterior model probabilities play an important role in both the design of the optimal discriminatory experiment (Eq. (4.18)) and the model evaluation, which simply consists of comparing the model probabilities. In the frequentist approach, discussed in the following section, this is not the case, and an optimal discriminatory experiment is designed and performed first, after which the adequacy of the model is evaluated using one of the approaches discussed in Section 2.8.

Despite the fact that the procedure proposed by Box and Hill (1967) has proved to be effective in practice and popular among researchers in a reasonable range of problems, there are several issues with this criterion (Hill, 1978). These are briefly discussed below.

- Buzzi-Ferraris and Forzatti (1983) showed that results obtained with the criterion of Box and Hill (1967) depend on how experimental observations are ordered because of the recursive law that is used to update model probabilities (Eq. (4.15)). This is unacceptable, as model probabilities should depend on the available information and not on the particular order used to present the data.

- In addition (and possibly related to the previous issue), some authors have seen that the model probabilities may oscillate considerably from iteration to iteration in the sequential strategy (for instance, Froment and Mezaki (1970)). The informal stopping rule suggested by Box and Hill (1967) to stop when the posterior model probabilities indicate that one of the rival models is clearly superior to the other ones, must therefore be applied very cautiously and a model should not be readily accepted on the basis of a small number of discriminatory experiments. In addition, Buzzi-Ferraris and Manenti (2009) state that if all rival models are bad, the use of model probabilities forces the selection of the *least bad* model.

- Buzzi-Ferraris and Forzatti (1983) also noted that the structure of Eq. (4.19) makes it possible that experiments are selected such that the difference in the model pre-

diction variance is large (first term in Eq. (4.19)) rather than the difference of model predictions (second term in Eq. (4.19)). Such experiments are obviously not desired, as they are not expected to have any discriminatory potential.

- The fact that the approach proposed by Box and Hill (1967) maximizes an upper bound of the expected change in entropy, may also lead to problems. Indeed, an increased upper bound does not guarantee that the expected change in entropy itself has increased (accordingly). Hsiang and Reilly (1971) also noted this problem and derived different approximations to the expected entropy change, but the experiments obtained with these design criteria were not significantly different from the ones obtained using Eq. (4.19). For a detailed discussion hereon, the reader is referred to the cited paper. Note, however, that also in optimal experimental design for parameter estimation (discussed in Chapter 6) experiments are designed by maximizing an upper bound.

### 4.4.6 Modified design criterion of Hunter and Reiner (1965)

To incorporate the uncertainty on the measurements, the differences in the model predictions are weighed such that a high measurement error decreases the contribution to the value of $T_{ij}(\boldsymbol{\xi})$. This results in

$$T_{ij}(\boldsymbol{\xi}) = \sum_{l=1}^{n_{sp}} \Delta\hat{\boldsymbol{y}}_{ij}(\boldsymbol{\xi}, t_l)' \cdot \boldsymbol{\Sigma}(\boldsymbol{\xi}, t_l)^{-1} \cdot \Delta\hat{\boldsymbol{y}}_{ij}(\boldsymbol{\xi}, t_l) \ , \tag{4.20}$$

where $\boldsymbol{\Sigma}(\boldsymbol{\xi}, t_l)$ represents the measurement error covariance matrix at time $t_l$ of experiment $\boldsymbol{\xi}$. Note that the approach followed here to incorporate the uncertainty of the measurement is similar to the one used in Eq. (2.4) for parameter estimation (Espie and Macchietto, 1989).

### 4.4.7 Design criterion proposed by Buzzi-Ferraris et al. (1984)

The design criterion proposed by Buzzi-Ferraris and co-workers (Buzzi-Ferraris et al., 1984) builds further on the modified version of Hunter and Reiner's design criterion (Eq. (4.20)) and also incorporates the uncertainty on the model predictions. This is done by weighing the difference in the predicted outcomes of an experiment, denoted as $\Delta\hat{\boldsymbol{y}}_{ij}(\boldsymbol{\xi}, t_l)$, with the uncertainty associated with it. As discussed in Section 4.2, this uncertainty originates from two sources: the uncertainty on the model predictions as such, as well as the uncertainty on

the measurements. In contrast to the modified design criterion of Hunter and Reiner (1965) discussed above, that only considers the uncertainty on the measurements (denoted as $\boldsymbol{\Sigma}(\boldsymbol{\xi}, t_l)$), the design criterion of Buzzi-Ferraris et al. (1984) also considers the uncertainty on the model predictions (denoted as $\boldsymbol{\Omega}(\boldsymbol{\xi}, t_l)$).

When $\boldsymbol{\Sigma}(\boldsymbol{\xi}, t_l)$ and $\boldsymbol{\Omega}(\boldsymbol{\xi}, t_l)$ are assumed to be independent, the uncertainty on the predicted outcome of an experiment can be estimated as $\boldsymbol{\Sigma}(\boldsymbol{\xi}, t_l) + \boldsymbol{\Omega}(\boldsymbol{\xi}, t_l)$. Although this assumption is not entirely valid, it is a reasonable one in this context because the objective function is practically useful and helps to identify the most appropriate model. Now, under a similar assumption of independence, the uncertainty on the difference between the predicted outcomes of an experiment $\boldsymbol{\xi}$ by model $i$ and $j$, denoted as $\boldsymbol{\Psi}_{ij}(\boldsymbol{\xi}, t_l)$, is given by

$$\begin{aligned} \boldsymbol{\Psi}_{ij}(\boldsymbol{\xi}, t_l) &= \boldsymbol{\Sigma}(\boldsymbol{\xi}, t_l) + \boldsymbol{\Omega}_i(\boldsymbol{\xi}, t_l) + \boldsymbol{\Sigma}(\boldsymbol{\xi}, t_l) + \boldsymbol{\Omega}_j(\boldsymbol{\xi}, t_l) \\ &= 2 \cdot \boldsymbol{\Sigma}(\boldsymbol{\xi}, t_l) + \boldsymbol{\Omega}_i(\boldsymbol{\xi}, t_l) + \boldsymbol{\Omega}_j(\boldsymbol{\xi}, t_l) \ . \end{aligned} \tag{4.21}$$

The objective function thus becomes

$$T_{ij}(\boldsymbol{\xi}) = \sum_{l=1}^{n_{sp}} \Delta\hat{\boldsymbol{y}}_{ij}(\boldsymbol{\xi}, t_l)' \cdot \boldsymbol{\Psi}_{ij}(\boldsymbol{\xi}, t_l)^{-1} \cdot \Delta\hat{\boldsymbol{y}}_{ij}(\boldsymbol{\xi}, t_l) \ , \tag{4.22}$$

where $\boldsymbol{\Psi}_{ij}(\boldsymbol{\xi}, t_l)$ represents the uncertainty on the difference between the predicted outcomes of an experiment by models $i$ and $j$ at time $t_l$.

### 4.4.8 Design criterion proposed by Schwaab et al. (2006)

A noteworthy extension of the design criterion of Buzzi-Ferraris et al. (1984) is described in Schwaab et al. (2006). As stated in Section 4.4.5, they remarked that the fact that the design criterion of Box and Hill (1967) focuses on the most probable models, is actually an interesting one. This is not included in the design criterion of Buzzi-Ferraris et al. (1984), and they therefore proposed to include this as follows. For each model $m_i$, the model probability, denoted as $\rho_i$, was defined by exploiting the characteristic of the WSSE to be a sample from a $\chi^2$-distribution with $n - n_p$ degrees of freedom, where $n$ represents the total amount of data points from which the $n_p$ model parameters were estimated. The probability associated with model $m_i$ is calculated as

$$\rho_i = 1 - \mathrm{P}\left[\chi^2_{n-n_p} \leqslant \mathrm{WSSE}\right] , \tag{4.23}$$

from which the relative model probability, denoted as $\pi_i$, is calculated as

$$\pi_i = \frac{\rho_i}{\sum_{k=1}^{m} \rho_k} . \tag{4.24}$$

The basic idea behind the use of Eq. (4.23) is that *bad* models are likely to exhibit large values of WSSE, and consequently a low value of $\rho_i$ (and $\pi_i$). These model probabilities are then used in a similar way as in Eq. (4.19), which results in the following equation for $T_{ij}$:

$$T_{ij}\left(\boldsymbol{\xi}\right) = \left(\pi_i \cdot \pi_j\right)^z \cdot \sum_{l=1}^{n_{sp}} \Delta\hat{\boldsymbol{y}}_{ij}\left(\boldsymbol{\xi}, t_l\right)' \cdot \boldsymbol{\Psi}_{ij}\left(\boldsymbol{\xi}, t_l\right)^{-1} \cdot \Delta\hat{\boldsymbol{y}}_{ij}\left(\boldsymbol{\xi}, t_l\right) . \tag{4.25}$$

The parameter $z$, introduced in Eq. (4.25), can be used to emphasize the most probable models. Indeed, when $z$ is is chosen larger than one, the (difference in the) relative model probabilities are accentuated and the sequential procedure is biased towards the discrimination among the best models. When one chooses values of $z$ that are smaller than one, the effect of the model probabilities is diminished and is totally neglected when $z$ equals zero. It is clear that high values of $z$ correspond to a high fidelity that the best model is one of those with the highest probability. If so, this may lead to model discrimination in fewer experiments, but experimental effort may be wasted if this was not the case afterall.

The model discrimination procedure stops when one of the model relative model probability that is higher than a predefined value (for instance, 0.95). Note that, differently from the design criterion proposed by Box and Hill (1967), the model discrimination procedure does not depend on the particular sequence in which the experiments are performed, which was a serious shortcoming of the design criterion of Box and Hill (1967). This is because the model probabilities are not updated using Bayes' theorem (Eq. (4.16)).

## 4.5 Anticipatory approach to optimal experimental design for model discrimination

From a conceptual point of view, the design criterion proposed by Buzzi-Ferraris and co-workers is superior to the other ones because of the importance it gives to the uncertainty with regard to model discrimination. In addition, this design criterion has been successfully applied by others (for instance, Burke et al. (1995), Burke et al. (1996), Kremling et al. (2004) and Schwaab et al. (2006)). The so-called anticipatory approach proposed in this section also uses this design criterion, but the methodological framework is modified.

In the original methodology of Buzzi-Ferraris, the parameter estimation error covariance matrix, needed to calculate $\mathbf{\Omega}(\boldsymbol{\xi}, t_l)$ in Eq.(4.21) from Eq. (2.24), is estimated from the information present in the experiments that have already been performed, and thus from the corresponding Fisher information matrix (as explained in Section 2.5). This results in the following equation:

$$\mathbf{\Phi}^{-1} = \sum_{k=1}^{n_e} \mathbf{FIM}(\boldsymbol{\xi}_k) \ . \tag{4.26}$$

In this equation, $n_e$ represents the number of experiments performed prior to the experimental design step, and $\mathbf{FIM}(\boldsymbol{\xi}_k)$ represents the Fisher information matrix associated with experiment $\boldsymbol{\xi}_k$. By doing so, the information that is to be gathered when performing the $(n_e+1)^{th}$ experiment (and that would eventually be used when the model adequacy is tested) is simply ignored.

In our modified version of Buzzi-Ferraris's methodology, the parameter estimation error covariance matrix is recalculated for each proposed experiment by also including the expected **FIM** associated with this new experiment. In this way, the expected information content of the newly designed experiment is accounted for, even before the experiment is performed (anticipatory design). This can be formalized as follows

$$\mathbf{\Phi}^{-1} = \sum_{k=1}^{n_e} \mathbf{FIM}(\boldsymbol{\xi}_k) + \mathbf{FIM}(\boldsymbol{\xi}_{n_e+1}) \ , \tag{4.27}$$

where the expected information content of the newly designed experiment is represented by $\mathbf{FIM}(\boldsymbol{\xi}_{n_e+1})$, and calculated using Eq. (2.21).

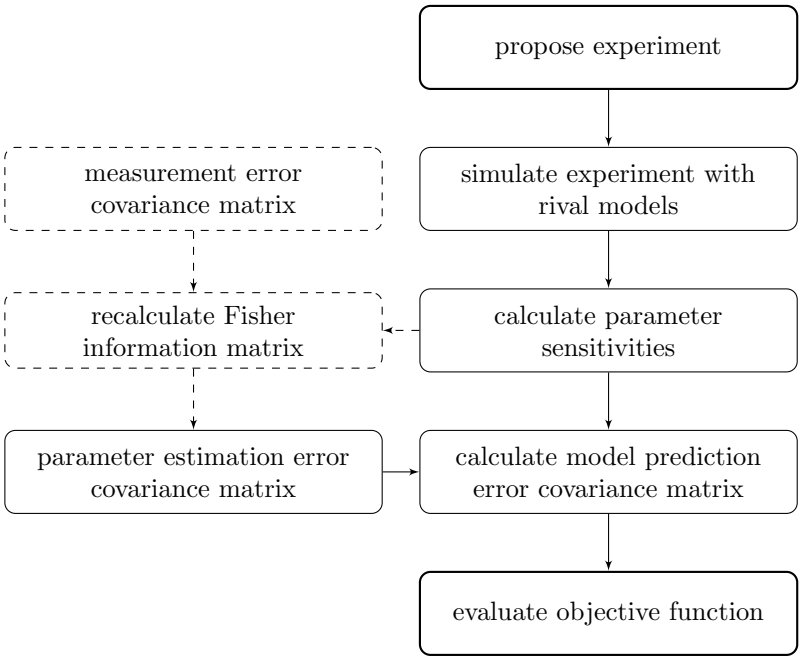



**Figure 4.3:** The full-line boxes in this figure represent the different steps currently performed when designing an optimal discriminatory experiment according to the Buzzi-Ferraris methodology. In the modified Buzzi-Ferraris methodology a number of additional steps are performed, indicated by the dashed-line boxes.

Because it may not be easy to extract the different steps that have to be performed from the equations given above, the sequence of steps is depicted in Fig. 4.3. The full-line boxes in Fig. 4.3 represent the different steps performed when designing an optimal discriminatory experiment according to the methodology of Buzzi-Ferraris, while the dashed-line boxes represent the additional steps to be performed when applying the modified version of Buzzi-Ferraris's methodology proposed in this work (which will be called the anticipatory approach in the following). The original Buzzi-Ferraris methodology starts from an estimate of the parameter estimation error covariance matrix, that is calculated prior to the experimental design step (from Eqs. (4.26) and (2.20)) and remains the same throughout the experimental design. In the anticipatory approach, the parameter estimation error covariance matrix is recalculated for each proposed experiment (using Eq. (4.27)), and is then used in the calculation of $T_{ij}$ through Eqs. (2.24) and Eqs. (4.21)–(4.22).

## 4.6 Similarity between anticipatory approach and optimal experimental design for parameter estimation

As will be discussed in more detail in Chapter 6, the **FIM** is also used in optimal experimental design for parameter estimation (OED/PE), the aim of which is to obtain more accurate parameter estimates with a minimal additional experimental effort. Several scalar measures of the **FIM** (for instance, its determinant or trace) are used as design criteria and, as with the anticipatory approach, the **FIM** is calculated from Eq. (4.27). Taking into account the information of the designed experiment is thus already common practice in OED/PE, but (to our knowledge) this was never done in the context of OED/MD until it was introduced here. Note, however, that the rational behind the anticipatory approach was developed simultaneously and independently by Schwaab et al. (2008).

However, although the parameter estimation error covariance matrix is approximated in the same way as with the anticipatory approach for OED/MD presented in the previous section, one should not expect that performing the experiment obtained using the anticipatory approach to OED/MD automatically results in more accurate parameter estimates. Indeed, although the **FIM** plays a significant role in the calculation of $T_{ij}\left(\boldsymbol{\xi}\right)$ (anticipatory approach), the latter is primarily driven by the difference in the model predictions.

This can be seen from the equation used to calculate the design criterion proposed by Buzzi-Ferraris et al. (1984) (Eq. (4.22)), which is repeated here for clarity:

$$T_{ij}\left(\boldsymbol{\xi}\right) = \sum_{l=1}^{n_{sp}} \Delta\hat{\boldsymbol{y}}_{ij}\left(\boldsymbol{\xi}, t_l\right)' \cdot \boldsymbol{\Psi}_{ij}\left(\boldsymbol{\xi}, t_l\right)^{-1} \cdot \Delta\hat{\boldsymbol{y}}_{ij}\left(\boldsymbol{\xi}, t_l\right) \ , \tag{4.28}$$

where

$$\boldsymbol{\Psi}_{ij}\left(\boldsymbol{\xi}, t_l\right) = 2 \cdot \boldsymbol{\Sigma}\left(\boldsymbol{\xi}, t_l\right) + \boldsymbol{\Omega}_i\left(\boldsymbol{\xi}, t_l\right) + \boldsymbol{\Omega}_j\left(\boldsymbol{\xi}, t_l\right) \ . \tag{4.29}$$

That the **FIM** plays an important role in the calculation of $T_{ij}\left(\boldsymbol{\xi}\right)$ becomes apparent when the equation to calculate $\boldsymbol{\Omega}_i\left(\boldsymbol{\xi}, t_l\right)$ is rewritten as follows (a similar equation can be written for $\boldsymbol{\Omega}_j\left(\boldsymbol{\xi}, t_l\right)$)

$$\boldsymbol{\Omega}_i\left(\boldsymbol{\xi}, t_l\right) = \left( \left. \frac{\partial \hat{\boldsymbol{y}}_i\left(\boldsymbol{\xi}, \boldsymbol{\theta}, t_l\right)}{\partial \boldsymbol{\theta}} \right|_{\hat{\boldsymbol{\theta}}_i} \right) \cdot \mathbf{FIM}_i\left(\boldsymbol{\xi}, t_l\right)^{-1} \cdot \left( \left. \frac{\partial \hat{\boldsymbol{y}}_i\left(\boldsymbol{\xi}, \boldsymbol{\theta}, t_l\right)}{\partial \boldsymbol{\theta}} \right|_{\hat{\boldsymbol{\theta}}_i} \right)' \ . \tag{4.30}$$

From these equations, one can see that a larger $\mathbf{FIM}_i(\boldsymbol{\xi})$ results in a smaller $\boldsymbol{\Omega}_i(\boldsymbol{\xi})$ (Eq. (4.30)), which results in a smaller $\boldsymbol{\Psi}_{ij}(\boldsymbol{\xi})$ (Eq. (4.29)) and thus in a larger $T_{ij}(\boldsymbol{\xi})$ (Eq. (4.28)). However, from Eq. (4.28), one can see that the effect of the $\mathbf{FIM}_i(\boldsymbol{\xi})$ will act together with the effect of the difference in the model predictions, denoted as $\Delta\hat{\boldsymbol{y}}_{ij}(\boldsymbol{\xi})$. Indeed, $\boldsymbol{\Psi}_{ij}(\boldsymbol{\xi})$ merely acts as a weighing matrix for the difference in the model predictions, and one could state that the latter has a steering effect on the experimental design. Note, however, that this should be interpreted from a conceptual point of view, as one deals with matrices and not with scalar values. Still, the reasoning described above helps us to interpret the logic with which the different equations are connected.

Nevertheless, both the design criteria used in OED/PE and the anticipatory approach to OED/MD benefit from the larger information content obtained by considering the designed experiment, and thus a larger **FIM**. Therefore, one can state that an experiment designed using the anticipatory approach to OED/MD is more likely to result in more precise parameter estimates than an experiment designed using the original approach of Buzzi-Ferraris. Although this was observed by Schwaab et al. (2008) in the case studies they have performed, it may not be entirely correct to state that the anticipatory approach as such tackles the dual problem of model discrimination and parameter estimation because a mixed design criterion is optimized that may result in a compromise. The performance of the anticipatory approach with regard to model discrimination, as well as its connection with the precision of the parameter estimates, will be investigated and further discussed in Section 4.11.

## 4.7 Criterion to evaluate the discriminatory potential of the designed experiment

When discussing the general procedure for model discrimination in Section 2.7, it was already mentioned that a situation may arise where (further) model discrimination appears to be impossible (given the experimental setup). In such a situation, it would be very interesting if the experimenter could be informed about this. In other words, a criterion that evaluates the discriminatory potential of the designed experiment such that the experimenter would be informed whether model discrimination can be expected after performing the experiment or not, would definitely be a merit.

Such a criterion was proposed by Buzzi-Ferraris and Forzatti (1983), where the design criterion formalized in Eq. (4.22) was described for the first time for the special case where an experiment consisted of one single measurement, that is, one measured state variable and one sampling time (or, $n_m = 1$ and $n_{sp} = 1$). In such experiments, the sample is usually taken at the end of the experiment, when steady state conditions are reached. For such applications, it was stated that the experiment may not result in model discrimination (and may not be worth performing) when the value of $T_{ij}(\boldsymbol{\xi}) < 1$. The rationale behind this criterion is that a $T_{ij}(\boldsymbol{\xi})$-value smaller than one indicates that the variance of the difference between the model predictions can be explained in terms of the measurement error ($\boldsymbol{\Sigma}$) and the uncertainty on the model predictions ($\boldsymbol{\Omega}_i + \boldsymbol{\Omega}_j$). In other words, one expects that no significant difference in the model predictions will be observed after performing the designed experiment.

This criterion was extrapolated by Buzzi-Ferraris et al. (1984) for applications where more state variables are measured ($n_m \geq 1$), by stating that the $T_{ij}(\boldsymbol{\xi})$-value associated with the designed experiment should be larger than the number of measured state variables ($T_{ij}(\boldsymbol{\xi}) > n_m$). This rationale was further extrapolated in Chen and Asprey (2003) for experiments where several samples are taken during the course of the experiment ($n_m \geq 1$ and $n_{sp} \geq 1$). They state that the discriminatory potential of the experiment is sufficiently high to expect that model discrimination will occur, when the $T_{ij}(\boldsymbol{\xi})$-value is larger than the total number of samples taken, which is equal to $n_m \times n_{sp}$. So, when adopting their criterion, the designed experiment is performed when $T_{ij}(\boldsymbol{\xi}) > n_m \times n_{sp}$.

Although these extrapolations have been used in literature (for instance, Burke et al. (1995), Schwaab et al. (2006) and Schwaab et al. (2008)), in our opinion, both extrapolations of the original evaluation criterion are problematic in their suggested forms. Indeed, the evaluation of the discriminatory potential of the designed experiment is based on the overall (or average) discriminatory potential of the different measurements, but the inability of one of the rival models to describe one or some of the measurements (or measured state variables) may be sufficient to discriminate between them. Suppose, for instance, a situation where one sample is taken at the end of the experiment (as with the original formulation of the evaluation criterion) and suppose that the $T_{ij}(\boldsymbol{\xi})$-value for this experiment (or measurement) is significantly larger than one. In other words, one expects that this experiment (or measurement) will enable us to identify the most probable model. After adding a sufficient number of uninformative measurements (that is, without any discriminatory potential), the experiment will be rejected by the extrapolated versions of the

evaluation criterion, although it does contain the information required to identify the most appropriate model.

Our doubts about the correctness of these evaluation criteria is confirmed by the results presented in Schwaab et al. (2008). In one of the case studies they performed, model discrimination was achieved even though the $T_{ij}(\boldsymbol{\xi})$-value was below the required threshold. The authors stated that the evaluation criterion appeared to be too pessimistic and indicated this may be related to the fact that it does not take into account that the model parameters will change when the designed experiment is actually performed. Unfortunately, they simply noted this problem and did not give an alternative criterion to evaluate the discriminatory potential of the designed experiments.

## 4.8 Computational costs of the different approaches

In the search for an optimal discriminatory experiment, many experiments are proposed by the optimization algorithm for which the discriminatory potential has to be assessed. The latter requires some prior calculations of which the computational costs differ for the presented approaches.

When the design criterion of Hunter and Reiner (1965) is used (Eq. (4.2)), simulating the experiment with the rival models brings forth all information necessary to calculate the trajectory of the objective function. When its modified version is used (Eq. (4.20)), the measurement uncertainties are needed as well. Although these are calculated from the simulation results in the *in silico* examples described in this dissertation, in practical applications they need to be determined from the experimental setup and are thus not calculated. Hence, the computational costs associated with these design criteria are identical in real applications, and comparable in *in silico* examples.

When the parameter and model prediction uncertainty have to be estimated, which is the case when the design criterion of Box and Hill (1967) or Buzzi-Ferraris et al. (1984) is used (Eqs. (4.19) and (4.22), respectively), the model is extended with additional ordinary differential equations to calculate the parameter sensitivities (as explained in Section 2.4). This will obviously result in a larger computational cost compared to the approach of Hunter and Reiner (1965) and its modified version. The additional computational cost caused by the fact that the parameter estimation error covariance matrix has to be recalculated in the anticipatory approach is minor and only involves some extra matrix manipulations, as can be concluded from Eqs. (2.20) and (4.27).

Note that in the special case where only the measurement times are to be optimized (as for instance in the case study presented in Section 4.10), the use of an optimization algorithm is only necessary when the anticipatory approach is used. This is because the choice of the sampling times has no effect on the model predictions, nor on the model prediction uncertainties. In other words, the trajectory of $T_{ij}$ is calculated once, and the optimal sampling times are easily determined from this trajectory. For the anticipatory approach, this is not the case, and a new trajectory of $T_{ij}$ has to be calculated for each set of the sampling times. However, the extra computational costs are relatively small, because the values of the parameter sensitivities at the proposed sampling times can be obtained by sampling from the (detailed) parameter sensitivity profiles calculated prior to the experimental design exercise.

The approaches proposed by Atkinson and Fedorov (1975) and Munack (1992), described in Sections 4.4.2, 4.4.3 and 4.4.4, are computationally demanding because two parameter estimation exercises have to be performed before the proposed experiment can be evaluated. Indeed, it is reasonable to assume that the data set generated by one particular model can only be described by its rival model after a new estimation of its parameters. However, the search for the new parameter estimates might be very efficient, as the new parameter estimates may be similar to the original ones. On the other hand, because of the large number of optimizations that have to be performed, the chances of ending in a local optimum will obviously increase. If this occurs, an experiment will be designed based on non-optimal values for the parameter estimates. In addition, the approach proposed by Munack (1992) requires the calculation of the parameter sensitivities, but the computational costs associated with this are negligible compared to the ones of the parameter estimation. Because of the requirement to re-estimate the model parameters, these approaches are computationally demanding for OED/MD purposes and they may not be a feasible option in many applications (Vanrolleghem and Dochain, 1998).

## 4.9 Selection of the most promising approaches

Given the theoretical aspects outlined in Section 4.4 and the discussion of the computational costs of the different approaches in the previous section, four approaches are selected for the design of optimal discriminatory experiments. To simplify the discussion of the case studies presented in Sections 4.10 and 4.11, a (shorter) notation is also introduced here.

**Table 4.1:** Overview of the selected approaches to design optimal discriminatory experiments (denoted as $T_a$, $T_b$, $T_c$ and $T_d$).

| | $T_a$ | $T_b$ | $T_c$ | $T_d$ |
|---|---|---|---|---|
| experimental design driven by the difference in the model predictions | × | × | × | × |
| uncertainty on the measurements taken into account | - | × | × | × |
| uncertainty on the model predictions taken into account | - | - | × | × |
| information content of the designed experiment taken into account | - | - | - | × |

The approach proposed by Hunter and Reiner (1965) (Section 4.4.1) is selected for its simplicity and because it represents the basic idea behind OED/MD. In a way, this approach can be seen as a reference approach with which the effects of the conceptual improvements incorporated in the other approaches can be compared. Also, the modified version of this approach (Section 4.4.6) is selected because this approach is the only one that takes into account the measurement error without considering the uncertainty on the model predictions as well. The notations $T_a$ and $T_b$ will be used to indicate these approaches, which, for clarity, use the objective functions given by Eqs. (4.2) and (4.20).

The approach proposed by Buzzi-Ferraris et al. (1984), described in Section 4.4.7, is also selected because it incorporates both the uncertainty on the measurements and on the model predictions, and because it has already been used frequently in literature (for instance in Burke et al. (1995, 1996); Kremling et al. (2004); Schwaab et al. (2006)). The objective function used in this approach (Eq. (4.22)), is the same as the one used in the anticipatory approach described in Section 4.5, which is the fourth approach that is selected. These approaches will respectively be referred to as $T_c$ and $T_d$, in the discussion below.

As discussed in Section 4.4.5, the approach of Box and Hill (1967) has a number of problems associated with it. Because of these issues, this approach was not considered in the following. The design criteria proposed by Atkinson and Fedorov (1975) and Munack (1992), respectively presented in Sections 4.4.2 and 4.4.4, were not withheld because of the computational costs associated with them.

## 4.10 Case study I: Improving the insight in the selected approaches to design discriminatory experiments

In this section, the experimental design concepts introduced in the previous sections will be illustrated by applying them on the working example discussed in Section 3. However, to facilitate the discussion of the obtained results, only two of the nine models will be considered.

### 4.10.1 Objective of this case study

The aim of this case study is to gain more insight in the selected approaches to design optimal discriminatory experiments. By applying approaches to a case study with only two rival models, the effect of the conceptual improvements of the design criterion of Hunter and Reiner (1965) will be investigated. Because the aim is to increase our insight into the different approaches, the experimental design exercise was limited to finding the optimal sampling times, while keeping the other experimental degrees of freedom fixed. This should make the discussion of the obtained results easier to understand.

### 4.10.2 Preliminary experiment

To initiate the case study, a preliminary experiment was defined and performed *in silico*. For this experiment, the volume of the reaction vessel was set to 10 mL and the initial glucokinase concentration was set such that 5 units were present in the reaction mixture. Further, it was assumed that no G6P, ADP and PEP were present at the start of the experiment, and the initial concentrations of glucose and ATP were set to 1.5 mM and 0.5 mM, respectively.

During the experiment, two pulses were given, both with a pulse volume of 1 mL. The first pulse was given five minutes after the start of the experiment, and contained only ATP. The ATP concentration in the pulse was chosen such that the ATP concentration in the reaction mixture was raised to 1.5 mM. The second pulse, given ten minutes after the start of the experiment, contained glucose and PEP, and their concentrations in the pulse were chosen such that the resulting concentrations in the reaction mixture were 1.5 mM and 0.1 mM, respectively.

The experiment stopped after 20 minutes, and ten measurements of glucose, ATP, G6P and ADP were taken in duplicate at arbitrarily chosen times (see Fig. 4.4). The minimal

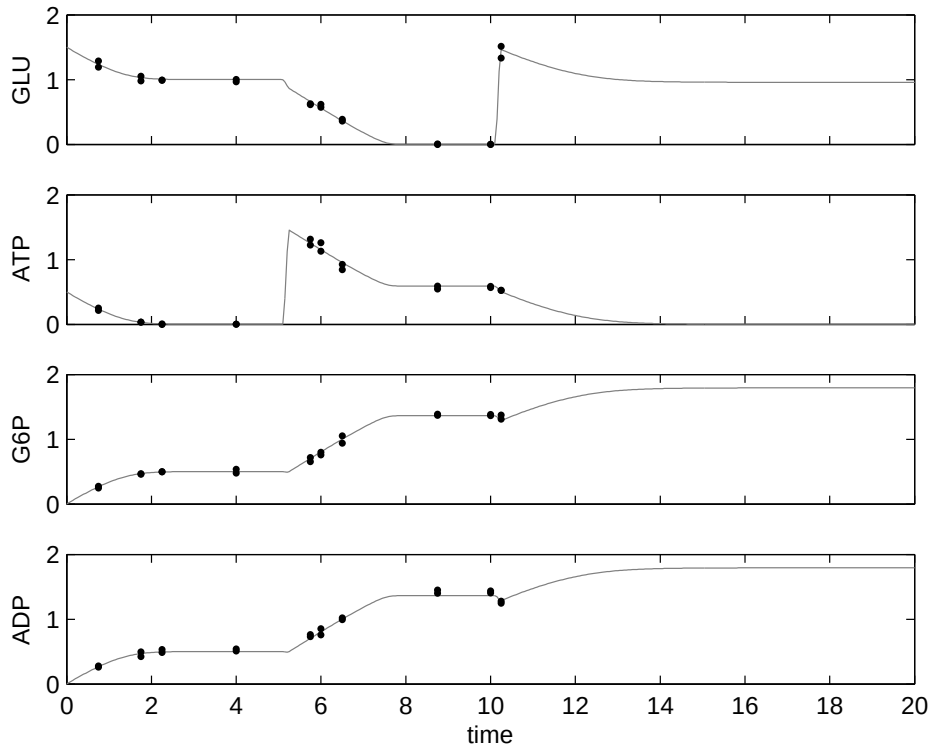



**Figure 4.4:** Preliminary experiment simulated with the *true* model ($m_5^\star$) and the experimental data derived from it (represented by the • symbols).

relative errors ($\varsigma$) were set to 0.05 for all measured state variables, and the lower accuracy bounds on the measurements were defined as 0.1 mM.

### 4.10.3 Parameter estimation

The parameters of the rival models were estimated using the data from the preliminary experiment (Fig. 4.10.2) and using the optimization algorithm described in Section 2.11. Since negative parameter values would not make any sense, the lower bounds were set to zero. The upper bounds were arbitrarily set to 1000 U/mg for parameter $k$, 2 mM for parameter $K_{\mathrm{GLU}}$, and 25 mM for both parameters $K_{\mathrm{ATP}}$ and $K_{\mathrm{PEP}}$.

The results of this parameter estimation exercise are shown in Table 4.2, and Fig. 4.5 shows how both models describe the experimental data after estimating their parameters. The figure also shows the 95 percent confidence intervals on these model predictions. As one can see from Table 4.2, the uncertainty on the parameter estimates is very large,

**Table 4.2:** Parameters of the real model ($m_5^\star$) that were used to generate experimental data, and the parameter estimates obtained after fitting the rival models ($m_2$ and $m_5$) to the data from the preliminary experiment, together with the 95% confidence intervals and the corresponding WSSE values.

| model | $k$ | $K_{\mathrm{GLU}}$ | $K_{\mathrm{ATP}}$ | $K_{\mathrm{PEP}}$ | WSSE |
|---|---|---|---|---|---|
| $m_5^\star$ | 312.00 | 0.1500 | 0.1300 | 0.1000 | – |
| $m_2$ | 308.05 ± 82.86 | 0.01 ± 0.15 | 0.14 ± 0.06 | 2.87 ± 855.20 | 57.1080 |
| $m_5$ | 309.81 ± 34.42 | 0.24 ± 1.41 | 0.12 ± 0.09 | 0.03 ± 0.18 | 56.9285 |

which indicates that the information content of the preliminary experiment with regard to these parameters is considerably low. One can also see that these confidence intervals sometimes become negative, which is of course unrealistic and is due to the fact that the model prediction uncertainties were obtained by linearly propagating the uncertainty on the parameter estimates (as explained in Section 2.6).

### 4.10.4 Model adequacy testing

The good agreement between the data and the model predictions is obviously reflected in the value of the weighted sum of squared errors (WSSE). As shown in Table 4.2, the WSSE values for models $m_2$ and $m_5$ were 57.1080 and 56.9285, respectively. According to the model adequacy test outlined in Section 2.8, these values have to be compared to the critical $\chi^2$ value taken from a $\chi^2_{76}$ distribution ($n = 80$ and $n_p = 4$) with a chosen confidence level of 95 percent. The latter is equal to 97.3510, so one can conclude that both models adequately describe the data from the preliminary experiment.

### 4.10.5 Design of optimal discriminatory experiments

Since both models passed the model adequacy test, the model discrimination procedure asks for an experiment that is designed to discriminate between them. As already stated in the introduction of the section, the experimental design was restricted to only optimizing the measurement times, while keeping the other experimental degrees of freedom fixed. In this way, the differences between the selected approaches for OED/MD should be easier to explain and interpret. In addition, this restriction led to a drastic decrease of required computation time, as discussed in Section 4.8.

For this experimental design exercise, a new experiment was defined and the optimal sampling times were determined using the different approaches ($T_a$, $T_b$, $T_c$ and $T_d$). In this experiment, three pulses were given, each with a volume of 1 mL. The first pulse was given five minutes after the start of the experiment, and contained ATP and PEP. The ATP concentration in the pulse was chosen such that the ATP concentration in the reactor was raised to 1.5 mM. The concentration of PEP in the pulse was set to 0.5 mM, which was also the case for the other pulses. The second pulse, given ten minutes after the start of the experiment, contained glucose and PEP, and the glucose concentration was chosen such that the resulting concentration was 1.5 mM. The third pulse contained ATP and PEP, and the concentration of the former was set to 3 mM. The experiment stopped after 20 minutes, and only glucose and ADP were measured. These measurements were taken in duplicate, and the minimal relative errors ($\varsigma$) were set to 0.015 for the measured state variables, and the lower accuracy bounds on the measurements were defined as 0.1 mM. Further, the minimum time interval between two measurements was set to 15 seconds.

Using the parameters tabulated in Table 4.2, the two models predict this experiment as shown in Fig. 4.6. To gain insight into the different methodologies, ten scenarios were defined with the number of sampling times ranging from one to ten. The results of the scenario in which ten optimal sampling times were determined will be discussed in Section 4.10.6, and the other scenarios will be discussed in Section 4.10.7.

### 4.10.6 Discussion of the experiment with ten sampling times

The results of the scenario in which ten optimal sampling times were determined are shown in Fig. 4.7. For each of the presented approaches ($T_a$, $T_b$, $T_c$ and $T_d$), the discriminatory potential at a given point in time can be calculated using the appropriate objective function (Eq. (4.2), (4.20) or (4.22)), by assuming that one only samples at that time. The trajectories of the objective functions obtained in this way are shown in Fig. 4.7, as well as the optimal sampling times obtained from them.

The upper graph of Fig. 4.7 shows the trajectory of the objective function used in the $T_a$ approach. This objective function (calculated using Eq. (4.2)) represents the difference between the two model predictions. From the graph, one can see that the difference between the models is rather small, with the largest difference occurring around 12 minutes. So, the optimal sampling times are predominantly located in this region when only the model predictions are considered ($T_a$ approach).

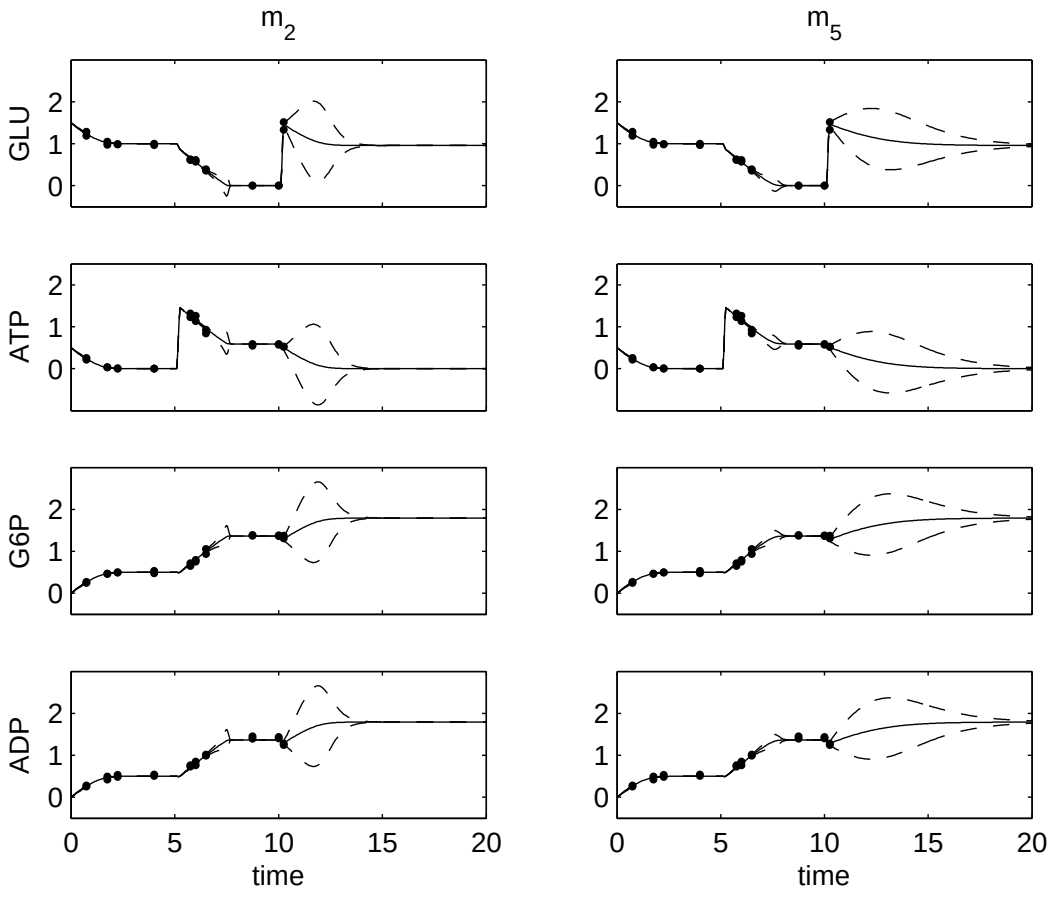



**Figure 4.5:** Preliminary experiment, predicted by model $m_2$ and $m_5$ after estimating their parameters. The dashed lines represent the 95 percent confidence intervals on the model predictions.

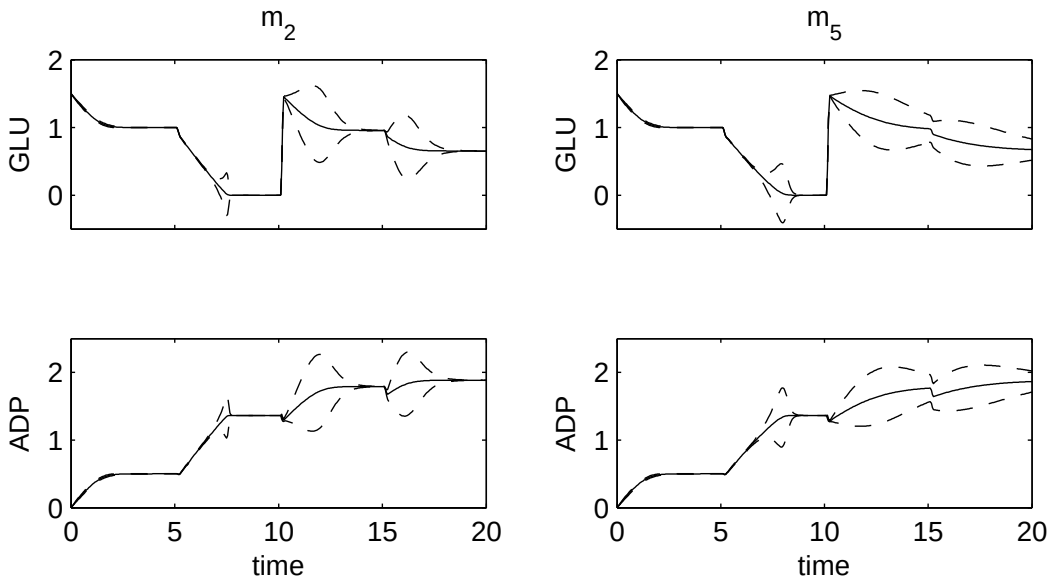



**Figure 4.6:** Rival models predicting the experiment for which the sampling times were optimized. The dashed lines represent the 95 percent confidence intervals on the model predictions.

The second graph of Fig. 4.7 shows the trajectory of the objective function used in the $T_b$ approach, which differs from the $T_a$ approach by the fact that the measurement uncertainty is accounted for. The latter is more or less proportional to the value of the corresponding measured state variables (see Eq. (3.24)), which explains the obtained results. Indeed, from Figs. 4.6 and 4.7, one can see that the largest difference in the model prediction occurs at relatively high values of the measured state variables, which are thus associated with relatively high measurements errors. The optimal sampling times are therefore partly shifted to the regions around 7 minutes and 17 minutes, where the measurement errors are smaller.

In the $T_c$ approach, the model prediction uncertainties (shown in Fig. 4.6) are also taken into account. These uncertainties are quite high, especially at time instants where a large difference between the model predictions is observed. This results in very low values of the objective function (Eq. (4.22)), and thus in a low discriminatory potential of the experiment. In fact, the value of $T_{ij}$, which is obtained as the sum of the values of the individual sampling times, is merely equal to 1.14. According to Chen and Asprey (2003), the experiment is only worth performing when this value is larger than $n_{sp} \cdot n_m$, which is 40 in our example (two measured state variables, two repetitions and ten sampling times).

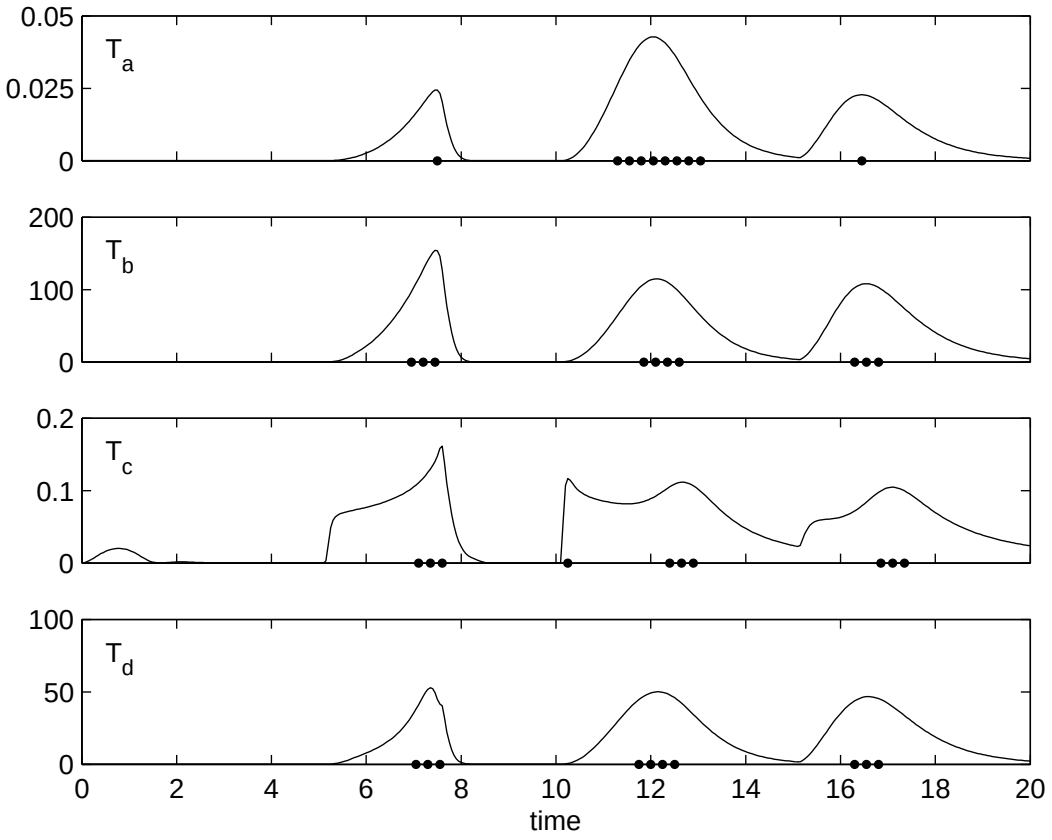



**Figure 4.7:** From top to bottom, these graphs show the trajectory of the objective function obtained when, respectively, the approaches $T_a$, $T_b$, $T_c$ and $T_d$ were applied to determine the optimal sampling times (•) in the scenario with ten samples.

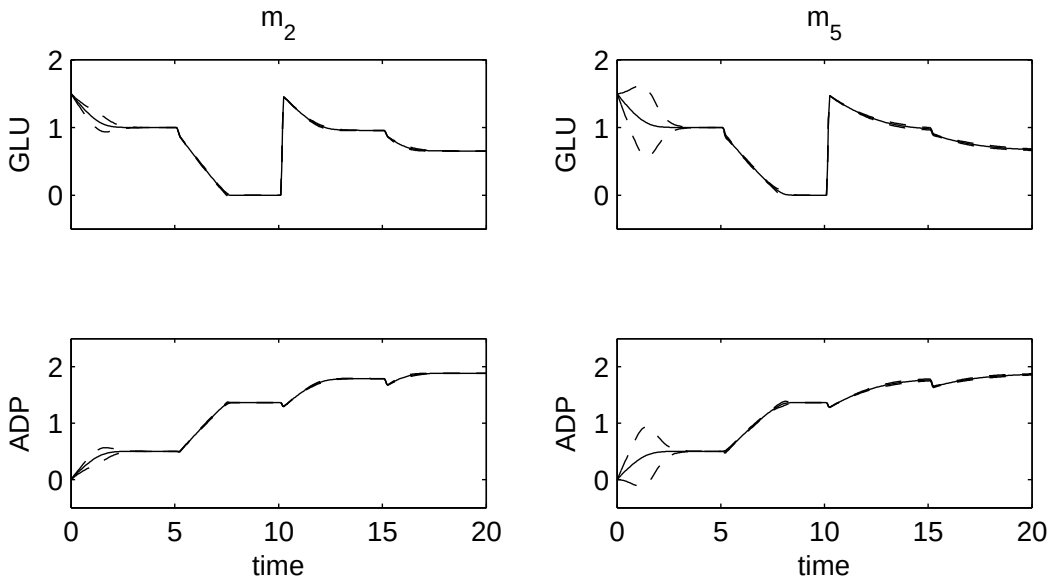



**Figure 4.8:** Rival models predicting the experiment for which the sampling times were optimized. The dashed lines represent the 95 percent confidence intervals on the model predictions as they were used when applying the anticipatory approach.

When applying the anticipatory approach ($T_d$ approach), the parameter and model prediction uncertainties are recalculated for each experiment the optimization algorithm proposes. The optimal sampling times that were obtained in this way are spread out over the three regions where the difference in the model predictions is significant (see Fig. 4.7), which was also the case for the $T_b$ and $T_c$ approach. The model prediction uncertainties associated with this optimal experiment, which were also used when calculating the objective function (Eq. (4.22)), are shown in Fig. 4.8. As can be seen from a comparison with Fig. 4.6, it appears to be possible to significantly reduce the model prediction uncertainty by anticipating for the new data, and an experiment was obtained that resembles the one found with the $T_b$ approach.

### 4.10.7 Apparent similarity between $T_b$ and $T_d$

As discussed in the previous section, the $T_d$ approach seemed to bring forth a similar experiment as the one obtained by the $T_b$ approach, although the latter differs from the anticipatory approach by the fact that it does not consider the uncertainty on the model predictions. The reason why both approaches result in a similar experiment is that the information content of the designed experiment is large enough to decrease the prediction

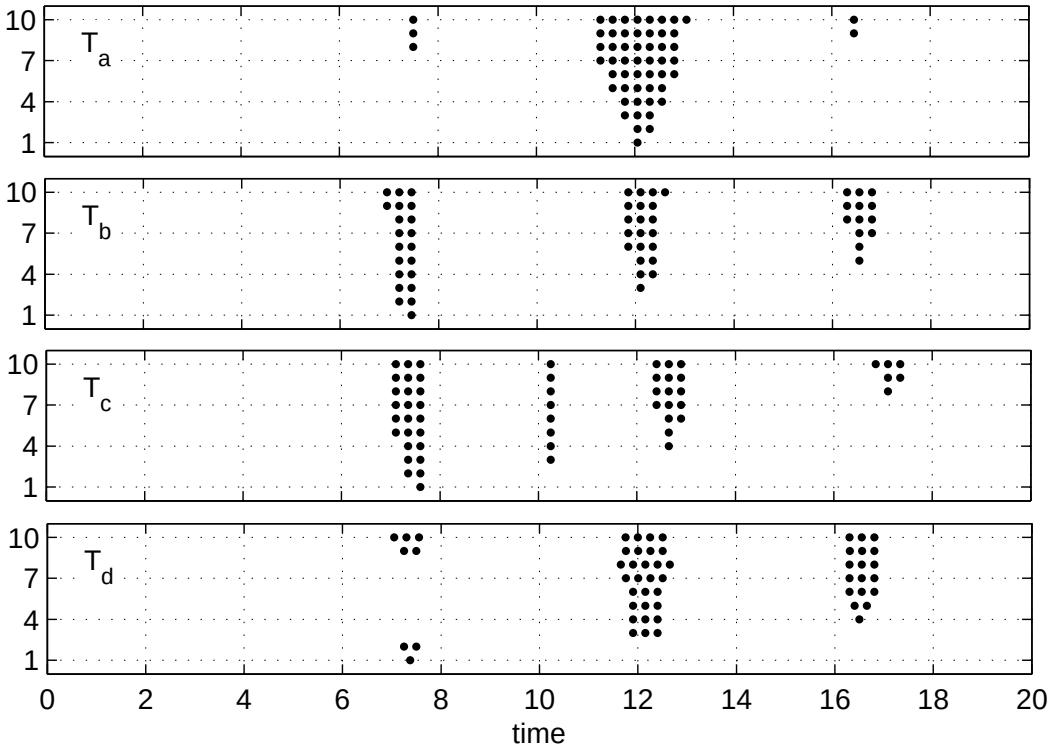



**Figure 4.9:** Optimal sampling times obtained after applying the $T_a$, $T_b$, $T_c$ and $T_d$ approaches for the ten different scenarios in which the number of sampling times ranges from one to ten.

uncertainty to a level where its impact on the experimental design becomes rather limited. To investigate this similarity in more detail, nine additional experimental design exercises were performed, where the number of samples to be optimized ranged from one to nine. The results are shown in Fig. 4.9, together with the scenario with ten samples.

One clearly sees that the similarity between the experiments designed using the $T_b$ and the $T_d$ approach does not always hold. In fact, some interesting observations can be made from Fig. 4.9 that clearly illustrate the anticipatory nature and the power of the $T_d$ approach. For instance, when only one or two samples are optimized, the region with the largest difference in the model predictions (around 12 minutes) cannot be exploited because the uncertainty in this region cannot be sufficiently reduced. However, when three and more samples are optimized, this is no longer the case. In addition, three sampling times appear to be enough, since the fourth, fifth and sixth samples are used to reduce the uncertainty in the region around 17 minutes. Another example that illustrates the difference between

the $T_b$ and the $T_d$ approach is found when comparing the scenarios with seven, eight and nine samples. These show that it appears to be beneficial to place the eighth sample in the region around 12 minutes, rather than placing it in the region around 7 minutes. However, when nine samples are optimized, two samples are placed in this region. Apparently, one sample cannot achieve a sufficient decrease in model prediction uncertainty, while two samples can.

From the discussion above, it is clear that the $T_b$ and the $T_d$ approaches do not always lead to the same experiments. In fact, this is only the case when the information content of the designed experiment is high enough to reduce the model prediction uncertainty beyond a certain level. If that is the case, the impact of $\mathbf{\Omega}_i$ and $\mathbf{\Omega}_j$ in Eq. (4.21) will become so small, that the experimental design will be dominated by the measurement error ($\mathbf{\Sigma}$), and thus a similar experiment will be found as with the $T_b$ approach. It is of course impossible to foresee this, and one cannot know beforehand whether the experiments designed using these approaches will be similar or not. The premise that both approaches lead to similar experiments is thus not a valid one.

### 4.10.8 A note on the estimate of the uncertainty on the model predictions

As advocated in Section 4.4, the uncertainty on the model predictions is important with regard to model discrimination, and more specifically for the design of discriminatory experiments. In principle, both the $T_c$ and the $T_d$ approach will select the experimental degrees of freedom such that the expected model prediction uncertainty is small at the times when samples are taken, thereby increasing the discriminatory potential of the resulting experiment. However, the anticipatory nature of the $T_d$ approach is likely to result in more reliable estimates of the model prediction uncertainties that will eventually be obtained after performing the designed experiment and re-estimating the model parameters.

This is illustrated in Fig. 4.10. Here, the uncertainties on the model predictions of glucose are shown (the results for ADP are similar). For clarity, these are the values that are added to/subtracted from the model predictions to obtain the confidence intervals in figures like Fig. 4.6 and 4.8. Figure 4.10 compares the model prediction uncertainties as they are used in the $T_c$ approach (dashed lines in the upper graphs) and in the $T_d$ approach (dashed lines in the lower graphs) to those obtained after the experiment is performed and the model parameters were re-estimated from the actual data (dotted lines). From this figure, one can clearly see that the model prediction uncertainties used in the experimental design (dashed lines) are closer to the dotted lines when the $T_d$ approach is used. Also note the

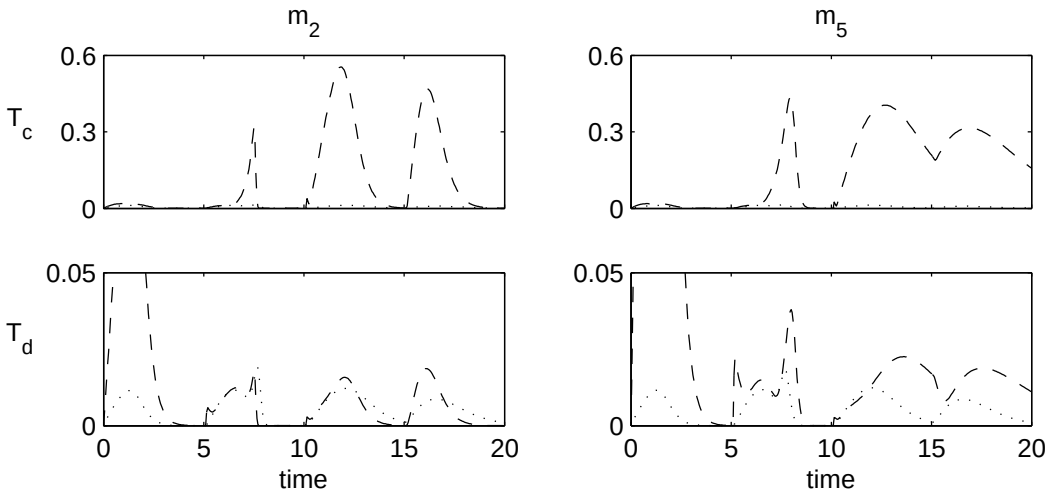



**Figure 4.10:** The 95 percent confidence intervals associated with the model predictions of the glucose concentration as used when designing an experiment according to the $T_c$ approach (dashed lines in the upper graphs), and according to the $T_d$ approach (dashed lines in the lower graphs). The dotted lines represent the confidence intervals obtained after performing the designed experiment and after re-estimating the parameters from the actual data.

difference in the order of magnitude between the confidence intervals (the Y-axes) obtained for the two approaches.

## 4.11 Case study II: Performance of the selected approaches to design discriminatory experiments

In the previous section, a case study was described, the aim of which was to improve our insight into the (four) selected approaches to design optimal discriminatory experiments ($T_a$, $T_b$, $T_c$ and $T_d$). In this section, these four approaches will be applied to a case study where the sequential model discrimination procedure presented in Section 2.7 will be performed until the best model from a set of nine rival models (the ones that were described in Chapter 3) is identified, or until all models are rejected.

### 4.11.1 Objective of this case study

The objective of this case study is to determine whether the four selected approaches differ in their ability to bring forth a series of (informative) discriminatory experiments. In the

evaluation of their performance, four aspects are considered. The first aspect (which will be discussed in Section 4.11.6) has to do with the outcome of the model discrimination procedure, that is, how the procedure ends (most appropriate model identified or all rival models rejected). The second aspect (discussed in Section 4.11.7) is related to the number of additional experiments that have to be (designed and) performed before the most appropriate model can be identified. It is clear that this is very important, as one obviously wants to minimize the number of additional experiments. The third aspect (discussed in Section 4.11.8) is related to the quality of the parameter estimates of the model that is eventually identified as the best one (if any). The fourth and last aspect of the performance evaluation is related to the rate at which the inadequate models are identified and is discussed in Section 4.11.9.

### 4.11.2 Design of the case study

To evaluate the performance of the different approaches, the model discrimination procedure was performed for five different scenarios. The only difference between these scenarios is the preliminary experiment used to initiate the model discrimination procedure (as shown in Fig. 4.11). These preliminary experiments are denoted as $\xi_1^i$, with $i = 1, \ldots, 5$, and are described in Section 4.11.3. This was done because the information content of the preliminary experiment determines the quality of the parameter estimates at the start of the model discrimination exercise. The latter may have an influence on the performance of the different approaches, and this has to be taken into account.

As in the previous case study, the experimental data were generated by simulating the experiment with the real model ($m_5^\star$) and the measurement error was simulated by adding random noise according to Eq. (3.24). Although the first discriminatory experiment that is designed in the model discrimination procedure is the same in each repetition of the procedure, the data obtained from this experiment are different due to these randomly generated measurement errors. Consequently, the discriminatory experiments designed in the following iterations of the procedure are different as well, because the differences in the experimental data sets lead to differences in the parameter estimates. To account for this, each model discrimination exercise was repeated thirty times (as indicated in Fig. 4.11).

In total, the model discrimination procedure discussed in Section 2.7 is performed $5 \times 4 \times 30 = 600$ times. Indeed, there are 5 scenarios (each with a different preliminary experiment), 4 different approaches for OED/MD ($T_a$, $T_b$, $T_c$ and $T_d$) and the model discrimination procedure was repeated 30 times. Note that each application of the model

discrimination procedure consists of an number of iterations, which differs from one repetition to another. In every iteration, an experiment is designed and performed, and the number of iterations is equal to the number of additional experiments that have to be performed before the most appropriate model can be identified. The data obtained from this simulation study will be presented and discussed in the following sections.

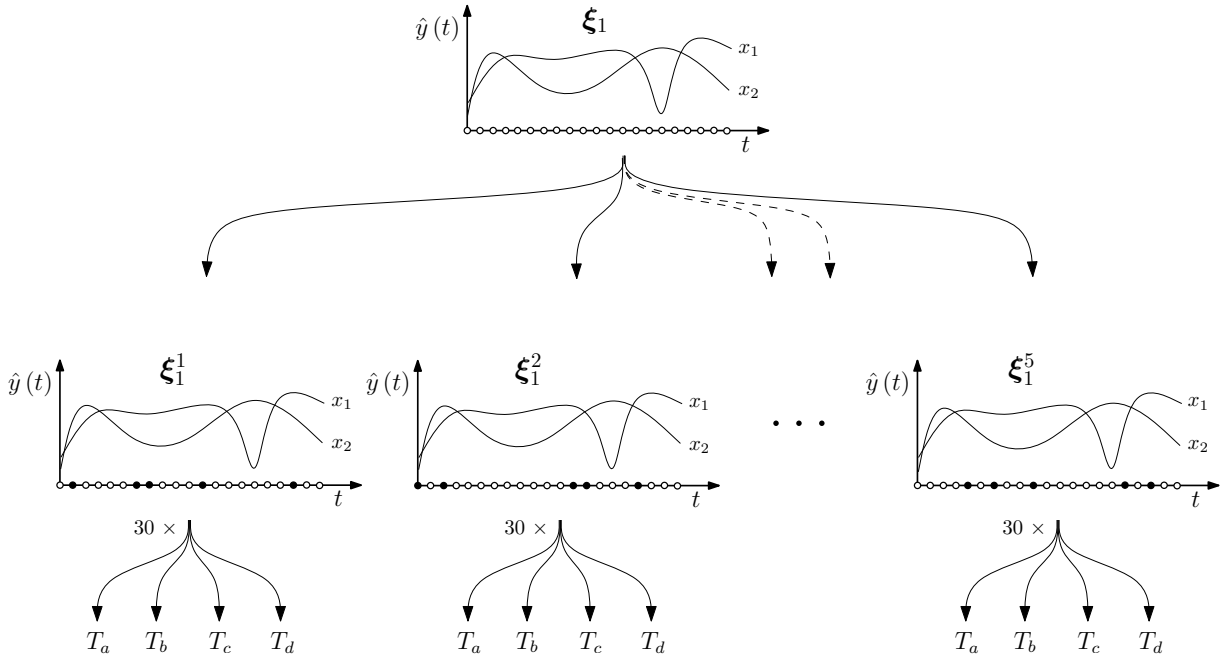



**Figure 4.11:** Illustration of the design of the case study (Section 4.11), where the aim was to evaluate and compare the performance to discriminate among rival models.

### 4.11.3 Preliminary experiments

The five preliminary experiments were all identical to the one used in the previous case study (Fig. 4.4), except for the ten sampling times which were chosen randomly. This resulted in five experiments with a different information content with regard to the parameters of the individual rival models, as assessed by calculating the so-called D-optimality criterion value. This D-optimality criterion is discussed in detail in Chapter 6, but for now it is sufficient to know that a large D-optimality criterion value indicates that the experiment contains a lot of information on the parameters of the corresponding model. An experiment with a large D-optimality criterion value is thus preferred. The criterion values associated with the five preliminary experiments are shown in Fig. 4.12. For the in-

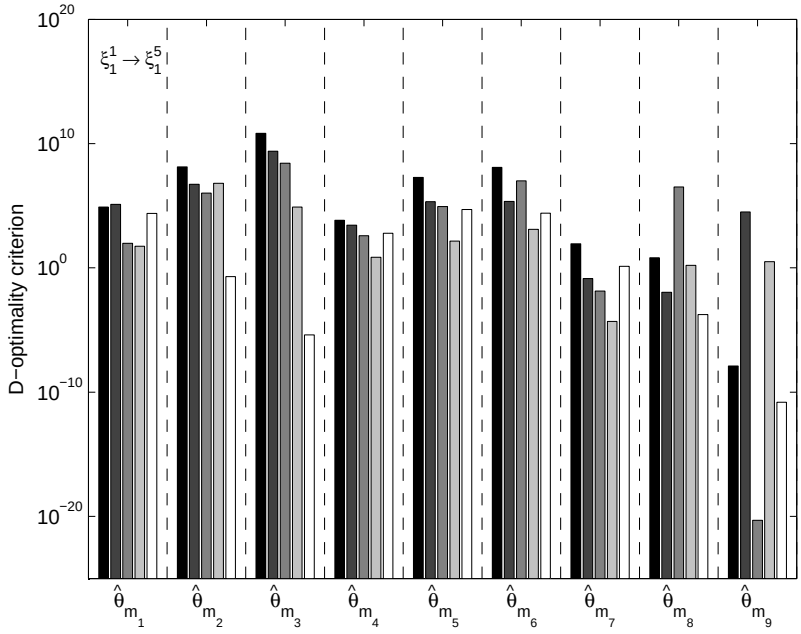



**Figure 4.12:** D-optimality criterion values for the different rival models for the five preliminary experiments ($\boldsymbol{\xi}_1^1$ till $\boldsymbol{\xi}_1^5$).

terpretation of the result depicted in this figure, one should be aware that the D-optimality criterion values reflect the information content with regard to the parameters of one particular model, and the criterion values obtained for different models can thus not simply be compared.

### 4.11.4 Parameter estimation

The parameters of the rival models were estimated using the data from the preliminary experiments, using the optimization algorithm described in Section 2.11. As in the previous case study, the lower bounds were set to zero and the upper bounds were set to 1000 U/mg for parameter $k$, 2 mM for parameter $K_{\mathrm{GLU}}$, and 25 mM for parameters $K_{\mathrm{ATP}}$ and $K_{\mathrm{PEP}}$.

### 4.11.5 Design of optimal discriminatory experiments

For the design of the optimal discriminatory experiments, the experimental degrees of freedom were the same as in the preliminary experiment except for the sampling times and the initial concentrations of glucose, ATP and PEP, which were optimized. The initial concentrations were allowed to take values between 0 mM and 2 mM, and ten optimal

sampling times were determined with the constraint that the time interval between two subsequent samples was 15 seconds or more. As in the previous case study (Section 4.10.4), the $\chi^2$ lack-of-fit test described in Section 2.8.1 was used to evaluate the adequacy of the rival models.

The design criteria described in Section 4.4 were developed for model discrimination problems with two rival models. However, when the number of rival mathematical models is larger than two (as in this case study), several strategies are thinkable to steer the model discrimination procedure (Buzzi-Ferraris et al., 1990; Schwaab et al., 2006), irrespective of which design criterion is chosen. For this case study, the so-called pairwise strategy is chosen. Here, an optimal discriminatory experiment is designed for each model pair, and the experiment with the largest $T_{ij}$ value is eventually performed.

### 4.11.6 Outcome of the model discrimination procedure

Ideally, the model discrimination procedure ends when one of the rival models is identified as the most appropriate one. In this case study, the experimental data were generated using model $m_5^\star$ and it can thus be expected that model $m_5$ is identified as the most appropriate model in the majority of the runs, regardless of the approach used to design the discriminatory experiments. However, the possibility that another model is identified as the most appropriate one cannot be excluded. A second possible outcome of the model discrimination procedure is that all models appear to be inadequate. Indeed, the adequacy of the rival models is evaluated based on the WSSE value, and even for the true model ($m_5^\star$) this WSSE value can in some situations be larger than the reference value ($\chi^2_{n-n_p}$) because of the (simulated) error on the measurements. Note that a third possibility, where the discriminatory potential of the designed experiment is conceived as too low to enable further discrimination among the remaining rival models, is not considered here as the correctness of the currently used criterion to evaluate the discriminatory potential of an experiment is questionable (as discussed in Section 4.4.7).

The results obtained in this case study are presented in Table 4.3. For the $T_d$ approach, for instance, one can see that model $m_5$ was identified as the most appropriate one in 134 of the 150 applications (or runs) of the model discrimination procedure (5 scenarios with a different preliminary experiment and 30 repetitions of each scenario). The results indicate that also for the other approaches the *true* model ($m_5^\star$) is found in most of the runs. From the results obtained for the $T_d$ approach, on can also see that another model was identified as the best model in 7 percent of the runs of the model discrimination

procedure. For brevity, it is not indicated which of the other rival models is eventually identified as the best model, but, in the majority of the runs where this occurred, model $m_2$ was selected. This is not surprising because models $m_2$ and $m_5$ only differ by the fact that the former assumes a random binding mechanism, whereas the latter assumes an ordered reaction mechanism. In other words, the models are very similar. In the other runs of the procedure, all rival models were rejected. It is, however, noteworthy that the true model was always identified as the most appropriate one when the $T_c$ approach was used. Although a profound explanation for this observation cannot be given, it might be the result of the conservative character of the $T_c$ approach, which will be discussed in the following.

### 4.11.7 Required number of experiments to achieve model discrimination

The number of additional experiments that have to be performed before the most appropriate model can be identified is an important aspect that should be taken into account when evaluating the performance of a certain OED/MD approach. Before starting the discussion of the obtained results, it is interesting to note that even with fifty randomly generated experiments it was not possible to identify the most appropriate model, while model discrimination could be achieved in far less experiments with any of the selected OED/MD approaches. The fifty random experiments were generated by randomly choosing the sampling times (between 0 and 20 min) and the initial concentrations of glucose, ATP and PEP (between 0 and 2 mM). This result clearly illustrates the necessity or at least the importance of performing experiments that are designed in a rational way, that is, designed with the aim to achieve model discrimination.

As explained earlier (Section 4.11.2), the model discrimination procedure was initiated with one of the five preliminary experiments and each of these was repeated thirty times to account for the influence of the measurement error. The number of experiments that were required in the different model discrimination runs are presented as boxplots in Fig. 4.13. This figure contains five subfigures with a white background (entitled $\boldsymbol{\xi}_1^i$, with $i = 1, \ldots, 5$) and one subfigure with a gray background (entitled $\boldsymbol{\xi}_1^1 - \boldsymbol{\xi}_1^5$). The former presents the results obtained for the simulations where the model discrimination procedure was initiated with the preliminary experiment indicated in the title of the corresponding subfigure, whereas the one with the gray background gives an overall picture of the number of required experiments and presents the values of all model discrimination runs ($4 \times 150 = 600$ in total). Note that in these figures, the preliminary experiment corresponds to experiment

**Table 4.3:** Overview of the observed outcomes of the 150 runs of the model discrimination procedure for the different approaches ($T_a$, $T_b$, $T_c$ and $T_d$) and the five scenarios (each with a different preliminary experiment, $\boldsymbol{\xi}_1^i$ with $i = 1, \ldots, 5$).

| | | $T_a$ | $T_b$ | $T_c$ | $T_d$ |
|---|---|---|---|---|---|
| model $m_5$ | $\boldsymbol{\xi}_1^1$ | 30 | 30 | 30 | 27 |
| | $\boldsymbol{\xi}_1^2$ | 30 | 27 | 30 | 29 |
| | $\boldsymbol{\xi}_1^3$ | 28 | 27 | 30 | 25 |
| | $\boldsymbol{\xi}_1^4$ | 30 | 20 | 30 | 27 |
| | $\boldsymbol{\xi}_1^5$ | 28 | 24 | 30 | 26 |
| | $\boldsymbol{\xi}_1^1 - \boldsymbol{\xi}_1^5$ | **146** *97%* | **128** *85%* | **150** *100%* | **134** *89%* |
| other model | $\boldsymbol{\xi}_1^1$ | 0 | 0 | 0 | 0 |
| | $\boldsymbol{\xi}_1^2$ | 0 | 3 | 0 | 1 |
| | $\boldsymbol{\xi}_1^3$ | 0 | 3 | 0 | 4 |
| | $\boldsymbol{\xi}_1^4$ | 0 | 5 | 0 | 3 |
| | $\boldsymbol{\xi}_1^5$ | 1 | 2 | 0 | 2 |
| | $\boldsymbol{\xi}_1^1 - \boldsymbol{\xi}_1^5$ | **1** *1%* | **13** *9%* | **0** *0%* | **10** *7%* |
| all models rejected | $\boldsymbol{\xi}_1^1$ | 0 | 0 | 0 | 3 |
| | $\boldsymbol{\xi}_1^2$ | 0 | 0 | 0 | 0 |
| | $\boldsymbol{\xi}_1^3$ | 2 | 0 | 0 | 1 |
| | $\boldsymbol{\xi}_1^4$ | 0 | 5 | 0 | 0 |
| | $\boldsymbol{\xi}_1^5$ | 1 | 4 | 0 | 2 |
| | $\boldsymbol{\xi}_1^1 - \boldsymbol{\xi}_1^5$ | **3** *2%* | **9** *6%* | **0** *0%* | **6** *4%* |

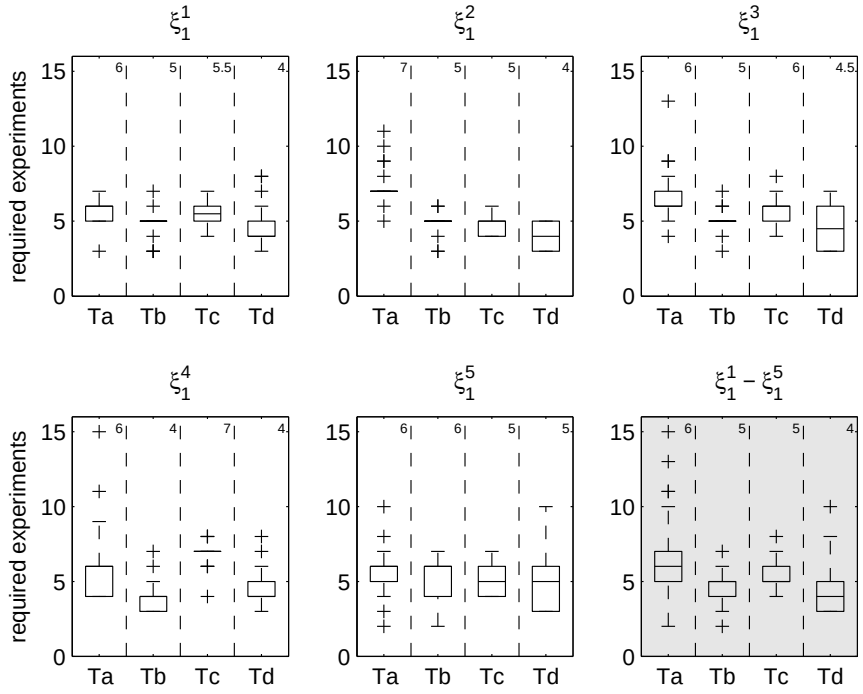



**Figure 4.13:** Boxplots showing the number of experiments that are required to achieve model discrimination starting from each of the five preliminary experiments ($\boldsymbol{\xi}_1^1$ till $\boldsymbol{\xi}_1^5$) (white background). The subfigure with the gray background gives an overall picture of the number of required experiments and were made using the results obtained from all starting situations. The numbers in the upper right corners represent the median of the number of required experiments.

number one. In other words, Fig. 4.13 shows the number of required experiments, and not the number of required *additional* experiments. Also note that the median of the number of required experiments, which will be used frequently in the discussion below, can not always be determined unambiguously from these boxplots (more precisely, when the horizontal line that indicates the median coincides with one of the edges of the box). Therefore, the median of the number of required experiments is also given in the upper right corners.

From the results shown in Fig. 4.13, one can see that the highest number of required experiments occurs when the $T_a$ approach is used, regardless of the information content of the preliminary experiment. As discussed earlier, the $T_a$ approach is the most naive one. Therefore, the discriminatory potential of the proposed experiments is often misjudged, and it is not surprising that the $T_a$ approach is the worst performing one when it comes to

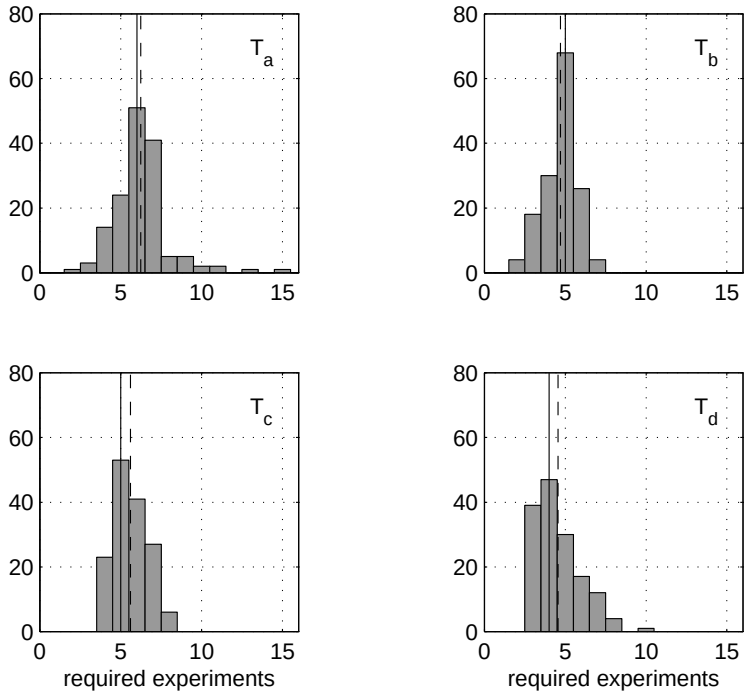



**Figure 4.14:** Histograms showing the number of experiments required to achieve model discrimination for each of the selected approaches ($T_a$, $T_b$, $T_c$ and $T_d$), thereby using the results obtained from all starting situations (gray figure from Fig. 4.13). The solid vertical lines indicate the median of the number of required experiments, while the dashed lines indicate the mean number of required experiments.

the required number of experiments. In addition, one can see that the number of required experiments varies significantly among the repetitions. The latter can be observed for each of the starting situations, except for $\boldsymbol{\xi}_1^1$. This large variability can be explained by the fact that the $T_a$ approach allows to take samples when the measurement error is high. Indeed, although the first discriminatory experiment is the same for each of the runs that were initiated with a certain preliminary experiment, the generated data sets of the other experiments can differ significantly because of this measurement error. Since the model parameters are estimated from these data sets, large differences in these parameter estimates can be expected. These obviously affect the model discrimination runs that follow and lead to the observed variability among the different repetitions. This variability also reflects the importance to include the information on the measurement errors when designing the discriminatory experiment.

For both the $T_b$ and the $T_c$ approaches, the median of the number of required experiments equals five and one could conclude that both approaches perform equally well. That these approaches perform better than the $T_a$ approach can be seen as an illustration of the importance of considering the measurement errors in the design of the experiments, and it confirms some of the conclusions drawn from the results for the $T_a$ approach discussed above. However, a close(r) investigation of the boxplots in Fig. 4.13 and especially of the histograms in Fig. 4.14 indicates that the $T_b$ approach is preferred over the $T_c$ approach. Although the median of the number of required experiments is the same, the distributions are clearly different and in favor of the $T_b$ approach. Knowing that the $T_c$ approach has frequently been applied in literature (for instance by Burke et al. (1995, 1996); Kremling et al. (2004) and Schwaab et al. (2006)) and that it was originally introduced as a conceptual improvement of the $T_b$ approach, this result is somewhat surprising and may be related to the fact that the information content of the designed experiment in not fully considered during the design. This will be further discussed below, together with the results of the $T_d$ approach.

The anticipatory approach ($T_d$) does take the information content of the to-be-performed experiment into account and performs better than the other approaches, regardless of the information content of the preliminary experiment. Indeed, for each of the starting situations, model discrimination was achieved with the least amount of experimental effort. However, one can see that the variability among the different repetitions is slightly larger than the variability observed for the $T_c$ approach. The latter can be explained as follows. Both approaches use the currently available parameter estimates to predict the outcome of the proposed experiment and the uncertainty associated with it, but the $T_c$ approach is more conservative than the $T_d$ approach because it only uses the information of the already performed experiments to evaluate the proposed experiment for its discriminatory potential. In other words, the $T_d$ approach is more sensitive to the accuracy of the available parameter estimates, but when the available parameter estimates are close to their actual values, the discriminatory potential of the designed experiment is assessed in a better way compared to the other approaches (as discussed in the previous case study). However, when the parameter estimates used in the experimental design differ significantly from the ones obtained after performing the designed experiment, the discriminatory potential of the experiment may not be as good as one expected. If this occurs, the designed experiment may not significantly contribute to the discriminatory potential of the experiments performed so far, and, in the end, an increased number of required experiments will be required before the most appropriate model can be identified. This might explain why a large tail can be

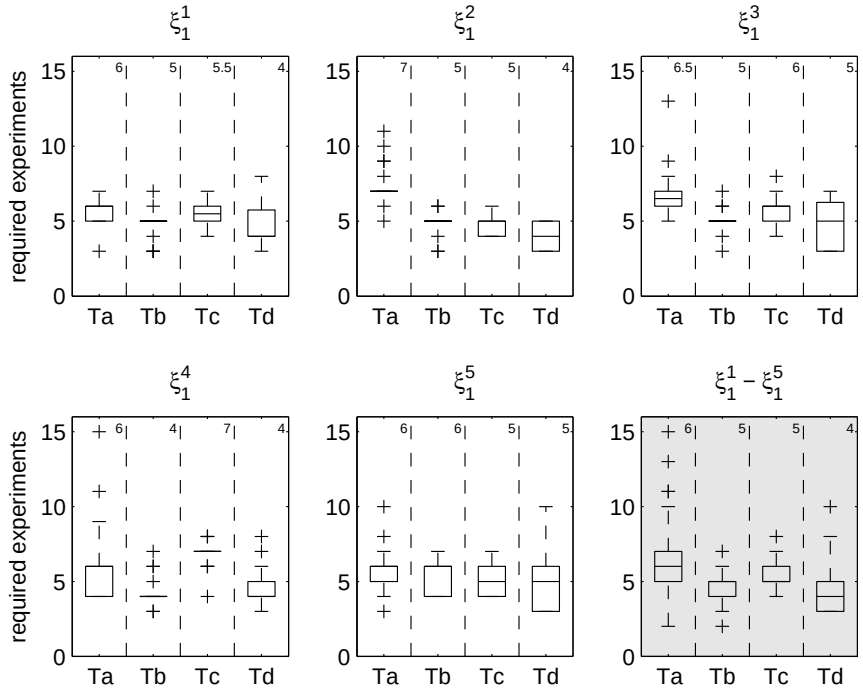



**Figure 4.15:** Boxplots showing the number of experiments that are required to identify the most appropriate model starting from each of the five preliminary experiments ($\boldsymbol{\xi}_1^1$ till $\boldsymbol{\xi}_1^5$) (white background). The subfigure with the gray background gives an overall picture of the number of required experiments and were made using the results obtained from all starting situations. The numbers in the upper right corners represent the median of the number of required experiments.

observed for the histogram of the $T_d$ approach. Nevertheless, the results in Figs. 4.13 and 4.14 show that the $T_d$ approach generally results in faster model discrimination compared to the $T_c$ approach, which may indicate that, at least in this case study, the parameter estimates did not change considerably during the model discrimination procedure.

To conclude this discussion, note that Figs. 4.13 and 4.14 represent the number of experiments that are performed until the model discrimination procedure stops, as explained in Section 4.11.6, regardless of the outcome. In this respect, one could argue that only those runs should be considered in the evaluation where model $m_5$ was identified as the most appropriate model, but the results nor the discussion are significantly influenced when doing so (as shown in Figs. 4.15 and 4.16). In other words, the runs in which, for instance, all rival models are rejected are not systematically those for which a high or low number of experiments are required.

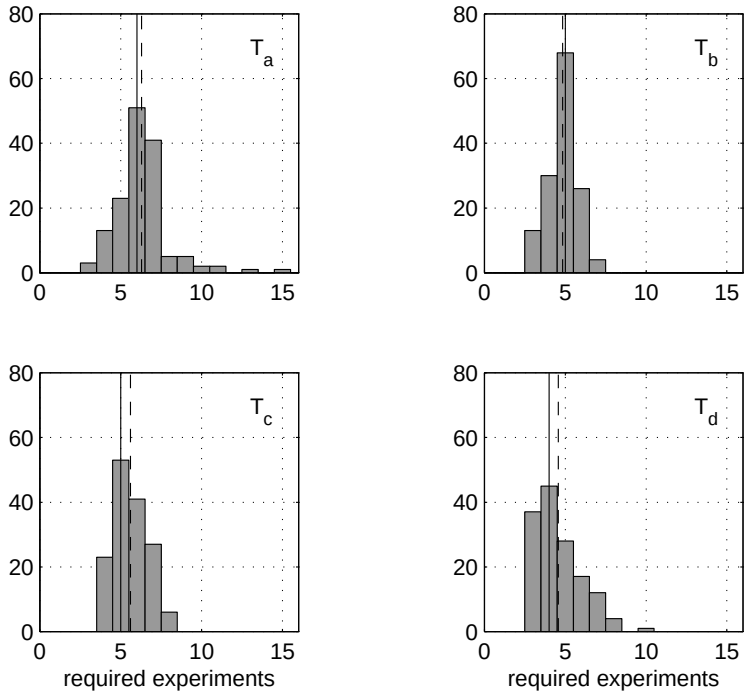



**Figure 4.16:** Histograms showing the number of experiments required to identify the most appropriate model for each of the selected approaches ($T_a$, $T_b$, $T_c$ and $T_d$), thereby using the results obtained from all starting situations (gray figure from Fig. 4.15). The solid vertical lines indicate the median of the number of required experiments, while the dashed lines indicate the mean number of required experiments.

### 4.11.8 Evaluation of the quality of the parameter estimates

A third aspect that has to be considered in the evaluation of the approaches is the quality of the parameter estimates, and especially the quality of the parameter estimates of the model that is eventually identified as the most appropriate one. Indeed, model discrimination is only one step of a more general model building procedure, and once an appropriate model is identified through model discrimination, the quality of its parameter estimates often has to be improved before the model can actually be applied for its intended use. This is important because inaccurate parameter estimates result in inaccurate (or uncertain) model predictions, which are obviously not desired. To increase the quality of the parameter estimates, dedicated experiments have to be designed (using the experimental design techniques (OED/PE) explained in Chapter 6) and performed. In this respect, it would be interesting to see whether there is a difference among the approaches with regard

to the evolution of the quality of the parameter estimates throughout the model discrimination procedure because this may have an influence on the overall required number of experiments. In other words, if the parameter estimates are already of a high quality when the model discrimination procedure ends, less additional experiments will be required to refine the parameter estimates afterwards.

The value of the D-optimality criterion, briefly described in Section 4.11.2, is used to evaluate the quality of the parameter estimates. This criterion is frequently used in optimal experimental design for parameter estimation (OED/PE) to quantify the information content of an experiment with regard to the parameters of a particular model. Its value is inversely proportional to the volume of the confidence region of the parameter estimates, and experiments that are characterized by a large criterion value are thus expected to bring forth more accurate parameter estimates than experiments with a small criterion value. For more information on the D-optimality criterion, the reader is referred to Chapter 6 where this and other design criteria for OED/PE purposes are described in more detail.

In Fig. 4.17, the evolution of the D-optimality criterion values of model $m_5$ are shown in gray for each of the (thirty) runs that were initiated with experiment $\boldsymbol{\xi}_1^1$ and where the $T_d$ approach was used to design the experiments. At first sight, it may seem strange that the criterion values are not (always) monotonically increasing with the number of performed experiments. Indeed, the D-optimality criterion value represents the information content of the set of experiments and one would expect that this information content can only increase when new information is collected from an additional experiment. However, the decreases in the criterion values can be explained by the fact that the model parameters are re-estimated after performing the new experiment. Since the D-optimality criterion value is dependent on these parameter estimates, a non-monotonic profile can be obtained.

From Fig. 4.17 and from the discussion held in the previous section, it is clear that the number of required experiments differs among the runs. When the results are represented as in Fig. 4.17, their interpretation would be hampered and it would be difficult to compare the results obtained for the different starting situations and with the different approaches to OED/MD. Therefore, the results will be presented differently in the following. The median of the D-optimality criterion values will be used to visualize how the quality of the parameter estimates changes during the model discrimination procedure. The number of criterion values from which the median was determined will be indicated and will also be reflected in the size of a bullet symbol (•). The results obtained when the model discrimination procedure was initiated with $\boldsymbol{\xi}_1^1$ and $\boldsymbol{\xi}_1^4$ are presented accordingly in Figs. 4.18 and

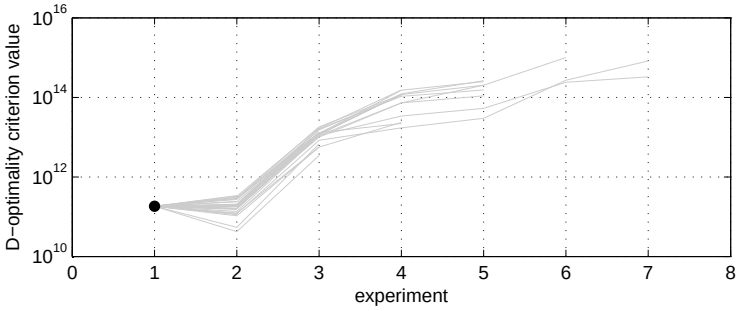



**Figure 4.17:** Evolution of the D-optimality criterion values of model $m_5$ (in gray) for each of the (thirty) runs of the model discrimination procedure that was initiated with experiment $\boldsymbol{\xi}_1^1$.

4.19, respectively. The results obtained starting from the other preliminary experiments are similar and are shown in Chapter A. In these figures, the median of the D-optimality criterion values for a certain approach to OED/MD is shown in black, while the values of the other approaches are shown in gray to facilitate their mutual comparison.

From these figures, one can conclude that the $T_a$, $T_b$ and $T_d$ approaches perform better than the $T_c$ approach, in the sense that the rate at which the D-optimality criterion values (or the quality of the parameter estimates) increases is faster. Indeed, one can see that, except for the runs starting from $\boldsymbol{\xi}_1^2$, the $T_c$ approach is the worst performing one, and its performance is significantly worse (Figs. 4.18 and 4.19) when the model discrimination procedure is initialized with preliminary experiments $\boldsymbol{\xi}_1^1$ and $\boldsymbol{\xi}_1^4$. The similarity in the performance of the $T_a$, $T_b$ and $T_d$ approaches agrees with what was observed and concluded in the previous case study. Indeed, it indicates that the information one expects to collect by performing the designed experiment will reduce the uncertainty on the model predictions such that the experimental design becomes primarily driven by the difference in the model predictions (and the uncertainty on the measurements). Still, one can state that the $T_d$ approach performs slightly better than the other ones, or, in other words, the $T_d$ approach generally results in experiments with a larger information content with regard to the parameter estimates compared to the other approaches.

That the $T_c$ approach can result in a poor performance with regard to the quality of the parameter estimates, is in agreement with the concepts from which it is derived. As the $T_c$ approach seeks a balance between the difference in the model predictions and the uncertainty associated with it, it will obviously avoid to take samples where the uncertainty

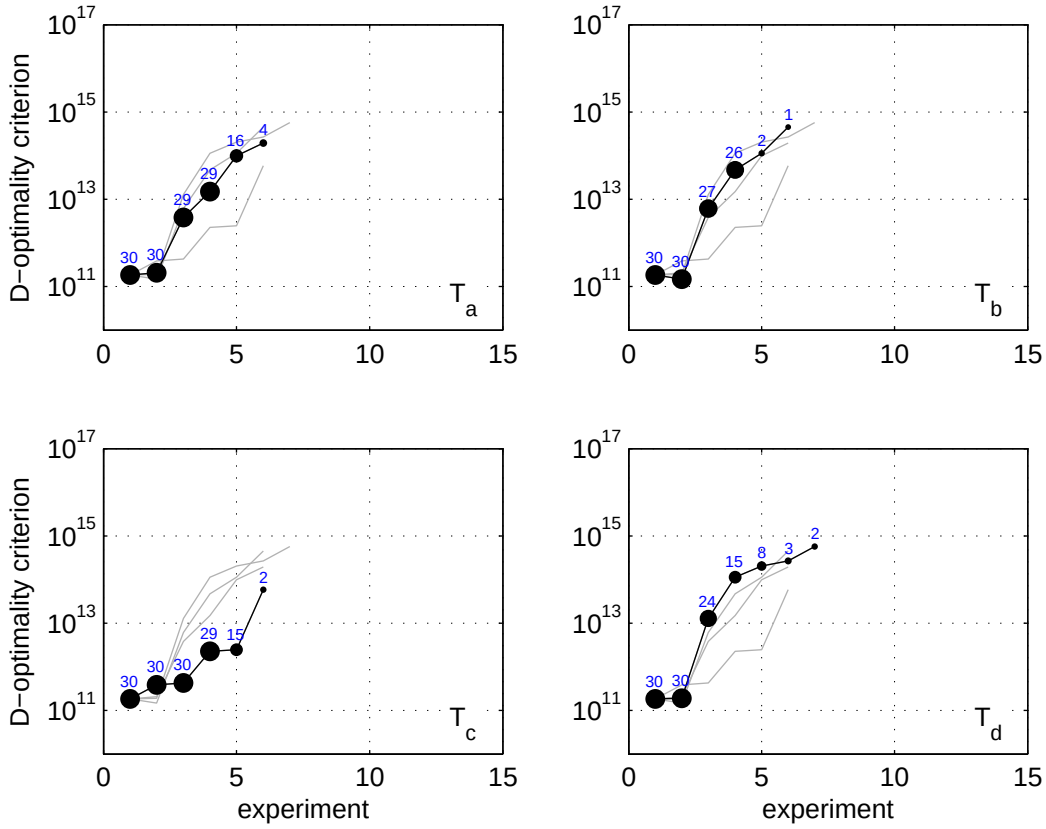



**Figure 4.18:** Evolution of the median D-optimality criterion values of model $m_5$ for the (thirty) runs of the model discrimination procedure that was initiated with experiment $\boldsymbol{\xi}_1^1$, for each of the selected approaches for OED/MD. The evolution of the median criterion values of the other approaches are shown in gray to ease the comparison. The number of criterion values from which the median was determined will be indicated by the size of a bullet symbol and the corresponding integer.

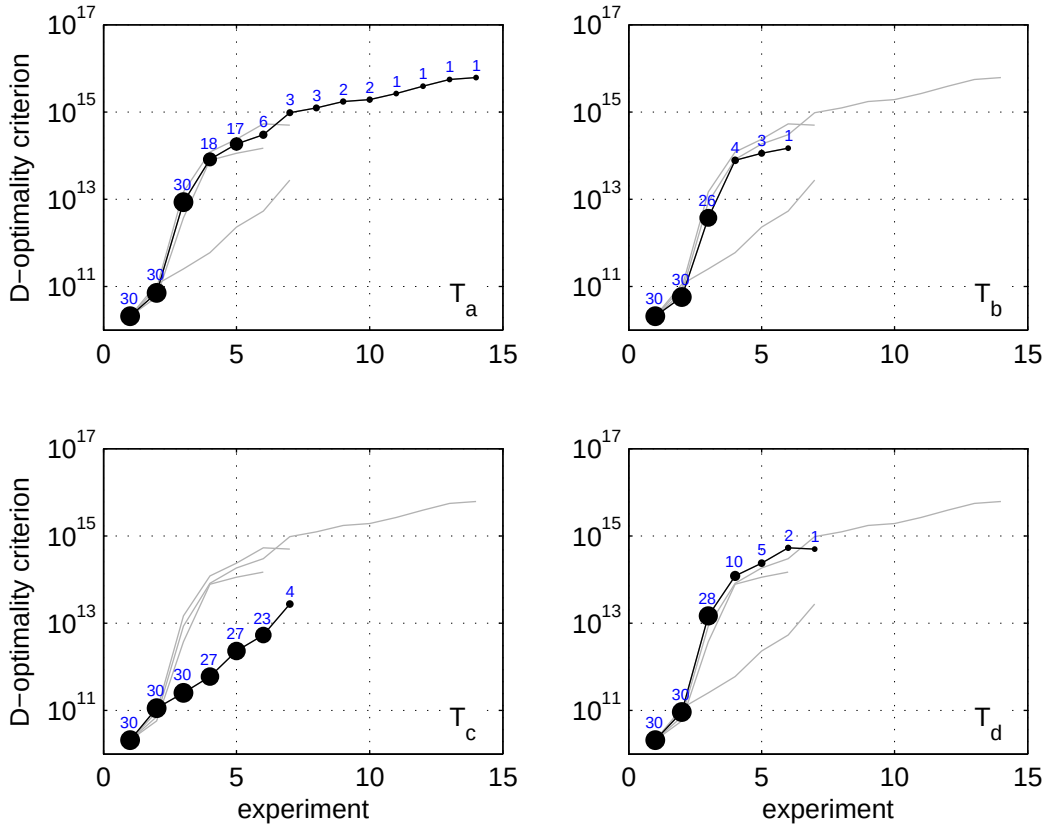


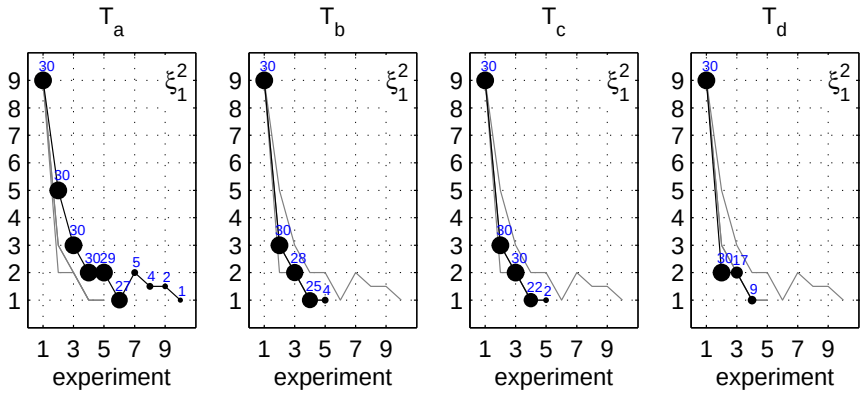

**Figure 4.19:** Evolution of the median D-optimality criterion values of model $m_5$ for the (thirty) runs of the model discrimination procedure that was initiated with experiment $\boldsymbol{\xi}_1^4$, for each of the selected approaches for OED/MD. The evolution of the median criterion values of the other approaches are shown in gray to ease the comparison. The number of criterion values from which the median was determined will be indicated by the size of a bullet symbol and the corresponding integer.

on the model predictions is (too) large. However, since the latter is to a large extent determined by (and often even correlates with) the sensitivities of these predictions to the values of the model parameters, this also has an impact on the information content of the designed experiment. Indeed, as will be explained in more detail in Chapter 6, the highest information content with regard to the model parameters are found where these sensitivities are large. In other words, the $T_c$ approach will not exploit the information present in regions where the model prediction uncertainty is large, unless the difference in the model predictions is significantly larger.

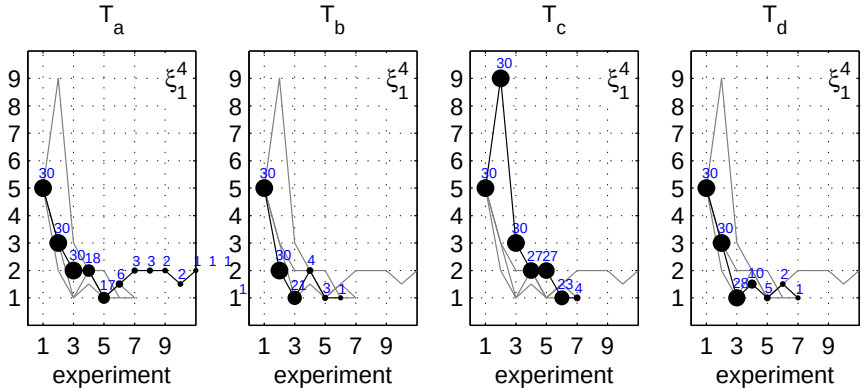
When the $T_d$ approach is used, the information that will be collected on the model parameters when performing the designed experiment is already considered in the evaluation of its discriminatory potential. This (small) conceptual difference with the $T_c$ approach has important consequences (as already briefly stated in Section 4.6). Indeed, experiments that are informative with regard to the model parameters will indirectly contribute to a reduction of the uncertainty on the model predictions. Therefore, the balance between the difference in model predictions and the uncertainty associated with it will shift towards the former compared to the $T_c$ approach. In other words, the regions where the information content with regard to the model parameters is highest, will more likely be exploited by the $T_d$ approach, whereas they will be avoided by the $T_c$ approach.

### 4.11.9 Rate at which inadequate models are identified

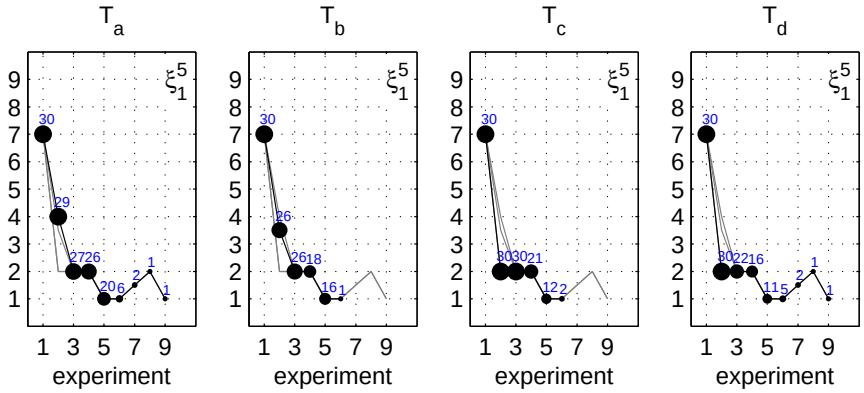
In the discussion above, it was assumed that both time and money were available to perform experiments until the most appropriate model was identified. However, in practice, these resources may be limited and the model discrimination procedure must be stopped after a particular number of experiments. In this respect, it is important to look at the rate at which inadequate models are identified. Indeed, when the model discrimination procedure is stopped before the most appropriate model is identified, it has to be selected from the remaining adequate models. It is clear that model selection (as explained in Section 2.9) is less challenging when the number of models to choose from is limited. To investigate this aspect of the performance evaluation, the median value of the number of adequate models obtained for the different runs of the model discrimination procedure are shown in Figs. 4.20a, 4.20b and 4.20c for the scenarios with preliminary experiments $\boldsymbol{\xi}_1^2$, $\boldsymbol{\xi}_1^4$ and $\boldsymbol{\xi}_1^5$, respectively. Note that the results obtained when the model discrimination procedure is initialized with the other preliminary experiments are similar and can be found in Appendix A.




(a) starting from preliminary experiment $\boldsymbol{\xi}_1^2$




(b) starting from preliminary experiment $\boldsymbol{\xi}_1^4$




(c) starting from preliminary experiment $\boldsymbol{\xi}_1^5$

**Figure 4.20:** The median values of the number of adequate models as a function of the number of experiments that have been performed, starting from preliminary experiment $\boldsymbol{\xi}_1^2$ (a), $\boldsymbol{\xi}_1^4$ (b) and $\boldsymbol{\xi}_1^5$ (c). The evolution of these median values shows the rate at which the number of adequate models decreases for the different approaches to design optimal discriminatory experiments ($T_a$, $T_b$, $T_c$ and $T_d$). The number of values (runs) from which the median was determined will be indicated by the size of a bullet symbol and the corresponding integer.

**Table 4.4:** The median value of the number of adequate models obtained after performing the first discriminatory experiment designed using the different approaches ($T_a$, $T_b$, $T_c$ and $T_d$). To increase the interpretability of these results, the lowest median values for a given preliminary experiment are indicated in bold.

| preliminary experiment | $T_a$ | $T_b$ | $T_c$ | $T_d$ |
|---|---|---|---|---|
| $\boldsymbol{\xi}_1^1$ | **2** | **2** | **2** | **2** |
| $\boldsymbol{\xi}_1^2$ | 5 | 3 | 3 | **2** |
| $\boldsymbol{\xi}_1^3$ | 3 | **2** | **2** | **2** |
| $\boldsymbol{\xi}_1^4$ | 3 | **2** | 9 | 3 |
| $\boldsymbol{\xi}_1^5$ | 3 | 3.5 | **2** | **2** |

From the results shown in Fig. 4.20, one can clearly see that, in general, the largest decrease in the number of adequate models is achieved after performing the first optimal discriminatory experiment. Therefore, the median of the number of adequate models after performing the first designed experiment is tabulated in Table 4.4 and will be used to facilitate the discussion. Although the results presented in Fig. 4.20 and Table 4.4 are rather inconclusive, one can still observe that the $T_a$ approach is the worst performing approach in most of the cases (although its performance is not bad). The performance of the $T_b$ and the $T_c$ approaches is comparable, while the $T_d$ approach performs slightly better than the other ones.

Note that an interesting remark can be made when looking at the results presented in Fig. 4.20b. Here, one can see that the number of adequate models increases from five to nine after performing the first discriminatory experiment designed according to the $T_c$ approach. Although this observation only has a minor influence on the performance evaluation of the different approaches, it is an interesting one because it clearly indicates the need to reconsider all models when new experimental data becomes available and illustrates that one should be aware of the possibility that a good model can accidentally be appointed as inadequate. It is therefore recommended to reconsider all models after performing new(ly designed) experiments.

## 4.12 Discussion of the work of Schwaab et al. (2008)

As already noted in Section 4.6, the idea of the anticipatory approach (that is, taking into account the information content of the newly designed experiment when evaluating its discriminatory potential) was developed simultaneously and independently by Schwaab et al. (2008) and applied to a number of case studies taken from the context of chemical engineering. Although their approach is identical to the one presented here and the conclusions drawn from their case studies, in general, agreed with ours, some remarks are given below.

However, before starting the discussion, it is important to note that the type of models considered in the work of Schwaab et al. (2008) differs from the ones considered here. Indeed, in this dissertation, dynamic models were considered. Such models describe the modelled process by a set of (coupled) differential and algebraic equations (as discussed in Section 2.2) and make it possible to describe the evolution of the modelled process over time, as already shown in the case studies described above. The models considered in Schwaab et al. (2008), on the other hand, are so-called static (or steady state) models. Such models are not intended to describe the dynamic behavior of the process, and can, in fact, be considered as a subclass of the type of models considered here (they are only described by algebraic equations, as in Eq. (2.2)). Although the experimental design methods are the same for both types of models, it should be noted that the computational requirements to simulate an experiment with dynamic models are much larger since a numerical solver is necessary to obtain the model predictions (as discussed in Section 2.2).

A first remark is related to the interpretation of the $T_{ij}(\boldsymbol{\xi})$-values. Schwaab et al. (2008) state that the main advantage of the anticipatory approach is that the $T_{ij}(\boldsymbol{\xi})$-value obtained with the original approach of Buzzi-Ferraris et al. (1984) (the $T_c$ approach) is lower than the one obtained with the anticipatory approach (the $T_d$ approach), which indicates an "*easier discrimination among the rival models*". Although we agree with the authors that the experiments obtained with the anticipatory approach typically have a higher discriminatory potential (as discussed in the previous section/case study), we do not agree that this can be concluded by comparing the $T_{ij}(\boldsymbol{\xi})$-values obtained with the two approaches. From the discussion held in Section 4.6, it is clear that one can expect a smaller value of $T_{ij}(\boldsymbol{\xi})$ when the information of the designed experiment is taken into account. The following thought experiment illustrates that the statement of Schwaab et al. (2008) is incorrect. Suppose that both approaches result in exactly the same discriminatory experiment. Since

the $T_{ij}(\boldsymbol{\xi})$-value calculated as prescribed by the anticipatory approach will be larger than with the original approach of Buzzi-Ferraris et al. (1984), one would conclude that the anticipatory approach results in an experiment with a larger discriminatory potential. But this is not true, as the experiment is identical.

A second remark has to do with the similarity between the anticipatory approach and optimal experimental design for parameter estimation. This similarity was already discussed in Section 4.6, and the results from the case study discussed in Section 4.11 confirmed that regions with the highest information content about the model parameters will more likely be exploited by the anticipatory approach, whereas they will be avoided by the approach proposed by Buzzi-Ferraris et al. (1984). Schwaab et al. (2008) have also noted this and state in their discussion that "*the use of [the anticipatory approach] allows for simultaneous improvement of model discrimination and parameter estimation, as pursued by many researches in the field.*" In their conclusions, they state that "*... [the anticipatory approach] leads to simultaneous improvement of the variances of the parameter estimates along the experimental design, as reduction of the variances of the parameter estimates also contributes to the increase of the model discrimination power*". However, although these statements are in agreement with what was observed in this work, they should not be seen as an absolute truth. Indeed, although the uncertainty on the parameter estimates plays its role in the design of discriminatory experiments, the latter is primarily driven by the difference in the model predictions. It may thus well be that the informative regions with regard to the parameters do not coincide with the regions that are interesting with regard to model discrimination. In such cases, the optimal discriminatory experiment may not be informative at all with respect to the parameters and no significant improvement of the parameter estimates would be observed. In addition, it should be kept in mind that the experiments are designed based on the predictions of both rival models. The regions (or experiments) that are informative for model $m_i$ may not be informative at all for model $m_j$, and also in this case their statements may not be valid.

## 4.13 Summary and conclusions

Conceptually, the design of optimal discriminatory experiments comes down to finding that experiment that maximizes the difference between the model predictions, preferably taking into account the uncertainty associated with this difference. This uncertainty originates from two sources, that is, the uncertainty on the measurements and the uncertainty on the model predictions. Hunter and Reiner (1965) proposed a design criterion that did not con-

sider any of these uncertainties, but their design criterion could be easily modified to take into account the uncertainty on the measurements. The design criterion proposed by Buzzi-Ferraris et al. (1984) does consider both sources of uncertainty and is thus the superior one from a conceptual point of view. However, in the original approach of Buzzi-Ferraris et al. (1984), the discriminatory potential of the designed experiment is evaluated using the information content of the already performed experiments. In this chapter, this approach was modified by taking into account the information content of the newly designed experiment when evaluating its discriminatory potential, and was therefore called the anticipatory approach to optimal experimental design for model discrimination (OED/MD). These approaches to OED/MD were applied in two case studies in which two and nine of rival models, respectively, were proposed to describe the kinetics of an enzyme.

In a first case study, the experimental design exercise consisted of determining the optimal sampling times for a given dynamic profile of the manipulatory variables. The results showed that the anticipatory approach arranges the experimental degrees of freedom such that the expected model prediction uncertainty is small at the times when samples are taken, thereby increasing the discriminatory potential of the resulting experiment. The results also showed that applying the anticipatory approach can result in an experiment that is similar to the one found with the simpler and less computationally expensive approach that only considers the uncertainty on the measurements. This was the case when the information content of the designed experiment was large enough to decrease the model prediction uncertainty to a level where its impact on the experimental design was rather limited, and the experimental design became dominated by the uncertainty due to the measurement error. In addition, the results showed that, compared to the original approach, the anticipatory approach led to more reliable estimates of the model prediction uncertainties that are eventually obtained after performing the designed experiment and re-estimating the model parameters. This is important because when this estimate is far from the one eventually obtained after performing the designed experiment, its discriminatory potential may not be as high as expected during the experimental design.

In a second case study, the performance of the selected approaches was evaluated by looking at four aspects: (1) the outcome of the model discrimination procedure, (2) the number of experiments that were required before the model discrimination procedure ended, (3) the evolution of the uncertainty on (or the quality of) the parameter estimates during the model discrimination procedure, and (4) the rate at which the number of adequate models decreases. The results indicated that it definitely makes sense to design discriminatory ex-

periment (regardless of the approach used to design them), as model discrimination could not be achieved from random experiments. One could also conclude that the approach proposed by Buzzi-Ferraris et al. (1984) appeared to be a rather conservative one. The true model was always identified as the most appropriate one when this approach was used, but, on average, more experiments were required compared to the other approaches. In addition, the information content (with regard to the parameter estimates) of the experiments designed using this approach was often lower than those obtained from the other approaches. With the anticipatory approach, on the other hand, model discrimination was achieved in the lowest number of experiments, and it generally resulted in experiments with a larger information content compared to the other approaches. The reason why the anticipatory approach performs better than the original approach of Buzzi-Ferraris et al. (1984) is related to the uncertainty on the parameter estimates, and can be explained by the similarity between the anticipatory approach and optimal experimental design for parameter estimation (both design criteria benefit from a larger information content of the designed experiment with regard to the model parameters). In addition, the rate at which the inadequate models were identified was largest for the anticipatory approach. Based on the results obtained in this case study, one can conclude that the anticipatory approach to design optimal discriminatory experiments is preferred.

# CHAPTER 5

## Evaluation of experimental design strategies to discriminate among several rival models

*"However beautiful the strategy, you should occasionally look at the results."*

Winston Churchill, former Prime Minister of the United Kingdom

## Abstract

Because the experimental design methods developed to discriminate among rival mathematical models are model-based, the optimal discriminatory experiments depend on the currently available parameter estimates. As a consequence of this dependence on the parameter estimates, model discrimination has to be seen as an iterative procedure. In every iteration, an optimal discriminatory experiment is designed using only those models that are considered adequate at that stage. When the number of rival models is larger than two, several design strategies are possible to steer the model discrimination exercise, irrespective of which design criterion is chosen. In this chapter, a number of design strategies that are interesting from a conceptual point of view are formalized and applied to a case study. The results indicate that the design strategy in which the average discriminatory potential of the designed experiment is maximized, is to be preferred.

## 5.1 Introduction

When more than one mathematical model is proposed for one and the same system, one is confronted with a model discrimination problem. Typically, one is interested only in that model that describes the process under study in the most appropriate way, and additional experiments are necessary to identify this most appropriate model. For this purpose, one of the design criteria described in the previous chapter can be used to design so-called optimal discriminatory experiments. Basically, the design comes down to finding that experiment that maximizes the difference between the model predictions, preferably taking into account the uncertainty associated with this difference.

In this respect, it is important to be aware of the fact that the experimental design is model-based and thus dependent on the currently available parameter estimates. It can therefore not be guaranteed that one or more rival models will fail to describe the newly collected data. Indeed, the parameter estimates used for the experimental design may not reflect their *true* values and may change considerably when re-estimating them using the new information that is collected from the process being studied. Nevertheless, given the information available at the experimental design stage, performing this discriminatory experiment is the best one can do.

As a consequence of this dependency on the parameter estimates, model discrimination has to be seen as an iterative procedure. This model discrimination procedure has already been described in Section 2.2, but for clarity this is briefly repeated here (Fig. 5.1). In every iteration, an optimal discriminatory experiment is designed using only those models that are considered adequate at that stage. After performing the designed experiment, the parameters of all models (and not only the adequate ones) are re-estimated (Buzzi-Ferraris et al., 1990) and the adequacy of the models is evaluated after taking the newly collected information (or experimental data) into account. By reconsidering all models during the model evaluation phase, one accounts for the possibility that one (or more) models were wrongly rejected by chance in a previous iteration (Buzzi-Ferraris et al., 1990).

When the number of rival mathematical models is larger than two, several design strategies are thinkable to steer the model discrimination exercise (Buzzi-Ferraris et al., 1990; Schwaab et al., 2006), irrespective of which design criterion is chosen. Indeed, one can opt for a design strategy that tries to eliminate as many models as possible in each iteration, or one can choose a design strategy that maximizes the expectation that (at least) one of the rival models becomes inadequate after performing the experiment. In this chapter,

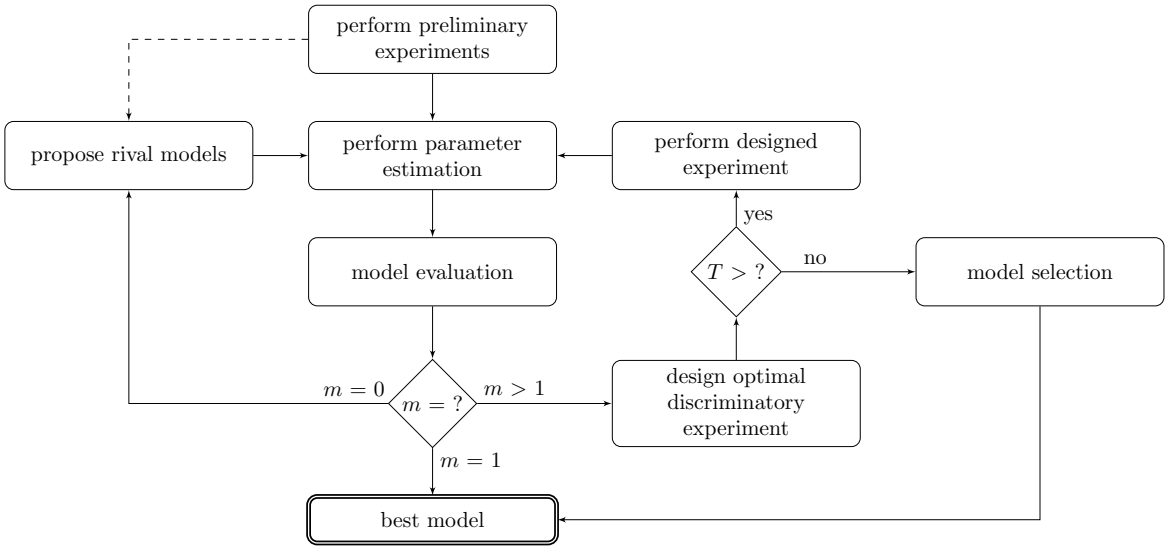



**Figure 5.1:** General procedure to discriminate among $m$ rival models (adapted from Chen and Asprey (2003) and Schwaab et al. (2006)).

a number of design strategies that are interesting from a conceptual point of view are formalized and applied to a case study.

## 5.2 OED/MD as an optimization problem

Before discussing the different design strategies for model discrimination, the general idea of optimal experimental design for model discrimination is briefly repeated. In general, optimal experimental design is an optimization problem, where the optimum of a well-defined objective function, denoted as $T$, is sought by varying the experimental degrees of freedom. This can be formalized as follows

$$\boldsymbol{\xi}^{\star} = \arg\max_{\boldsymbol{\xi}\in\boldsymbol{\Xi}} \; T\left(\boldsymbol{\xi}\right) \; . \tag{5.1}$$

The experimental degrees of freedom, $\boldsymbol{\xi}$, are restricted by a number of constraints that define a set of possible experiments, denoted as $\boldsymbol{\Xi}$. These constraints are determined by the experimental setup and are specified before the start of the experimental design exercise. So, the optimal discriminatory experiment is thus that experiment for which $T\left(\boldsymbol{\xi}\right)$ is maximal.

## 5.3 Strategies for the design of optimal discriminatory experiments

In the following, three design strategies will be described. Each of these design strategies originates from a particular view on how the discriminatory potential of the designed experiment should be defined. Indeed, one can, for instance, design an experiment by maximizing the expectation that at least one of the rival models becomes inadequate after performing the experiment. By adopting this rationale of designing the discriminatory experiments, one tries to avoid situations where the number of adequate models stays the same after performing the experiment. This design strategy corresponds to the so-called pairwise design strategy discussed in Section 5.3.1. Two other design strategies will be discussed below.

### 5.3.1 Pairwise design strategy

As a first design strategy (Buzzi-Ferraris et al., 1990; Schwaab et al., 2006), an optimal discriminatory experiment is designed for each model pair. When $m$ models are proposed, this means that $\frac{m!}{(m-2)!\cdot 2!}$ experiments have to be designed. From these experiments, the one associated with the highest $T_{ij}(\boldsymbol{\xi})$ value, represented as $\boldsymbol{\xi}^\star$ in the following, will eventually be performed. This design strategy can be formalized as follows:

$$\boldsymbol{\xi}^\star = \arg \max_{\substack{i=1,\ldots,m-1 \\ j=i+1,\ldots,m}} T_{ij}\left(\boldsymbol{\xi}_{ij}^\star\right) , \tag{5.2}$$

where $T_{ij}\left(\boldsymbol{\xi}_{ij}^\star\right)$ represents the $T_{ij}(\boldsymbol{\xi})$ value associated with the optimal discriminatory experiment to discriminate between model $m_i$ and model $m_j$, which is denoted as $\boldsymbol{\xi}_{ij}^\star$ and calculated as

$$\boldsymbol{\xi}_{ij}^\star = \arg \max_{\boldsymbol{\xi} \in \boldsymbol{\Xi}} \quad T_{ij}(\boldsymbol{\xi}) . \tag{5.3}$$

From a conceptual point of view, one could state that this design strategy maximizes the expectation that one of the rival models becomes inadequate after performing the experiment and one could state that one tries to avoid the situation in which the number of adequate models does not change after performing the experiment.

The experiment that is eventually obtained using this design strategy may be very good to discriminate between model $m_i$ and $m_j$, but not for the other model pairs. Indeed, besides

the fact that the discriminatory potential of the optimal discriminatory experiments for the individual model pairs is mutually compared (as formalized in Eq. (5.2)), the other model pairs are not considered in the design of experiment $\boldsymbol{\xi}_{ij}^{\star}$, as can be seen from Eq. (5.3). As the focus of the experimental design is primarily on that model pair for which model discrimination is most likely, this design strategy might not lead to a fast reduction of the number of adequate models and many iterations might be required before the most appropriate model can be identified from the $m$ rival models available at the start of the model discrimination procedure.

### 5.3.2 Averaged design strategy

The fact that the pairwise design strategy ignores the discriminatory potential of the experiment with regard to the other model pairs, may have as a consequence that more experiments are required to achieve model discrimination (as discussed above). This disadvantage is less pronounced in the averaged design strategy, where the optimal discriminatory experiment is found after optimizing the average of the design criterion values obtained for each of the model pairs (Schwaab et al., 2006; Burke et al., 1997; Dumez, 1977). This design strategy can be formalized as follows

$$\boldsymbol{\xi}^{\star} = \arg\max_{\boldsymbol{\xi}\in\boldsymbol{\Xi}} \quad \sum_{i=1}^{m-1}\sum_{j=i+1}^{m} T_{ij}\left(\boldsymbol{\xi}\right) . \tag{5.4}$$

Note that, in principle, one should divide the right-hand side of Eq. (5.4) by the number of model pairs to get a true average. However, this was omitted for simplicity because it does not have an influence on the optimum of the design criterion, and thus the same optimal discriminatory experiment is found.

### 5.3.3 Maximin design strategy

For both design strategies presented above, a situation may occur where the discriminatory potential of the designed experiment is very small for one or more model pairs. That this may occur was already mentioned and discussed for the pairwise design strategy, but it might also occur for the averaged design strategy. Indeed, it is possible that the value of $\sum_{i=1}^{m-1}\sum_{j=i+1}^{m} T_{ij}\left(\boldsymbol{\xi}\right)$ is dominated by one or more model pairs, while the discriminatory potential of the proposed experiment is (too) small for the other model pairs. This can be avoided by determining the optimal discriminatory experiment as follows:

$$\boldsymbol{\xi}^{\star} = \arg\max_{\boldsymbol{\xi}\in\boldsymbol{\Xi}} \quad \min_{\substack{i=1,\ldots,m-1 \\ j=i+1,\ldots,m}} T_{ij}\left(\boldsymbol{\xi}\right) \ . \tag{5.5}$$

In this so-called maximin design strategy, the $T_{ij}$-value is calculated for each model pair and the smallest one is maximized. In this way, one should be protected from a situation where the discriminatory potential of the designed experiment is too small for one or more model pairs.

### 5.3.4 Focus on the best model pairs

To complement the design strategies discussed above, it is interesting to note that one can also choose to focus the experimental design on the best models. Indeed, one can assume that the worst performing models will eventually be found inadequate as the number of performed experiments increases. Following this rationale, the optimal discriminatory experiment is predominantly determined by the most probable models and may lead to a reduction of the number of required experiments since no experimental effort is spent on the discrimination among rival models that will drop out anyway.

Such weighing can be performed using the model probabilities introduced by Schwaab et al. (2006) (described in Section 4.4.5, and more specifically in Eq. (4.16)). For instance, the design criterion used in the averaged design strategy then becomes

$$\boldsymbol{\xi}^{\star} = \arg\max_{\boldsymbol{\xi}\in\boldsymbol{\Xi}} \quad \sum_{i=1}^{m-1}\sum_{j=i+1}^{m} \pi_i \cdot \pi_j \cdot T_{ij}\left(\boldsymbol{\xi}\right) \ , \tag{5.6}$$

where $\pi_i$ represents the model probability associated with model $m_i$. Apart from mentioning this possibility, this will not be dealt with in more detail in this dissertation.

## 5.4 Case study III: Evaluating the performance of the strategies for OED/MD

In this section, the discriminatory potential of the experiments designed by applying the different design strategies described above will be examined. Therefore, the model discrimination procedure depicted in Fig. 5.1 will be applied to the working example described in Chapter 3, where nine rival models are proposed to describe the kinetics of glucokinase.

### 5.4.1 Objective of this case study

The objective of this case study is to compare the performance of the three strategies in the design of optimal discriminatory experiments. In particular, the objective is to find out whether the rationale that is incorporated in their corresponding mathematical formulations (or equations) is also reflected in the performance of the designed experiments. The latter will be evaluated based on (1) the rate at which model discrimination is achieved (as in the previous case study, described in Section 4.11) and (2) on the overall discriminatory potential of the designed experiment.

### 5.4.2 Preliminary experiments

As in the case study described in Section 4.11, the model discrimination procedure was performed starting from the five preliminary experiments, denoted as $\xi_1^i$, with $i = 1, \ldots, 5$, and described in Section 4.11.3. As in the previous case study, this was done to account for the difference in the information content of the preliminary experiments. The latter has an influence on the quality of the parameter estimates at the start of the model discrimination procedure, and thus on the model discrimination itself.

### 5.4.3 Design of optimal discriminatory experiments

For the design of the optimal discriminatory experiments, the experimental degrees of freedom were the same as in the preliminary experiments except for the sampling times and the initial concentrations of glucose and ATP, which were to be optimized. The initial concentrations were allowed to take values between 0 mM and 2 mM, and ten optimal sampling times were determined with the constraint that the time interval between two subsequent samples was 15 seconds or more. However, for reasons that will become clear during the discussion of the results, the duration of the designed experiments was limited here to 6 minutes, whereas this is 20 minutes in the other case studies that are performed in this dissertation. For the same reasons, the initial concentration of PEP was set to 0 mM, but the liquid used for the pulse contained PEP with a concentration of 1.2 mM. Since the anticipatory approach showed the best performance in the case studies discussed in the previous chapter, this approach was used to design the discriminatory experiments.

### 5.4.4 Rate at which inadequate models are identified

The objective of optimal experimental design for model discrimination is to identify the most appropriate model with as few additional experiments as possible. A design strategy that identifies the inadequate models faster (that is, in less experiments) is thus preferred and the rate at which the inadequate models are identified is investigated in this section. Note that this aspect becomes especially relevant when the resources (for instance, time and money) are scarce and only a limited number of experiments can be performed, as already discussed in Section 4.11.9. In such a case, the model discrimination procedure must be halted before the most appropriate model is identified and one of the remaining adequate models is selected. The latter is obviously less problematic when the number of adequate models is small.

To investigate the rate at which model discrimination is achieved, the number of adequate models is presented as a function of the number of (designed) experiments that have been performed. As one can expect that the difference between the design strategies is primarily apparent in the beginning of the model discrimination procedure (and thus when there are still a large number of adequate rival models), the procedure was stopped after five iterations. Similar to the case study described in Section 4.11, the model discrimination procedure was repeated thirty times to account for the uncertainty on the measurements (as discussed in Section 4.11.2).

Fig. 5.2 shows the evolution of the number of adequate models (using histograms) obtained for the scenario with preliminary experiments $\boldsymbol{\xi}_1^1$, $\boldsymbol{\xi}_1^2$ and $\boldsymbol{\xi}_1^3$ and for which the experiments are designed according to the pairwise design strategy. To make a comparison among the design strategies, the median value of the number of adequate models obtained for the different runs of the model discrimination procedure are shown in Fig. 5.3 for the scenarios with preliminary experiments $\boldsymbol{\xi}_1^1$, $\boldsymbol{\xi}_1^2$ and $\boldsymbol{\xi}_1^3$ and for the three design strategies. In these figures, experiment number one corresponds with the preliminary experiment. Note that the results obtained with the other preliminary experiments are similar and are therefore not discussed here for brevity. They can, however, be found in Appendix B.

From these results, one can see that some of the rival models already appear to be unable to describe the data obtained in this preliminary experiment. This is, for instance, the case in Fig. 5.3a, where the number of adequate models is reduced to only four (out of nine) after performing the preliminary experiment. For the interpretation of the results depicted in Fig. 5.3, it is important to keep in mind that the number of values from which

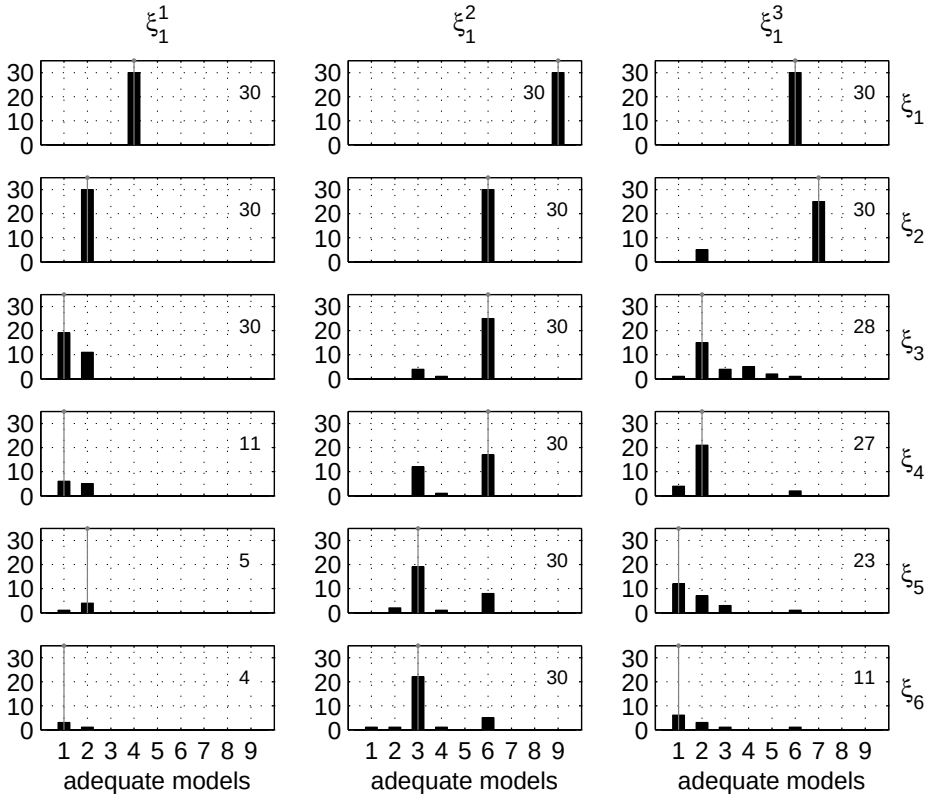


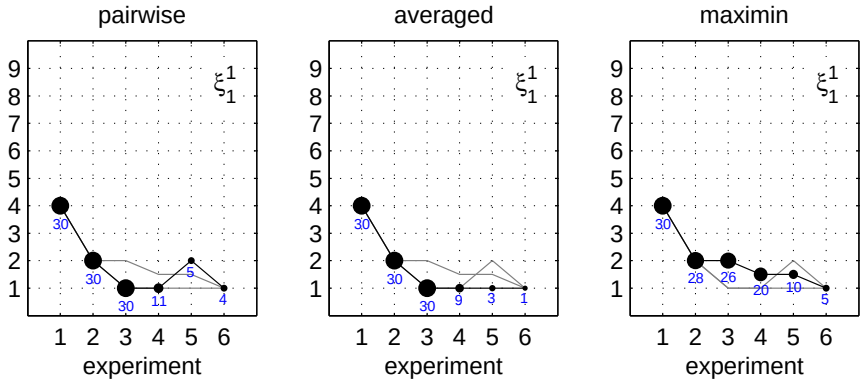

**Figure 5.2:** Evolution of the number of adequate models as experimentation progresses (top to bottom) by means of histograms determined from the (thirty) runs of the model discrimination procedure that were initiated with experiments $\boldsymbol{\xi}_1^1$ (left), $\boldsymbol{\xi}_1^2$ (middle) and $\boldsymbol{\xi}_1^3$ (right), and for which the experiments are designed according to the pairwise design strategy. The number of runs in which model discrimination has not been achieved yet is indicated in the upper right corner of each graph and the median of the number of adequate models is represented by the gray line.

the median was determined, differs (as indicated by the size of the bullet and the integer accompanying it). This is because the number of experiments that is required to achieve model discrimination varies between the different repetitions of the model discrimination procedure (similar to the results reported in Fig. 4.17).

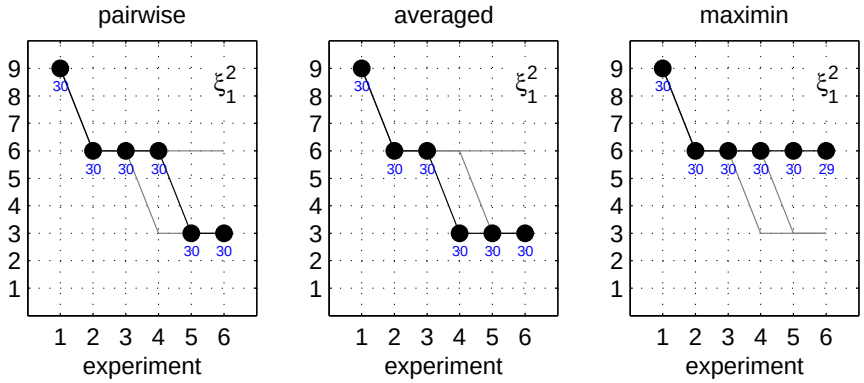
From the results presented in Fig. 5.3, one can also see that the number of adequate models decreases slightly faster when the averaged design strategy is applied instead of the two other design strategies, although the difference between the pairwise and the averaged design strategy is minor. The maximin design strategy seems to be the worst performing one, which is most apparent when looking at the results obtained with preliminary experiment $\boldsymbol{\xi}_1^2$ (Fig. 5.3b). However, one can observe that these results do not allow to draw a definite conclusion.

In Fig. 5.3a, one can see that the median value increases from one to two after performing the fifth experiment in the scenario where the pairwise strategy is used for the experimental design. However, even though it is possible that the number of adequate models increases after performing a new experiment, this was not the case here and the observation can be explained by the fact that the number of values from which the median is determined differs as experimentation progresses. This phenomenon can more easily be understood from Fig. 5.2 where the evolution of the number of adequate models is presented by means of histograms. An example where the number of models does increase after performing a new experiment is found in Figs. 5.3c and 5.2 (right column), for the scenario where the pairwise design strategy was used. After performing the first designed experiment, the median of the number of adequate models increases from six to seven even though none of the thirty repetitions of the model discrimination procedure ended yet.

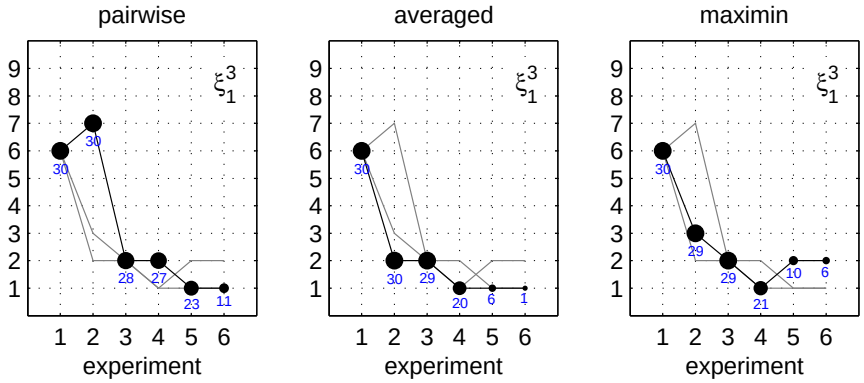
Note that the results presented in Fig. 5.3a do not contribute to the study of the difference between the three design strategies. Indeed, regardless of the design strategy used, the number of adequate models drops from four to two after performing the first designed experiment. When the number of rival models is equal to two, the three design strategies are equivalent. This example clarifies why the experimental degrees of freedom used for the design of the optimal discriminatory experiments were different from the ones used in the other case studies that are discussed in this dissertation (as mentioned in Section 5.4.3). When the same exercise as described above is performed with the original experimental degrees of freedom (described in Section 4.11.5), the obtained results are similar to the ones shown in Fig. 5.3a and did not contribute to the study of the design strategies. Therefore,




(a) starting from preliminary experiment $\boldsymbol{\xi}_1^1$




(b) starting from preliminary experiment $\boldsymbol{\xi}_1^2$




(c) starting from preliminary experiment $\boldsymbol{\xi}_1^3$

**Figure 5.3:** The median of the number of adequate models, determined from the (thirty) repetitions of the model discrimination procedure that were initiated with experiment $\boldsymbol{\xi}_1^1$ (a), $\boldsymbol{\xi}_1^2$ (b) and $\boldsymbol{\xi}_1^3$ (c). In each of the subfigures, the results of one of the design strategies are presented in black, while the evolution of the median values of the other design strategies are shown in gray to ease their mutual comparison. The number of values (runs) from which the median is determined are indicated by the size of a bullet symbol and the corresponding integer (in blue).

the experimental degrees of freedom were defined in a different way, in order to impose differences between the results obtained with the different design strategies.

### 5.4.5 Overall discriminatory potential of the designed experiment

A second aspect that is considered in the evaluation is the overall discriminatory potential of the designed experiment. The latter gives insight in the different design strategies and helps us to interpret the observed evolution of the number of adequate models as more discriminatory experiments are performed.

Regardless of the design strategy used, the outcome of the experimental design step is a single well-defined experiment. For the pairwise design strategy, for instance, an optimal discriminatory experiment is designed for each model pair, and the experiment for which the corresponding $T_{ij}$-value is largest, is chosen. As already discussed in Section 5.3.1, the experiment is (designed and) expected to discriminate between these two models only, although it is possible that after performing this experiment other models are found to be inadequate too. The latter is, however, not guaranteed since the other models are not considered in the design of this experiment. From a conceptual point of view, one can thus expect that only a limited number of models will be rejected after performing an experiment designed using the pairwise design strategy. In other words, the overall discriminatory potential of the designed experiment could, in principle, be limited.

To investigate the overall discriminatory potential of the experiment designed using the different design strategies, the following approach was adopted. For each design strategy, the discriminatory potential of the first designed experiment was calculated from Eqs. (4.22) and (4.27) (anticipatory approach) for each of the model pairs. Since this experiment is performed right after the preliminary experiment, it is in fact the second experiment that is eventually performed. Therefore, it will be denoted as $\boldsymbol{\xi}_2^\star$, and the corresponding discriminatory potential will be denoted as $T_{ij}\left(\boldsymbol{\xi}_2^\star\right)$. When starting the model discrimination procedure with one particular preliminary experiment, this experiment (and its corresponding $T_{ij}\left(\boldsymbol{\xi}_2^\star\right)$-value) should in theory be the same in each of the repetitions of the procedure. Indeed, when designing the first experiment, the parameter estimates are identical for each repetition. However, when this experiment is performed, the simulated experimental data will not be the same in the different repetitions of the procedure, and the parameter estimates obtained from these data will be different as well. Therefore, the experiments designed in the second iteration of the model discrimination procedure will be different from one another and cannot simply be compared. The evaluation of the

discriminatory potential discussed below is thus restricted to experiment $\boldsymbol{\xi}_2^\star$, and thus to the $T_{ij}\,(\boldsymbol{\xi}_2^\star)$-values only.

As stated above, the $T_{ij}\,(\boldsymbol{\xi}_2^\star)$-values were calculated for all model pairs. These are shown in Figs. 5.4 and 5.5, where the $T_{ij}\,(\boldsymbol{\xi}_2^\star)$-values are presented for the scenarios with preliminary experiments $\boldsymbol{\xi}_1^1$ and $\boldsymbol{\xi}_1^2$, respectively. Note that the results obtained for the other scenarios are similar. They are not shown here for brevity, but can be found in Appendix B. As some of the models can already be rejected after performing the preliminary experiment (as described in the previous section), the number of model pairs can be different for scenarios that are initiated with another preliminary experiment. Indeed, as explained when discussing the model discrimination procedure in Section 2.7, only the adequate models are considered to design the optimal discriminatory experiment, which explains why the number of model pairs is different in Figs. 5.4 and 5.5.

The overall performance of the designed experiment is evaluated by looking at the mean of all $T_{ij}\,(\boldsymbol{\xi}_2^\star)$-values obtained for that particular design strategy. The latter is indicated in Figs. 5.4 and 5.5 by the vertical line. From these lines, is clear that the overall discriminatory potential is more or less the same for the pairwise and the averaged design strategies, but significantly lower for the maximin design strategy. This is in agreement with the discussion held in the previous section, where the rate at which inadequate models are identified was found to be (slightly) lower for the maximin design strategy than for the other design strategies.

However, one can also see that the minimum $T_{ij}\,(\boldsymbol{\xi}_2^\star)$-value is significantly larger for the maximin design strategy than the minimum value obtained with the other design strategies. This is in accordance with the conceptual idea from which this design strategy was derived and its importance should not be underestimated. Indeed, it may well be that the experiment obtained using the pairwise (or the averaged) design strategy allows to discriminate among those model pairs for which the corresponding $T_{ij}\,(\boldsymbol{\xi}_2^\star)$-value is larger than $10^2$, but that it is not possible to discriminate between the model pairs for which this is not the case. If so, a second experiment has to be designed to further discriminate among these remaining models. However, because the minimum $T_{ij}\,(\boldsymbol{\xi}_2^\star)$-value is significantly larger for the maximin design strategy, it might be possible to discriminate among all model pairs in one single experiment. In this respect, it would be good if there existed a workaround that combines the idea of maximizing the minimal Tij-value with the one of maximizing the overall (or average) Tij-values. One idea to achieve this would be to use a strategy where the product of the Tij-values is maximized (instead of the sum in

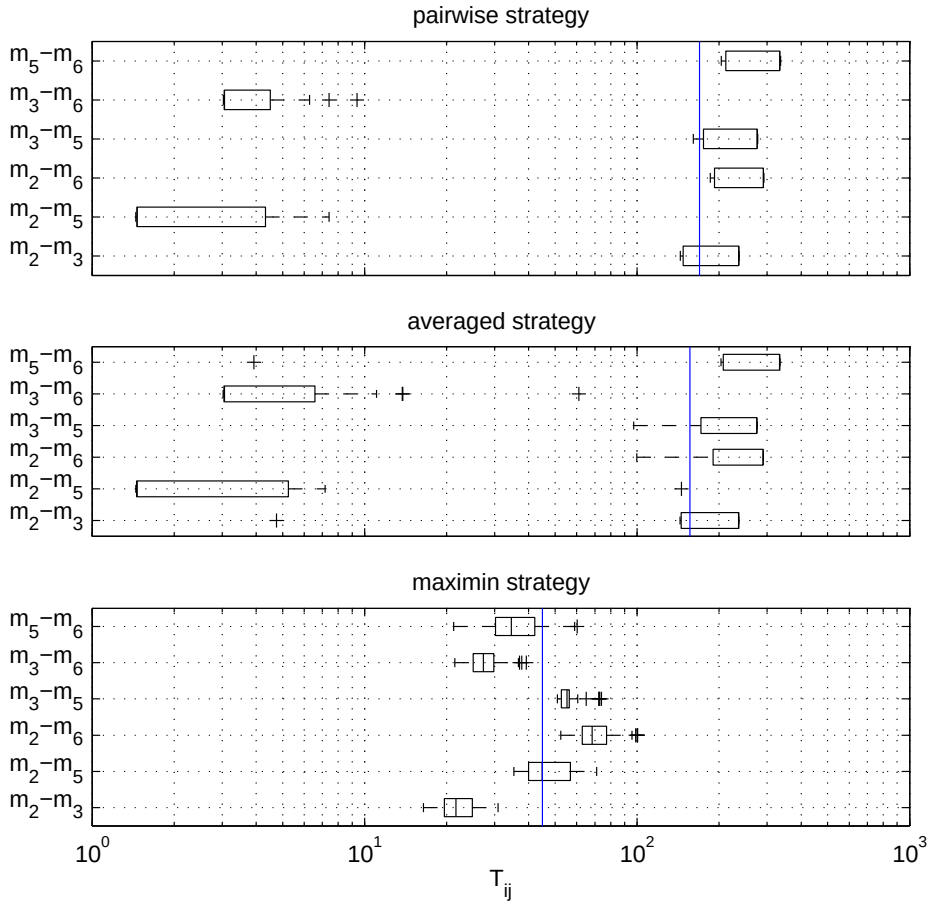



**Figure 5.4:** Boxplots of the $T_{ij}$ $(\boldsymbol{\xi}_2^\star)$-values calculated for each of the model pairs, where $\boldsymbol{\xi}_2^\star$ refers to the first experiment that is designed when applying the model discrimination procedure, and is performed after preliminary experiment $\boldsymbol{\xi}_1^1$. The mean of these $T_{ij}$-values is indicated by the vertical line and gives an idea of the overall discriminatory potential of the designed experiment.

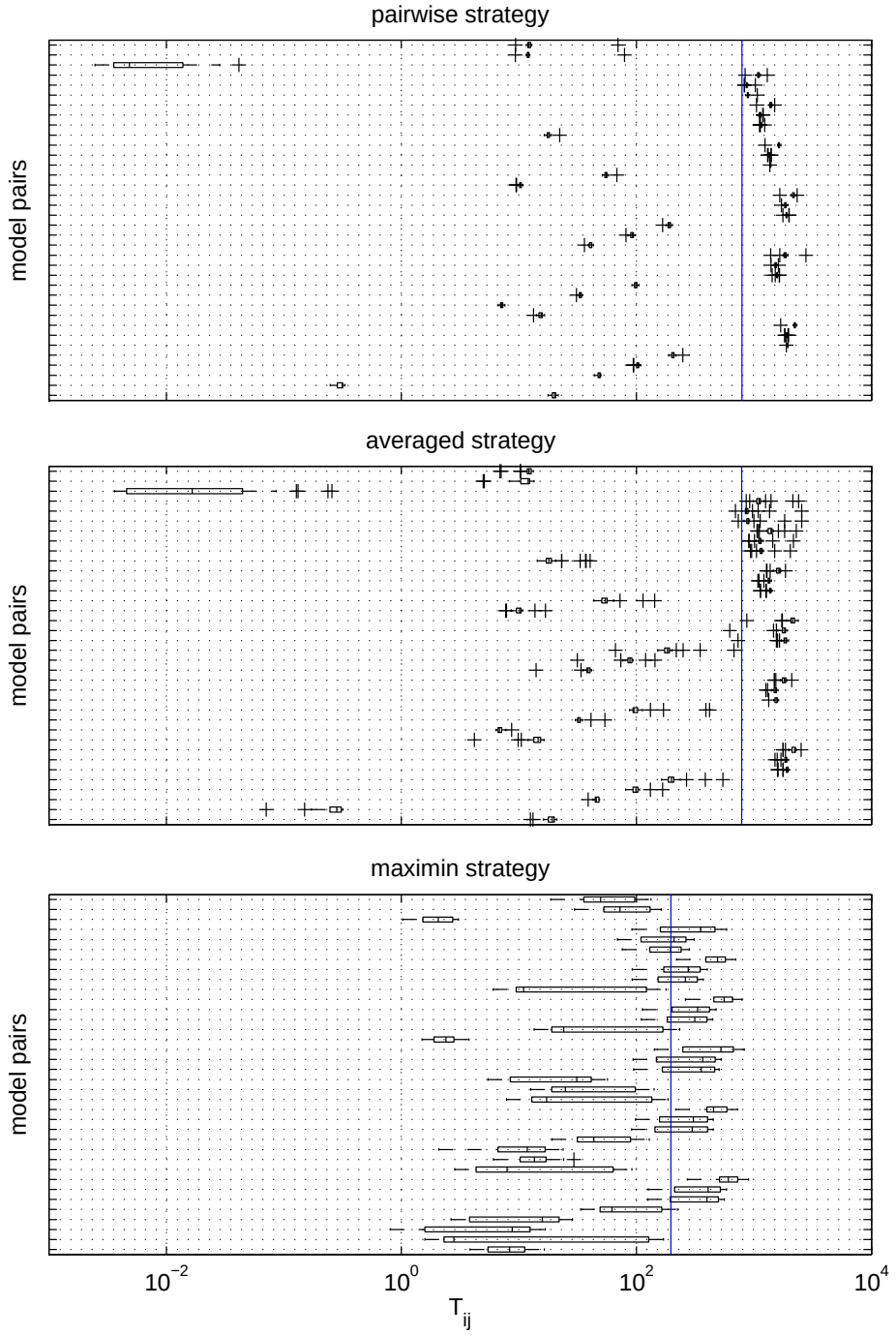



**Figure 5.5:** Boxplots of the $T_{ij}$ ($\boldsymbol{\xi}_2^\star$)-values calculated for each of the model pairs, where $\boldsymbol{\xi}_2^\star$ refers to the first experiment that is designed when applying the model discrimination procedure, and is performed after preliminary experiment $\boldsymbol{\xi}_1^2$. The mean of these $T_{ij}$-values is indicated by the vertical line and gives an idea of the overall discriminatory potential of the designed experiment. Note that due the large number of model pairs, their labels were omitted for brevity.

the averaged design criterion). This would prevent situations where one or several of the Tij-values become too small and the Tij-values of all model pairs are still considered as such in the objective function.

Despite the fact that the $T_{ij}\left(\boldsymbol{\xi}_2^\star\right)$-values for a given model pair should in theory be identical for each of the thirty repetitions (as discussed above), one can see from the results in Figs. 5.4 and 5.5 that this is not the case in practice. Although one should not forget to take into account that a logarithmic scale was used to represent the $T_{ij}\left(\boldsymbol{\xi}_2^\star\right)$-values, one can clearly observe a significant variability among the $T_{ij}\left(\boldsymbol{\xi}_2^\star\right)$-values. This variability can, however, be explained by the fact that the optimal discriminatory experiments are found using an optimization algorithm to solve the problem formalized by Eq. (5.1). Given the complexity of this optimization problem, it is not realistic to expect that the optimal solution is always the same, nor that the global optimum is always found. Note that this issue is partly dealt with by repeating each model discrimination procedure thirty times.

## 5.5 Conclusions

Model discrimination has to be seen as an iterative procedure, because the designed experiments depend on the currently available parameter estimates. In every iteration, an optimal discriminatory experiment is designed using only those models that are considered as adequate models at that stage. When the number of rival models is larger than two, several design strategies are thinkable to steer the model discrimination exercise, irrespective of which design criterion is chosen. In this chapter, some design strategies that are interesting from a conceptual point of view were formulated and applied to a case study. The performance of the design strategies was evaluated by investigating (1) the rate at which the number of adequate models decreased as the model discrimination procedure progresses, and (2) the overall discriminatory potential of the designed experiment. Although the results are not conclusive, they did suggest that the design strategy in which the average discriminatory potential of the designed experiment is maximized, might be preferred over the other design strategies. However, although the overall discriminatory potential of the experiment obtained with the maximin design strategy was lower than with the other design strategies, discrimination among a larger number of rival models can be expected as it focuses on that model pair for which model discrimination is most challenging.

# PART III

## INTEGRATION OF EXPERIMENTAL DESIGN FOR PARAMETER ESTIMATION AND MODEL DISCRIMINATION

# CHAPTER 6

# Optimal experimental design for parameter estimation and its relation with model discrimination

*"Without deviation from the norm, progress is not possible."*

Frank Zappa, musician

## Abstract

A mathematical model contains model parameters, the value of which has to be estimated from experimental data. The quality of the parameter estimates is highly dependent on the information content of the available experiments. Therefore, experimental design methods have been developed to design experiments with a maximal information content with regard to the model parameters. These methods are described in this chapter and discusses its relevance for model discrimination.

## 6.1 Introduction

In most modelling exercises, the aim is to identify a mathematical model which describes the studied process with the desired accuracy. When such a model is available, it can be used as a surrogate for the actual process, making it possible to investigate the process behavior under various input conditions both rapidly and inexpensively, and without necessarily tampering with the actual process (Ogunnaike and Ray, 1994). This makes models very useful tools for process design, control and optimization, but they can also be of great value to increase the insight in very complex processes.

The optimal experimental design techniques described in the previous chapters are used to identify the most appropriate model from a set of rival models. Here, the term *model* refers to the model structure. However, as explained earlier, the model structure contains model parameters that have to be determined from experimental data. In order to obtain an accurate and reliable prediction of the behavior of the studied process, it is not sufficient to have an adequate model structure. Indeed, it is also important that the quality of the parameter estimates is sufficiently good, as this determines the uncertainty on the model predictions.

This chapter deals with the experimental design techniques that are used to increase the quality of the parameter estimates. The latter are often called optimal experimental design for parameter estimation (Vanrolleghem and Dochain, 1998) or parameter precision (Franceschini and Macchietto, 2008). For clarity, the theoretical considerations on the estimation of the model parameters are briefly explained before describing the experimental design techniques themselves. Its relation with model discrimination will be shortly discussed in Section 6.8.

## 6.2 Parameter estimation

The values of the model parameters, which by definition do not change during the course of the simulation, have to be determined from experimental data. This process is called parameter estimation, and typically consists of minimizing the weighted sum of squared errors (WSSE) functional by optimal choice of the parameters $\boldsymbol{\theta}$. The optimal parameter estimates, denoted as $\hat{\boldsymbol{\theta}}$, are thus given by

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \text{WSSE}(\boldsymbol{\theta}) \ . \tag{6.1}$$

Here, $\mathrm{WSSE}(\boldsymbol{\theta})$ is calculated as

$$\mathrm{WSSE}(\boldsymbol{\theta}) = \sum_{k=1}^{n_e} \sum_{l=1}^{n_{sp_k}} \Delta\hat{\boldsymbol{y}}(\boldsymbol{\xi}_k, \boldsymbol{\theta}, t_l)' \cdot \mathbf{Q} \cdot \Delta\hat{\boldsymbol{y}}(\boldsymbol{\xi}_k, \boldsymbol{\theta}, t_l) \tag{6.2}$$

and

$$\Delta\hat{\boldsymbol{y}}(\boldsymbol{\xi}_k, \boldsymbol{\theta}, t_l) = \boldsymbol{y}(\boldsymbol{\xi}_k, t_l) - \hat{\boldsymbol{y}}(\boldsymbol{\xi}_k, \boldsymbol{\theta}, t_l) \tag{6.3}$$

represents the difference between the vector of the $n_m$ measured response variables and the model predictions at time $t_l$ ($l = 1, \ldots, n_{sp_k}$) of experiment $\boldsymbol{\xi}_k$ ($k = 1, \ldots, n_e$). Further, $n_e$ represents the number of experiments from which data are used for estimating the parameters, $n_{sp_k}$ represents the number of samples in experiment $\boldsymbol{\xi}_k$, and $\mathbf{Q}$ is an $n_m$-dimensional matrix of user-supplied weighting coefficients. Typically, $\mathbf{Q}$ is chosen as the inverse of the measurement error covariance matrix $\boldsymbol{\Sigma}$ (Marsili–Libelli et al., 2003; Omlin and Reichert, 1999; Vanrolleghem and Dochain, 1998). In this way, the measurement uncertainty is incorporated in the WSSE.

## 6.3 Fisher information matrix

In order to have accurate parameter estimates, the minimum of the WSSE functional has to be clearly defined, meaning that a parameter set that is slightly different from the optimal one should result in a WSSE that is significantly higher than the minimum WSSE value. In that case, there will be little doubt about the parameter values that minimize WSSE (Eq. (6.2)). Mathematically, this can be formalized as follows (Munack, 1991; Vanrolleghem and Dochain, 1998).

Suppose one is estimating the parameters of a model using data from $n_e$ experiments $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{n_e}$. The expected value of the WSSE for a parameter set that is slightly different ($\delta\boldsymbol{\theta}$) from the optimal one can be written as (Munack, 1989)

$$\mathrm{E}\left[\mathrm{WSSE}\left(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{n_e}, \hat{\boldsymbol{\theta}} + \delta\boldsymbol{\theta}\right)\right] =$$

$$\sum_{k=1}^{n_e} \sum_{l=1}^{n_{sp_k}} \Delta\hat{\boldsymbol{y}}\left(\boldsymbol{\xi}_k, \hat{\boldsymbol{\theta}} + \delta\boldsymbol{\theta}, t_l\right)' \cdot \boldsymbol{\Sigma}^{-1} \cdot \Delta\hat{\boldsymbol{y}}\left(\boldsymbol{\xi}_k, \hat{\boldsymbol{\theta}} + \delta\boldsymbol{\theta}, t_l\right), \tag{6.4}$$

where E stands for the expectation operator and $\boldsymbol{\xi}_k$ stands for the $k$-th experiment. Now, one can linearize the model with respect to the parameters (Taylor series approximation), which results in the following equation

$$\hat{\boldsymbol{y}}(\boldsymbol{\xi}_k, \hat{\boldsymbol{\theta}} + \delta\boldsymbol{\theta}, t_l) \approx \hat{\boldsymbol{y}}(\boldsymbol{\xi}_k, \hat{\boldsymbol{\theta}}, t_l) + \left.\frac{\partial \hat{\boldsymbol{y}}}{\partial \boldsymbol{\theta}}(\boldsymbol{\xi}_k, \boldsymbol{\theta}, t_l)\right|_{\hat{\boldsymbol{\theta}}} \cdot \delta\boldsymbol{\theta}. \tag{6.5}$$

Introducing Eq. (6.5) into Eq. (6.4), one finds that Eq. (6.4) can be approximated as

$$\mathrm{E}\left[\mathrm{WSSE}\left(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{n_e}, \hat{\boldsymbol{\theta}} + \delta\boldsymbol{\theta}\right)\right] \approx$$

$$\mathrm{E}\left[\mathrm{WSSE}\left(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{n_e}, \hat{\boldsymbol{\theta}}\right)\right] + \sum_{k=1}^{n_e}\sum_{l=1}^{n_{sp_k}} \delta\boldsymbol{\theta}' \cdot \mathbf{FIM}\left(\boldsymbol{\xi}_k, \hat{\boldsymbol{\theta}}, t_l\right) \cdot \delta\boldsymbol{\theta}, \tag{6.6}$$

where $\mathbf{FIM}\left(\boldsymbol{\xi}_k, \hat{\boldsymbol{\theta}}, t_l\right)$ represents the so-called Fisher information matrix, calculated as

$$\mathbf{FIM}\left(\boldsymbol{\xi}_k, \hat{\boldsymbol{\theta}}, t_l\right) = \left(\left.\frac{\partial \hat{\boldsymbol{y}}(\boldsymbol{\xi}_k, \boldsymbol{\theta}, t_l)}{\partial \boldsymbol{\theta}}\right|_{\hat{\boldsymbol{\theta}}}\right)' \cdot \boldsymbol{\Sigma}^{-1} \cdot \left(\left.\frac{\partial \hat{\boldsymbol{y}}(\boldsymbol{\xi}_k, \boldsymbol{\theta}, t_l)}{\partial \boldsymbol{\theta}}\right|_{\hat{\boldsymbol{\theta}}}\right). \tag{6.7}$$

In order to get a clearly defined minimum for WSSE, it is necessary that the difference between $\mathrm{WSSE}(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{n_e}, \hat{\boldsymbol{\theta}})$ and $\mathrm{WSSE}(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{n_e}, \hat{\boldsymbol{\theta}} + \delta\boldsymbol{\theta})$ is maximized (as illustrated in Fig. 6.1). From Eq. (6.6) one can see that this can be done by maximizing the Fisher information matrix, which will briefly be called **FIM** in the following.

As can be seen from Eq. (6.7), the **FIM** is composed of two components, the parameter sensitivities ($\partial\hat{\boldsymbol{y}}/\partial\boldsymbol{\theta}$) and the measurement error covariance matrix ($\boldsymbol{\Sigma}$). The sensitivity of a certain state variable with respect to a parameter expresses how much that state variable will change when this parameter is slightly perturbed. A state variable that is highly sensitive to a certain parameter will therefore contain a lot of information about this parameter, while a variable that is insensitive to the parameter does not contribute to the information content for that parameter. The role of the measurement error covariance matrix in the calculation of the **FIM** is rather straightforward, since it is obvious that a measurement associated with a large measurement error will contribute less to the information content than a measurement with a small measurement error.

## 6.4 Central rationale behind optimal experimental design for parameter estimation

In general, optimal experimental design is an optimization problem, where the optimum of a well-defined objective function is sought by varying the experimental degrees of freedom. The experimental degrees of freedom, $\boldsymbol{\xi}$, are restricted by a number of constraints that define a set of possible experiments, denoted as $\boldsymbol{\Xi}$. These constraints are determined by the experimental setup and are specified before the start of the experimental design exercise.

The Fisher information matrix described in the previous section expresses the information content of the $n_e$ experiments with regard to the model parameters, and its maximization is the central rationale behind optimal experimental design for parameter estimation (Asprey and Macchietto, 2000; Goodwin and Payne, 1977; Ljung, 1999; Mehra, 1974; Munack, 1991; Shirt et al., 1994; Vanrolleghem and Dochain, 1998; Walter and Pronzato, 1997). The $(n_e + 1)$th experiment is obtained as

$$\boldsymbol{\xi}^{\star}_{n_e+1} = \arg \max_{\boldsymbol{\xi} \in \boldsymbol{\Xi}} \mathbf{FIM}\left(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{n_e+1}, \hat{\boldsymbol{\theta}}_{n_e}\right) , \tag{6.8}$$

with

$$\mathbf{FIM}\left(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{n_e+1}, \hat{\boldsymbol{\theta}}_{n_e}\right) = \sum_{k=1}^{n_e} \mathbf{FIM}\left(\boldsymbol{\xi}_k, \hat{\boldsymbol{\theta}}_{n_e}\right) + \mathbf{FIM}\left(\boldsymbol{\xi}_{n_e+1}, \hat{\boldsymbol{\theta}}_{n_e}\right) . \tag{6.9}$$

The information content of the proposed $(n_e + 1)$th experiment, which is represented by $\mathbf{FIM}(\boldsymbol{\xi}_{n_e+1}, \hat{\boldsymbol{\theta}}_{n_e})$, is thus maximized, given the information content of the already performed experiments (denoted as $\mathbf{FIM}(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{n_e}, \hat{\boldsymbol{\theta}}_{n_e})$) and the parameter values derived from these experiments (denoted as $\hat{\boldsymbol{\theta}}_{n_e}$). For simplicity, $\mathbf{FIM}(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{n_e+1}, \hat{\boldsymbol{\theta}}_{n_e})$ will be denoted as **FIM** in the following.

The effect of the maximization of the **FIM** on the WSSE is illustrated in Fig. 6.1 for a hypothetical example in which only one parameter has to be estimated. The figure shows the trajectories of the WSSE functional for a parameter estimation based on data from two experiments, a non-optimal experiment (outer curve) and an optimal experiment (inner curve). The parameter estimation based on the data from the optimal experiment, which was obtained after maximization of the **FIM**, results in a smaller confidence interval and thus in a more accurate parameter estimate.

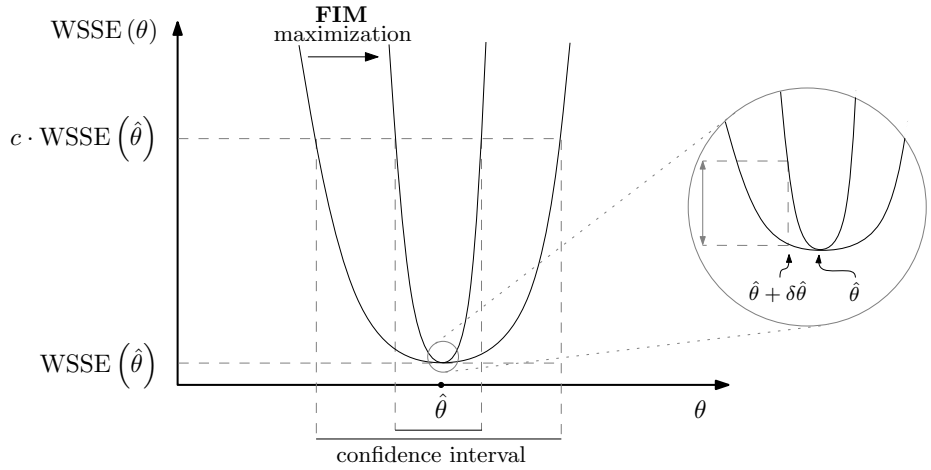



**Figure 6.1:** Effect of **FIM** maximization on the trajectory of the WSSE $(\theta)$ functional for a single parameter estimation problem and the corresponding confidence intervals, constructed based on a cut-off value of $c \cdot \text{WSSE}\left(\hat{\theta}\right)$ (see Eq. (2.19)).

## 6.5 Experimental design criteria based on the FIM

Since the **FIM** is a matrix, it cannot be maximized as such. Therefore, several criteria/objective functions have been proposed based on the **FIM**, all of which exploit the inversely proportional relationship between the **FIM** and the parameter estimation error covariance matrix (Atkinson and Donev, 1992; Munack, 1991; Petersen, 2000; Vanrolleghem and Dochain, 1998). This relationship is dictated by the Cramér-Rao inequality (Ljung, 1999; Walter and Pronzato, 1997), which states that under certain conditions (that is, uncorrelated white measurement noise), the inverse of the **FIM** gives a lower bound of the parameter estimation error covariance matrix. In this way, properties of the **FIM** determine the size, shape and orientation of the confidence region of the parameter estimates, and thus their precision. Some of the criteria are briefly described below (partly adapted from Petersen (2000)).

**A-optimality design criterion**: $\min_{\boldsymbol{\xi}\in\boldsymbol{\Xi}} \; tr\left(\mathbf{FIM}^{-1}\right)$

With this criterion, the trace of the inverse of the **FIM** is minimized, which is equivalent to minimizing the sum of the variances of the parameter estimates. In other words, this criterion minimizes the arithmetic average of the variances of the parameter estimate. Because this criterion is based on an inversion of the **FIM**, numerical problems will arise when the **FIM** is close to singular.

**Modified A-optimality design criterion**: $\max_{\boldsymbol{\xi}\in\boldsymbol{\Xi}} \; tr\,(\mathbf{FIM})$
This criterion is similar to the A-optimal design criterion, but the trace of the **FIM** is maximized instead of the trace of the inverse of the **FIM**. Since this criterion does not require an inversion of the **FIM**, the numerical problems that might occur with the A-optimal design criterion do not occur here. However, this advantage is also a disadvantage because when an unidentifiable experiment is evaluated (that is a case where the **FIM** is singular and thus the confidence region goes to infinity in a certain direction because one of the eigenvalues of the **FIM** is zero) the trace can still be optimized and the problem of unidentifiability will not be noticed (Goodwin and Payne, 1977). This is less of a problem with the A-optimal design criterion since an inversion of the **FIM** will not be possible and the problem of unidentifiability will thereby be exposed.

**D-optimality design criterion**: $\max_{\boldsymbol{\xi}\in\boldsymbol{\Xi}} \; \det(\mathbf{FIM})$
Here, the idea is to maximize the determinant of the **FIM** (Box and Lucas, 1959). The latter is inversely proportional to the volume of the confidence region of the parameter estimates, and this volume is thus minimized when maximizing $\det(\mathbf{FIM})$. In other words, one minimizes the geometric average of the variances of the parameter estimates. Moreover, D-optimal experiments possess the property of being invariant with respect to any rescaling of the parameters (Petersen, 2000; Seber and Wild, 1989). According to Walter and Pronzato (1997), the D-optimal design criterion is the most used criterion. However, several authors have pointed out that this criterion tends to give excessive importance to the parameter which is most influential (Franceschini and Macchietto, 2008; Pinto et al., 1990).

**E-optimality design criterion**: $\max_{\boldsymbol{\xi}\in\boldsymbol{\Xi}} \; \lambda_{min}\,(\mathbf{FIM})$
The E-optimal design criterion maximizes the smallest eigenvalue of the **FIM** and thereby minimizes the length of the largest axis of the confidence ellipsoid. Thus, these designs aim at minimizing the largest parameter estimation variance and thereby at maximizing the distance from the singular, unidentifiable case.

**ModE-optimality design criterion**: $\min_{\boldsymbol{\xi}\in\boldsymbol{\Xi}} \; \frac{\lambda_{max}(\mathbf{FIM})}{\lambda_{min}(\mathbf{FIM})}$
With this criterion, the focus is on the minimization of the condition number, which is the ratio between the largest and the smallest eigenvalue, or, in other words, the ratio of the shortest and the longest ellipsoid axes. The minimum of this ratio is one, which corresponds to the case where the shape of the confidence ellipsoid is a (hyper)sphere.

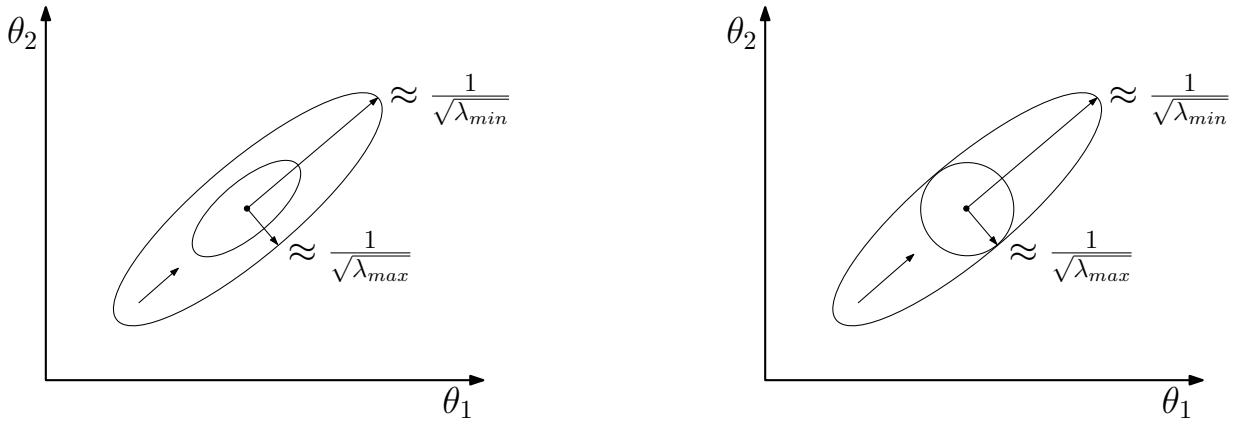



**Figure 6.2:** Illustration of (a) the D-optimal design criterion that causes the volume of the confidence region to decrease, and (b) the modified E-optimal design criterion that causes the shape of the confidence region to become as circular as possible (both figures adopted from De Pauw (2005); De Pauw and Vanrolleghem (2006a)).

The effect on the confidence region of the D-optimal design criterion and the modE-optimal design criterion is illustrated in Fig. 6.2 for an estimation problem with two parameters ($\theta_1$ and $\theta_2$). The size, shape and orientation of the confidence region, which is an ellipse in the case of two parameters, are determined by the eigenvalues and eigenvectors of the **FIM**. The largest axis of the confidence ellipse is inversely proportional to the square root of the smallest eigenvalue ($\lambda_{min}$), while the smallest axis is inversely proportional to the square root of the largest eigenvalue ($\lambda_{max}$).

Note that some other design criteria are also described in literature. For a detailed discussion of other, less frequently used design criteria such as G-, L-, C- and $D_s$-optimality, the reader is referred to the work of Atkinson and Donev (1992), Mehra (1974), Sidoli et al. (2004) and Walter and Pronzato (1990).

## 6.6 On the approximation of the parameter estimation error covariance matrix

The common approaches to determine the parameter and model prediction uncertainties are based on linear propagations of uncertainties (as discussed in Chapter 2). Consequently, they should only be considered as approximate estimates when working with (highly)

nonlinear models. Nevertheless, it is common practice to use these approximations for experimental design purposes. When designing experiments to increase the precision of the parameter estimates, for instance, scalar functions of the **FIM** are maximized (as discussed above), and also the design of optimal discriminatory experiments often relies on linear approximations of the parameter estimates (and model prediction uncertainties), as discussed in Chapter 4.

Because the use of the linear approximations of the parameter estimation error covariance matrix can result in poor experimental designs, Benabbas et al. (2005) proposed to take into account the curvature of the WSSE functional in the experimental design. Their approach to design the optimal experiment is the same as the one described above (Eq. (6.8)), but a constraint is added which assures that the observed curvature is below a predefined acceptable level of curvature. In other words, the linear approximation of the parameter estimation error covariance matrix (or the **FIM**) is used, but only if it can be trusted and used. The optimal experimental design problem is thus given by

$$\boldsymbol{\xi}^{\star}_{n_e+1} = \arg \max_{\boldsymbol{\xi}\in\boldsymbol{\Xi}} \mathbf{FIM}\left(\hat{\boldsymbol{\theta}}, \boldsymbol{\xi}\right) , \tag{6.10}$$

subject to

$$\zeta\left(\hat{\boldsymbol{\theta}}, \boldsymbol{\xi}\right) < \varepsilon , \tag{6.11}$$

where $\zeta\left(\hat{\boldsymbol{\theta}}, \boldsymbol{\xi}\right)$ represents the curvature and $\varepsilon$ represents the acceptable level of curvature. To calculate this curvature, the second-order sensitivities are required. For detailed information on the calculation of these sensitivities and the curvature itself, the reader is referred to the cited paper, but it is important to realize that this requires an additional $n_s \times n_p \times n_p$ differential equations, on top of the $n_s$ differential equations used to define the model as such (Eq. (2.1)) and the $n_s \times n_p$ additional differential equations used to calculate the first-order sensitivities (Eq. (2.17)). It is clear that this is a serious drawback, since these additional differential equations drastically increase the computational burden.

Alternatively, Joshi et al. (2006) proposed a bootstrap method to obtain a better approximation of (possibly) asymmetric confidence regions for parameter estimates. In their approach, the mean and variance of the available experimental data are assumed to be known, and they are used to define a probability density function of the measured data. By sampling from this distribution, a number of new experimental data sets is generated,

from each of which the model parameters are estimated. This results in an equal number of parameter estimates, and the variability among these estimates gives an idea about the uncertainty associated with them. Apart from the fact that this approach strongly depends on the distribution of the measured data, a great number of parameter estimation exercises are required to obtain a representative estimate of the uncertainty on the parameter estimates.

As already indicated above, these approaches are computationally intensive, which currently makes them less suited for optimal experimental design purposes (Franceschini and Macchietto, 2008; Joshi et al., 2006). It should be noted however that, although the experimental designs based on a linear approximation of the parameter estimation error covariance matrix can be less informative than expected (Benabbas et al., 2005), they are still very useful from a practical point of view.

## 6.7 Locally versus robust optimal experimental designs

Next to the fact that the common approaches to optimal experimental design for parameter estimation are based on an approximation of the parameter estimation error covariance matrix, a second and probably more important issue has not been discussed so far. Indeed, since the Fisher information matrix (**FIM**) is calculated using the currently available parameter estimates, the optimal experimental designs are dependent on them and can thus only be considered optimal in their proximity. In principle, they should thus be called locally optimal experimental designs, as noted by several authors including Atkinson and Donev (1992), Balsa-Canto et al. (2007), Bernaerts et al. (2000), De Pauw (2005), Versyck et al. (1999) and Vanrolleghem and Dochain (1998). The effectiveness of the design thus depends on how well these parameter estimates match those of the actual process, which is problematic since the ultimate aim of optimal experimental design for parameter estimation is exactly to find the true parameter values, or at least to approximate them with a reasonable precision.

The most often used approach to deal with this problem is to design the experiments in a sequential way by alternating parameter estimation and experimental design (Walter and Pronzato, 1997; Vanrolleghem and Dochain, 1998). After each parameter estimation the knowledge of the process parameters improves and this is reflected in an increased quality of the experiment designed subsequently. Many authors acknowledge the usefulness of this approach (for instance, Ford et al. (1989), Walter and Pronzato (1990) and

Atkinson (2003)), but many iterations (and thus experiments) may be required before the parameters converge to their *true* values. However, it may not be possible to perform that many experiments on the same system due to limitations in time or resources, or simply because the experimental setup does not allow it. In addition, it is not guaranteed that the parameters converge to the true values.

For this reason, several authors have emphasized the need of experimental design methods that are less sensitive (or more robust) to the starting values of the parameter estimates. Several approaches have been described in literature to deal with the robustness of experimental design for parameter estimation (Asprey and Macchietto, 2002; De Pauw, 2005; Franceschini and Macchietto, 2008; Walter and Pronzato, 1997). One of these approaches, the so-called maximin approach, will be briefly discussed below.

With the maximin approach, the experiment is selected that optimizes the worst possible performance for any value of $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ (Pronzato and Walter, 1988). Using the D-optimality design criterion, for instance, this approach is given by

$$\boldsymbol{\xi}^{\star}_{n_e+1} = \arg \max_{\boldsymbol{\xi} \in \boldsymbol{\Xi}} \min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left| \mathbf{FIM}\left(\boldsymbol{\theta}, \boldsymbol{\xi}\right) \right| . \tag{6.12}$$

In other words, for each experiment that is proposed by the optimization algorithm, those parameters are determined for which the D-optimality design criterion value is lowest and this value is returned to the optimization algorithm and is eventually maximized. For this approach, the necessary prior information on the possible parameter values is limited to the upper and lower bounds, but the computational burden introduced by the nested optimization limits its application.

## 6.8 Integration of OED/PE and model discrimination

The experimental design techniques discussed so far are appropriate when the experimenter either wants to discriminate among rival models (Chapter 4) or to estimate the model parameters (this chapter). However, from the discussion held in the introduction of this chapter, it is clear that both the model structure and (the quality of) its parameters are important. This is obviously reflected in the classical strategy for building a mathematical model, which basically consists of two steps (Asprey and Macchietto, 2000; Box and Hill, 1967; Buzzi-Ferraris, 1999; Espie and Macchietto, 1989; Vanrolleghem and Dochain, 1998). First, the adequate model structure is identified, and then the quality of its parameters is

(further) improved to permit reliable model predictions. This sequential strategy can thus be achieved by first applying the model discrimination procedure discussed in Chapter 4 to select the best model, and then using a parameter estimation procedure on this model.

From the discussion held in Chapters 4 and 5, however, it is known that the uncertainty on the model predictions is of crucial importance for model discrimination. This is because the experimental design methods are model-based, and high model prediction uncertainties obviously hamper the efficacy and efficiency of the model discrimination procedure. These model prediction uncertainties are determined by the quality of the available data, since low quality data will result in poorly estimated parameters, which on their turn result in uncertain model predictions. The discrimination among several rival models may thus become more efficient and effective if this uncertainty could be reduced prior to the start of the model discrimination procedure. In other words, from a conceptual point of view, it may be interesting to perform optimal experimental design for parameter estimation first, before proceeding with model discrimination. This rationale is challenged in the following chapters.

# CHAPTER 7

# Design of compromise experiments to simultaneously estimate the parameters of several rival models

*"Ik heb altijd gelijk."*

Isabelle Gheysen, mijn lief kleintje

## Abstract

When several mathematical models are proposed for one and the same process, experimental design techniques are available to design optimal discriminatory experiments. However, because these design techniques are model-based, it is important that the required model predictions are not too uncertain. This uncertainty is determined by the quality of the already available data, since low quality data results in poorly estimated parameters, and subsequently in uncertain model predictions, which prevent that significant differences between the models can be found. Therefore, model discrimination may become more efficient and effective if this uncertainty is reduced first. This can be achieved by performing dedicated experiments, designed to increase the precision of the parameter estimates. However, performing such an additional experiment for each rival model may undermine the overall goal of optimal experimental design, which is to minimize the experimental effort. Therefore, this chapter is dedicated to the design of so-called compromise experiments, which are defined as experiments that are sufficiently informative to improve the overall precision of the parameters of all rival models in a single experiment. For this purpose, two methods are proposed: the kernel-based method and the ideal point method. The ability of both methods to design such compromise experiments is evaluated after applying them to a case study where nine rival models are defined to describe the kinetics of an enzymatic reaction (glucokinase).

## 7.1 Introduction

Many experimental studies are performed (1) to determine the model structure that adequately describes the process under study (often called model discrimination), or (2) to obtain (more) accurate estimates of the model parameters. For both problems, experimental design methods have been developed that help the experimenter to plan the experiments. For the problem of model discrimination, the methods described in Chapter 4 can be used, while the ones described in Chapter 6 can be used to design experiments that result in an increased precision of the parameter estimates. Common to these experimental design methods is the overall goal to maximize the information content of the designed experiments, and thus to minimize the experimental effort needed.

As discussed before, the most intuitive approach to address the problems of model discrimination and accurate parameter estimation is to deal with them successively (Hill et al., 1968; Walter and Pronzato, 1997). First, experiments are designed and performed to choose between the rival model structures, and then, once the most promising model structure has been selected, experiments are designed and performed to accurately estimate its parameters. Alternatively, one could deal with both problems simultaneously. For this purpose, a joint criterion has been described by Hill et al. (1968) (see Section 8.3) where the basic design strategy is to emphasize model discrimination when there is considerable doubt as to which model is best, and gradually shift the emphasis to parameter estimation as experimentation progresses, and model discrimination becomes possible. Both approaches thus deal with model discrimination first, and then the focus is (gradually) shifted to parameter estimation.

From the discussion held in Chapters 4 and 5, however, it is known that the uncertainty on the model predictions is of crucial importance for model discrimination. This is because the experimental design methods are model-based, and high model prediction uncertainties obviously hamper the efficacy and efficiency of the model discrimination procedure. These model prediction uncertainties are determined by the quality of the available data, since low quality data will result in poorly estimated parameters, which in turn result in uncertain model predictions. The discrimination among several rival models may thus become more efficient and effective if this uncertainty could be reduced prior to the start of the model discrimination procedure.

Reducing the uncertainty on the model predictions can be achieved by designing and performing experiments dedicated to reducing the uncertainty on the parameter estimates.

However, performing an additional experiment for each rival model may undermine the overall goal of optimal experimental design, since this would require at least as many experiments as the number of rival models. Therefore, this chapter investigates the possibility to design a compromise experiment, which is not optimal for one or more of the individual rival models, but is sufficiently informative to improve the overall precision of the parameters of all rival models.

In Section 7.2 of this chapter, a kernel-based method is presented to design a compromise experiment for experimental design problems where only the sampling times are to be optimized. The idea of designing a compromise experiment is further explored in Section 7.3, by treating it as a multi-objective problem. The so-called ideal point method proposed in this section can be used for experimental design problems with experimental degrees of freedom of any type (manipulations, initial conditions and sampling times), whereas this would be difficult or even impossible with the kernel-based method.

Both the kernel-based method and the ideal point method are illustrated on the case study (described in Chapter 3 and used throughout this dissertation), where nine models are proposed to describe the kinetics of the enzyme glucokinase. For each method, the ability to design compromise experiments is evaluated (Sections 7.5, 7.6 and 7.7) and the conclusions drawn from these results are listed in Section 7.8.

## 7.2 A kernel-based method to determine compromise sampling times

As stated above, this chapter investigates the possibility to design a compromise experiment, that is, an experiment which may not be optimal for each individual rival model, but sufficiently informative to improve the overall precision of the parameters of all rival models. In this section, a first method is presented to determine such a compromise experiment. It is inspired by what is called kernel density estimation or the Parzen window approach (Jenssen et al., 2006; Rosenblatt, 1956).

### 7.2.1 Introduction to kernel density estimation

Kernel density estimation (Rosenblatt, 1956) is a non-parametric method for estimating the probability density function (pdf) of a random variable from its independent and identically distributed (i.i.d.) samples (Duda et al., 2001; Jenssen et al., 2006). The term

non-parametric refers to the fact that it is not necessary to assume a particular model for the pdf prior to the density estimation exercise (Duda et al., 2001).

Suppose $n$ samples of a random variable $(x)$ are drawn i.i.d. according to the (unknown) probability function $p(x)$. The kernel density estimation of this probability density function, denoted as $\hat{p}(x)$, is given by

$$\hat{p}(x) = \frac{1}{n \cdot h} \cdot \sum_{i=1}^{n} \kappa \left( \frac{x - x_i}{h} \right) \tag{7.1}$$

where $x_i$ represents the $i$th sample, $\kappa$ represents the so-called kernel function (or Parzen window), and $h$ represents the smoothing parameter (or bandwidth parameter). Quite often, $\kappa$ is taken to be a standard Gaussian function with zero mean and a variance equal to one, given by

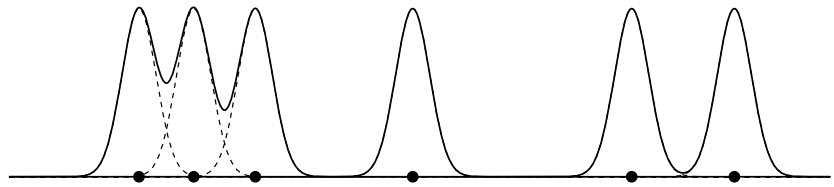
$$\kappa(u) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2} \cdot u^2} . \tag{7.2}$$

Apart from this Gaussian kernel function, other kernel functions have been proposed, but the choice of the kernel function seems to be less important than the choice of the smoothing parameter (Jenssen et al., 2006).

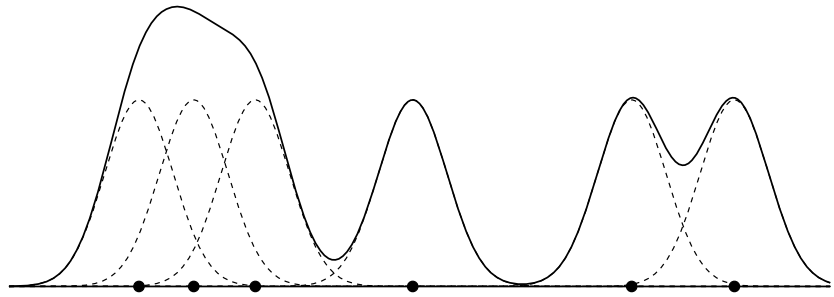
The principle of kernel density estimation and the importance of the smoothing parameter are illustrated in Fig. 7.1, where it is applied to an illustrative example for three different values of the smoothing parameter $h$. The dots represent six samples from an unknown probability density function, the dashed lines represent the individual kernel functions, and the full line represents the estimated probability density function that is calculated as the sum of the individual kernel functions.

### 7.2.2 Kernel-based method for experimental design

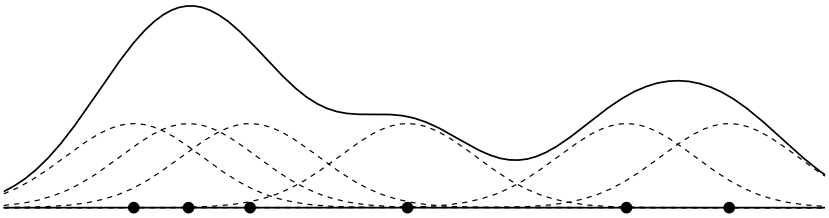
This section explains how the concepts of kernel density estimation described above can be useful in an experimental design context, and focuses on its application for those cases where only the optimal sampling times are determined. The possibility to extend the kernel-based method for applications where experimental degrees of freedom of all types (manipulations, initial conditions and sampling times) are considered, is further discussed in Section 7.2.5.



(a) $h = 0.20$



(b) $h = 0.50$



(c) $h = 1.00$

**Figure 7.1:** Kernel density estimation applied to an illustrative example for three different values of the smoothing parameter $h$. The dots represent six samples from an unknown probability density function, the dashed lines represent the individual kernel functions, and the full line represents the estimated probability density function that is calculated as the sum of the individual kernel functions.

Suppose $n_{sp}$ optimal sampling times were determined for each of the $m$ rival models by optimizing one of the optimal design criteria described in Chapter 6. Then, similarly to what is done in kernel density estimation, a function is defined, given by

$$\hat{p}\,(t) = \sum_{i=1}^{m} \sum_{j=1}^{n_{sp}} \kappa \left( \frac{t - t_{ij}}{h} \right) \tag{7.3}$$

where $t_{ij}$ represents the $j$th sampling time that was found to be optimal for model $i$, $h$ represents the smoothing parameter, and $\kappa$ represents the kernel function, for which a Gaussian-like function is chosen, given by

$$\kappa(u) = e^{-u^2} . \tag{7.4}$$

The reason why the factors $1/(n \cdot h)$, $1/\sqrt{2\pi}$ and $1/2$ were omitted from Eqs. (7.1) and (7.2) to form Eqs. (7.3) and (7.4), is that in kernel density estimation $\hat{p}(x)$ represents an estimate of a probability density function for which $\int_{-\infty}^{+\infty} p(x)\,dx$ has to be equal to one. Since this is not required when applying this method for optimal experimental design purposes, these factors were omitted for simplicity. Note that this does not influence the resulting experiment.

Now, to select $n_{sp,c}$ compromise sampling times from the $n_{sp} \cdot m$ optimal sampling times ($n_{sp,c} \leq m \cdot n_{sp}$), the following approach is adopted. The compromise sampling times are those that maximize $\hat{p}(t)$, under the constraint that a minimum time interval between two sampling times is required by the experimental setup. Basically, the compromise sampling times correspond to those points of $\hat{p}(t)$ that a horizontal line through $\max(\hat{p}(t))$ encounters while going down, taking into account the requirement of a minimum time interval. This is illustrated in Fig. 7.2, where three compromise sampling times are determined. In this figure, the black and gray dots represent optimal sampling times for two rival models. The compromise sampling times correspond to those points that maximize $\hat{p}(t)$ (represented by the full line) under the constraint that a minimum time interval between two sampling times is maintained, and their location is indicated by the white dots. Note that, in principle, the presented approach allows the compromise sampling times to coincide. Indeed, one can always select (some of) the compromise sampling times more than once, at the expense of sampling times with lower $\hat{p}(t)$-values that were originally selected as compromise sampling times.

### 7.2.3 Choice of the smoothing parameter

An interesting feature of the proposed method is the fact that neighboring samples intensify each other, because the associated kernel functions overlap. The extent to which this intensifying effect occurs not only depends on the time interval between the individual sampling times, but also on the smoothing parameter $h$ (as illustrated in Fig. 7.1). The intensifying effect is hardly present for small values of $h$ (as in Fig. 7.1a), where an erratic and noisy estimate of $p(x)$ is found. For too large values of $h$, a very smooth and out-of-focus estimate of $p(x)$ is obtained, as in Fig. 7.1c. The latter is obviously explained by the fact that higher values of the smoothing parameter result in broader kernel functions

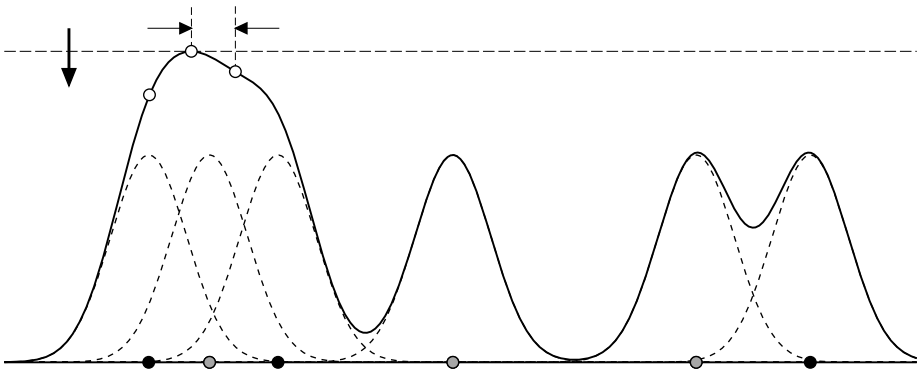


**Figure 7.2:** The black dots represent the optimal sampling times for model $i$, whereas the gray ones represents those for model $j$. The compromise sampling times correspond to those points that maximize $\hat{p}(t)$ (represented by the full line) under the constraint that a minimum time interval between two sampling times is maintained. Their location is indicated by the white dots.

(dashed lines in Fig. 7.1), which overlap with those of other samples. For the sharp kernel functions obtained with the lower value of the smoothing parameter, no significant overlap occurs.

Note that techniques are available to determine an optimal smoothing parameter in the context of kernel density estimation, but for this the reader is referred to Jenssen et al. (2006), Jones et al. (1996) or Turlach (1993), and the references therein. However, in this context, we prefer to link this parameter to the minimum time interval, denoted as $t_{min}$. Since $t_{min}$ is dictated by the experimental setup, the choice of the smoothing parameter is then straightforward, general and objective. More specifically, we suggest to define $h$ as $0.25 \cdot t_{min}$.

This suggestion is based on a compromise. On the one hand, a high value for $h$ is desired because the intensifying effect becomes more apparent (as explained above), but on the other hand, the value for $h$ must not be too high. The latter is illustrated in Fig. 7.3, where two sampling times are shown that are located as close to each other as allowed by the minimum time interval (the black dots). Since both sampling times are equally important, they should both be selected as compromise sampling times (indicated by the white dots). This is the case when a value of $0.25 \cdot t_{min}$ is chosen for the smoothing parameter (see Fig. 7.3a), whereas higher values of $h$ result in compromise sampling times that are different from the optimal ones and lower values of $h$ will decrease the intensifying effect (see Fig. 7.3d).

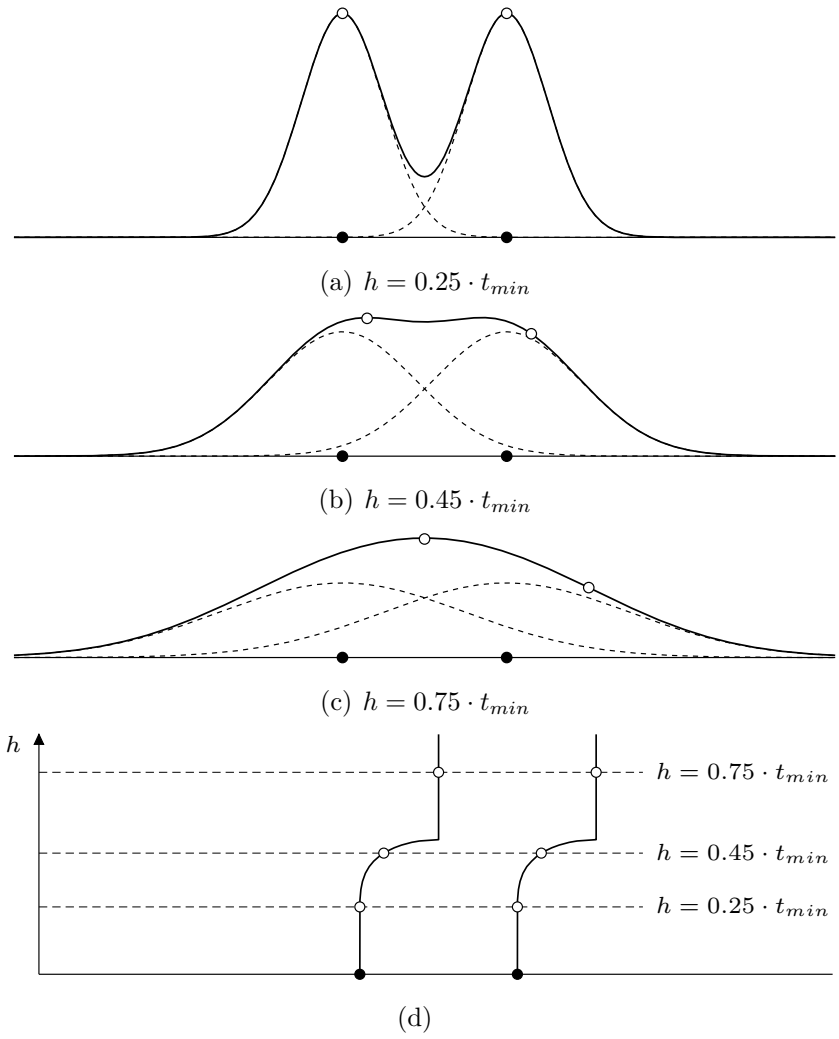



**Figure 7.3:** Kernel density estimation applied to an illustrative example for three different values of the smoothing parameter $h$. The time interval between the two sampling times (black dots) is equal to the minimum time interval ($t_{min}$), and the location of the compromise sampling times is indicated by the white dots. The lower graph (d) indicates where the compromise sampling times will be located for values of $h$ between 0 and 1 times the minimum time interval.

### 7.2.4 Weighing of the sampling times

The contribution of the individual sampling times to the information content of the experiment varies, and some sampling times are thus more important/informative than others. With the presented method, it is possible to give weights to the $n_{sp}$ individual sampling times, representing their importance. The higher their contribution to the information content of the experiment, the higher their weight. For this, Eq. (7.3) can be easily modified into

$$\hat{p}\,(t) = \sum_{i=1}^{m} \sum_{j=1}^{n_{sp}} w_{ij} \cdot \kappa \left( \frac{t - t_{ij}}{h} \right), \tag{7.5}$$

where $w_{ij}$ represents the weight of the $j$th sampling time of the optimal experiment for model $i$ ($t_{ij}$).

The weight of an individual sampling time is defined relative to the extent to which a chosen optimality criterion (see Section 6.5) diminishes when that sampling time is removed from the set of optimal sampling times. Below, an approach is presented to determine the weights for two cases, one with the D-optimality design criterion and one with the modE-optimality design criterion. Similar approaches can be formulated for the other design criteria that were described in Section 6.5.

In the case where the D-optimality design criterion is used (maximized), the weight can be determined as follows. Suppose $\boldsymbol{\xi}_i^\star$ represents the optimal experiment for model $i$, and $\boldsymbol{\xi}_{i\setminus t_j}$ represents the same experiment, but without sampling time $t_j$. Then, the weight of this sampling time can be determined as

$$w_{ij} = \frac{\mathrm{D}\left(\boldsymbol{\xi}_i^\star\right) - \mathrm{D}\left(\boldsymbol{\xi}_{i\setminus t_j}\right)}{\mathrm{D}\left(\boldsymbol{\xi}_i^\star\right)}, \quad 0 \leq w_{ij} \leq 1\,, \tag{7.6}$$

where $\mathrm{D}\left(\boldsymbol{\xi}\right)$ represents the D-optimality design criterion value associated with experiment $\boldsymbol{\xi}$. When sampling time $t_j$ is not important with regard to the information content of the experiment, $\det\left(\mathbf{FIM}\left(\boldsymbol{\xi}_{i\setminus t_j}\right)\right)$ will only be slightly smaller than $\det\left(\mathbf{FIM}\left(\boldsymbol{\xi}_i^\star\right)\right)$ and $w_{ij}$ will be close to zero. Besides, it would not be realistic that $\det\left(\mathbf{FIM}\left(\boldsymbol{\xi}_{i\setminus t_j}\right)\right)$ is larger than $\det\left(\mathbf{FIM}\left(\boldsymbol{\xi}_i^\star\right)\right)$ since information is lost when a sampling time is discarded. Taking into account that the eigenvalues and thus the determinant of the **FIM** cannot take negative values (the **FIM** is a positive definite matrix (Walter and Pronzato, 1997)), the value of $w_{ij}$ cannot be negative.

Contrary to the D-optimality design criterion, the modE-optimality design criterion has to be minimized. A similar equation for the calculation of the weights can be formulated as follows:

$$w_{ij} = \max\left(0, \frac{\text{modE}\left(\boldsymbol{\xi}_{i\setminus t_j}\right) - \text{modE}\left(\boldsymbol{\xi}_i^\star\right)}{\text{modE}\left(\boldsymbol{\xi}_{i\setminus t_j}\right)}\right), \quad 0 \leq w_{ij} \leq 1\,, \tag{7.7}$$

where $\text{modE}\left(\boldsymbol{\xi}\right)$ represents the modE-optimality design criterion value associated with experiment $\boldsymbol{\xi}$. When sampling time $t_j$ does not have a significant influence on the value of the modE-optimality design criterion, the weight will be close to zero. In this case, however, it cannot be guaranteed that discarding an additional sampling time leads to a worse (larger) value of the modE-optimality design criterion. Therefore, $w_{ij}$ can in principle take negative values. For the case studies described in the following, such a situation did not occur, but if it occurs one can set the weight of the corresponding sampling time to zero. The latter is achieved using the $\max\left(0, \cdot\right)$ operation in Eq. (7.7).

### 7.2.5 Extending the kernel-based method for all types of experimental degrees of freedom

The kernel-based method presented above was developed to determine compromise sampling times. However, it is often the case that more experimental degrees of freedom are available to the experimenter, for instance when the initial conditions of one or more state variables can be chosen and/or when certain process inputs can be manipulated during the course of the experiment. When additional experimental degrees of freedom become available, a wider palette of process conditions can be reached and experiments with a higher information content may be (designed and) performed. Therefore, it would be interesting to know if the kernel-based method could be extended such that it could also be used for such experimental design problems.

At first sight, it may seem straightforward to extend the method for more complicated experimental design problems. Indeed, the black and gray dots from Fig. 7.2, in which the rationale of the kernel-based method is presented, represent the optimal sampling times of two rival models, but they can also be seen as the optimal initial values of a certain state variable, the optimal timings to alter the value of one of the process inputs, or the optimal values of any other experimental degree of freedom available to the experimenter.

As explained in Section 7.2.2, the kernel-based method requires the choice of a so-called smoothing parameter that determines the width of the kernel functions. For the case where only the sampling times are optimized, this smoothing parameter can be linked to and calculated from the required minimum time interval between two measurements (as explained in Section 7.2.5). Since the latter is dictated by the experimental setup, this smoothing parameter can be determined in a systematic and objective manner. For other experimental degrees of freedom, such as the initial conditions of certain process variables or the timing of a pulse, such an approach is not (readily) available.

However, the biggest challenge (or problem) may arise when several types of experimental degrees of freedom are combined or when the manipulations are optimized. Suppose, for instance, that one is dealing with an experimental design exercise where both the sampling times and the initial conditions are to be optimized. The kernel-based method determines the compromise values for these experimental degrees of freedom from the optimal values obtained for each of the individual rival models prior to the application of the kernel-based method. However, one cannot guarantee that the compromise sampling times determined with the kernel-based method will be informative in combination with the compromise values obtained for the initial conditions. A similar reasoning holds when one or more manipulatory variables (or process inputs) are optimized. For instance, when one of the manipulatory variables (the feed rate or the concentration of one of the process variables) is set at another value a certain time after the start of the experiment, the task of the experimental design then consists of finding the optimal new value for the manipulatory variable and the optimal time to change it. It is clear that both experimental degrees of freedom are dependent of each other and the optimal value for the one depends on the values given to the other. They can thus not simply be determined separately, as done when the kernel-based method is applied for one experimental degree of freedom. In a way, one could state that as more experimental degrees of freedom become available, the optimal experiments will become more and more specific for the corresponding (rival) models, and the rationale of the kernel-based method may be too simple to cope with this specificity. Note that this will be further investigated and discussed in Section 7.6, where the kernel-based method is applied to an experimental design problem where the initial conditions are optimized for fixed sampling times.

## 7.3 Ideal point method for the design of compromise experiments

As discussed above, the use of the kernel-based method is limited to experimental design problems where the sampling times are the only experimental degrees of freedom. Therefore, an alternative method is presented in this section, that is able to cope with experimental design problems where experimental degrees of freedom of all types are considered (manipulations, initial conditions, sampling times and combinations thereof). Since the optimal experiment for model $m_i$ may not be optimal for model $m_j$, the experimental design problem can be seen as a multi-objective problem, where the aim is to systematically and simultaneously optimize (that is, maximize or minimize) a number of possibly conflicting objectives, each of which is translated into an objective function. Without loss of generality, it is assumed in the following that the aim of the optimization exercise is to minimize these objective functions.

### 7.3.1 Multi-objective optimization problems

A general multi-objective optimization problem is posed as (Deb, 2001; Marler and Arora, 2004):

$$\min_{\boldsymbol{x}\in\boldsymbol{X}} \quad \boldsymbol{F}(\boldsymbol{x}) \equiv \left[f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \ldots, f_d(\boldsymbol{x})\right]', \tag{7.8}$$

where $\boldsymbol{x} = [x_1, x_2, \ldots, x_q]' \in \boldsymbol{X}$ represents the $q$-dimensional design or decision vector, $\boldsymbol{X}$ represents the feasible design or decision space, and $\boldsymbol{F}(\boldsymbol{x})$ represents the $d$-dimensional vector containing the individual objective function values $f_i(\boldsymbol{x})$. The feasible criterion space, denoted as $\boldsymbol{Z}$, is defined as the set $\{\boldsymbol{F}(\boldsymbol{x}) \mid \boldsymbol{x} \in \boldsymbol{X}\}$ (Deb, 2001; Marler and Arora, 2004).

In this respect, it is important to realize that each point in the design space maps to a point in the criterion space (see Fig. 7.4), but the reverse may not be true. For instance, in general, there does not exist a vector $\boldsymbol{x}$ for which each objective function is minimal. The point in criterion space that is formed by the individual minima of the different objective functions is called the ideal point (Deb, 2001), denoted as $\boldsymbol{f}^\circ$ and defined as follows:

$$\boldsymbol{f}^\circ = \left[\min_{\boldsymbol{x}\in\boldsymbol{X}} f_1(\boldsymbol{x}), \min_{\boldsymbol{x}\in\boldsymbol{X}} f_2(\boldsymbol{x}), \ldots, \min_{\boldsymbol{x}\in\boldsymbol{X}} f_d(\boldsymbol{x})\right]'. \tag{7.9}$$

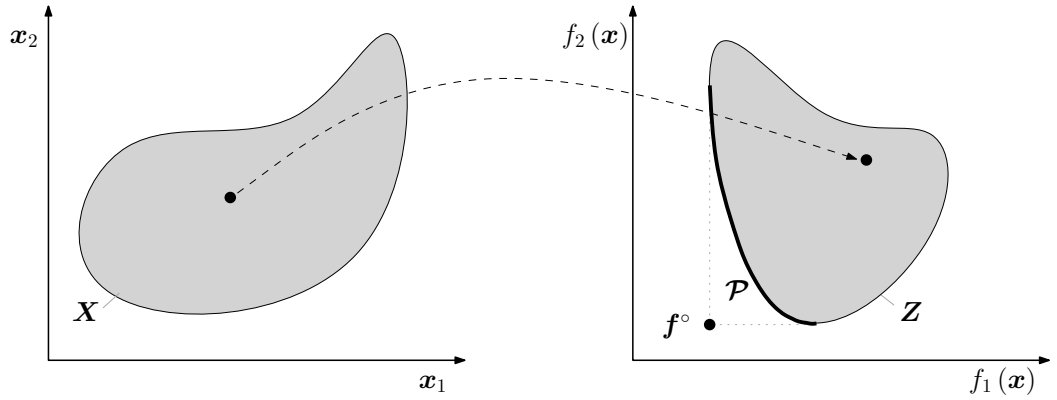



**Figure 7.4:** Feasible design space ($\boldsymbol{X}$) and feasible criterion space ($\boldsymbol{Z}$) for a hypothetical multi-objective problem with two design variables ($x_1$ and $x_2$) where the aim is to simultaneously minimize two objective functions ($f_1$ and $f_2$). In addition, the Pareto-optimal front ($\boldsymbol{\mathcal{P}}$) and the ideal point ($\boldsymbol{f}^\circ$) are shown.

In an experimental design context, $\boldsymbol{x}$ represents an experiment, denoted as $\boldsymbol{\xi}$, and $\boldsymbol{X}$ represents the set of all possible experiments, denoted as $\Xi$. The objective functions correspond to the D-optimality design criteria associated with each of the rival models. So, the number of objective functions equals the number of rival models. Note that also the other design criteria that were described in Section 6.5 could have been used (for instance, the A-optimality criterion or the modE-optimality criterion), but one could also use the cost or the duration of the experiment as a design criterion. For more information on these design criteria, the reader is referred to Chapter 6 and the work of Atkinson and Donev (1992), Munack (1991), Petersen (2000) and Vanrolleghem and Dochain (1998).

### 7.3.2 Solving multi-objective optimization problems

In contrast to single-objective optimization, a solution to a multi-objective problem is more a concept than a definition. In the case of conflicting objective functions, the resulting multi-objective optimization problem gives rise to a set of points that all fit a predetermined definition of an optimum (Deb, 2001). The predominant concept in defining an optimal point is that of Pareto-optimality, which is best explained and defined through the concept of dominance (Deb, 2001; Marler and Arora, 2004; Smith et al., 2004).

**Definition 7.1**:
*A point, $\boldsymbol{a} \in \boldsymbol{X}$, dominates another point, $\boldsymbol{b} \in \boldsymbol{X}$, if $f_i(\boldsymbol{a}) \leq f_i(\boldsymbol{b})$ for all $i = 1, \ldots, d$; and $f_i(\boldsymbol{a}) < f_i(\boldsymbol{b})$ for at least one $i$.*

Often, the notation $\boldsymbol{a} \prec \boldsymbol{b}$ is used to indicate that $\boldsymbol{a}$ dominates $\boldsymbol{b}$. It is clear that this relationship is not a total order and two points can be mutually non-dominating if neither dominates the other. Pareto-optimality is closely related to dominance, and is defined as follows:

**Definition 7.2**:
*A point, $\boldsymbol{x}^\star \in \boldsymbol{X}$, is Pareto-optimal if there does not exist another point, $\boldsymbol{x} \in \boldsymbol{X}$, such that $\boldsymbol{f}(\boldsymbol{x}) \leq \boldsymbol{f}(\boldsymbol{x}^\star)$, and $f_i(\boldsymbol{x}) < f_i(\boldsymbol{x}^\star)$ for at least one objective function.*

In words, a point is Pareto-optimal if there is no other point that improves at least one objective function without worsening another objective function. Or, a point is said to be Pareto-optimal if no other point dominates it. The set of all Pareto-optimal or non-dominated points is known as the Pareto-front, denoted as $\boldsymbol{\mathcal{P}}$ and shown in Fig. 7.4. Each point located on this front may thus in a sense be considered as optimal. In practice, one or some of the Pareto-optimal points will eventually be selected by the decision maker.

However, although several optimization algorithms are described in literature to determine the Pareto front (Deb, 2001), it often appears to be a difficult and computationally demanding task (Deb, 2001; Goel et al., 2007). This is especially true for the optimal experimental design applications focused on in this work, where the evaluation of an experiment proposed by the optimization algorithm involves several model simulations. In addition, the problem of finding the Pareto front may become prohibitively complex as the number of objectives increases, and visualizing the Pareto front is difficult for problems with more than three dimensions. The latter is important because the Pareto front will eventually be used as a reference by the decision maker/experimenter, who has to choose which experiment will be performed.

### 7.3.3 Ideal point method (for experimental design)

To overcome the issues raised above, multi-objective problems are often translated into single-objective problems. Such an approach is also proposed here, where the optimal solution is defined as the point that is as close as possible to the ideal point (defined above), and for which all individual objective functions are thus as close as possible to their corresponding minima (Deb, 2001; Marler and Arora, 2004). The presented method is often called the ideal point method (Deb, 2001), and this terminology will be used in the following.

To define closeness, different mathematical measures of distance can be used, such as the $\ell_p$ distance function (or Minkowski distance function) (Deb, 2001). The $\ell_p$ distance of any point in criterion space $\boldsymbol{x}$ from the ideal point $\boldsymbol{f}^\circ$ can be calculated as

$$\ell_p(\boldsymbol{x}) = \left( \sum_{i=1}^{d} |f_i(\boldsymbol{x}) - f_i^\circ|^p \right)^{1/p}, \tag{7.10}$$

where $p$ can take any value between 1 and $+\infty$. In this work, only the $\ell_1$, the $\ell_2$ and the $\ell_\infty$ distance functions are considered.

When the $\ell_1$ distance function ($p = 1$) is used, the multi-objective problem reduces to a single-objective problem where the different objective functions are simply summed. Note that this $\ell_1$ distance function is also known as the taxicab or Manhattan distance function. When the $\ell_2$ distance function ($p = 2$) is used, the Euclidean distance between the ideal point and any point in criterion space is minimized. For larger values of $p$, the largest term of Eq. (7.10) will dominate the value of $\ell_p(\boldsymbol{x})$ more and more, and the distance function associated with $p = +\infty$ (also called the Chebyshev distance function) eventually becomes

$$\ell_\infty(\boldsymbol{x}) = \max_{i=1,\ldots,d} |f_i(\boldsymbol{x}) - f_i^\circ| \,. \tag{7.11}$$

The multi-objective problem thus reduces to a problem where the maximal deviation from the ideal point is minimized. The working principle of the ideal point method for each of these three distance functions discussed above is shown in Figs. 7.5, 7.6 and 7.7, respectively. In these figures, the feasible criterion space is indicated by the gray area and the Pareto front is represented by the thick black line. The dotted gray lines indicate those points that are equally far from the ideal point ($\boldsymbol{f}^\circ$) according to the corresponding distance functions and the optimal solution is indicated by the ○ symbol.

In an experimental design context, the ideal point is defined by the optimal experiments for the individual models. When the D-optimality design criterion is used, the compromise experiment, denoted as $\boldsymbol{\xi}_c$, is found and defined as follows:

$$\boldsymbol{\xi}_c = \arg \min_{\boldsymbol{\xi} \in \boldsymbol{\Xi}} \left( \sum_{i=1}^{d} |\mathrm{D}(m_i, \boldsymbol{\xi}_i^\star) - \mathrm{D}(m_i, \boldsymbol{\xi})|^p \right)^{1/p}. \tag{7.12}$$

Here, $\mathrm{D}(m_i, \boldsymbol{\xi})$ represents the D-optimality design criterion value for model $m_i$ associated with experiment $\boldsymbol{\xi}$, $\mathrm{D}(m_i, \boldsymbol{\xi}_i^\star)$ represents the D-optimality design criterion value for model

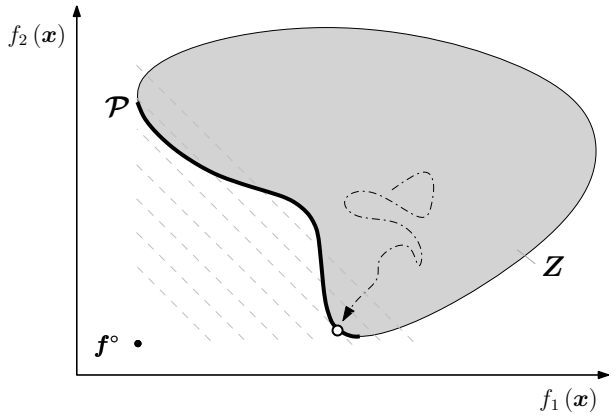



**Figure 7.5:** Illustration of the ideal point ($\boldsymbol{f}^\circ$) method for a hypothetical multi-objective problem with two design variables ($\boldsymbol{x}_1$ and $\boldsymbol{x}_2$), where the aim is to simultaneously minimize two objective functions ($f_1$ and $f_2$) and where the $\ell_1$ distance function is used (Manhattan distance).

$m_i$ associated with its corresponding D-optimal experiment ($\boldsymbol{\xi}_i^\star$), and $p$ is equal to 1, 2 or $+\infty$.

## 7.4 Evaluating the capability to design compromise experiments

To evaluate the presented method on its capability to design a compromise experiment, the following approach was adopted. Since an optimal experiment was designed for each model, each of these experiments could have been performed instead of the compromise experiment. The information that is lost or gained when doing so, is used for the evaluation.

Because the information content of an experiment is reflected by the value of the design criterion, the basis of the evaluation lies in the comparison of these criterion values. In this respect, it is important to realize that the information content or the quality of an experiment with regard to the parameters of a particular model can be compared to that of another experiment, but it is not meaningful to compare design criterion values of different models.

When the D-optimality criterion is used, the criterion values are calculated for each of the D-optimal experiments ($\boldsymbol{\xi}_j^\star$, with $j = 1, \ldots, m$), and these are compared to the criterion value associated with the compromise experiment ($\boldsymbol{\xi}_c$). The ratio between these criterion values, denoted as $\Gamma_{\mathrm{D}_{ij}}$, is eventually used for the evaluation, and is calculated as

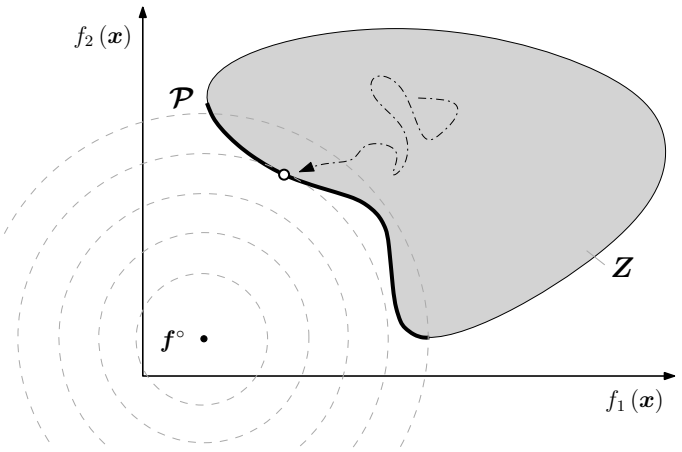



**Figure 7.6:** Illustration of the ideal point ($\boldsymbol{f}^\circ$) method for a hypothetical multi-objective problem with two design variables ($\boldsymbol{x}_1$ and $\boldsymbol{x}_2$), where the aim is to simultaneously minimize two objective functions ($f_1$ and $f_2$) and where the $\ell_2$ distance function is used (Euclidean distance).

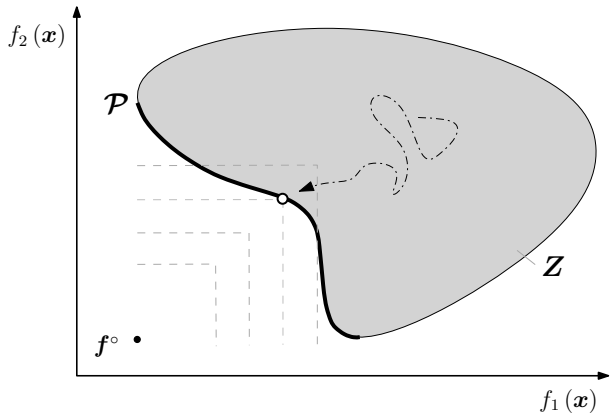



**Figure 7.7:** Illustration of the ideal point ($\boldsymbol{f}^\circ$) method for a hypothetical multi-objective problem with two design variables ($\boldsymbol{x}_1$ and $\boldsymbol{x}_2$), where the aim is to simultaneously minimize two objective functions ($f_1$ and $f_2$) and where the $\ell_\infty$ distance function is used (Chebyshev distance).

$$\Gamma_{\mathrm{D}_{ij}} = \frac{\mathrm{D}\left(m_i, \boldsymbol{\xi}_c\right)}{\mathrm{D}\left(m_i, \boldsymbol{\xi}_j^\star\right)}, \tag{7.13}$$

where $\mathrm{D}\left(m_i, \boldsymbol{\xi}_c\right)$ represents the D-optimality criterion value for model $m_i$ associated with experiment $\boldsymbol{\xi}_c$ and $\mathrm{D}\left(m_i, \boldsymbol{\xi}_j^\star\right)$ represents the D-optimality criterion value for model $m_i$ associated with experiment $\boldsymbol{\xi}_j^\star$.

Since a higher information content is represented by a higher value of the D-optimality design criterion, it holds that $\Gamma_{\mathrm{D}_{ij}} > 1$ when the compromise experiment contains more information with regard to the parameters of model $m_i$ than the optimal experiment for model $m_j$ ($\boldsymbol{\xi}_j^\star$). In other words, when $\Gamma_{\mathrm{D}_{ij}} > 1$, the estimates of the parameters of model $m_i$ should be more accurate when the compromise experiment is performed instead of experiment $\boldsymbol{\xi}_j^\star$.

For the modE-optimality design criterion, which has to be minimized, smaller criterion values are associated with better experiments. The expression used to quantify the designed compromise experiment is therefore given by

$$\Gamma_{\mathrm{modE}_{ij}} = \frac{\mathrm{modE}\left(m_i, \boldsymbol{\xi}_j^\star\right)}{\mathrm{modE}\left(m_i, \boldsymbol{\xi}_c\right)}, \tag{7.14}$$

where $\mathrm{modE}\left(m_i, \boldsymbol{\xi}\right)$ represents the value of the modE-optimality design criterion for experiment $\boldsymbol{\xi}$ with regard to the parameters of model $m_i$. Since a lower modE-optimality design criterion corresponds to a better experiment, $\Gamma_{\mathrm{modE}_{ij}} > 1$ when the compromise experiment ($\boldsymbol{\xi}_c$) is preferred to the optimal experiment for model $j$ ($\boldsymbol{\xi}_j^\star$) with regard to the estimation of the parameters of model $m_i$.

## 7.5 Case study IV: Kernel-based method to determine optimal sampling times for the simultaneous estimation of the parameters of rival mathematical models

In this case study, the kernel-based method will be used for the design of a compromise experiment in a case where the sampling times are the only experimental degrees of freedom available to the experimenter. As described in Section 7.2, this method was developed to determine those sampling times that are optimal for the simultaneous estimation of the parameters of rival models, and in the following the ability of the kernel-based method to design such a compromise experiment will be evaluated.

### 7.5.1 Objective of this case study

The objective of this case study is to investigate whether compromise sampling times can be determined using the kernel-based method. In other words, the capability of the kernel-based method to design a compromise experiment will be evaluated for cases where the sampling times are optimized.

### 7.5.2 Preliminary experiment

To initiate the case study, a preliminary experiment was defined and performed *in silico*. For this experiment, the volume of the reaction vessel was set to 10 mL, and the initial glucokinase concentration was set such that 5 units were present in the reaction mixture. Further, it was assumed that no G6P, ADP or PEP were present at the start of the experiment, and the initial concentrations of glucose and ATP were set to 1.5 mM and 0.5 mM, respectively.

Two pulses were given during the course of the experiment, both with a volume of 1 mL. The first pulse was given five minutes after the start of the experiment, and only contained ATP. The ATP concentration was chosen such that the ATP concentration in the reaction mixture was raised to 1.5 mM. The second pulse, given ten minutes after the start of the experiment, contained glucose and PEP, and their concentrations were chosen such that the resulting concentrations were 1.5 mM and 0.1 mM, respectively.

The experiment stopped after 20 minutes, and ten measurements of GLU, ATP, G6P and ADP were taken in duplicate (see Fig. 7.8). To mimic the error on the measurements, the approach described in Section 3.6 was used. For this purpose, the minimal relative errors ($\varsigma$) were arbitrarily set to 0.05 for all measured state variables, and the lower accuracy bounds on the measurements were defined as 0.1 mM.

### 7.5.3 Parameter estimation

The parameters of the rival models were estimated from the data of the preliminary experiment (Fig. 7.8) using the optimization algorithm described in Section 2.11. Since negative parameter values would not make any sense, the lower bounds were set to zero. The upper bounds were set to 1000 U/mg for parameter $k$, 2 mM for parameter $K_{\mathrm{GLU}}$, 50 mM for parameter $K_{\mathrm{ATP}}$, and 25 mM for parameter $K_{\mathrm{PEP}}$. The results of this parameter estimation exercise are shown in Table 7.1. From these results, one can conclude that the precision

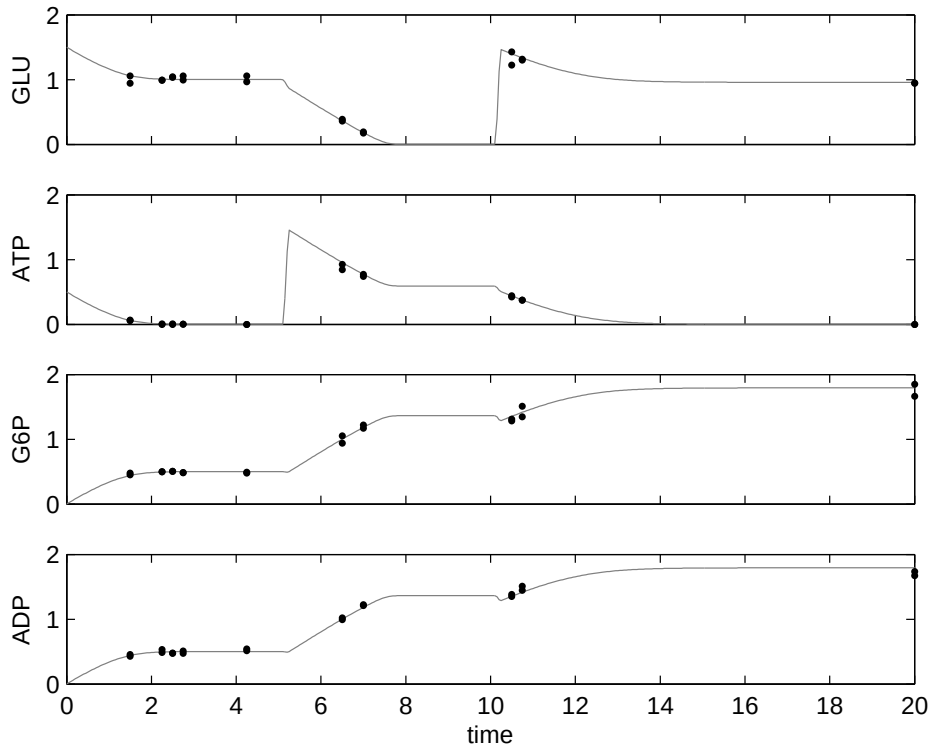



**Figure 7.8:** Preliminary experiment simulated with the *true* model ($m_5^\star$) and the experimental data derived from it (represented by the • symbols).

of the parameter estimates is quite low, indicating that it may be beneficial to perform a compromise experiment to increase the precision of the parameter estimates prior to the start of the model discrimination procedure.

### 7.5.4 Optimal experimental design for parameter estimation

For each rival model, an experiment was designed to accurately estimate its parameters. The experimental degrees of freedom were chosen as in the preliminary experiment, except for the ten sampling times, which were optimized. This experimental design exercise was performed both for the case where the D-optimality design criterion was optimized and the case where the modE-optimality design criterion was optimized. The results of these experimental design exercises are shown in Figs. 7.9 and 7.10, respectively. Then, for both cases, a compromise experiment is designed based on the corresponding optimal experiments, the results of which are discussed in the following.

**Table 7.1:** Parameters of the real model ($m_5^\star$) that were used to generate experimental data, and the parameter estimates obtained after fitting the rival models to the data from the preliminary experiment, together with the 95% confidence intervals and the corresponding WSSE values.

| model | $k$ | $K_{\mathrm{GLU}}$ | $K_{\mathrm{ATP}}$ | $K_{\mathrm{PEP}}$ | WSSE |
|---|---|---|---|---|---|
| $m_5^\star$ | 312 | 0.15 | 0.13 | 0.10 | – |
| $m_1$ | 314.13 ± 90.48 | 0.0173 ± 0.1135 | 0.1407 ± 0.0694 | – | 61.5287 |
| $m_2$ | 336.14 ± 107.66 | 0.0451 ± 0.1341 | 0.1533 ± 0.0772 | 0.1466 ± 0.2198 | 57.1080 |
| $m_3$ | 317.21 ± 93.38 | 0.0191 ± 0.1162 | 0.1412 ± 0.0705 | 0.0091 ± 0.0544 | 56.9125 |
| $m_4$ | 307.64 ± 49.17 | 0.1299 ± 0.8481 | 0.1245 ± 0.0441 | – | 61.2821 |
| $m_5$ | 312.41 ± 51.28 | 0.2011 ± 0.9320 | 0.1207 ± 0.0461 | 0.1261 ± 0.2145 | 56.9285 |
| $m_6$ | 319.87 ± 55.23 | 0.3112 ± 1.0616 | 0.1182 ± 0.0491 | 0.1076 ± 0.3577 | 57.1491 |
| $m_7$ | 412.58 ± 180.30 | 0.0099 ± 0.1706 | 27.9603 ± 464.53 | – | 94.2223 |
| $m_8$ | 428.11 ± 236.59 | 0.0148 ± 0.2146 | 19.4047 ± 265.68 | 8.7458 ± 127.33 | 77.5805 |
| $m_9$ | 543.60 ± 438.34 | 0.1102 ± 0.3893 | 3.6812 ± 9.2598 | 0.0127 ± 0.0327 | 88.2185 |

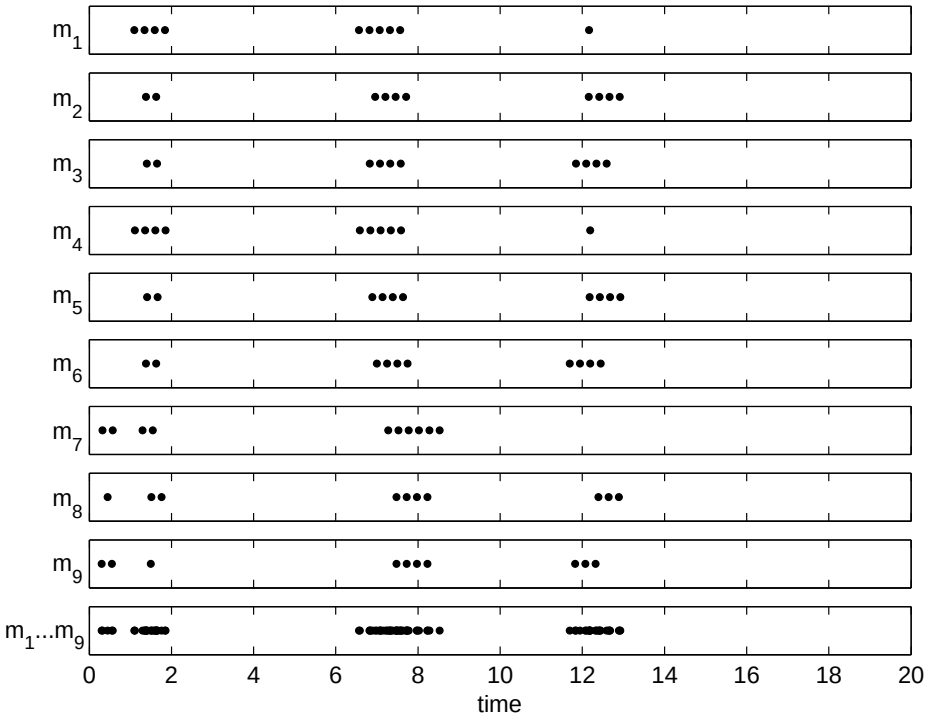



**Figure 7.9:** Optimal sampling times (•) found for the case where the D-optimality design criterion is optimized. The graph at the bottom is obtained by plotting the ten optimal sampling times of the nine individual models on the same axis.

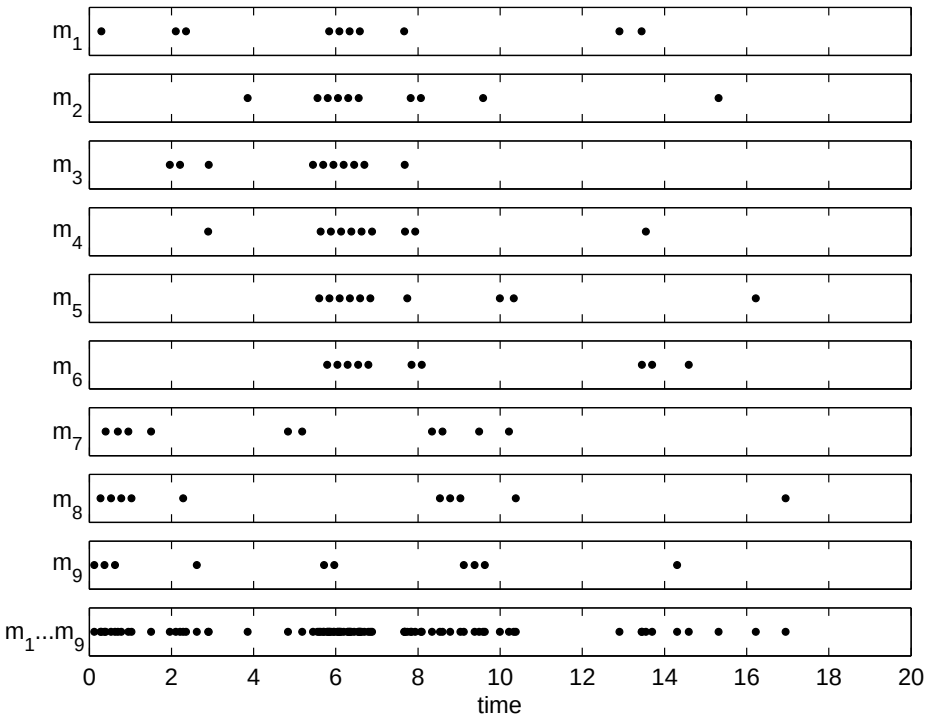



**Figure 7.10:** Optimal sampling times (•) found for the case where the modE-optimality design criterion is optimized. The graph at the bottom is obtained by plotting the ten optimal sampling times of the nine individual models on the same axis.

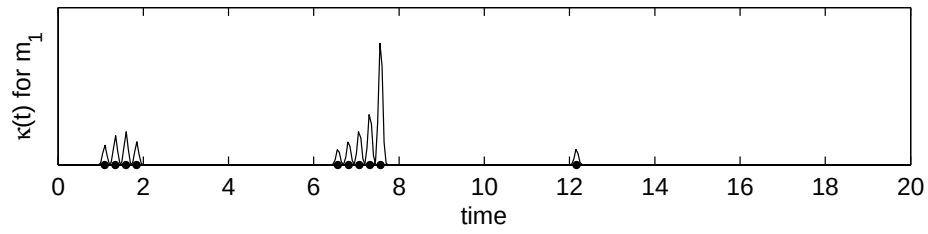



**Figure 7.11:** Weighted kernel functions associated with the D-optimal sampling times obtained for model $m_1$.

### 7.5.5 Compromise experiment obtained from D-optimal experimental designs

This section describes the results for the case where the D-optimality design criterion is optimized (maximized). To clearly illustrate the different steps of the methodology, these results will be discussed in more detail than the results for the case where the modE-optimality design criterion is used (Section 7.5.6).

Based on the D-optimal experiments corresponding to the individual models (shown in Fig. 7.9), the weighted kernel functions were determined after weighing the sampling times as explained in Section 7.2.4. As an example, the kernel functions associated with the D-optimal sampling times for model $m_1$ are shown in Fig. 7.11. This figure clearly illustrates that the contribution of the individual sampling times to the information content of the experiment differs, and that it makes sense to weigh the different sampling times.

The weighted kernel functions associated with the different models are used to calculate $\hat{p}(t)$ according to Eq. (7.5), and the resulting trajectory of $\hat{p}(t)$ is shown in Fig. 7.12. The compromise sampling times are determined from this trajectory as explained in Section 7.2.2, and their location is indicated by the white dots. One can see that the optimal sampling times are located in four groups, and that the compromise sampling times are (not surprisingly) spread over these groups as well. An interesting observation in Fig. 7.9 is that the optimal experiments for models $m_7$ to $m_9$ contain sampling times in the first minute of the experiment, while this is not the case for models $m_1$ to $m_6$. The fact that this region is important for one third of the models is reflected in the compromise experiment, where one of the sampling times is put in this region. This example clearly illustrates that the presented method considers the optimal sampling times for each of the models when designing the compromise experiment.

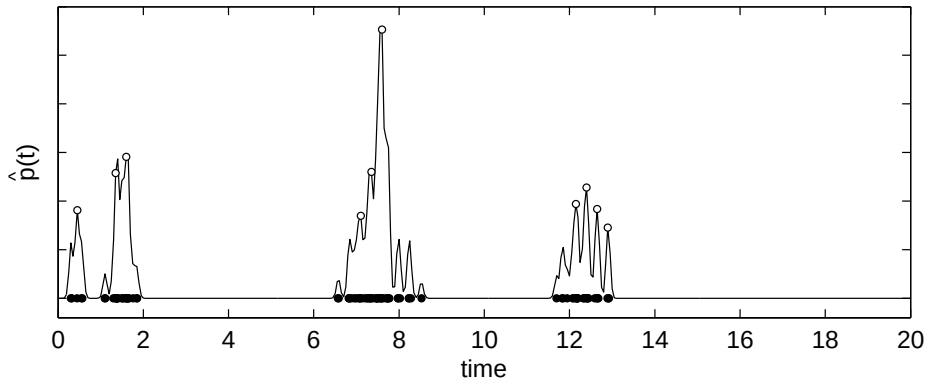



**Figure 7.12:** Trajectory of $\hat{p}\,(t)$ for the case where the D-optimality design criterion is applied, and illustration of how the compromise sampling times ($\circ$) are obtained from it. The optimal sampling times for the different models are represented by the black dots ($\bullet$), and are the same as the ones presented in Fig. 7.9 (bottom).

The capability of the presented method to design a compromise experiment was evaluated as explained in Section 7.4. The results of this evaluation are presented in Fig. 7.13. For each model $m_i$, the value of $\Gamma_{\mathrm{D}_{i1}}$ is represented by the black bar ($\boldsymbol{\xi}_1^\star$), and the bars become increasingly white as the model number increases ($\boldsymbol{\xi}_1^\star \rightarrow \boldsymbol{\xi}_9^\star$). To present the results in a systematic and easily interpretable form, the values of $\Gamma_{\mathrm{D}_{ij}}$ are represented on a logarithmic scale. In this way, it is easy to see when $\Gamma_{\mathrm{D}_{ij}} > 1$. Some interesting observations are discussed below.

The results for model $m_1$, for example, show that $\Gamma_{\mathrm{D}_{1j}} < 1$ for experiments $\boldsymbol{\xi}_1^\star$, $\boldsymbol{\xi}_3^\star$ and $\boldsymbol{\xi}_4^\star$, which indicates that the experimental designs to optimally estimate the parameters of models $m_1$, $m_2$ and $m_3$, respectively, contain more information with regard to the parameters of model $m_1$ than the compromise experiment. For the other optimal experiments, this is not the case and the compromise experiment is preferred. If one would perform $\boldsymbol{\xi}_4^\star$ instead of the compromise experiment, the information content would indeed be higher for model $m_1$, but it would be lower for the other models (except for model $m_4$, of course). The latter can be seen when comparing the bars corresponding to $\boldsymbol{\xi}_4^\star$ for the different models. Similar observations can be made for the other models/optimal experiments, which clearly shows the ability of the proposed method to design an experiment with the characteristics of a compromise experiment. That the compromise experiment is not optimal for the individual models is a direct result of the fact that the timings of the optimal sampling times are different for the individual models (see Fig 7.9). Yet, the compromise experiment seems to be sufficiently informative to improve the overall precision of the parameter estimates.

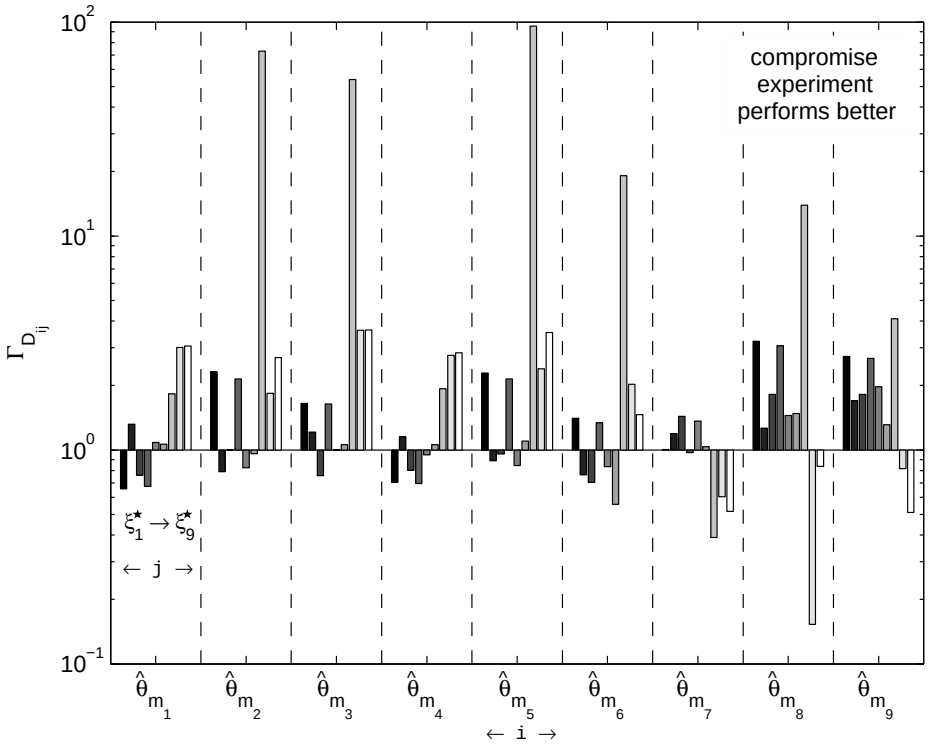



**Figure 7.13:** Comparison of the D-optimality design criterion values obtained when performing the compromise experiment, with those that would be obtained when the optimal experiments for the individual models were performed instead. In the figure, the ratio between these criterion values is shown for each model, and this for the optimal experiments associated with models $m_1$ (black bars) to $m_9$ (white bars).

### 7.5.6 Compromise experiment obtained from modE-optimal experimental designs

This section describes the results for the case where the modE-optimality design criterion is optimized (minimized). From Fig. 7.10, one can see that the optimal sampling times are not located in distinct groups as was the case for the D-optimality design criterion (see Fig. 7.9). This makes this exercise more challenging than the previous one.

The trajectory of $\hat{p}(t)$ and the compromise sampling times derived from it are shown in Fig. 7.14. The results of the evaluation are presented in Fig. 7.15 and indicate that the capability of the presented method to design a compromise experiment is at least as good as for the case where the D-optimality design criterion was used. Here too, it is clear that the presented method leads to an experiment with the characteristics of a compromise experiment. For each model, some of the optimal experiments are preferred to the compromise experiment ($\Gamma_{\mathrm{modE}_{ij}} < 1$), and vice verse ($\Gamma_{\mathrm{modE}_{ij}} > 1$). The results also show that for some of the models, a significant amount of information is lost when an experiment that is optimal for another model is performed instead of the compromise experiment. For instance, if experiments $\boldsymbol{\xi}_7^\star$, $\boldsymbol{\xi}_8^\star$ or $\boldsymbol{\xi}_9^\star$ were performed, a substantial amount of information would be lost with regard to the parameters of models $m_1$ to $m_6$, while the gain in information for models $m_7$ to $m_9$ is not that large.

### 7.5.7 Dependence of the experimental designs on the parameter estimates

The values of the D-optimality design criterion shown in Fig. 7.13 represent the expected information content of the designed experiments. As explained in Section 6.4, the latter is assessed based on the **FIM**. For linear models, the parameter sensitivities ($\partial\hat{\boldsymbol{y}}/\partial\boldsymbol{\theta}$) are, by definition, independent of the parameters to be estimated. Hence, the **FIM** is independent of the parameter estimates as well (Eq. (2.21)). However, for nonlinear models (as the ones used in this dissertation), this is not the case, and the **FIM**, as well as the design criteria derived from it, are dependent on the values of the parameter estimates available at the experimental design step. Thus, at the start of a modelling exercise, when the parameter estimates are still uncertain, the designed experiment may be less informative/optimal than expected, and the resulting parameter estimates may not be that accurate after all. Therefore, the designed experiments are often called locally optimal instead of optimal (Atkinson and Donev, 1992).

This effect of the parameter uncertainty is illustrated for the case where the D-optimality design criterion is used. First, the ten different experiments were performed *in silico* ($\boldsymbol{\xi}_c$ and the nine $\boldsymbol{\xi}_j^\star$), and using the data from these experiments the parameters of each of the nine models were re-estimated. Then, a figure similar to Fig. 7.13 is created (Fig. 7.16), but now with the better estimates of the parameters. As expected from the discussion above, the results in Fig. 7.16 show that the values for $\Gamma_{\mathrm{D}_{ij}}$ are not entirely the same as the ones presented in Fig. 7.13, which indicates that the parameter estimates changed after re-estimating them.

Nevertheless, the overall precision of the parameter estimates of all individual models has improved by performing the compromise experiment. This can be concluded from Table 7.2, in which the parameter estimates and their corresponding 95% confidence intervals are reported and from Table 7.3 where the procentual improvement of these 95% confidence intervals are reported. One should note, however, that the covariance or correlation between the parameter estimates is not considered in the calculation of the confidence intervals and may give an incomplete picture. Since the D-optimality criterion values are proportional to the volume of the confidence region of the parameter estimates (as explained in Section 6.5), they can also be interpreted as an overall measure for the precision of the parameter estimates. Therefore, a better picture of the uncertainty or precision of the parameter estimates can be obtained from Fig. 7.16.

## 7.6 Case study V: (In)ability of the kernel-based method to determine optimal initial conditions for the simultaneous estimation of the parameters of rival mathematical models

The kernel-based method was originally developed for experimental design exercises where one is interested in finding the optimal sampling times (Section 7.2). In Section 7.2.5, it was advocated that extending the kernel-based method so that is also able to produce compromise experiments in experimental design exercises with other (or more) experimental degrees of freedom is not straightforward. In this case study, the kernel-based method will be applied to an experimental design problem in which the initial values of the state variables are optimized. As the ideal point method was introduced as an alternative method, and should be thus able to cope with such experimental design problems. Although the results obtained with the ideal point method are discussed in detail in the following section (Section 7.7), they will also be presented here as a reference.

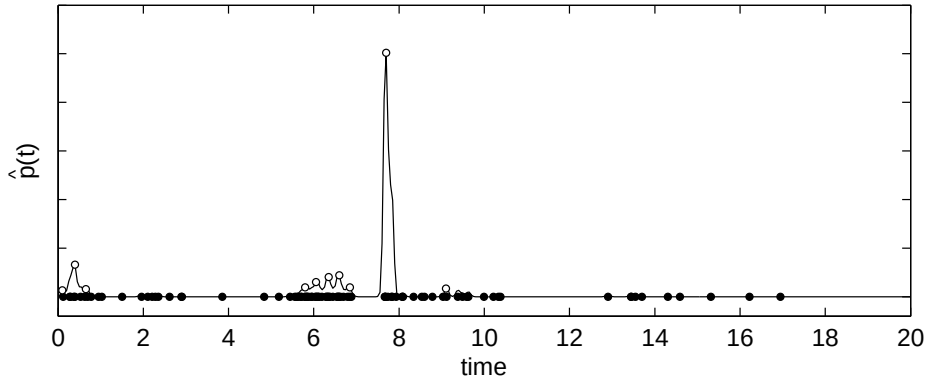



**Figure 7.14:** Trajectory of $\hat{p}\,(t)$ for the case where the modE-optimality design criterion is applied, and illustration of how the compromise sampling times ($\circ$) were obtained from it. The optimal sampling times for the different models are represented by the black dots ($\bullet$).

**Table 7.2:** Parameters of the real model ($m_5^\star$) that were used to generate experimental data, and the parameter estimates obtained after fitting the rival models to the data from both the preliminary experiment and the compromise experiment, together with the 95% confidence intervals and the corresponding WSSE values.

| model | $k$ | $K_{\mathrm{GLU}}$ | $K_{\mathrm{ATP}}$ | $K_{\mathrm{PEP}}$ | WSSE |
|---|---|---|---|---|---|
| $m_5^\star$ | 312 | 0.15 | 0.13 | 0.10 | – |
| $m_1$ | 356.75 ± 26.76 | 0.0383 ± 0.0191 | 0.2275 ± 0.0304 | – | 506.14 |
| $m_2$ | 328.35 ± 22.16 | 0.0332 ± 0.0177 | 0.1504 ± 0.0245 | 0.1182 ± 0.0273 | 128.50 |
| $m_3$ | 323.27 ± 20.49 | 0.0252 ± 0.0163 | 0.1458 ± 0.0230 | 0.0059 ± 0.0038 | 126.58 |
| $m_4$ | 341.31 ± 18.64 | 0.1930 ± 0.0891 | 0.1894 ± 0.0189 | – | 503.37 |
| $m_5$ | 317.09 ± 14.99 | 0.2063 ± 0.1097 | 0.1264 ± 0.0172 | 0.0945 ± 0.0219 | 128.78 |
| $m_6$ | 322.97 ± 15.62 | 0.1759 ± 0.1043 | 0.1351 ± 0.0252 | 0.0293 ± 0.0141 | 137.27 |
| $m_7$ | 583.38 ± 123.24 | 0.0095 ± 0.0618 | 49.9430 ± 315.47 | – | 1362.41 |
| $m_8$ | 358.86 ± 36.15 | 0.0034 ± 0.0329 | 40.4657 ± 394.03 | 17.6839 ± 174.29 | 630.45 |
| $m_9$ | 459.68 ± 66.39 | 0.0406 ± 0.0489 | 5.7254 ± 6.5114 | 0.0026 ± 0.0028 | 817.53 |

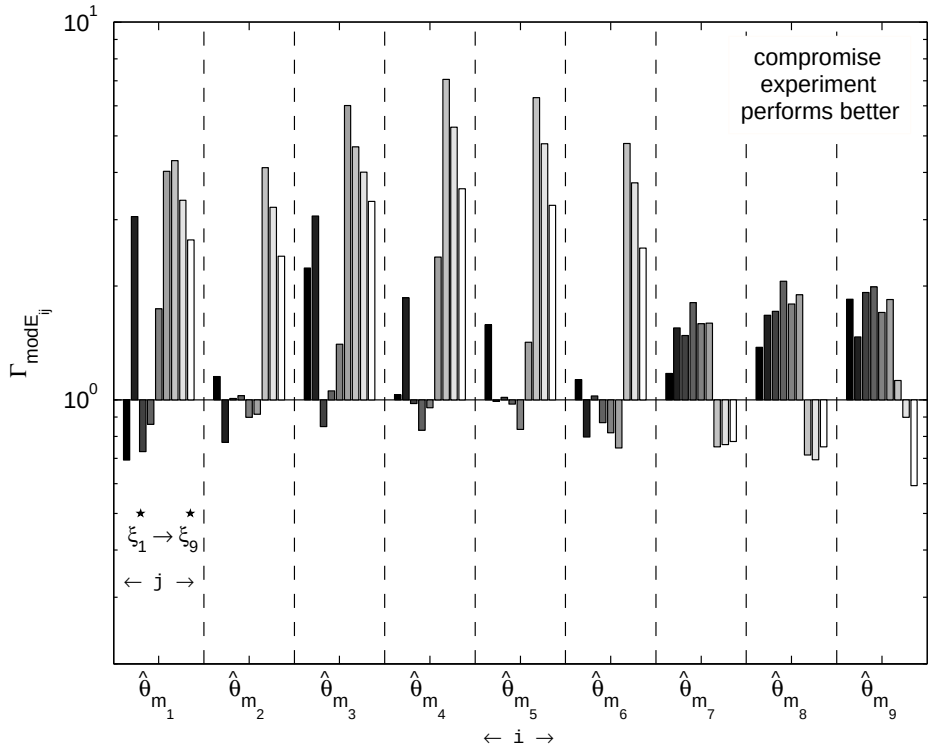



**Figure 7.15:** Comparison of the modE-optimality design criterion values obtained when performing the compromise experiment, with those that would be obtained when the optimal experiments for the individual models were performed instead. In the figure, the ratio between these criterion values is shown for each model, and this for the optimal experiments associated with models $m_1$ (black bars) to $m_9$ (white bars).

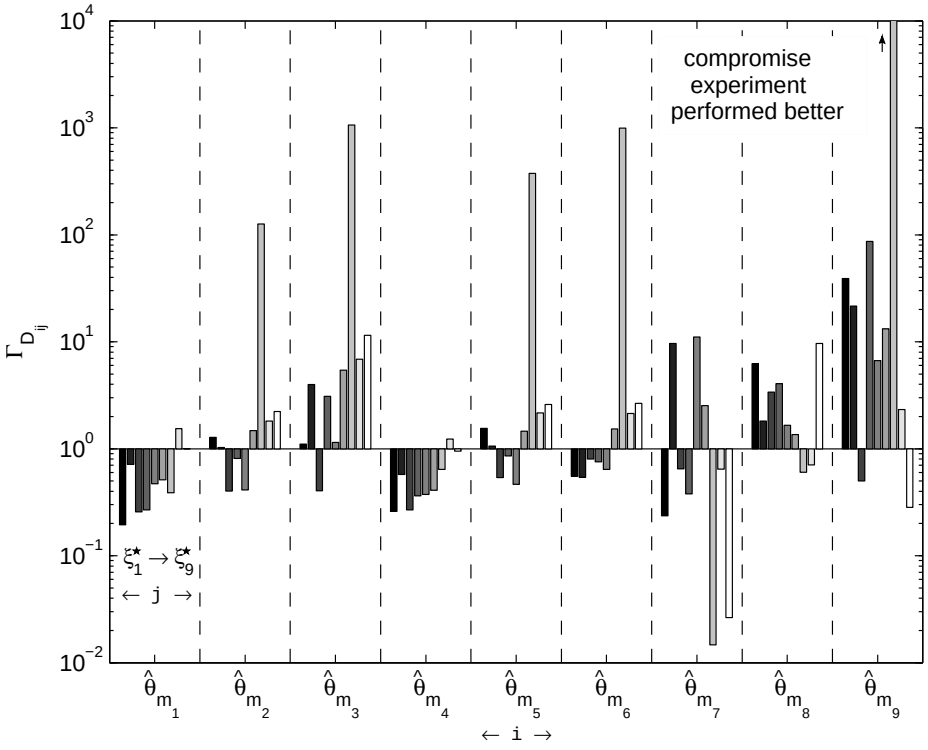



**Figure 7.16:** Comparison of the D-optimality design criterion values obtained after performing the compromise experiment and re-estimating the parameters, with those that were obtained when the optimal experiments for the individual models were performed instead. In the figure, the ratio between these criterion values is shown for models $m_1$ (black bars) to $m_9$ (white bars), and this for each optimal experiment.

**Table 7.3:** Procentual improvement of the 95% confidence intervals of the parameter estimates obtained after fitting the rival models to the data from both the preliminary experiment and the compromise experiment.

| model | $k$ | $K_{\mathrm{GLU}}$ | $K_{\mathrm{ATP}}$ | $K_{\mathrm{PEP}}$ |
|---|---|---|---|---|
| $m_1$ | 70% | 83% | 56% | – |
| $m_2$ | 79% | 87% | 68% | 88% |
| $m_3$ | 78% | 86% | 67% | 93% |
| $m_4$ | 62% | 89% | 57% | – |
| $m_5$ | 71% | 88% | 63% | 90% |
| $m_6$ | 72% | 90% | 49% | 96% |
| $m_7$ | 32% | 64% | 32% | – |
| $m_8$ | 85% | 85% | 148% | 137% |
| $m_9$ | 85% | 87% | 30% | 91% |

### 7.6.1 Objective of this case study

The objective of this case study is to explore the potential of the kernel-based method to design compromise experiments for experimental design problems where the initial conditions are to be optimized. The ultimate aim is, however, to investigate whether the kernel-based method, or at least the rationale behind it, can also be useful for cases where not only the sampling times are available for optimization, but also other experimental degrees of freedom such as the initial conditions and the manipulations.

### 7.6.2 Preliminary experiment and parameter estimation

To initiate the case study, a preliminary experiment was defined and performed *in silico*. The experiment was identical to the preliminary experiment from the case study discussed in the previous section (Fig. 7.8). Obviously, also the parameter estimates obtained from this preliminary experiment are identical to the ones presented earlier (Table 7.1).

### 7.6.3 Optimal experimental design for parameter estimation

The optimal values for the initial concentrations obtained after optimizing the D-optimality design criterion are shown in Fig. 7.17. The sampling times and the manipulations (the pulses) were fixed to the ones from the preliminary experiment, but the initial concentra-

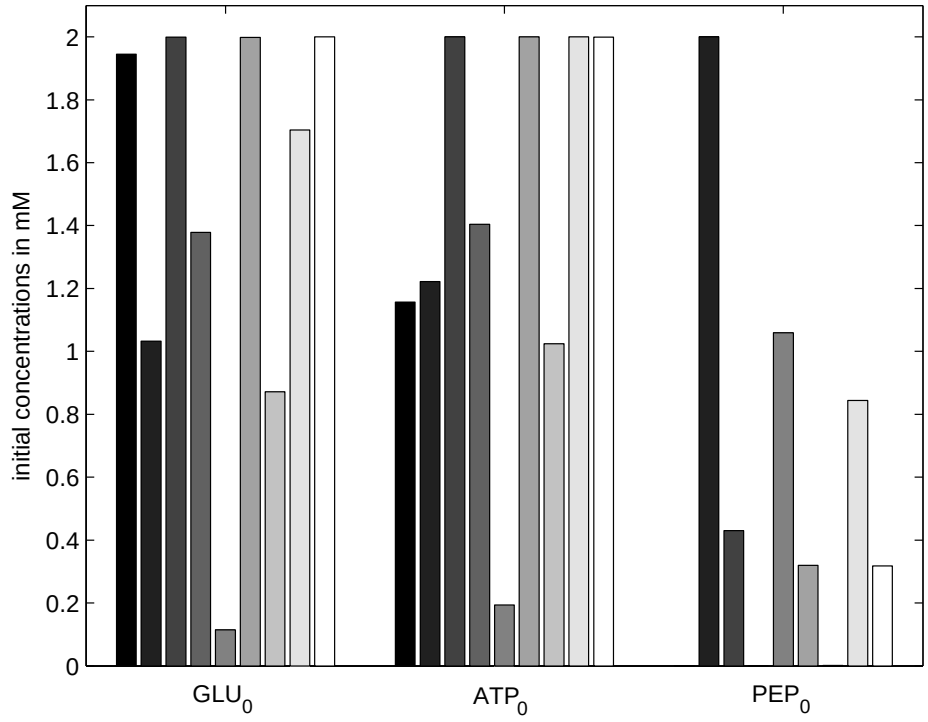



**Figure 7.17:** Optimal initial concentrations of glucose, ATP and PEP associated with the D-optimal experiments of the nine rival models ($m_1$ corresponds to the black bar, and the bars become increasingly white as the model number increases).

tions of glucose, ATP and PEP were optimized. For this, the lower bounds were set to 0 mM and the upper bounds to 2 mM. Because both the characteristics of the two pulses (timing and concentrations of glucose, ATP and PEP) and the sampling times are fixed, the initial conditions are chosen such that the information with regard to the parameters of the individual models is maximal at the given sampling times. This explains why the optimal experiments are very different for the individual rival models.

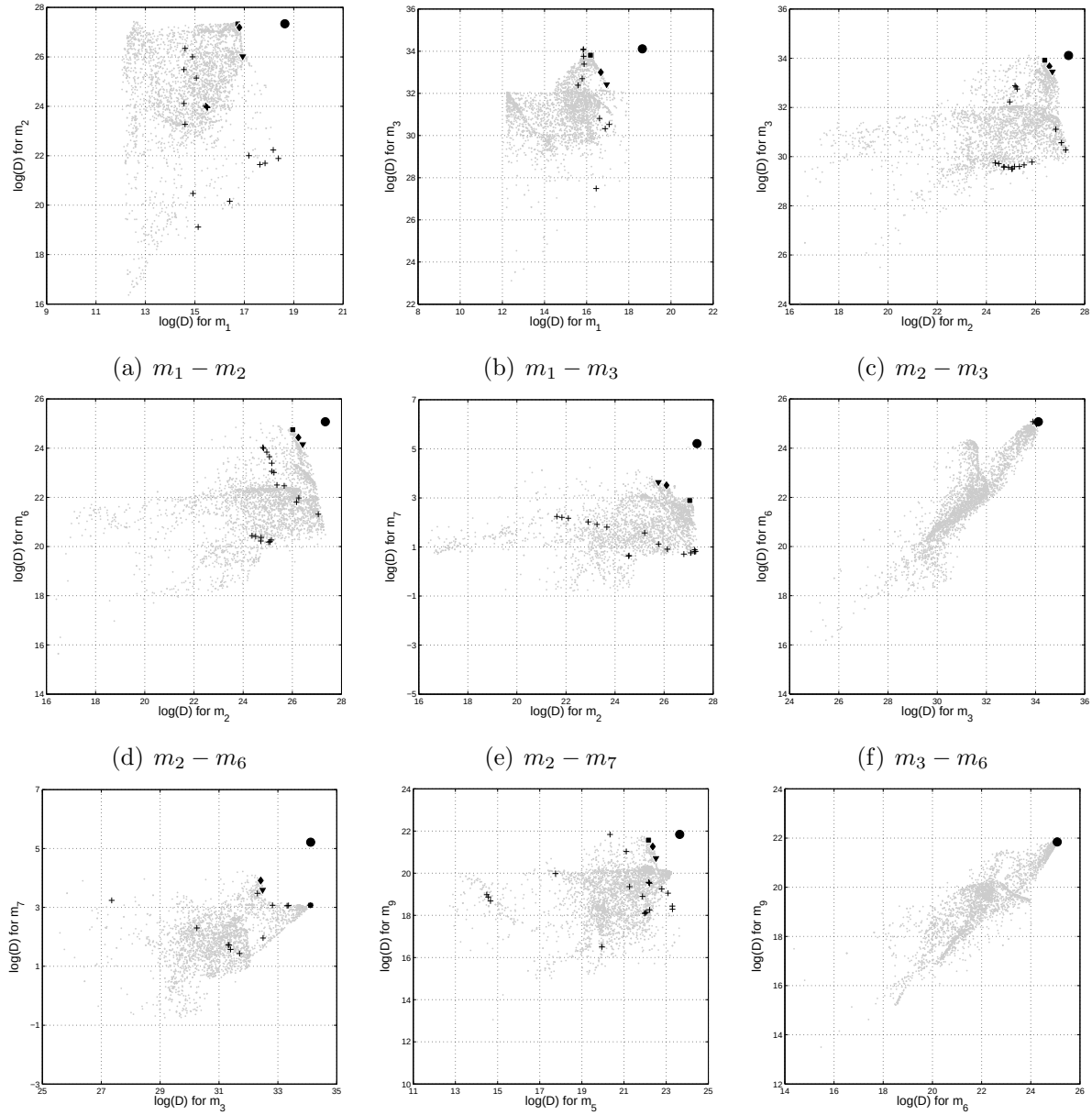
### 7.6.4 Design of the compromise experiment for the case with two rival models

In this section, the results that were obtained are shown for the case where only two rival models are considered in the experimental design exercises performed in this case study. From the nine rival models that were described in Chapter 3, 36 different model pairs can be defined and for each of these model pairs a compromise experiment was designed. It is true that one can discuss about the usefulness of designing a compromise experiment when there are only two rival models, but it is very difficult, if not impossible, to visualize

the results from a nine-dimensional optimization problem. The results obtained for nine of these model pairs are shown in Fig. 7.18 and discussed in Section 7.6.5.

As discussed in Section 7.2.5, one problem encountered when applying the kernel-based method for experimental design problems where the initial conditions are to be optimized, is the absence of an approach to determine the smoothing parameter in a systematic and objective manner. Therefore, a compromise experiment was designed for a range of smoothing parameters ($h = 0.05, 0.10, \ldots, 0.95, 2$) and the D-optimality design criterion values obtained for each of these experiments are shown in Fig. 7.18 for the corresponding models (+ symbols). The compromise experiments obtained with the ideal point method using the $\ell_1$, $\ell_2$ and $\ell_\infty$ distance functions are indicated by the ■, ♦ and ▼ symbols, respectively. The ideal point itself is indicated by the • symbol. Because these compromise experiments are determined after an optimization exercise, a whole range of experiments are evaluated before the compromise experiment is found. Because these experiments give an idea of the feasible criterion space (see Fig. 7.4), they are also indicated on this figure as gray dots.

From these results, one can see that the D-optimality design criterion values associated with the compromise experiments found using the kernel-based method (indicated by the + symbols) are significantly different from the ones found using the ideal point method, regardless of the distance function used (indicated by the ■, ♦ and ▼ symbols). This is not the case for the model pairs $m_3 - m_6$ and $m_6 - m_9$ (Figs. 7.18(f) and 7.18(i), respectively), but these models are very similar to each other (ATP is the first binding substrate) and the optimal values for the initial concentrations of glucose, ATP and PEP are more or less the same (see Fig. 7.17). It is therefore not surprising that an experiment is obtained that is optimal for both rival models, and that both methods result in the same experiment. Nevertheless, the results obtained for the other model pairs clearly indicate that the kernel-based method is less suited than the ideal point method when it comes to the design of compromise experiments for cases where the initial conditions are to be optimized. Indeed, regardless of the value of the smoothing parameter, the points in criterion space that correspond with the experiments found using this method are far from the so-called ideal point, which is associated with a generally non-existing experiment that is (D-)optimal for each of the rival models. An experiment that is close to this ideal point is thus likely to have the characteristics of a compromise experiment, as the information content of the experiment is large for both rival models (the D-optimality design criterion values are large). In addition, it is important to note that no value of the smoothing parameter could




(a) $m_1 - m_2$ (b) $m_1 - m_3$ (c) $m_2 - m_3$

(d) $m_2 - m_6$ (e) $m_2 - m_7$ (f) $m_3 - m_6$

(g) $m_3 - m_7$ (h) $m_5 - m_9$ (i) $m_6 - m_9$

**Figure 7.18:** D-optimality criterion values of the experiments found by the kernel-based method for a range of smoothing parameters (+ symbols), and of the experiments found by the ideal point method using the $\ell_1$, $\ell_2$ and $\ell_\infty$ distance functions (respectively represented by the ■, ♦ and ▼ symbols). The ideal point itself is represented by the • symbol. The D-optimality criterion values associated with the experiments that were encountered during the application of the ideal point method (optimization) are represented by the gray dots.

be observed that systematically resulted in better experiments, which obviously limits the applicability of this method.

### 7.6.5 Design of the compromise experiment for the case with nine rival models

As discussed in Section 7.2.5, the choice of the smoothing parameter ($h$) is one problem that arises when applying the kernel-based method to experimental design problems where not (only) the sampling times are to be optimized, but also other experimental degrees of freedom. The results presented in the previous section, where only two rival models were considered, already suggested that the ideal point method performs better than the kernel-based method for experimental design problems where the initial conditions are to be optimized. Here, the kernel-based method and the ideal point method are applied to the same case study, but all nine rival models are considered.

To investigate the performance of the kernel-based method, a range of smoothing parameter values ($h = 0.1, 0.2, \ldots, 2.4, 2.5$) were defined and for each of them a compromise experiment was determined. The $\Gamma_{\mathrm{D}_{ij}}$-values were calculated for each of the nine models and these are presented to the right of the dotted vertical line in Fig. 7.19 using boxplots. The same experimental design exercise was performed for the ideal point method and the $\Gamma_{\mathrm{D}_{ij}}$-values obtained with the three distance functions are shown on the left of the dotted vertical line in the same figure (from left to right: $\ell_1$, $\ell_2$ and $\ell_\infty$). Note that the results obtained with the ideal point method will be discussed in detail in Section 7.7.4 and are only shown here as a reference. To interpret the results presented in Fig. 7.19, it is important to know that a median value larger than one indicates that the compromise experiment performs better with regard to the parameters of the corresponding model than the majority of the D-optimal experiments (obtained for the other models).

From these results one can conclude that the performance of the kernel-based method is quite good, but not as good as the performance of the ideal point method. Indeed, despite the fact that the median of the $\Gamma_{\mathrm{D}_{ij}}$-values is predominantly larger than one, no value for the smoothing parameter could be identified for which this is the case for each of the rival models. For instance, for models $m_1$, $m_4$ and $m_7$, the best experiments were obtained for small values of $h$ while these experiments were the worst performing ones for the other models (for which higher values of $h$ are preferred). In other words, there is no value for the smoothing parameter for which the median values are all above one. This, as well as the lack of a systematic approach to determine the smoothing parameter, obviously limits the applicability of the kernel-based method.

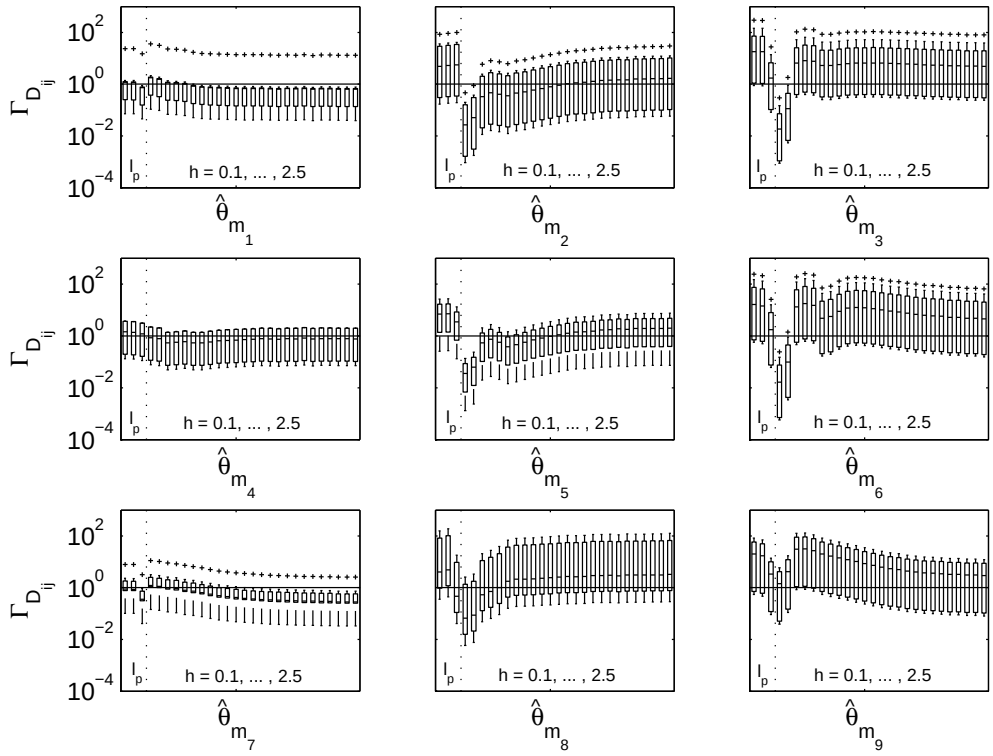



**Figure 7.19:** $\Gamma_{D_{ij}}$-values for the scenario in which the initial conditions are optimized. The $\Gamma_{D_{ij}}$-values obtained when the ideal point method is used are shown on the left of the dotted vertical line. On the right of this line, the ones obtained with the kernel-based method are shown for a range of values for the smoothing parameter (denoted as $h$).

## 7.7 Case study VI: Ideal point method to determine compromise experiments for the simultaneous estimation of the parameters of rival mathematical models

In this case study, the ideal point method is applied to design a compromise experiment for three different scenarios: optimization of the sampling times, optimization of the initial conditions of the experiment, and optimization of both the sampling times and the initial conditions. Because the preliminary experiment is the same as the one used in the case study described in Section 7.5, the results for the scenario where the sampling times are available for optimization obtained using the ideal point method can be compared with those obtained with the kernel-based method (discussed in Section 7.5).

### 7.7.1 Objective of this case study

With these examples, a number of research questions are addressed. First of all, the case study should demonstrate the capability or incapability of the ideal point method to design compromise experiments. In addition, it should indicate whether the choice of the distance function ($\ell_1$, $\ell_2$ and $\ell_\infty$) matters, and, if so, which of these metrics is preferred. Finally, it would be interesting to compare the performance of the ideal point method to that of the kernel-based method where possible.

### 7.7.2 Preliminary experiment and parameter estimation

To initiate the case study, a preliminary experiment was defined and performed *in silico*. The experiment was identical to the preliminary experiment from the case study discussed in section 7.5 (and depicted in Fig. 7.8). Obviously, also the parameter estimates obtained from this preliminary experiment are identical to the ones presented earlier (Table 7.1).

### 7.7.3 Optimization of sampling times

This section describes the results for the scenario in which ten sampling times were optimized. The initial concentrations and the characteristics of the two pulses that are given during the course of the experiment are fixed to the ones of the preliminary experiment. Further, it is assumed that a minimum time interval of 15 seconds is required by the experimental setup between two subsequent sampling times. To clearly illustrate the different steps of the ideal point method, the results of this scenario will be discussed in more detail

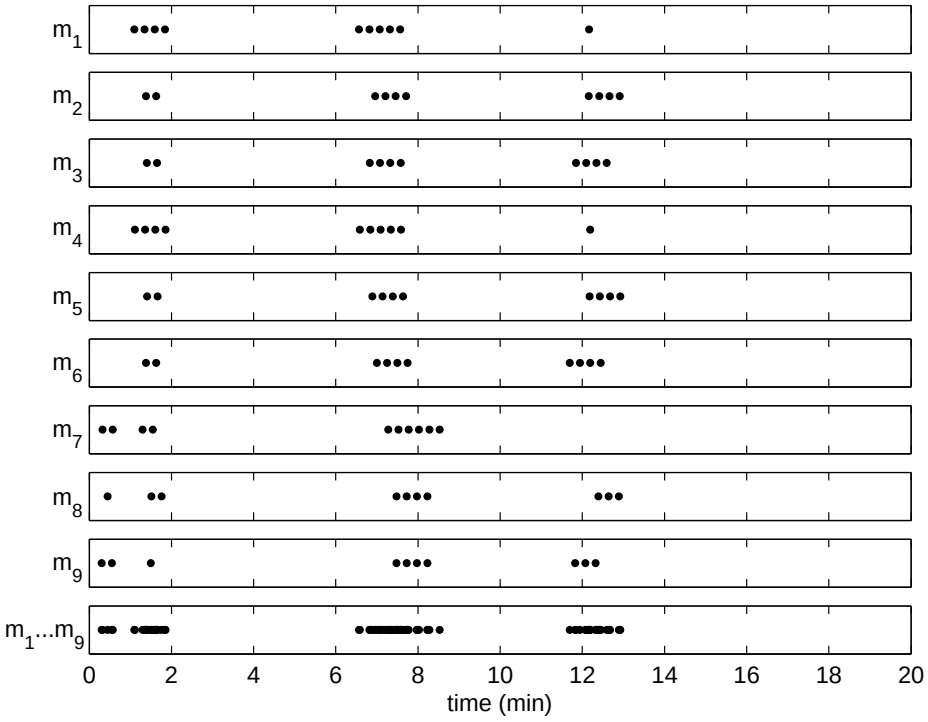



**Figure 7.20:** Optimal sampling times (•) found for the nine rival models for the case where the D-optimality design criterion was optimized by varying the sampling times. The graph at the bottom was obtained by plotting the ten optimal sampling times of the nine individual models on the same axis.

than the ones for the other scenarios. In addition, they will be compared to the results obtained and discussed in Section 7.5, where the kernel-based method was applied to the same case study.

The compromise experiment is found after minimizing the distance between the ideal point and the point in criterion space that corresponds with the experiment being proposed by the optimization algorithm (as explained in Section 7.3.2). Obviously, the ideal point has to be determined first. For this purpose, an experiment is designed for each of the rival models by optimizing the D-optimality design criterion (Eq. (7.9))). The results of these (nine) experimental design exercises are shown in Fig. 7.20.

Once the ideal point is determined, the compromise experiment is found by applying the optimization algorithm described in Section 2.11 to solve the optimization problem formalized in Eq. (7.9). This optimization exercise is done for each of the three distance functions described earlier ($\ell_1$, $\ell_2$ and $\ell_\infty$). The resulting compromise sampling times are shown in

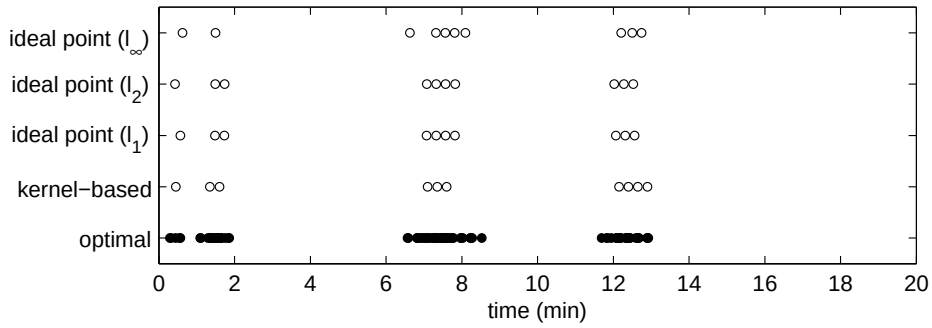



**Figure 7.21:** Optimal sampling times for the individual models (•) and the compromise sampling times (○) found using the kernel-based method, and the ideal point method using the different distance functions ($\ell_1$, $\ell_2$ and $\ell_\infty$).

Fig. 7.21, as well as the compromise sampling times found using the kernel-based method (Fig. 7.12). From these results, one can see that, except for the $\ell_\infty$ distance function, the compromise sampling times are similar for the different methods.

To evaluate the capability of the presented method to design a compromise experiment, the approach outlined in Section 7.4 is also adopted here. For this purpose, the $\Gamma_{\mathrm{D}_{i1}}$-values are calculated from Eq. (7.13) and presented as a barplot (see Fig. 7.22). For each model $m_i$, the value of $\Gamma_{\mathrm{D}_{i1}}$ is represented by the black bar ($\boldsymbol{\xi}_1^\star$), and the bars become increasingly white as the model number increases ($\boldsymbol{\xi}_1^\star \rightarrow \boldsymbol{\xi}_9^\star$). When $\Gamma_{\mathrm{D}_{ij}} > 1$, the estimates of the parameters of model $m_i$ should be more accurate when the compromise experiment is performed instead of experiment $\boldsymbol{\xi}_j^\star$. To present the results in a systematic and easily interpretable form, the values of $\Gamma_{\mathrm{D}_{ij}}$ are represented on a logarithmic scale. In this way, it is easy to see when $\Gamma_{\mathrm{D}_{ij}} > 1$.

For brevity, the results of this evaluation will only be discussed in detail for the case where the $\ell_2$ distance function was used. The results are shown in Fig. 7.22, and clearly illustrate the ability of the presented method to design an experiment with the characteristics of a compromise experiment. For instance, the $\Gamma_{\mathrm{D}_{i1}}$-values for model $m_1$ show that $\Gamma_{\mathrm{D}_{1j}} < 1$ for experiments $\boldsymbol{\xi}_1^\star$, $\boldsymbol{\xi}_3^\star$ and $\boldsymbol{\xi}_4^\star$, which indicates that these experiments contain more information with regard to the parameters of model $m_1$ than the compromise experiment. For the other optimal experiments, this is not the case and the compromise experiment is preferred. If one would perform $\boldsymbol{\xi}_4^\star$ instead of the compromise experiment, the information content would indeed be higher for model $m_1$, but it would be lower for the other models (except for model $m_4$, of course). The latter can be seen when comparing the bars corresponding to $\boldsymbol{\xi}_4^\star$ for the different models. That the compromise experiment is not optimal for the individual

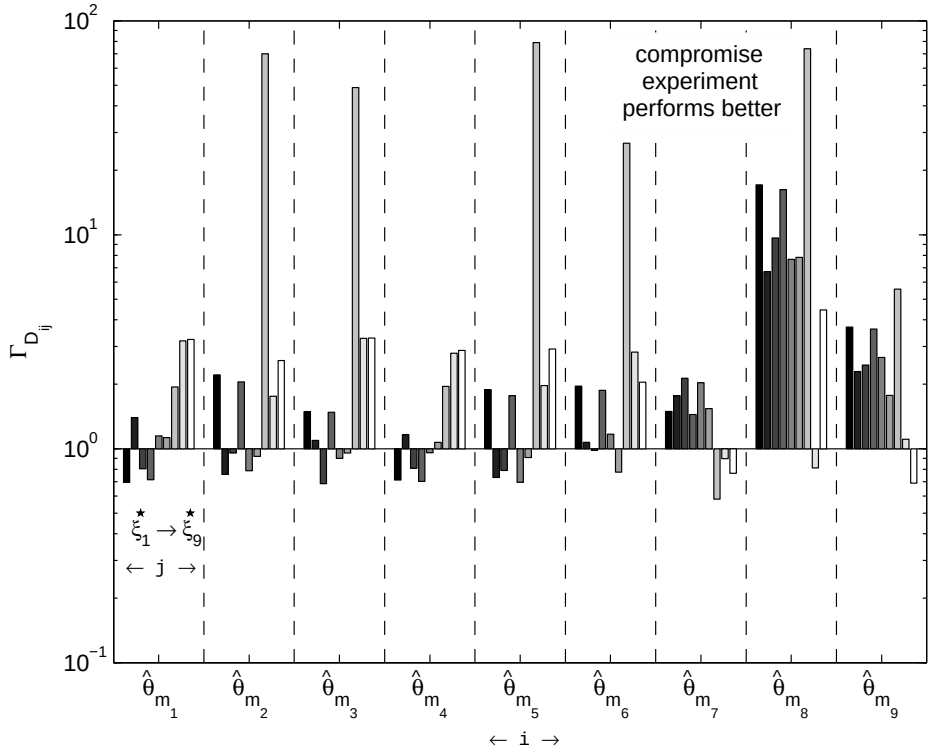



**Figure 7.22:** $\Gamma_{D_{ij}}$-values for the scenario in which the sampling times are optimized according to the $\ell_2$ distance function. The black bars correspond to the optimal experiments associated with model $m_1$, and the bars become increasingly white as the model number increases.

models is a direct result of the fact that the optimal sampling times are different for the individual models (see Fig. 7.20). Yet, the compromise experiment seems to be sufficiently informative to improve the overall precision of the parameter estimates.

Another, even more striking observation can be made from the results for the models $m_7$, $m_8$ and $m_9$. Apparently, each of the D-optimal experiments of the other models is significantly less informative with regard to their parameters compared to the compromise experiment, which performs quite well. This indicates that one or more compromise sampling times that are not present in the D-optimal experiments of the other models contain a lot of information on the parameters of models $m_7$, $m_8$ and $m_9$. Indeed, from Fig. 7.20 one can see that the D-optimal experiments for models $m_7$, $m_8$ and $m_9$ have one or two sampling times around 0.5 min, while this is not the case for models $m_1$ till $m_6$. Because one of the compromise sampling times is located in this time range (as shown in Fig. 7.21), the compromise experiment performs significantly better for these models $m_7$, $m_8$ and $m_9$

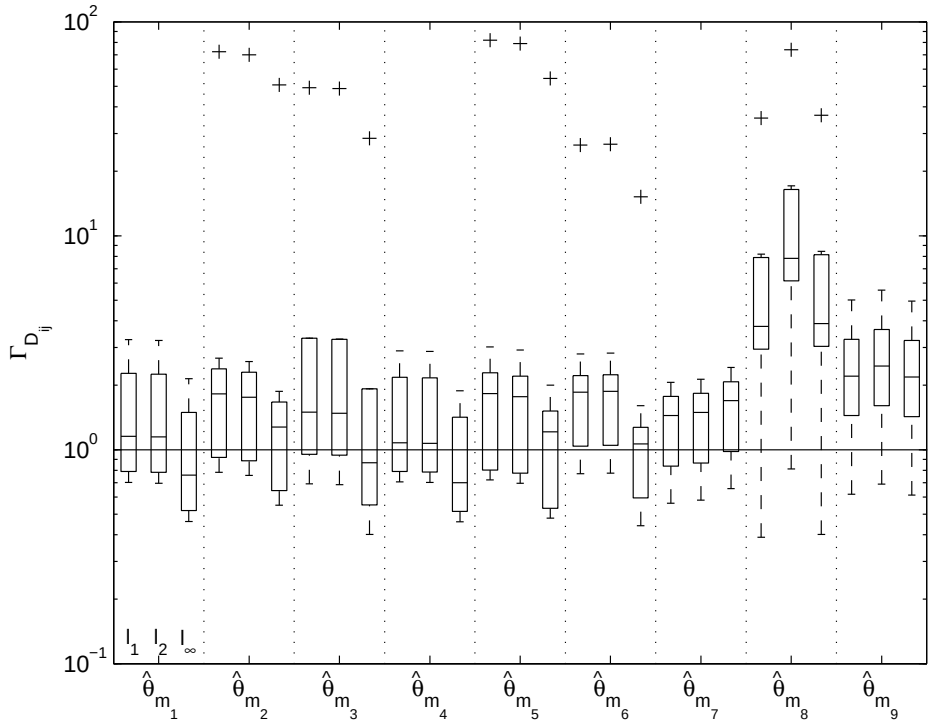



**Figure 7.23:** $\Gamma_{\mathrm{D}_{ij}}$-values for the scenario in which the sampling times are optimized, and where the $\ell_1$, $\ell_2$ and $\ell_\infty$ distance functions are used.

than the D-optimal experiments from the other models. In addition, one can see that the D-optimal experiments for models $m_7$ and $m_9$ have two sampling points at about 0.5 min, while the D-optimal experiment for model $m_8$ and the compromise experiment have only one. The latter explains why this phenomenon is slightly less pronounced for models $m_7$ and $m_9$.

To evaluate the results obtained with the other distance functions, the values of $\Gamma_{\mathrm{D}_{ij}}$ were calculated as well, but to facilitate the comparison between the distance functions, these values are presented in one figure using boxplots (Fig. 7.23). By presenting the $\Gamma_{\mathrm{D}_{ij}}$-values in this way, information is lost on which of the D-optimal experiments performs better or worse than the compromise experiment for the individual models, but this information is not essential for this purpose. In fact, the median is of great importance. If the median is above one, it indicates that the compromise experiment performs better than the majority of the D-optimal experiments.

That it is possible to design a compromise experiment using the $\ell_2$ distance function can of course be concluded from these boxplots as well. Indeed, the medians (indicated by the

horizontal lines in the box) are always larger than one, but on top of that, a larger part of the box is above this level. So, for each of the rival models, the compromise experiment performs better than the majority of the D-optimal experiments, which is in accordance with the required characteristics of a compromise experiment.

From Fig. 7.23, one can also observe that the results obtained with the $\ell_1$ distance function are very similar to the ones obtained with the $\ell_2$ distance function, which is not surprising given the fact that the compromise sampling times are nearly identical (Fig. 7.20). Indeed, situations can occur where the Pareto front is such that the solutions obtained with the different distance functions are the same. Although this is not the case for the hypothetical Pareto front shown in Figs. 7.5, 7.6 and 7.7, it would for instance be the case when the Pareto front looks like the one shown in Fig. 7.4. An example of this can also be found in Figs. 7.18(a), 7.18(f) and 7.18(i).

For the $\ell_\infty$ distance function, the compromise sampling times are slightly different. This is also reflected in the values of $\Gamma_{\mathrm{D}_{ij}}$, from which one can conclude that the majority of the D-optimal experiments perform slightly better than the designed compromise experiment for models $m_1$ to $m_6$. For models $m_7$, $m_8$ and $m_9$, which are the models where ATP is the first binding substrate and for which the model structures resemble each other, this is not the case. This suggests that for this example the $\ell_\infty$ distance function may not be the most suitable one to design a compromise experiment.

To conclude this section, the results obtained using the ideal point method are compared to the ones obtained with the kernel-based method. Although the compromise sampling times obtained with these methods are not identical (as shown in Fig. 7.21), the $\Gamma_{\mathrm{D}_{ij}}$-values are very similar (Figs. 7.13 and 7.22). From these results, one can conclude that the considered methods perform equally well for experimental design exercises where only the sampling times are optimized. Note, however, that the ideal point method is computationally more demanding, because next to the optimizations required to determine the ideal point, one additional optimization is needed to solve Eq. (7.9). This is not necessary when the kernel-based method is used, where the compromise sampling times are directly calculated from the optimal sampling times.

### 7.7.4 Optimization of initial conditions

This section describes the results for the scenario in which the sampling times and the manipulations (the pulses) were fixed to the ones of the preliminary experiment (described

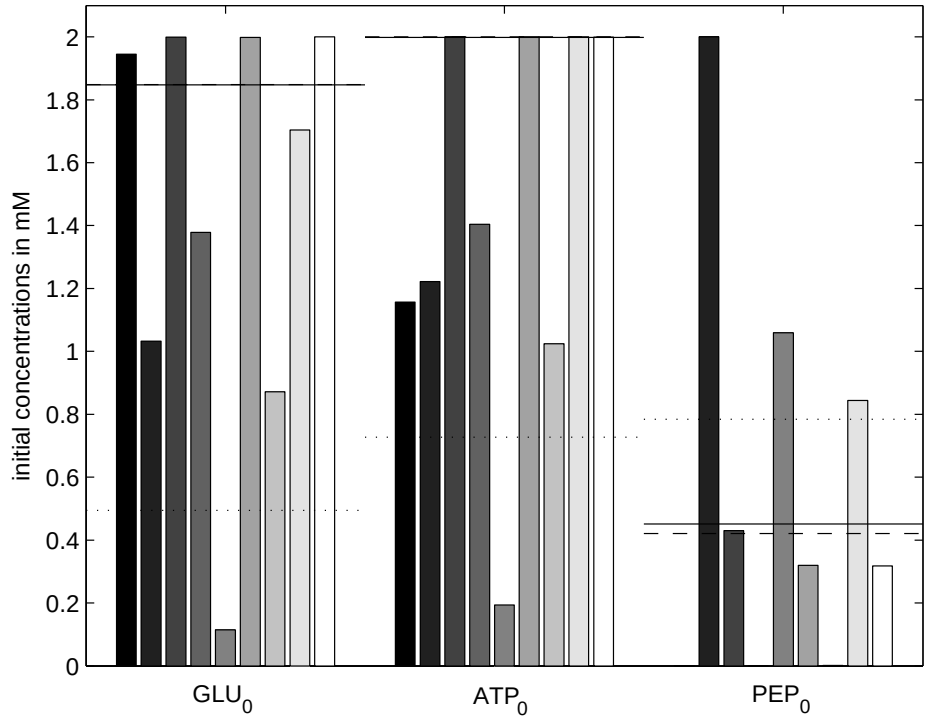



**Figure 7.24:** Initial concentrations of glucose, ATP and PEP associated with the D-optimal experiments of the rival models ($m_1$ corresponds to the black bar, and the bars become increasingly white as the model number increases), as well as those associated with the compromise experiment found using the ideal point method (horizontal lines) after using the $\ell_1$ (– –), the $\ell_2$ (—) and the $\ell_\infty$ ($\cdots$) distance function. Note that the lines for $\ell_1$ and $\ell_2$ coincide for $GLU_0$ and $ATP_0$.

in Section 7.7.2), but the initial concentrations of glucose, ATP and PEP were optimized. For this, the lower bounds were set to 0 mM and the upper bounds to 2 mM.

The optimal values for the initial concentrations are shown in Fig. 7.24 and are the same as the ones presented in the previous case study (Fig. 7.17). As explained in Section 7.6.3, optimal experiments are very different for the individual rival models because both the characteristics of the two pulses (timing and concentrations of glucose, ATP and PEP) and the sampling times are fixed, the initial conditions are chosen such that the information with regard to the parameters of the individual models is maximal at the given sampling times.

Based on the D-optimality criterion values associated with these optimal experiments, the ideal point was defined and a compromise experiment was designed using the three

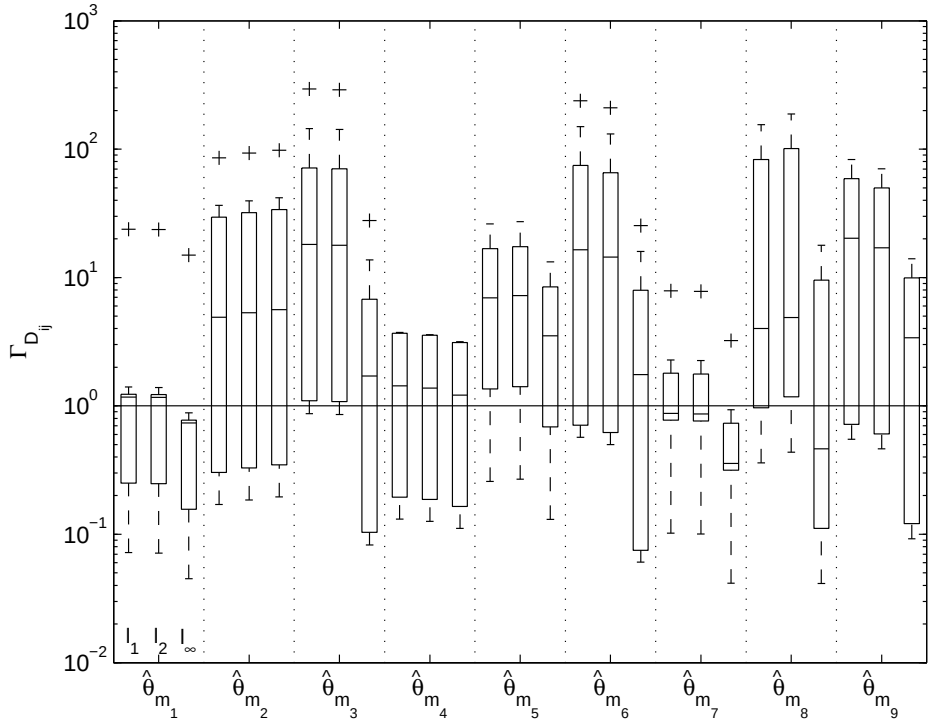



**Figure 7.25:** $\Gamma_{D_{ij}}$-values for the scenario in which the initial conditions are optimized, and where the $\ell_1$, $\ell_2$ and $\ell_\infty$ distance functions are used.

distance functions described above. In Fig. 7.24, the initial concentrations of the obtained compromise experiments (horizontal lines) are compared to the initial concentrations of the D-optimal experiments of the rival models. One can clearly see that the compromise experiments found using the $\ell_1$ and $\ell_2$ distance function are very similar, while a different experiment is found with the $\ell_\infty$ distance function.

The similarity between the compromise experiments found using the $\ell_1$ and the $\ell_2$ distance function is obviously reflected in the $\Gamma_{D_{ij}}$-values, shown in Fig. 7.25. For both cases, the boxplots indicate that the majority of these values are larger than one, indicating that it is advisable to perform the compromise experiment instead of the corresponding D-optimal experiments. Note, however, that the performance was not that good for models $m_1$ and $m_7$. For the case where the $\ell_\infty$ distance function is used, on the other hand, the barplots indicate that, although the majority of the medians is larger than one, some of the $\Gamma_{D_{ij}}$-values are significantly lower (for instance, for models $m_1$, $m_3$, $m_6$ and $m_7$). In conclusion, one can state that also for this scenario, the $\ell_\infty$ distance function seems to be the least suitable one as it does not really aim for a compromise (experiment).

### 7.7.5 Optimization of initial conditions and sampling times

The scenario in which both the sampling times and the initial conditions are optimized, is described in this section. As in the previous scenarios, the initial conditions were allowed to take values between 0 mM and 2 mM, ten samples were allowed to be taken and the minimum time between two samples was set to 15 seconds.

The D-optimal experiments and the compromise experiments found using the $\ell_1$, $\ell_2$ and $\ell_\infty$ distance functions are represented in Figs. 7.26, 7.27 and 7.28. Note that for several models the values of $\text{PEP}_0$ are zero, as well as those associated with the compromise experiments found using the $\ell_2$ and $\ell_\infty$ distance functions. These results indicate that, as more experimental degrees of freedom become available, the D-optimal experiments become more and more specific for the individual models. This is especially clear when considering the optimal sampling times shown in Fig. 7.26. This specificity obviously makes the design of a compromise experiment more challenging.

From the $\Gamma_{\text{D}_{ij}}$-values shown in Fig. 7.29, one can conclude that also for this scenario a compromise experiment is found when the $\ell_2$ distance function is used. However, the values of $\Gamma_{\text{D}_{ij}}$ obtained for the two other distance functions are systematically smaller than the ones obtained with the $\ell_2$ distance function. In addition, for several models, the medians are smaller than one. This means that for the majority of the D-optimal experiments more accurate parameter estimates can be obtained than if the designed compromise experiment would be performed. This indicates that the characteristics of the experiments designed using the $\ell_1$ and $\ell_\infty$ distance functions are not consistent with the ones of a compromise experiment.

To conclude this section, the reader is pointed to the fact that the specificity discussed above can also be observed from the $\Gamma_{\text{D}_{ij}}$-values shown in Fig. 7.29. These $\Gamma_{\text{D}_{ij}}$-values are generally larger than the ones obtained in the other scenarios (Figs. 7.23 and 7.25) and the variation among these values is much higher. This indicates that for some models, a significant amount of information on its parameters can be lost when performing an optimal experiment for another model instead of the compromise experiment. In other words, an optimal experiment for one model is often all but optimal for another one. Note that this specificity may also be the reason why the kernel-based method failed to provide compromise experiments for more complicated experimental design problems (as discussed in Section 7.2.5).

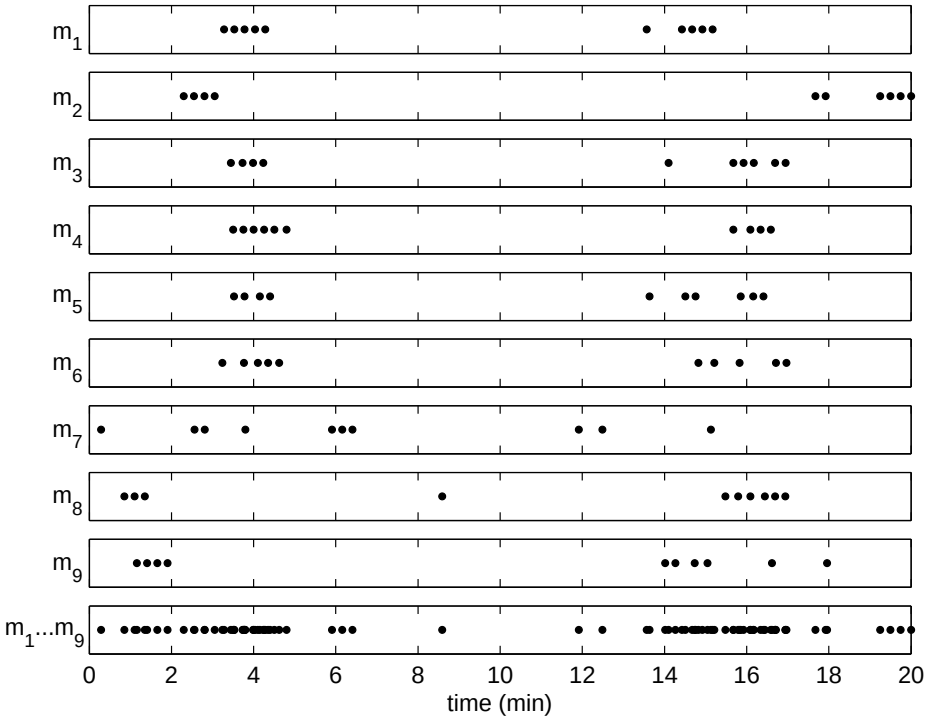



**Figure 7.26:** Optimal sampling times (•) found for the nine rival models for the case where the D-optimality design criterion is optimized by varying both the sampling times and the initial concentrations of glucose, ATP and PEP. The graph at the bottom is obtained by plotting the ten optimal sampling times of the nine individual models on the same axis.

### 7.7.6 Further discussion of the performance of the distance functions

From the results and discussion above, one can conclude that the $\ell_2$ distance function is the preferred one for this case study. However, the fact that the $\ell_1$ distance function performed well in the first two scenarios but was not the best option in the third scenario, indicates that the performance of a particular distance function is case specific. Indeed, the performance of the distance function depends on the shape of the Pareto front, which, unfortunately, is not known. And even it is was known, it cannot be clearly visualized for multi-objective problems with more than three objectives. Nevertheless, when the ideal point method is applied in another case study, the approach used here to evaluate the capability to design a compromise experiment can also be used to judge the experiments obtained.

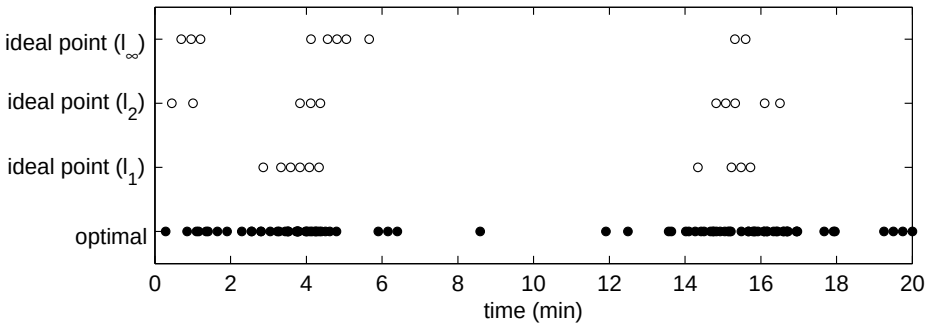



**Figure 7.27:** Optimal sampling times for the individual models (•) and the compromise sampling times (◦) found using the ideal point method using the different distance functions ($\ell_1$, $\ell_2$ and $\ell_\infty$).

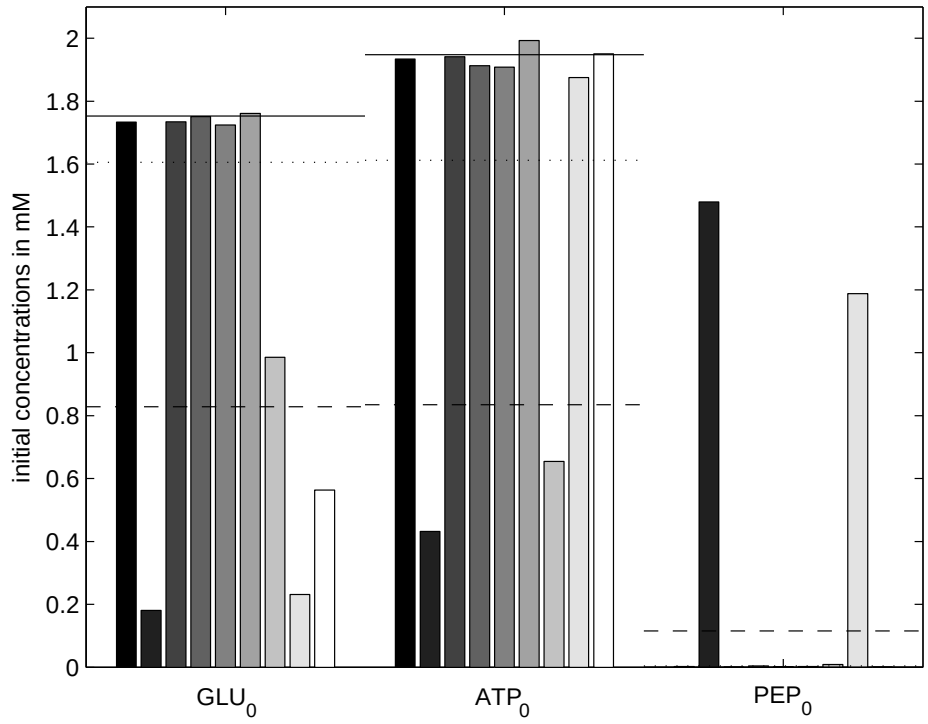



**Figure 7.28:** Initial concentrations of glucose, ATP and PEP associated with the D-optimal experiments of the rival models ($m_1$ corresponds to the black bar, and the bars become increasingly white as the model number increases), as well as those associated with the compromise experiment found using the ideal point method (horizontal lines) after using the $\ell_1$ (– –), the $\ell_2$ (—) and the $\ell_\infty$ ($\cdots$) distance function.

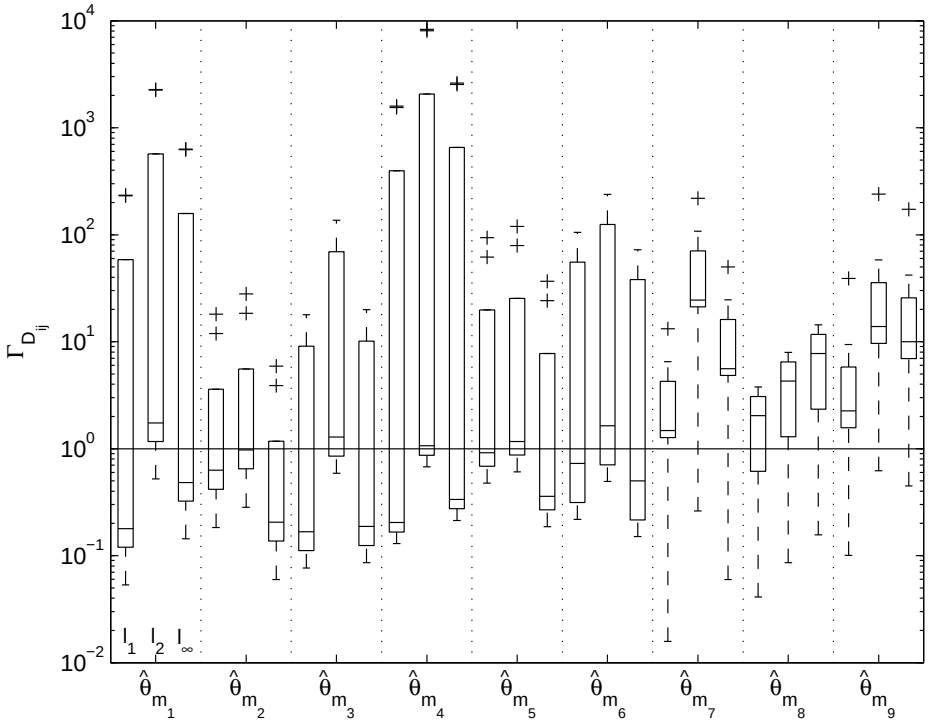



**Figure 7.29:** $\Gamma_{D_{ij}}$-values for the scenario in which both the initial conditions and the sampling times are optimized, and where the $\ell_1$, $\ell_2$ and $\ell_\infty$ distance functions are used.

## 7.8 Summary and conclusions

Model discrimination may become more efficient and effective if the uncertainty on the parameter estimates and consequently on the model predictions, is reduced first. This can be achieved by performing dedicated experiments, but performing such an additional experiment for each rival model may increase the experimental effort instead of minimizing it. Therefore, this chapter was dedicated to the design of a so-called compromise experiment, which is defined as an experiment that is sufficiently informative to improve the overall precision of the parameters of all rival models in a single experiment. For this purpose, two methods were proposed: the kernel-based method and the ideal point method.

The kernel-based method was developed to obtain the optimal sampling times to simultaneously improve the precision of the parameter estimates of several rival models. The method is inspired by kernel density estimation and uses the optimal sampling times for the individual models to design a compromise experiment. The fact that the contribution of the individual sampling times to the information content of the experiment varies, and

some sampling times are thus more important/informative than others, could be taken into account. To illustrate the kernel-based method, it was applied to our working example where nine rival models are defined to describe the kinetics of an enzyme-catalyzed reaction (glucokinase). The capability of the kernel-based method to design compromise experiments was evaluated, and the results of this evaluation clearly showed that the kernel-based method is capable to design compromise experiments.

Although the kernel-based method was originally developed to determine compromise sampling times, it would be desirable if this method could be extended for experimental design problem where more experimental degrees of freedom are available to the experimenter. To investigate this, the rationale of the kernel-based method was applied in an experimental design problem where the initial conditions were optimized. The results showed that the kernel-based method is less suited for the design of compromise experiments in such a case, and given the relative simplicity of the experimental design problem (only the initial conditions were optimized), one could conclude that the kernel-based method is not suited for more advanced experimental design exercises.

As an alternative to the kernel-based method, the ideal point method was developed to design an experiment to simultaneously improve the precision of the parameter estimates of several rival models. In the rationale of the ideal point method, the problem of designing a compromise experiment was approached as a multi-objective problem. This method searches for the experiment that is as close to possible to the optimal experiments of the individual rival models. Because closeness can be defined in several ways, the $\ell_1$, $\ell_2$ and $\ell_\infty$ distance functions were considered. The ideal point method can be applied to experimental design problems with experimental degrees of freedom of all types (manipulations, initial conditions and sampling times), which appeared to be difficult, if not impossible, with the kernel-based method. The ideal point method was applied to the same working example as with the kernel-based method. This was done for a scenario in which only the sampling times were optimized, one in which the initial conditions were optimized, and one in which both the initial conditions and the sampling times were optimized. The results showed that when more experimental degrees of freedom are available, the optimal experiments for the individual models become more and more specific, which makes the design of a compromise increasingly challenging. Nevertheless, the ideal point method proved to be capable of designing compromise experiments in each of the scenarios. The results also suggested that the use of the $\ell_2$ distance function is preferred over the use of the $\ell_1$ and $\ell_\infty$ distance functions.

# CHAPTER 8

## Integration of model discrimination and optimal experimental design for parameter estimation

## Abstract

To obtain a practically useful model, it is required that the model structure is adequate and that its parameters are estimated with a satisfactory level of precision. In this chapter, three approaches to integrate model discrimination and optimal experimental design for parameter estimation were investigated. In a first procedure, both aspects are dealt with sequentially, that is, the model discrimination procedure is performed first, and then the parameters of the selected model are further refined through the design of optimally informative experiments. The second procedure, is similar, except that a compromise experiment is designed and performed prior to the start of the model discrimination procedure to improve the quality of the parameter estimates. In the third procedure, both issues are dealt with simultaneously. For this purpose, the joint design criterion proposed by Hill et al. (1968) is modified such that the anticipatory approach can be used to quantify the discriminatory potential of the proposed experiments. Further, these three procedures are applied to a case study, the results of which showed that, although model discrimination was not achieved in less experiments compared to the sequential procedure without a compromise experiment, the quality of the parameter estimates improved faster when a compromise experiment was performed. The performance of the simultaneous procedure was worse compared to the performance of the sequential procedures, as a wrong model was selected more often than with the sequential procedures.

## 8.1 Introduction

To obtain a practically useful model, it is required that the model structure is adequate and that its parameters are estimated with a satisfactory level of precision. One can only exploit the potential of a mathematical model when it produces reliable predictions of the actual behavior of the modeled process. In the previous chapters (Part II), a procedure to identify the most appropriate model (structure) from a set of rival models was presented, in which the design of discriminatory experiments is of crucial importance. Indeed, when several models are able to describe the available experimental data, additional information has to be collected to further discriminate among these rival models. Once the most appropriate model is identified, experimental design methods are available that help the experimenter to design experiments that result in an increased precision of its parameter estimates.

The procedure described above corresponds to the classical, sequential procedure to build a mathematical model. Indeed, when building a mathematical model, experiments are typically designed and performed until the most appropriate model is identified, and then its parameters are refined. However, model discrimination and optimal experimental design for parameter estimation are closely related (as discussed in the previous chapters). Therefore, it is important to investigate and explore alternative approaches to integrate model discrimination and optimal experimental design for parameter estimation. In this chapter, some possibilities are proposed and applied to the working example used throughout this dissertation.

## 8.2 Sequential procedure to integrate model discrimination and optimal experimental design for parameter estimation

In essence, an experimental design procedure is said to be sequential if model discrimination and precise parameter estimation are treated separately. In this section, two possibilities to integrate optimal experimental design for parameter estimation (OED/PE) with the procedure to discriminate among rival models (MD) are described.

### 8.2.1 Classical sequential procedure

The classical sequential procedure for building a mathematical model basically consists of two steps (Asprey and Macchietto, 2000; Espie and Macchietto, 1989; Franceschini and

Macchietto, 2008; Vanrolleghem and Dochain, 1998) and is depicted in Fig. 8.1. First, the adequate model structure is identified, and then the quality of its parameters is (further) improved to enable reliable model predictions. This sequential procedure can thus be achieved by first applying the model discrimination procedure discussed in Chapters 2 and 4 to identify the most appropriate model. The optimal experimental design methods described in Chapter 6 can then be used to improve its parameter estimates until a desired level of precision is achieved.

Note that the model that is finally obtained has to be validated, that is, confronted with experimental data that has not been used during model building, nor during the estimation of its parameters. If the model, with its estimated parameters, is able to describe this new experimental data set, the model is said to be a valid model. Although model validation is only briefly discussed in Section 2.10, it should be noted that it is an important step in the model building procedure which should not be neglected in practical modeling applications.

### 8.2.2 Sequential procedure with a compromise experiment

As advocated in the previous chapters, the uncertainty associated with the parameter estimates is important with regard to the design of optimal discriminatory experiments. This is because the uncertainty on the parameter estimates will to a large extent determine the uncertainty on the model predictions. When the latter is high, it becomes difficult to design an experiment that enables the discrimination among the rival models. Following this rationale, the idea of designing a compromise experiment prior to the start of the model discrimination procedure was developed in Chapter 7. Performing this compromise experiment should improve the quality of the parameter estimates of the different models and thus decrease the uncertainty on their predictions, making the design of discriminatory experiments more effective. How such a compromise experiment can be obtained was explained in the previous chapter.

The procedure for building a model that arises from this rationale is depicted in Fig. 8.2. Although only one compromise experiment will be designed and performed in the examples discussed below, the procedure presented in this figure includes the possibility to design several compromise experiments before moving on to model discrimination (indicated by the "$D > ?$" decision node). Indeed, one could do this until a desired level of precision is reached, but a criterion to judge this is currently unavailable.

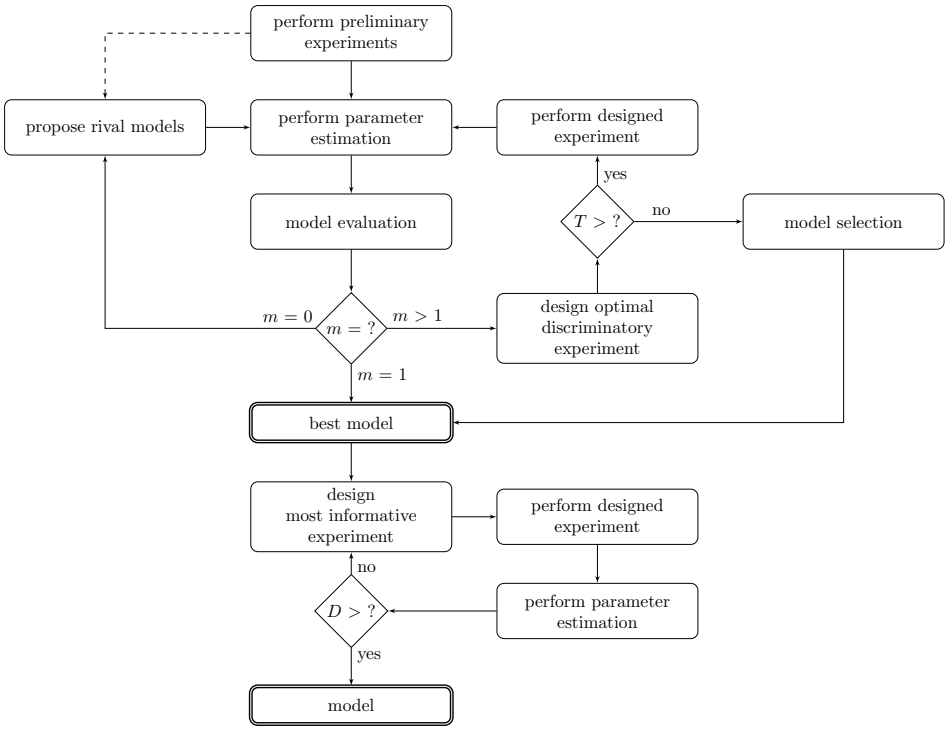



**Figure 8.1:** Classical, sequential procedure for building a mathematical model. First, the most appropriate model structure is identified through the design of (a) discriminatory experiment(s), and then the quality of its parameters is (further) improved through the design of experiments with a maximal information content with regard to the parameters.

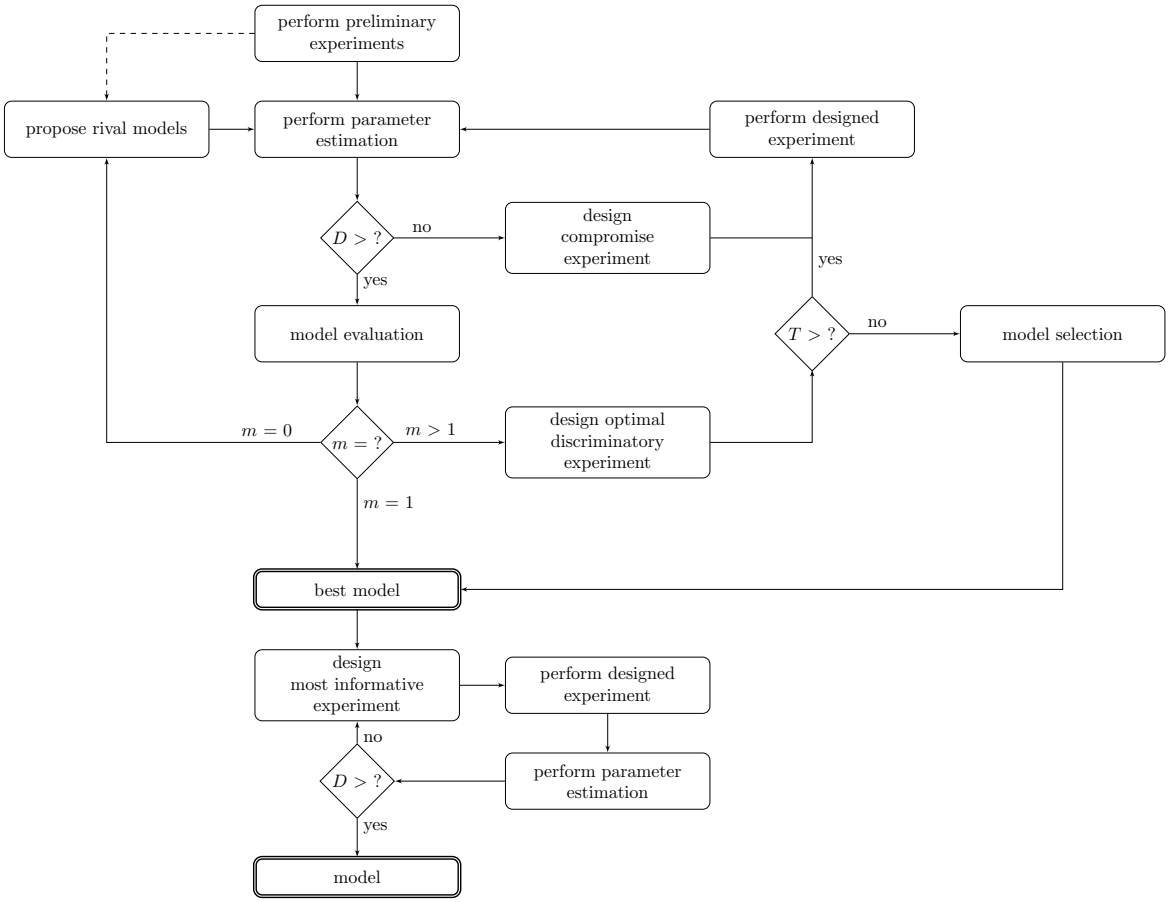



**Figure 8.2:** Sequential procedure for building a mathematical model. When the quality of the parameter estimates is insufficient, a compromise experiment is designed and performed before starting the model discrimination procedure. After the most appropriate model is identified, the quality of its parameters is (further) improved to enable reliable model predictions through the design of experiments with a maximal information content with regard to the parameters.

## 8.3 Simultaneous procedure to integrate model discrimination and optimal experimental design for parameter estimation

Instead of dealing with model discrimination and precise parameter estimation sequentially, one could also try to deal with both aspects simultaneously. A number of design criteria have been proposed that seek a balance between model discrimination and parameter estimation (for instance in Atkinson (2008), Hill et al. (1968) and Waterhouse et al. (2005)), but here, only the design criterion proposed by Hill et al. (1968) will be discussed. For the other design criteria the reader is referred to the cited papers.

Hill et al. (1968) stated that an experimental design procedure that deals with both aspects should emphasize model discrimination when there is substantial uncertainty as to which model is the most appropriate one and should emphasize parameter estimation when one of the rival models seems to be overwhelmingly superior to the others. The authors suggest that the same result (that is, an adequate model structure with precise estimates of its parameters) might be obtained in fewer experiments compared to the classical sequential procedure discussed in Section 8.2.1.

In addition, Hill et al. (1968) stated that it is more likely that the selected model is in fact the correct model if a joint design criterion is used that takes both objectives into account from the onset of the model discrimination exercise. A joint design criterion can be developed such that the experimenter is not likely to prematurely choose a wrong model and then concentrate on parameter estimation for this model. By taking both objectives into account from the beginning of a model discrimination exercise, one essentially starts by performing optimal discriminatory experiments. However, as the investigation proceeds, those experimental degrees of freedom are selected that provide for both discrimination among the important rival models at that stage and, hopefully, for precise estimates of the parameters of the model that will ultimately be selected as the correct one.

As stated by Hill et al. (1968), a logical choice for a design criterion of this kind, although by no means unique, is given by

$$\boldsymbol{\xi}^{\star} = \max_{\boldsymbol{\xi} \in \boldsymbol{\Xi}} \; \left( w_1 \cdot \Upsilon \left( \boldsymbol{\xi} \right) + w_2 \cdot \Delta \left( \boldsymbol{\xi} \right) \right) , \tag{8.1}$$

where $\Upsilon \left( \boldsymbol{\xi} \right)$ represents a measure of the discriminatory potential of experiment $\boldsymbol{\xi}$, $\Delta \left( \boldsymbol{\xi} \right)$ represents a measure of the parameter estimation precision, and $w_1$ and $w_2$ are non-negative

weights whose values evolve as experimentation progresses to reflect the relative emphasis that is given to each of the objectives. Below, the original formulation provided by Hill et al. (1968) is described, as well as a modified version that can be used when one prefers to use the design criterion of Buzzi-Ferraris et al. (1984).

### 8.3.1 Original formulation of the joint design criterion proposed by Hill et al. (1968)

In their original formulation of the joint design criterion to tackle the problems of model discrimination and precise parameter estimation, Hill et al. (1968) built further on the design criterion developed by Box and Hill (1967) to discriminate among rival models (discussed in Section 4.4.5), and $\Upsilon(\boldsymbol{\xi})$ is calculated as follows

$$\Upsilon(\boldsymbol{\xi}) = \frac{T(\boldsymbol{\xi})}{T(\boldsymbol{\xi}_T^\star)}, \tag{8.2}$$

where $T(\boldsymbol{\xi})$ is calculated from Eq. (4.19) and $\boldsymbol{\xi}_T^\star$ represents the optimal discriminatory experiment found after optimizing Eq. (4.19). To quantify the precision of the parameter estimates, the D-optimality design criterion is used, and $\Delta(\boldsymbol{\xi})$ is calculated as

$$\Delta(\boldsymbol{\xi}) = \sum_{i=1}^{m} \pi_i \cdot \frac{D_i(\boldsymbol{\xi})}{D_i(\boldsymbol{\xi}_D^\star)}. \tag{8.3}$$

Here, $\pi_i$ represents the relative probability that model $m_i$ is the true model (calculated from Eq. (4.16)), and $\boldsymbol{\xi}_D^\star$ represents the experiment that was found to be optimal to estimate the parameters of model $m_i$.

The weights, $w_1$ and $w_2$ in Eq. (8.1), reflect the relative emphasis that is given to model discrimination on the one hand, and precise parameter estimation on the other hand. These weights are calculated as follows

$$w_1 = \left( m \cdot \frac{1 - \max_{i \in \{1,\ldots,m\}} \pi_i}{m-1} \right)^{\lambda}, \tag{8.4}$$

$$w_2 = 1 - w_1. \tag{8.5}$$

From Eq. (8.5), one can see that when the model probabilities are all equal to $1/m$, $w_2$ equals zero and the experimental design is thus concentrated on the discrimination among

the $m$ rival models. On the other hand, when the probability of one of the models approaches one, $w_2$ will approach one as well and the focus is moved to the precise estimation of the parameters of the most probable model. The parameter $\lambda$ (with $0 < \lambda < +\infty$) is used to control the rate at which this transition of model discrimination to precise parameter estimation occurs. As illustrated in Fig. 8.3, a large $\lambda$-value results in a faster decrease of $w_1$ and thus in a faster transition to precise parameter estimation.

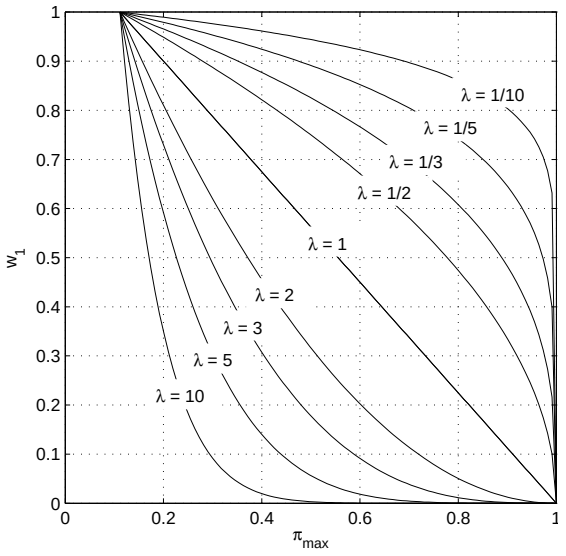



**Figure 8.3:** The weight $w_1$ as a function of the probability of the most probable rival model (indicated as $\pi_{max}$ and calculated as $\max_{i\in\{1,\ldots,m\}} \pi_i$) for different values of $\lambda$ and for $m = 10$. This weight indicates the relative emphasis that is given to model discrimination. The weight given to precise parameter estimation ($w_2$) is obtained as $w_2 = 1 - w_1$.

In this respect, it is important to note that the rationale behind the calculation of $\Delta(\boldsymbol{\xi})$ from Eq. (8.3) is similar to the rationale of the compromise experiment. Indeed, since the D-optimality criterion values of all rival models are considered in its calculation, it is expected that the resulting experiment, at least to some extent, has the characteristics of a compromise experiment. Also the fact that the $D_i(\boldsymbol{\xi})$-values are compared with the D-optimality design criterion value of the most informative experiment, denoted as $\boldsymbol{\xi}_D^\star$, resembles the concept of the ideal point method.

Nevertheless, both procedures are significantly different. The most striking difference between the simultaneous procedure and the sequential procedure with (a) compromise ex-

periment(s) is of course the fact that the latter focuses on the parameter estimates first, while in the former the emphasis is shifted to the parameter estimates as model discrimination procedure progresses. In addition, the relative model probabilities are considered in the calculation of $\Delta(\boldsymbol{\xi})$, while all rival models are treated equally in the approach proposed in Chapter 7 for the design of compromise experiments. Note, however, that the latter can easily be incorporated in the ideal point as proposed in Chapter 7.

### 8.3.2 Modified joint design criterion of Hill et al. (1968)

Contrary to the original formulation of the joint design criterion discussed above, the design criterion proposed by Buzzi-Ferraris et al. (1984) is used in this dissertation to quantify the potential of an experiment to discriminate among a number of rival models. Below, one way of modifying the joint design criterion formalized in Eq. (8.1) is proposed, such that the design criterion of Buzzi-Ferraris et al. (1984) can be used instead of the one of Box and Hill (1967).

In the original formulation of the joint design criterion, the discriminatory potential of the proposed experiment, denoted as $T(\boldsymbol{\xi})$, is based on the sum of the discriminatory potentials over the different model pairs (Eq. (4.19)). Therefore, it is proposed to calculate $T(\boldsymbol{\xi})$ in Eq. (8.2) as follows

$$T(\boldsymbol{\xi}) = \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} T_{ij}(\boldsymbol{\xi}) \tag{8.6}$$

and the optimal discriminatory experiment is thus given by

$$\boldsymbol{\xi}^{\star} = \arg\max_{\boldsymbol{\xi} \in \boldsymbol{\Xi}} \quad \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} T_{ij}(\boldsymbol{\xi}) \ . \tag{8.7}$$

To calculate $\Delta(\boldsymbol{\xi})$, the D-optimality design criterion is used (as in the original formulation), and the relative model probabilities are calculated as proposed by Schwaab et al. (2006). The latter was discussed in Section 4.4.7, and is repeated here for clarity. For each model $m_i$, the model probability, denoted as $\rho_i$, is defined by exploiting the characteristic of the WSSE to be a sample from a $\chi^2$-distribution with $n - n_p$ degrees of freedom, where $n$ represents the total number of data points from which the $n_p$ model parameters were estimated. The probability associated with model $m_i$ is calculated as

$$\rho_i = 1 - \mathrm{P}\left[\chi^2_{n-n_p} \leqslant \mathrm{WSSE}\right] , \tag{8.8}$$

from which the relative model probability, denoted as $\pi_i$, is calculated as

$$\pi_i = \frac{\rho_i}{\sum_{k=1}^{m} \rho_k} . \tag{8.9}$$

The basic idea behind the use of Eq. (8.8) is that *bad* models are likely to exhibit large values of WSSE, and consequently a low value of $\rho_i$ (and $\pi_i$).

## 8.4 Case study VII: Comparison of the sequential procedures to integrate model discrimination and OED/PE

This section describes the results of a case study in which the sequential procedures to integrate model discrimination and OED/PE (Section 8.2) are applied to the working example used throughout this dissertation. The procedure in which model discrimination and optimal experimental design for parameter estimation are considered simultaneously (Section 8.3.2) will be investigated in a separate case study (described in Section 8.5) because it is difficult to compare its performance with that of the sequential procedures (as will be discussed in more detail in Section 8.5).

### 8.4.1 Objective of this case study

The main objective of this case study is to investigate the performance of the sequential procedures presented in Section 8.2 to integrate model discrimination and optimal experimental design for parameter estimation. In addition, it will be investigated whether a difference in the performance can be observed between the procedure with and without a compromise experiment (respectively described in Sections 8.2.2 and 8.2.1) or, in other words, whether it is beneficial or useful to design and perform a compromise experiment before starting the model discrimination procedure or not.

### 8.4.2 Design of the case study

As the objective of this case study is to investigate the usefulness of performing a compromise experiment prior to the model discrimination procedure, the performance of the

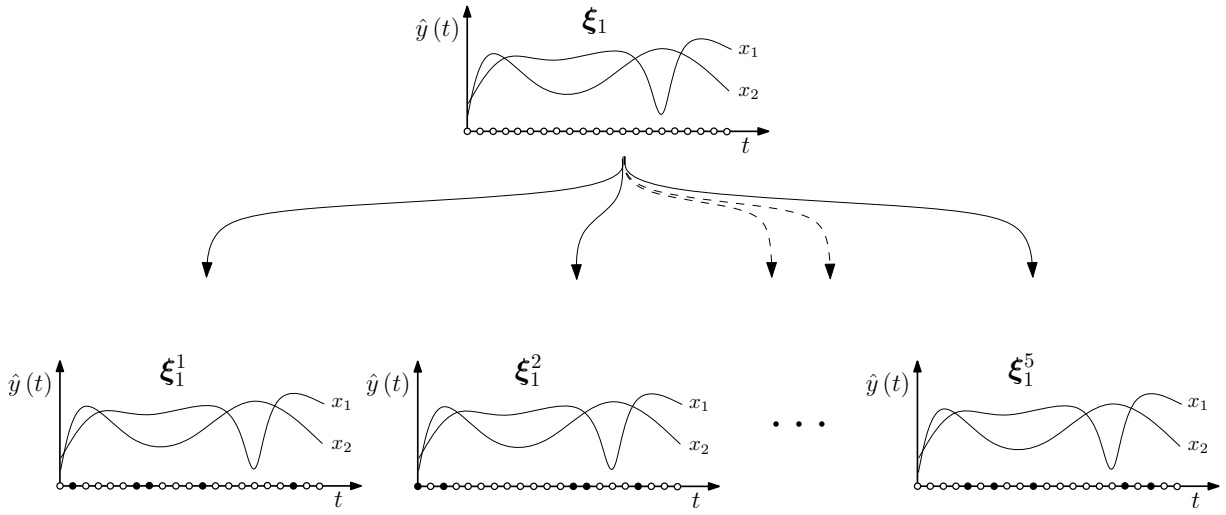



**Figure 8.4:** Illustration of the design of the case study (Section 8.4), where the aim was to evaluate and compare the performance to discriminate among rival models.

sequential procedure with and without compromise experiment has to be compared. Note that the classical sequential procedure to integrate OED/PE and MD described in Section 8.2.1 was already applied in a previous case study (the one described in Section 4.11). The results discussed in the mentioned case study will be confronted and compared with the ones obtained here, where the sequential procedure with compromise experiment is applied to the same case study.

Obviously, the case study presented in this section was designed in an identical manner as the case study of Section 4.11 (depicted in Fig. 8.4). Also in this case study, five scenarios are considered, in each of which the model discrimination procedure is performed starting from a different preliminary experiment. Each of these scenarios is then repeated thirty times to account for the effect of the measurement error on the performance of the experimental design.

### 8.4.3 Preliminary experiments

To allow a mutual comparison of the results obtained from both case studies, the preliminary experiments used to initiate the model discrimination procedure in the different scenarios were identical to the ones used in the case study of Section 4.11. For brevity, the characteristics of these preliminary experiments are not discussed here and the reader is referred to Section 4.11.3. Obviously, in the case study described here, the first experiment

that is performed after the preliminary experiment is the compromise experiment, while an optimal discriminatory experiment was performed in the case study described earlier.

### 8.4.4 Design of the compromise experiments

For each of the five scenarios, the compromise experiment was designed based on the information available after performing the preliminary experiment. As the information (for instance, the parameter estimates and the Fisher information matrix of the already performed (preliminary) experiment) is the same for each repetition of a particular scenario, the design of the compromise experiments will also be identical.

To design the compromise experiments, the experimental degrees of freedom were the same as in the preliminary experiment except for the sampling times and the initial concentrations of glucose, ATP and PEP, which were optimized. The initial concentrations were allowed to take values between 0 mM and 2 mM, and ten optimal sampling times were determined with the constraint that the time interval between two subsequent samples was at least 15 seconds. These experimental degrees of freedom are identical to the ones used in the case studies described in Section 4.11. As the results obtained in Chapter 7 indicated that the compromise experiment is preferably designed using the ideal point method and the $\ell_2$ distance function, this approach was also followed here.

Note that in this case study, only one compromise experiment is performed prior to the model discrimination procedure. In principle, several compromise experiments could be performed until the quality of the parameter estimates is sufficiently good, as depicted in the sequential procedure shown in Fig. 8.2. The reason why only one compromise experiment is performed here, is that no real criterion has been proposed yet to evaluate whether the overall quality of the parameter estimates of the rival models is too low (indicated as "$D > ?$" in Fig. 8.2).

### 8.4.5 Design of the optimal discriminatory experiments

For the design of the optimal discriminatory experiments, the experimental degrees of freedom were the same as for the design of the compromise experiments (discussed in the previous section). Since the anticipatory approach showed the best performance in the case studies discussed in Chapter 4, this approach was used to design the discriminatory experiments. In addition, the pairwise design strategy was adopted (as in the case study

of Section 4.11), in which an optimal discriminatory experiment is designed for each model pair, and that experiment is chosen for which the corresponding $T_{ij}$-value is largest.

### 8.4.6 Design of optimal experiments for parameter estimation

Also for the design of optimally informative experiments, the experimental degrees of freedom were the same as for the design of the compromise experiment. The experiments were performed as explained in Chapter 6 and the D-optimality design criterion was used to quantify the information content with regard to the model parameters.

### 8.4.7 Outcome of the model discrimination procedure

Ideally, the model discrimination procedure ends when one of the rival models is identified as the most appropriate one. In this case study, the experimental data were generated using model $m_5^\star$ and it can thus be expected that model $m_5$ is identified as the most appropriate model in the majority of the runs, regardless of the approach used to design the discriminatory experiments. However, the possibility that another model is identified as the most appropriate one cannot be excluded.

A second possible outcome of the model discrimination procedure is that all models appear to be inadequate. Indeed, the adequacy of the rival models is evaluated based on the WSSE value, and even for the true model ($m_5^\star$) this WSSE value can in some situations be larger than the reference value ($\chi^2_{n-n_p}$) because of the (simulated) error on the measurements.

In Table 8.1, the outcomes of the model discrimination procedure obtained with and without using a compromise experiment are shown for the five scenarios initiated with a different preliminary experiment. As one can see, no significant difference can be observed between the two sequential procedures and one can thus conclude that the use of one compromise experiment before starting the model discrimination procedure does not affect the outcome of the model discrimination procedure.

### 8.4.8 Required number of experiments to achieve model discrimination

The number of additional experiments that have to be performed before the most appropriate model can be identified is an important aspect that should be taken into account when evaluating the performance of a certain model discrimination procedure. Indeed, experiments can be time and money consuming, and limiting the experimental effort is the

**Table 8.1:** Overview of the observed outcomes of the 150 runs of the sequential procedures with and without performing a compromise experiment ($\boldsymbol{\xi}_c^\star$) prior to the start of the model discrimination procedure, and this for the five scenarios initiated with a different preliminary experiment, $\boldsymbol{\xi}_1^i$ with $i = 1, \ldots, 5$.

| | | $-$ | $\boldsymbol{\xi}_c^\star$ |
|---|---|---|---|
| model $m_5$ | $\boldsymbol{\xi}_1^1$ | 30 | 24 |
| | $\boldsymbol{\xi}_1^2$ | 29 | 29 |
| | $\boldsymbol{\xi}_1^3$ | 25 | 25 |
| | $\boldsymbol{\xi}_1^4$ | 27 | 27 |
| | $\boldsymbol{\xi}_1^5$ | 26 | 26 |
| | $\boldsymbol{\xi}_1^1 - \boldsymbol{\xi}_1^5$ | **134** *89%* | **131** *87%* |
| other model | $\boldsymbol{\xi}_1^1$ | 0 | 0 |
| | $\boldsymbol{\xi}_1^2$ | 1 | 0 |
| | $\boldsymbol{\xi}_1^3$ | 4 | 3 |
| | $\boldsymbol{\xi}_1^4$ | 3 | 2 |
| | $\boldsymbol{\xi}_1^5$ | 2 | 4 |
| | $\boldsymbol{\xi}_1^1 - \boldsymbol{\xi}_1^5$ | **10** *7%* | **9** *6%* |
| all models rejected | $\boldsymbol{\xi}_1^1$ | 3 | 6 |
| | $\boldsymbol{\xi}_1^2$ | 0 | 1 |
| | $\boldsymbol{\xi}_1^3$ | 1 | 2 |
| | $\boldsymbol{\xi}_1^4$ | 0 | 1 |
| | $\boldsymbol{\xi}_1^5$ | 2 | 0 |
| | $\boldsymbol{\xi}_1^1 - \boldsymbol{\xi}_1^5$ | **6** *4%* | **10** *7%* |

ultimate aim of each experimental design exercise. In this section, the effect of performing a compromise experiment prior to the start of the model discrimination procedure is investigated by looking at the number of experiments that have to be designed and performed until the model discrimination procedure ends.

The number of experiments required to achieve model discrimination is presented in Fig. 8.5 for the sequential procedures with and without performing a compromise experiment ($\boldsymbol{\xi}_c^\star$) prior to the model discrimination procedure. This figure contains five subfigures with a white background (entitled $\boldsymbol{\xi}_1^i$, with $i = 1, \ldots, 5$) and one subfigure with a gray background (entitled $\boldsymbol{\xi}_1^1 \; - \; \boldsymbol{\xi}_1^5$). The former present the results obtained for the simulations where the model discrimination procedure was initiated with the preliminary experiment indicated in the title of the corresponding subfigure, whereas the one with the gray background gives an overall idea of the number of required experiments and presents the values of all model discrimination runs ($30 \times 5 = 150$ in total). Note that in these figures, the preliminary experiment corresponds to experiment number one and, if applicable, the (single) compromise experiment corresponds to experiment number two.

From these results, one can conclude that the number of experiments that is required to achieve model discrimination does not significantly change when a compromise experiment is performed, and also the variability of the number of experiments needed according to the two procedures is very similar. Indeed, when the results of all scenarios are considered, the median of the number of required experiments is the same for both sequential procedures. In a way, one could even state that one additional experiment is required to achieve model discrimination when no compromise experiment is performed, although this statement is not entirely true as model discrimination can also occur after performing the compromise experiment. Note that, for the scenario with preliminary experiment $\boldsymbol{\xi}_1^4$, model discrimination was achieved in three experiments instead of four without the use of a compromise experiment, while for the scenario with preliminary experiment $\boldsymbol{\xi}_1^2$ (at least) one additional experiment was required. These observations could however not be explained or linked with, for instance, the information content of the mentioned preliminary experiments. As an overall conclusion, one can state that, on the basis of the results obtained in this case study, an equal number of experiments is required when a compromise experiment is performed before starting the model discrimination procedure.

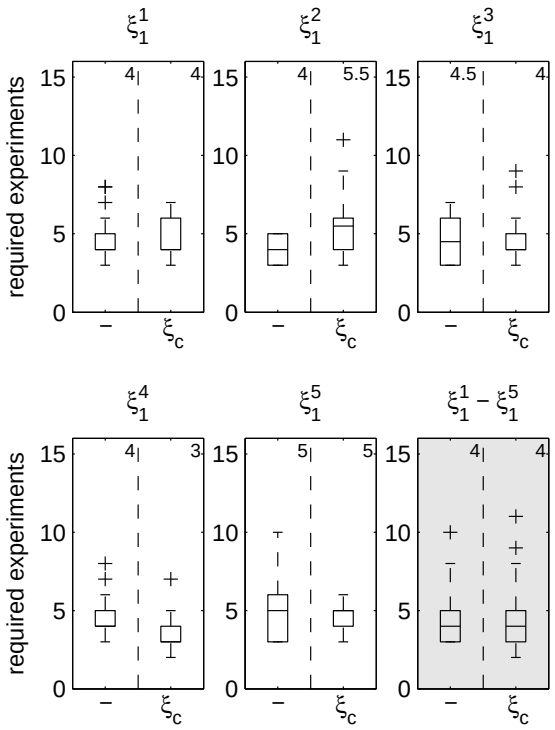


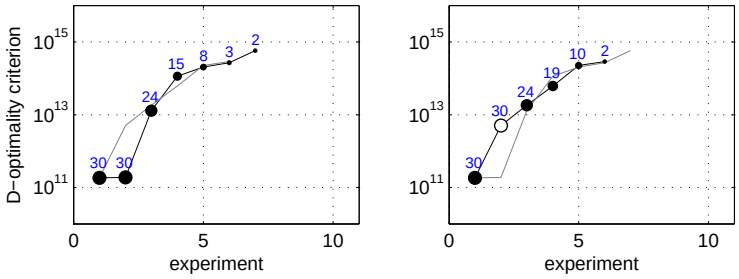

**Figure 8.5:** Boxplots showing the number of experiments (median in upper right corner) required to achieve model discrimination for the scenario where the sequential procedure is applied with (right) and without (left) the use of a compromise experiment ($\boldsymbol{\xi}_c$), and this starting from each of the five preliminary experiments ($\boldsymbol{\xi}_1^1$ till $\boldsymbol{\xi}_1^5$) (white background). The boxplots with the gray background give an overall idea of the number of required experiments and were made using the results obtained from all starting situations.

### 8.4.9 Evaluation of the quality of the parameter estimates during model discrimination

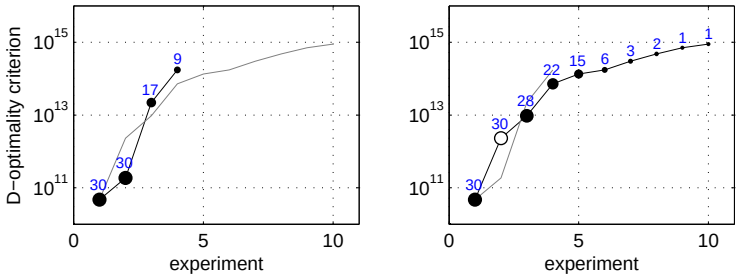
In this section, the quality of the parameter estimates is investigated, and especially the quality of the parameter estimates of the model that is eventually identified as the most appropriate one (see overall procedure of Fig. 8.1 and 8.2). Obviously, this is important because once an appropriate model is identified through model discrimination, the quality of its parameter estimates typically has to be improved before the model can actually be applied for its intended use (as illustrated and discussed in the following section).

To quantify the quality of the parameter estimates of model $m_5$, the D-optimality criterion values were calculated and their evolution as a function of the number of performed experiments is presented in Fig. 8.6(a-c) for the scenarios initiated with preliminary experiments $\boldsymbol{\xi}_1^1$, $\boldsymbol{\xi}_1^2$ and $\boldsymbol{\xi}_1^3$. The figures for the other scenarios are not shown here for brevity, but are shown in Appendix C. From these results, one can clearly see that the quality of the parameter estimates increases faster when a compromise experiment is performed. In a way, this was expected since the compromise experiment was especially designed for this purpose. However, one can also see that as the experimentation progresses, the quality of the parameter estimates obtained for the cases with and without the compromise experiment converges for the scenarios with preliminary experiment $\boldsymbol{\xi}_1^1$ (and $\boldsymbol{\xi}_1^4$, the results of which are shown in Fig. C.1). For the scenario with preliminary experiment $\boldsymbol{\xi}_1^2$ (shown in Fig. 8.6(b)) and with a compromise experiment, the D-optimality criterion values obtained after four experiments were even slightly lower than the ones obtained after an equal number of experiments for the scenario without a compromise experiment. As the model discrimination is not achieved after four experiments when a compromise was used, the procedure continues and consequently also the corresponding D-optimality criterion values further increase.

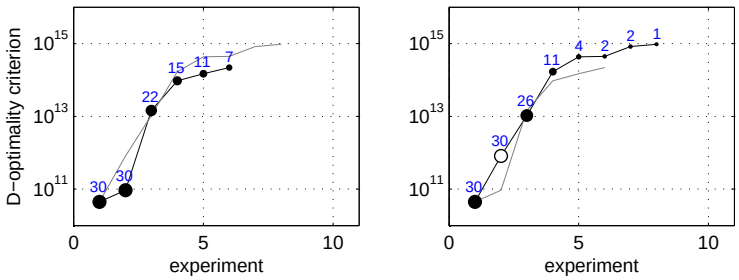
The observation that the quality of the parameter estimates increases faster in the beginning of the procedure should, however, not be underestimated. Indeed, as already discussed in the previous section, performing experiments may be time and money consuming. Because of this, the number of experiments that can be designed given the available resources (that is, time and/or money), may be limited and it may not be possible to perform experiments until complete model discrimination is achieved. In such a situation, it is obviously beneficial that the quality of the parameter estimates increases quickly as the model discrimination procedure progresses.




(a) starting from preliminary experiment $\boldsymbol{\xi}_1^1$




(b) starting from preliminary experiment $\boldsymbol{\xi}_1^2$



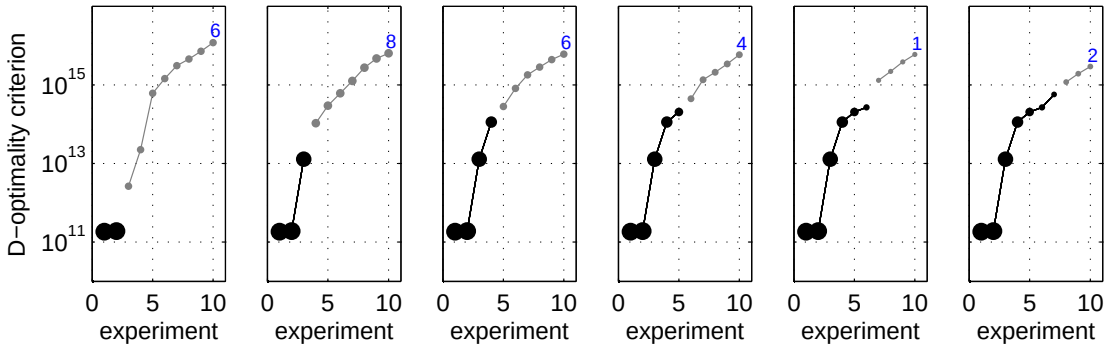

(c) starting from preliminary experiment $\boldsymbol{\xi}_1^3$

**Figure 8.6:** Evolution of the median of the D-optimality criterion values of model $m_5$ for the (thirty) runs of the model discrimination procedure initiated with (a) experiment $\boldsymbol{\xi}_1^1$, (b) experiment $\boldsymbol{\xi}_1^2$ and (c) experiment $\boldsymbol{\xi}_1^3$. The results obtained without the use of a compromise experiment are shown in the left subfigure and the results obtained with a compromise experiment are shown in the right subfigure. The evolution of the median criterion values of the other strategy is shown in gray to ease the comparison. The number of criterion values from which the median was determined is mentioned and also indicated by the size of the bullet symbol.

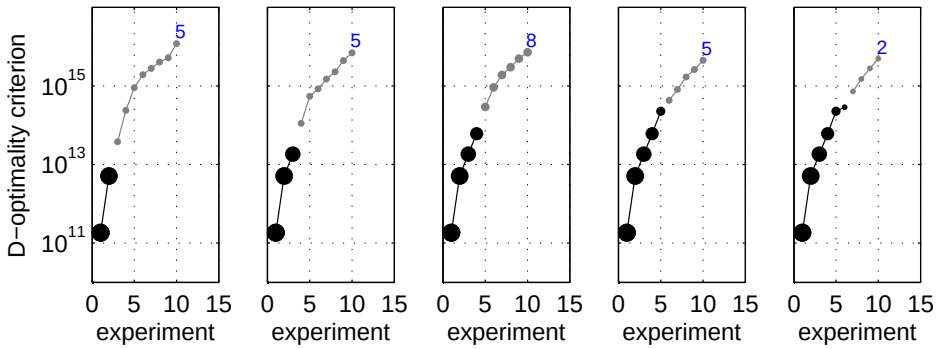
### 8.4.10 Further improvement of the parameter estimates of the most appropriate model

As discussed in Section 8.2, once the most appropriate model is identified, its parameters are typically further improved until a desired level of precision is reached, through the design of optimally informative experiments. For illustratory purposes, this is also done in this case study, where optimally informative experiments are designed and performed to improve the precision of the parameters of model $m_5$ (the *true* model) if it was selected. Below, the evolution of the D-optimality criterion values for model $m_5$ will be presented for the sequential procedures with and without the use of a compromise experiment. This application of optimal experimental design for parameter estimation is stopped after ten experiments (in total). Since the model discrimination procedure stops after a varying number of optimal discriminatory experiments, the number of optimally informative experiment that are designed and performed will also differ.

The results obtained for the scenario with preliminary experiment $\boldsymbol{\xi}_1^1$ are shown in Fig. 8.7 for the procedures (a) with and (b) without the use of a compromise experiment. These figures have to be interpreted as follows. The black lines and dots in the subfigures show the evolution of the median of the D-optimality criterion values as the number of performed optimal discriminatory experiments increases (model discrimination part of the sequential procedure). In the first subfigure of Fig. 8.7(a), for instance, one can see that model discrimination is achieved after performing the first designed discriminatory experiment (the second experiment that is performed) in six of the thirty repetitions or runs of the procedure. When model $m_5$ was identified as the most appropriate model, which was most often the case (as discussed in Section 4.11.6), its parameters were further refined by designing and performing a series of optimally informative experiments until, in total, ten experiments were performed. So, for this specific example, eight optimally informative experiments were designed and performed. The evolution of the median of the D-optimality criterion values obtained after performing these experiments (designed to obtain more precise parameter estimates) is represented by the gray line and dots. The size of these dots (and also of the black ones) indicates the number of criterion values from which the median was determined and corresponds to the number of repetitions where model $m_5$ was identified as the most appropriate model after one discriminatory experiment and is indicated by the number in the upper right corner of the corresponding subfigure.




(a) without compromise experiment




(b) with compromise experiment

**Figure 8.7:** Evolution of the median of the D-optimality criterion values of model $m_5$ for the (thirty) runs of the model discrimination procedure that was initiated with experiment $\boldsymbol{\xi}_1^1$ for the procedures with (a) and without the use of a compromise experiment (b). The black lines indicate the D-optimality criterion values obtained in the model discrimination procedure, whereas the gray lines shows how the D-optimality criterion values evolves after model $m_5$ has been identified as the most appropriate model.

In the second subfigure, the same information is presented, but for the repetitions of the model discrimination procedure where model $m_5$ was identified as the most appropriate model after the second discriminatory experiment. The interpretation of the other subfigures and the subfigures from Fig. 8.7(b) have to be interpreted in a similar way. The results obtained from the scenarios with the other preliminary experiments are shown in Appendix C.

From the results shown in Fig. 8.7, one can see that regardless of the number of discriminatory experiments that were needed to identify model $m_5$ as the most appropriate model, the D-optimality design criterion values converge to more or less the same value after performing ten experiments. This suggests that the experiments designed with this procedure are informative with regard to the parameters of model $m_5$, which illustrates the similarity between OED/PE and the anticipatory approach for optimal experimental design for model discrimination (as discussed in Chapter 4). Indeed, if the optimal discriminatory experiment were not informative with regards to the parameters (of model $m_5$), the D-optimality design criterion values obtained after performing ten experiments would be significantly higher for those runs in which model $m_5$ was identified in fewer discriminatory experiments, because in those runs more experiments would be performed that were especially designed to improve the precision of its parameters.

## 8.5 Case study VIII: Application of the simultaneous procedure to integrate model discrimination and OED/PE

In this case study, the simultaneous procedure to integrate model discrimination and optimal experimental design for parameter estimation presented in Section 8.3 is applied to our working example. To be able to compare (at least to some extent) the obtained results with those obtained with the sequential procedures, the modified version of the simultaneous procedure will be applied. Indeed, the original version uses the design criteria proposed by Box and Hill (1967) to design optimal discriminatory experiments, whereas the anticipatory approach (which is based on the design criterion proposed by Buzzi-Ferraris et al. (1984)) was used in the sequential procedures and in the modified joint design criterion.

### 8.5.1 Objective of this case study

The objective of this case study is to investigate the performance of the simultaneous procedure to identify the most appropriate model from a set of rival models and estimate

its parameters. In the original paper of Hill et al. (1968), it was argued that less experiments may be required to identify the most appropriate model with precise estimates of its parameters when applying the simultaneous procedure instead (of one) of the sequential procedure(s). Where possible, the results will be compared to those obtained with the sequential procedures with and without the use of a compromise experiment.

### 8.5.2 Design of the case study

Also for this case study, the procedure was performed for five scenarios (each with a different preliminary experiment) and each scenario was repeated thirty times. As the design is identical to the one used in previous case studies, the reader is referred to Section 4.11.2 for more details hereon.

### 8.5.3 Preliminary experiments and experimental design

The preliminary experiments used in this case study are identical to the ones used in the previous case studies and the reader is referred to Section 4.11.3 for a more detailed description of these experiments. Also for the design of the optimally informative and optimal discriminatory experiments, the experimental degrees of freedom were defined in the same way as the ones used in the previous case studies (see, for instance, Section 4.11.5). The experiments are designed as explained in Section 8.3.2, with $\lambda = 1$. The relative model probabilities at the start of the procedure were calculated from the WSSE value obtained after estimating their parameters from the data of the corresponding preliminary experiment (according to Eq. (8.8)).

### 8.5.4 Model evaluation

In the sequential procedure, the models are evaluated using a so-called model adequacy test. As explained in Section 2.8.1, the $\chi^2$-test can be used in this context since the measurement errors are known (and calculated from Eq. (3.24)). The models that are rejected by this test are considered to be inadequate and are not used for the design of discriminatory experiments. When the (modified version of the) simultaneous procedure is used, on the other hand, such a test is not performed and the adequacy of the rival model is evaluated based on their relative model probabilities. However, in principle, a bad model is never rejected as such, but a very low relative model probability will be associated with it. As a result, all models are considered in the design of the discriminatory experiments,

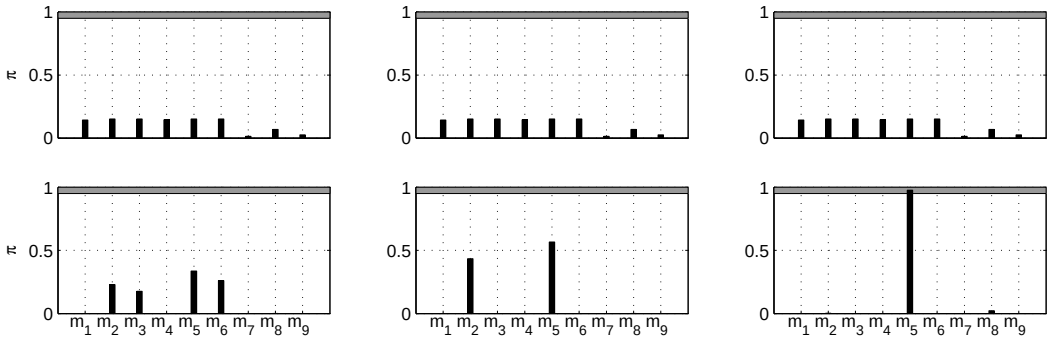



**Figure 8.8:** Relative model probabilities obtained for three (of the thirty) runs of the simultaneous procedure after performing preliminary experiment $\boldsymbol{\xi}_1^1$ (upper graph) and after performing the first designed experiment (lower graph).

together with their corresponding relative model probabilities, which will determine their relative contribution to the experimental design.

In this case study, the simultaneous procedure was stopped when the relative model probability of one of the models was larger than a predefined and arbitrarily chosen value of 0.95. Note, however, that Schwaab et al. (2006), who introduced the approach to calculate the relative probabilities (Section 4.4.8), considered a model inadequate when its relative model probability is below a certain value (for instance, 0.025) and the inadequate model is discarded from the experimental design. Although this is an interesting approach, it was not adopted in the following. In other words, all models were considered in the experimental design procedure until the end.

For illustratory purposes, the relative model probabilities obtained after performing preliminary experiment $\boldsymbol{\xi}_1^1$ and after performing the first designed experiment are shown in Fig. 8.8 for three (of the thirty) runs of the simultaneous procedure. As one can see, the relative model probabilities of most models have similar values after performing the preliminary experiment. This clearly indicates that at least six of the nine models are able to describe the data obtained from the preliminary experiment. After performing the first designed experiment (which is identical for the different repetitions or runs of the procedure, but not the data obtained from it) the model probabilities change considerably. In the first example (left), one can observe that four models still have a high relative model probability, whereas this is the case for only two models in the second example (middle). In the third example (right), it even appeared to be possible to identify the (correct) most appropriate model.

### 8.5.5 Outcome of the simultaneous procedure

It is clear that also with this procedure the ultimate aim is to identify the most appropriate model from the set of rival models. As explained in the previous section, the simultaneous procedure is stopped when the relative model probability of one of the models is larger than 0.95. To investigate the performance of the simultaneous procedure, the procedure was performed for the five different scenarios, and here too the procedure was repeated thirty times (as explained in Section 8.5.2). The results obtained in this way, are shown in Table 8.2.

From the result in Table 8.2, one can clearly see that the model that is identified as the most appropriate model is often different from model $m_5$ (more precisely, in 26% of the repetitions). It is clear that this is not desired and is in contrast with the statement of Hill et al. (1968), saying that the use of their joint design criterion in a way protects the experimenter to prematurely choose a wrong model and then concentrate on parameter estimation for this model (as discussed in Section 8.3.1). This observation may however be related with the criticism raised by several authors (including Froment and Mezaki (1970) and Buzzi-Ferraris and Forzatti (1983)), that a model should not be accepted too rapidly when (relative) model probabilities are used to evaluate the adequacy of the models. One of the reasons is that the model probabilities may oscillate considerably as experimentation progresses.

One can also see that in many cases, no model was selected. In principle, even if all rival models are wrong, there will always be a model for which the model probability is largest. In other words, in the end, there will always be one model that will be identified as the most appropriate one. However, as a result of numerical problems, this was not the case here. Indeed, when a model is inadequate, it will fail to describe the available experimental data and the WSSE will consequently be very large. It is thus possible that the WSSE is so large, that $\mathrm{P}\left[\chi^2_{n-n_p} \leqslant \mathrm{WSSE}\right] \approx 1$. If that is the case, the model probability of the corresponding model, denoted as $\rho$, will approximate zero (Eq. (8.8)). Although, in theory, $\rho$ can only be equal to zero when $\mathrm{WSSE} = \infty$, it occurs in practice for large values of WSSE due to rounding errors. This explains why the simultaneous procedure stopped in some of the repetitions without selecting one of the models as the most probable model. However, it should be noted that these large WSSE values might be caused by local optima in the parameter estimation step. Indeed, it can occur that the optimization algorithm fails to

**Table 8.2:** Overview of the observed outcomes of the 150 runs of the simultaneous procedure to integrate model discrimination and optimal experimental design for parameter estimation, and this for the five scenarios with a different preliminary experiment, $\boldsymbol{\xi}_1^i$ with $i = 1, \ldots, 5$.

| | | simultaneous |
|---|---|---|
| model $m_5$ | $\boldsymbol{\xi}_1^1$ | 16 |
| | $\boldsymbol{\xi}_1^2$ | 15 |
| | $\boldsymbol{\xi}_1^3$ | 21 |
| | $\boldsymbol{\xi}_1^4$ | 23 |
| | $\boldsymbol{\xi}_1^5$ | 18 |
| | $\boldsymbol{\xi}_1^1 - \boldsymbol{\xi}_1^5$ | **93** *62%* |
| other model | $\boldsymbol{\xi}_1^1$ | 8 |
| | $\boldsymbol{\xi}_1^2$ | 11 |
| | $\boldsymbol{\xi}_1^3$ | 4 |
| | $\boldsymbol{\xi}_1^4$ | 6 |
| | $\boldsymbol{\xi}_1^5$ | 10 |
| | $\boldsymbol{\xi}_1^1 - \boldsymbol{\xi}_1^5$ | **39** *26%* |
| all models rejected | $\boldsymbol{\xi}_1^1$ | 6 |
| | $\boldsymbol{\xi}_1^2$ | 4 |
| | $\boldsymbol{\xi}_1^3$ | 5 |
| | $\boldsymbol{\xi}_1^4$ | 1 |
| | $\boldsymbol{\xi}_1^5$ | 2 |
| | $\boldsymbol{\xi}_1^1 - \boldsymbol{\xi}_1^5$ | **18** *12%* |

locate the global optimum. For the evaluation of the procedure, this issue is partly dealt with by repeating each model discrimination procedure thirty times.

### 8.5.6 Required number of experiments

Obviously, the number of additional experiments that have to be performed before the most appropriate model can be identified is an important aspect of the evaluation, especially in view of the claim of Hill et al. (1968) regarding the efficiency of their approach. In this section, the results obtained for the simultaneous procedure will be presented and compared to the ones obtained with the sequential procedures (using the averaged design strategy). However, it should be kept in mind that the approaches to evaluate the adequacy of the rival models (that will determine when the procedure is stopped) are different.

The number of experiments required to achieve model discrimination are presented in Fig. 8.9. This figure contains five subfigures with a white background (entitled $\boldsymbol{\xi}_1^i$, with $i = 1, \ldots, 5$) and one subfigure with a gray background (entitled $\boldsymbol{\xi}_1^1 - \boldsymbol{\xi}_1^5$). The former present the results obtained for the simulations where the model discrimination procedure was initiated with the preliminary experiment indicated in the title of the corresponding subfigure, whereas the one with the gray background gives an overall picture of the number of required experiments and presents the values of all model discrimination runs ($5 \times 30 = 150$ in total). Note that in these figures, the preliminary experiment corresponds to experiment number one and, if applicable, the compromise experiment corresponds to experiment number two.

From the results in Fig. 8.9, one can see that no significant difference can be observed based on the overall number of required experiments. However, the results obtained for the individual scenarios (that is, for the different preliminary experiments) show that the simultaneous procedure is always (one of the) best performing one(s). However, it should be noted that, although no significant differences were observed between the two sequential procedures in this particular case study, it could very well be that the conceptual differences between both procedures become more apparent when they are applied to another case study.

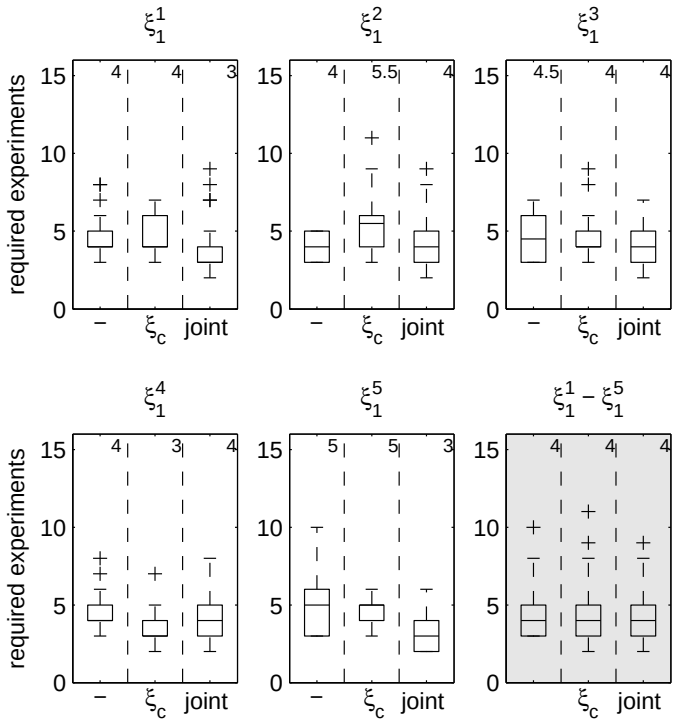



**Figure 8.9:** Boxplots showing the number of experiments (median in upper right corner) required to achieve model discrimination for the scenario where the where the sequential procedure is applied with (middle) and without (left) the use of a compromise experiment ($\boldsymbol{\xi}_c$) and where the simultaneous procedure is applied (right), and this starting from each of the five preliminary experiments ($\boldsymbol{\xi}_1^1$ till $\boldsymbol{\xi}_1^5$) (white background). The boxplots with the gray background give an overall idea of the number of required experiments and were made using the results obtained from all starting situations.

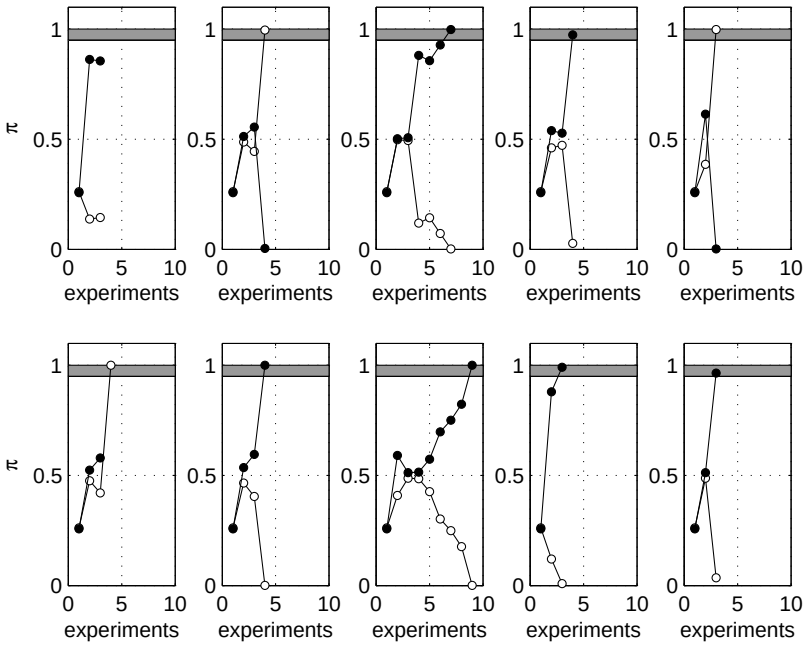



**Figure 8.10:** Evolution of the relative model probabilities ($\pi$) of models $m_2$ ($\circ$) and $m_5$ ($\bullet$) as the number of performed experiments increases for some of the repetitions of the simultaneous procedure starting with preliminary experiment $\boldsymbol{\xi}_1^1$.

### 8.5.7 Evolution of the relative model probabilities

As explained in Section 8.3.2, the (relative) model probabilities are recalculated when a new experiment is performed. The way in which the relative model probabilities ($\pi$) of models $m_2$ ($\circ$) and $m_5$ ($\bullet$) evolve as the number of performed experiments increases, is shown in Figs. 8.10 and 8.11 for ten (of the thirty) repetitions of the simultaneous procedure starting with preliminary experiments $\boldsymbol{\xi}_1^1$ and $\boldsymbol{\xi}_1^3$, respectively. As already stated above, the procedure stops when the relative probability of one of the rival models is above 0.95, which is indicated by the gray bar in the figures. Note that the results for the other models are not shown, because their relative probabilities decrease to zero very rapidly and do not really contribute to the evaluation of the simultaneous procedure.

From these results, one can see that the evolution of the relative model probabilities of models $m_2$ and $m_5$ differ considerably among the different repetitions of the procedure. However, the relative model probabilities of these models often approaches 0.5 after performing the first designed experiment. This indicates that both models describe the available experimental data equally well and that the other models (which were not shown for brevity)

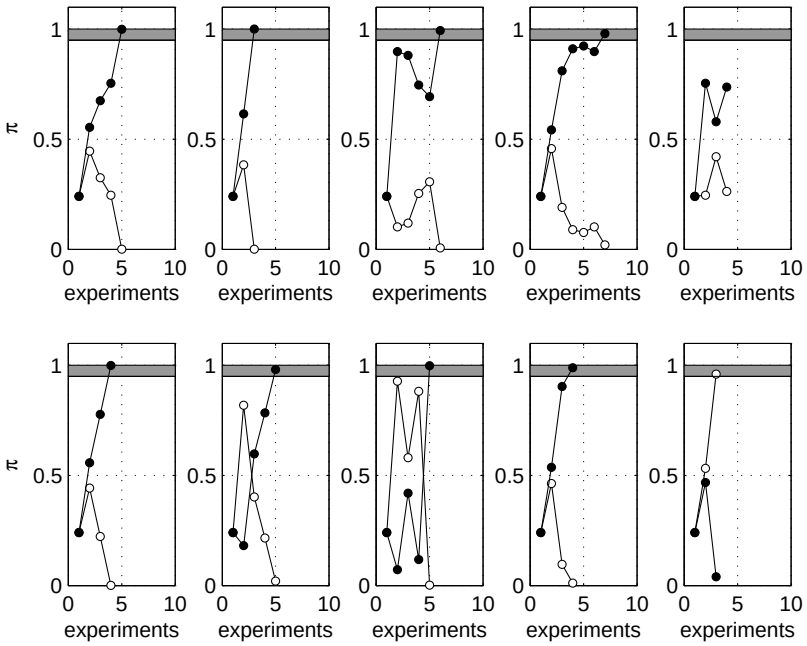



**Figure 8.11:** Evolution of the relative model probabilities ($\pi$) of models $m_2$ ($\circ$) and $m_5$ ($\bullet$) as the number of performed experiments increases for some of the repetitions of the simultaneous procedure starting with preliminary experiment $\boldsymbol{\xi}_1^3$.

do not describe the data in an adequate manner as $\sum_i \pi_i = 1$. In some cases, discrimination is achieved after one or two additional experiments, but sometimes more experiments are required before the relative model probability tends to one (see Section 8.5.6).

Note that the results shown in Figs. 8.10 and 8.11 may give the impression that the relative model probabilities are the same for each of the rival models at the start of the procedure (for instance, equal to $1/m$ as in the original formulation of the simultaneous procedure discussed in Section 8.3.1). This is, however, not the case. It would make less sense to do this when the model probabilities are calculated as suggested by Schwaab et al. (2006). Indeed, in the original formulation of Hill et al. (1968) the model probabilities are updated according to Bayes' theorem, while here the relative model probabilities are simply calculated from the corresponding WSSE values.

As already stated in Section 4.4.5, some authors have observed that the model probabilities may oscillate considerably from iteration to iteration in the sequential strategy (for instance, Froment and Mezaki (1970)). Although this behavior could also be observed in some of the repetitions of the procedure, it was generally not the case here. However, it should be noted this may be due to the fact that, in this case study, the (relative) model

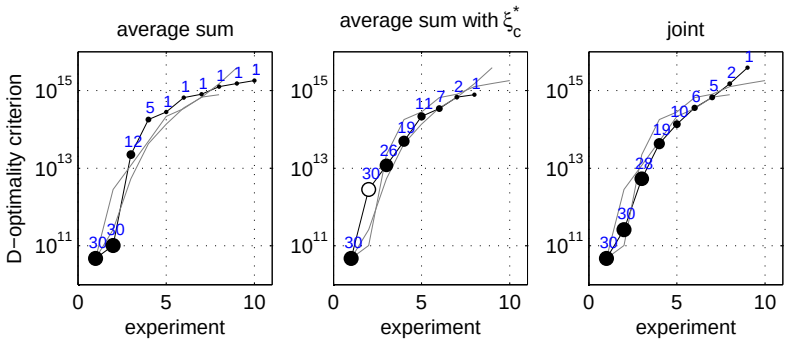
probabilities are not calculated using Bayes' theorem, where the current model probabilities (that is, the ones before performing the newly designed experiment) are used to (re)calculate the new ones after performing the designed experiment. One can also notice that the procedure sometimes stopped without selecting one of the models as the most appropriate one (for instance in the upper left subfigure of Fig. 8.10). The reason why this occurs has already been discussed in Section 8.5.5 and was related to numerical issues or possibly to problems with the optimization, which are inherent to this kind of simulation studies.

### 8.5.8 Evaluation of the quality of the parameter estimates

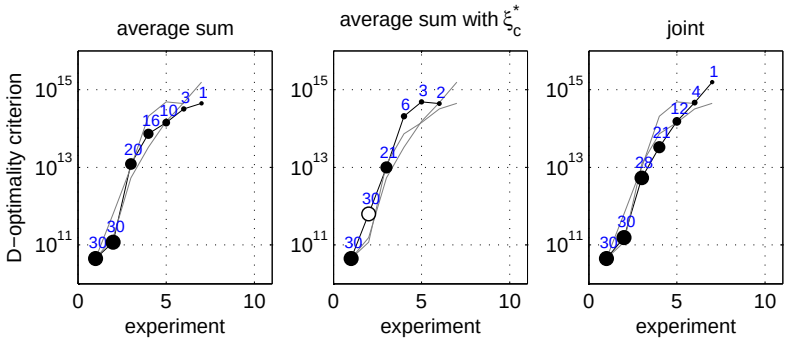
In the section, the quality of the parameter estimates of the model that is eventually identified as the most appropriate one is presented by means of the D-optimality criterion value. However, as the modified version of the simultaneous procedure uses the averaged design strategy to quantify the discriminatory potential of an experiment (Eq. (8.6)), the results obtained here cannot simply be compared with the ones obtained with the sequential procedures discussed in the previous section (which were obtained using the pairwise design strategy). Therefore, the case study was also performed for the sequential procedures where the averaged design strategy was used to design the optimal discriminatory experiments. In this way, the performance between the simultaneous procedure and the sequential procedures can be investigated.

The evolution of the median of the (thirty) D-optimality criterion values of model $m_5$ is shown in Fig. 8.12(a-c) for the scenarios starting from $\boldsymbol{\xi}_1^2$ (a), experiment $\boldsymbol{\xi}_1^3$ (b) and experiment $\boldsymbol{\xi}_1^4$ (c) and for the three procedures. In these figures, the results shown in black were obtained for the procedure indicated in the title of the subfigure, whereas the gray lines show the results obtained for the other procedures. The latter are shown to facilitate the comparison among the results. The results obtained for the other scenarios are shown in Appendix C and are similar to the ones shown in Fig. 8.12.

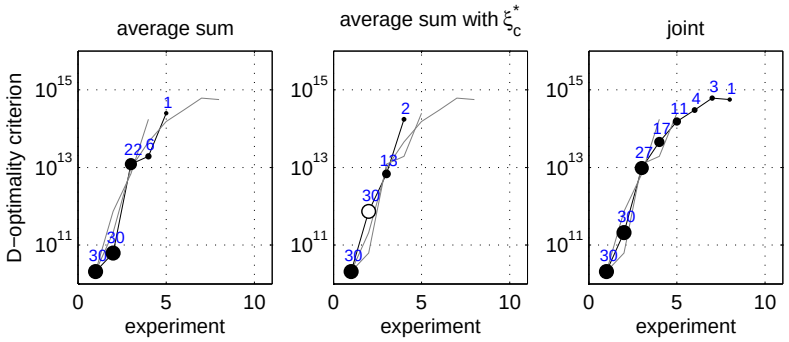
In Section 8.4.9, the effect of performing a compromise experiment was investigated for the sequential procedure where the pairwise design strategy was used to design the optimal discriminatory experiments. From these results presented here, one can see that also with the averaged design strategy, the quality of the parameter estimates increases faster when a compromise experiment is performed. When looking at the results obtained with the simultaneous procedure (joint design criterion), one can see that the D-optimality criterion values obtained after the performing the first designed experiment are slightly larger than




(a) starting from preliminary experiment $\boldsymbol{\xi}_1^2$




(b) starting from preliminary experiment $\boldsymbol{\xi}_1^3$




(c) starting from preliminary experiment $\boldsymbol{\xi}_1^4$

**Figure 8.12:** Evolution of the median of the D-optimality criterion values of model $m_5$ for the (thirty) repetitions of the model discrimination procedure that were initiated with experiment $\boldsymbol{\xi}_1^2$ (a), experiment $\boldsymbol{\xi}_1^3$ (b) and experiment $\boldsymbol{\xi}_1^4$ (c). The results obtained with and without the use of a compromise experiment are shown in the middle and left subfigures, respectively, whereas the ones obtained with the simultaneous procedure are shown in the right subfigure. The evolution of the median criterion values of the other strategy is shown in gray to ease the comparison. The number of criterion values from which the median was determined will be indicated by the size of a bullet symbol and the corresponding integer (in blue).

the ones obtained using the sequential procedure without a compromise experiment (but lower than the ones obtained with the compromise experiment). In a way, this is in agreement with the rationale of the joint design criterion. For instance, for the scenario with preliminary experiment $\boldsymbol{\xi}_1^1$, the value of $w_1$ is equal to 0.83 (Eq. 8.5), which indicates that the information content of the designed experiment with regard to the parameters of the rival models will already contribute in the designed experiment. As the relative probability of model $m_5$ is more or less equal to 0.25 after performing the preliminary experiment (see Figs. 8.10 and 8.11), it is clear that information content of the experiment with regard to its parameters is already taken into account when designing the first experiment.

However, as the number of performed experiments increases, the D-optimality criterion values obtained for the three procedures converge to more or less the same value and no systematic difference can thus be observed. Note that a similar behavior was observed in Section 8.4.10 (Fig. 8.7) where the results obtained with the two sequential procedures (using the pairwise design strategy) were compared.

### 8.5.9 Further improvement of the parameter estimates of the most appropriate model

As in the previous case study, the evolution of the D-optimality criterion values for model $m_5$ was determined from the moment it was identified as the most appropriate model and experiments could be designed only focusing on parameter estimation. For illustratory purposes, the evolution of the D-optimality criterion values for model $m_5$ is presented in Fig. 8.13 for the first ten experiments that are performed. For more information on the interpretation of this figure, the reader is referred to Section 8.4.10.

As with the sequential procedures discussed in Section 8.4, one can see that after ten experiments, the D-optimality design criterion values converge to more or less the same value (in Fig. 8.13). In other words, also for the simultaneous procedure it does not really matter how much effort is spent on model discrimination. As discussed in Section 8.4.10, this observation can be explained by the similarity between OED/PE and the anticipatory approach for optimal experimental design for model discrimination (as discussed in Chapter 4) as the discriminatory potential of the experiment was evaluated accordingly.

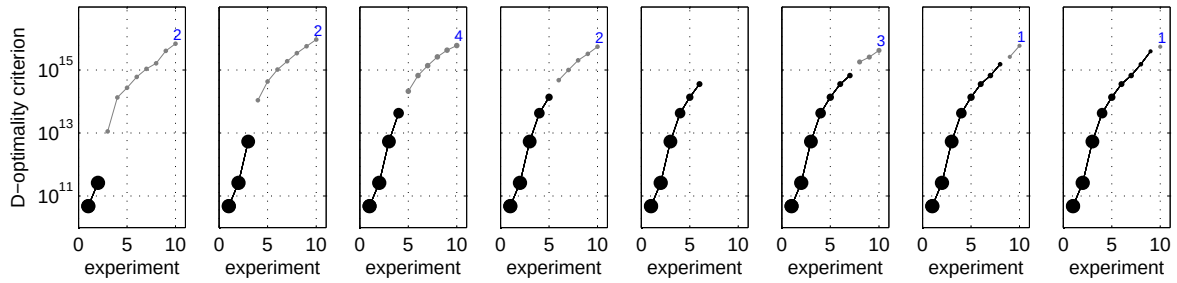



**Figure 8.13:** Evolution of the median of the D-optimality criterion values of model $m_5$ for the (thirty) runs of the model discrimination procedure that was initiated with experiment $\boldsymbol{\xi}_1^2$ for the simultaneous procedures. The black lines indicate the D-optimality criterion values obtained in the model discrimination procedure, whereas the gray lines shows how the D-optimality criterion values evolves after model $m_5$ has been identified as the most appropriate model.

## 8.6 Conclusions

To obtain a practically useful model, it is required that the model structure is adequate and that its parameters are estimated with a satisfactory level of precision. As methods are available to deal with both aspects individually (model discrimination and optimal experimental design for parameter estimation (OED/PE), respectively), it is important to investigate how these methods can be integrated in a more general procedure for model building. In this chapter, three possibilities were proposed and applied to the working example used throughout this dissertation.

In the first procedure, both aspects are dealt with sequentially, that is, the model discrimination procedure is performed first, and then the parameters of the selected model are further refined through the design of optimally informative experiments. The second procedure, is similar, except that a compromise experiment is designed and performed prior to the start of the model discrimination procedure to improve the quality of the parameter estimates. In the third procedure to integrate model discrimination and OED/PE, both aspects were dealt with simultaneously. For this purpose, the joint criterion proposed by Hill et al. (1968) was modified such that the anticipatory approach could be used to quantify the discriminatory potential of the proposed experiments. The basic design strategy of the joint design criterion is to emphasize model discrimination when there is considerable doubt as to which model is best, and gradually shift the emphasis to parameter estimation as experimentation progresses and model discrimination becomes possible.

The results obtained after applying the two sequential procedures to the working example showed that it is interesting to design and perform a compromise experiment before starting the model discrimination procedure. Even though model discrimination was not achieved in fewer experiments compared to the sequential procedure without a compromise experiment, the quality of the parameter estimates improved faster when a compromise experiment was performed first. As resources may be limited in practical applications, this should not be underestimated as it may not be possible to perform experiments until model discrimination is achieved. However, one could observe the quality of the parameter estimates converged as the number of performed experiments increased and the benefit of the compromise experiment disappeared.

The performance of the simultaneous procedure was also investigated by applying it to the working example, and it appeared to be worse than the performance of the sequential procedures. This was concluded from the observation that the wrong model was identified as the most appropriate one in 26% of the repetitions of the procedure, while this was much lower for the sequential procedures (6 - 7%). However, this may be due to the way in which the adequacy of the models is evaluated in the simultaneous procedure and may not necessarily be associated with the rationale of the joint design criterion used in the procedure.

In addition, the results also confirmed the similarity between the anticipatory approach developed to design optimal discriminatory experiments on the one hand, and optimal experimental design for parameter estimation on the other hand. No matter which procedure was applied, the D-optimality design criterion values converged to more or less the same value after performing ten experiments, regardless of the number of discriminatory experiments that were needed to identify the most appropriate model. In other words, it did not really matter how much effort is spent on model discrimination, since the discriminatory experiments were informative with regard to the model parameters as they were designed using the anticipatory approach.

# PART IV

## GENERAL CONCLUSIONS AND PERSPECTIVES

# CHAPTER 9

## General conclusions and suggestions for further research

## Abstract

In the opening chapter of this dissertation, the problem statement and the objectives of this research were described, and a number of research questions were formulated. In this final chapter, these research questions are revisited and an answer is provided based on the results obtained in this work. In addition, a number of suggestions are given for future research.

## 9.1 Problem statement and research objectives

The objectives of the research performed in the scope of this dissertation were formulated as follows:

1. obtain insight in the different aspects of the problem of model discrimination and how they influence each other,
2. propose a general procedure to discriminate among a set of rival models,
3. provide methods to design optimal discriminatory experiments,
4. position the model discrimination procedure in a more general procedure for building mathematical models.

To meet these research objectives, five research questions were formulated that allowed performing this research in a systematic way. These research questions will be repeated in the next sections and answers will be provided based on the results and experience gathered when performing this work.

## 9.2 Research question I

> *What is the general procedure to identify the most appropriate model from a set of rival models?*

A general procedure to discriminate among a number of rival models consists of four steps that are performed in an iterative manner until a stopping criterion is met (as depicted in Fig. 9.1). In the first step, the parameters of the rival models are estimated from all the data that is available at that time. A second step involves an evaluation of the adequacy of the models and is thus performed in order to find out which models are able to describe the available data in a reasonable manner and which ones do not. Models that pass this test are used in a third step, where an optimal discriminatory experiment is designed. This experiment is then performed in a fourth and last step, after which the loop is closed by re-estimating the parameters of all rival models using all data available at that time. By reconsidering all models in the model evaluation step, one accounts for the possibility that one (or more) models were wrongly rejected by chance in a previous iteration. The

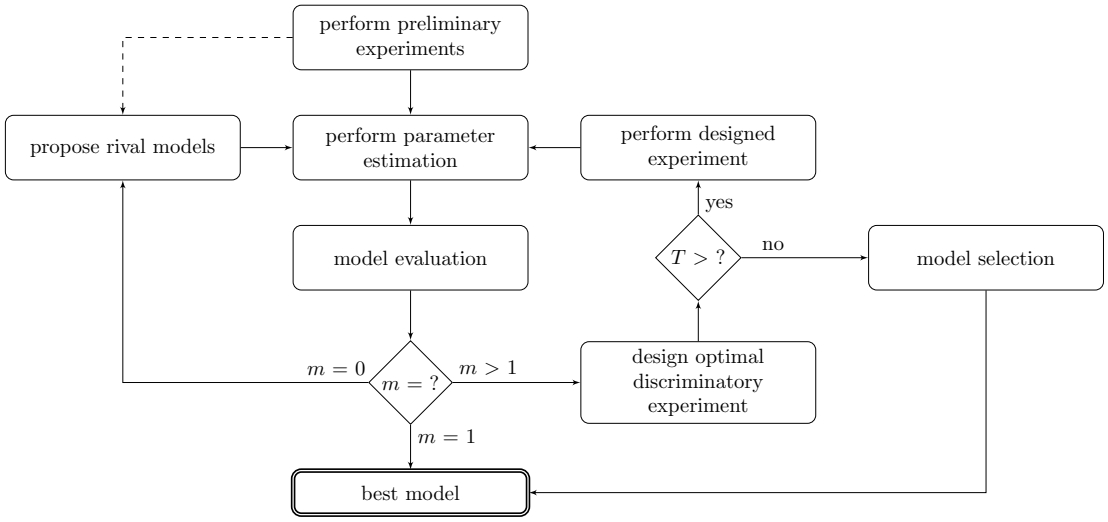



**Figure 9.1:** General procedure to discriminate among $m$ rival models (adapted from Chen and Asprey (2003) and Schwaab et al. (2006)).

iterative procedure described above continues until the best model is identified, all models appear to be inadequate and new models thus have to be proposed, or when discrimination among the remaining model candidates is no longer possible. The latter occurs when the discriminatory potential of the designed experiment is too low, and (further) model discrimination appears to be impossible given the experimental setup. In the latter case, the best model is selected from the remaining rival models as a trade-off between model fit, model parsimony and their parameter identifiability characteristics (Section 2.9).

## 9.3 Research question II

*How can the adequacy of a model be evaluated and (how) can this be translated into quantitative model evaluation criteria?*

In the model evaluation step of the model discrimination procedure, a model is considered adequate when it is able to describe the available experimental data in a reasonable manner. In other words, the model evaluation step deals with the model structure itself and one is interested in its ability to represent the available data or, put differently, in its inability to capture the dynamics of the studied process. Since the model parameters are estimated prior to the model evaluation step, the observed model fit is the best one possible given that particular model structure.

In case studies performed in this dissertation, experimental data was generated by adding noise to the simulation results obtained with one particular model, that was assumed to be the *true* model, and for given values of its parameters. Therefore, a statistical test ($\chi^2$ test, described in Section 2.8.1) could be used to evaluate whether the residual error between the measurements and the model predictions can be explained by the (simulated) measurement error. If not, the contribution to the residual error due to lack-of-fit is significant and the model is considered inadequate.

However, in practical applications, the processes being studied are often very complex and it is not realistic to assume that the true model can be found. Indeed, a mathematical model is an abstract representation of reality, and it can never mimic reality under all conditions. Therefore, a statistical test would often detect a significant lack-of-fit and reject useful models. After a literature review, two promising model evaluation criteria were suggested (in Section 2.8.2) but not further investigated: the modeling efficiency and the index of agreement. These model evaluation criteria were selected from the vast number of alternative model evaluation criteria because of their interpretability.

## 9.4 Research question III

> *If additional experimental data has to be collected to allow further model discrimination, how can experiments be designed such that model discrimination is achieved with a minimum of additional experimental effort?*

When several models are able to describe the available experimental data in an adequate manner and one wants to retain only the most appropriate one, new information about the process has to be collected and thus new experiments have to be performed. As performing experiments may be time and money consuming, it is advisable to design these experiments in a rational manner in order to minimize the experimental effort. For this purpose, experimental design methods have been developed (Chapter 4).

In general, the design of experiments is approached as an optimization problem, where the optimum of a well-defined design criterion is sought by varying the experimental degrees of freedom. When designing an optimal discriminatory experiment, the rival models themselves are used to evaluate an experiment for its discriminatory potential, which is basically determined by the difference between the model predictions (Hunter and Reiner, 1965), preferably by taking into account the uncertainty associated with this difference

(Buzzi-Ferraris et al., 1984). This uncertainty originates from two sources: the uncertainty on the measurements and the uncertainty on the model predictions.

These aspects are incorporated in the design criterion proposed by Buzzi-Ferraris et al. (1984), which was taken as a starting point because it is appealing from a conceptual point of view and because it was already used successfully by others. However, this design criterion only uses the information content of the already performed experiments to evaluate the discriminatory potential of the designed experiment and, therefore, a modified design criterion was proposed in this dissertation (Section 4.5), where the expected information content of the newly designed experiment is considered, even before the experiment is performed. Hence, this approach was called the anticipatory approach to optimal experimental design for model discrimination.

The results obtained after applying the original approach of Buzzi-Ferraris et al. (1984) and the anticipatory approach to a first case study showed that the anticipatory approach led to more reliable estimates of the model prediction uncertainties that are eventually obtained after performing the designed experiment and re-estimating the model parameters. In addition, the results showed that the anticipatory approach arranges the experimental degrees of freedom such that the expected model prediction uncertainty is small for the measurements taken, thereby increasing the discriminatory potential of the resulting experiment. One could also observe that when the information content of the designed experiment is large enough, the anticipatory approach results in an experiment that is similar to the one found with the simpler and less computationally expensive approach that only considers uncertainty on the measurements (and not the model prediction uncertainty). As this condition not always holds, it is not advised to design the optimal discriminatory experiments without considering the uncertainty on the model predictions.

In a second case study, the approaches were evaluated for their ability to bring forth a sequence of (informative) discriminatory experiments. From the results obtained in this dissertation, one could conclude that the approach proposed by Buzzi-Ferraris et al. (1984) appeared to be a rather conservative one. Even though the true model was always identified as the most appropriate one, more experiments were generally required compared to the anticipatory approach. In addition, the information content (with regard to the parameter estimates) of the experiments designed using this approach was often lower than the information content of the experiments obtained from the anticipatory approach. Based on the results obtained in this case study, one can conclude that the anticipatory approach to design optimal discriminatory experiments is to be preferred.

## 9.5 Research question IV

*What is the importance of the uncertainty on the parameter estimates with regard to the design of optimal discriminatory experiments?*

The uncertainty on the parameter estimates is determined by the information content of the already performed experiments. If this information content is low, the uncertainty on the parameter estimates will be high and this will propagate to the model predictions when the model is used to simulate the outcome of an experiment. This obviously hampers the design of optimal discriminatory experiments, as a better evaluation of the discriminatory potential of an experiment can be obtained when the model predictions are more reliable. As the uncertainty on the parameters plays a central role in the discrimination among rival models, considerable attention was given to this issue throughout this dissertation.

A first contribution that is related to the uncertainty on the parameter estimates and its importance for model discrimination, is the anticipatory approach to design discriminatory experiments. Indeed, the reason why the anticipatory approach performs better than the original approach of Buzzi-Ferraris et al. (1984) is related to the uncertainty on the parameter estimates, and can be explained by the similarity between the anticipatory approach and optimal experimental design for parameter estimation. Both the design criteria used in optimal experimental design for parameter estimation and the anticipatory approach to design discriminatory experiments take into account the information of the newly designed experiment. In addition, both design criteria benefit from a larger information content of the designed experiment with regard to the model parameters. Therefore, one can state that an experiment designed using the anticipatory approach is more likely to result in more precise parameter estimates than an experiment designed using the original approach of Buzzi-Ferraris et al. (1984).

A second contribution of this dissertation related to the uncertainty on the parameter estimates is the design of so-called compromise experiments (Chapter 7). Regardless of the design criterion used, it would be beneficial if the uncertainty of the parameter estimates of the different rival models could be reduced prior to the start of the model discrimination procedure. This can be achieved by performing dedicated experiments, designed using optimal experimental design methods. However, performing such an additional experiment for each rival model may increase the experimental effort instead of minimizing it. Therefore, the design of a compromise experiment was investigated, where a compromise experiment is defined as a single experiment that is sufficiently informative to improve the

overall precision of the parameters of all rival models. For this purpose, two methods were developed and applied: the kernel-based method (Section 7.2) and the ideal point method (Section 7.3). The kernel-based method was developed for the specific problem of optimally selecting sampling times to simultaneously improve the precision of the parameter estimates of several rival models, whereas the ideal point method proved to be capable of designing compromise experiments in experimental design problems with experimental degrees of freedom of all types (manipulations, initial conditions and sampling times).

## 9.6 Research question V

> *Can optimal experimental design for parameter estimation be integrated with the procedure for model discrimination, and is it beneficial to do so?*

In the classical sequential procedure to integrate model discrimination and optimal experimental design for parameter estimation, both aspects are dealt with sequentially, that is, the model discrimination procedure is performed first, and then the parameters of the selected model are further refined through the design of optimally informative experiments. However, two other approaches to integrate model discrimination and optimal experimental design for parameter estimation were described and applied in this dissertation (Chapter 8).

A first alternative procedure (Section 8.2.2) is similar to the classical sequential procedure, but a compromise experiment is designed and performed before starting the model discrimination procedure. Performing this compromise experiment should improve the quality of the parameter estimates of the different models and thus decrease the uncertainty on their predictions, making the design of discriminatory experiments more effective. In the second alternative procedure (Section 8.3), both aspects are dealt with simultaneously. For this purpose, the joint criterion proposed by Hill et al. (1968) was modified such that the anticipatory approach could be used to quantify the discriminatory potential of the proposed experiments. The basic design strategy of this joint design criterion is to emphasize model discrimination when there is considerable doubt as to which model is best, and gradually shift the emphasis to parameter estimation as experimentation progresses and model discrimination becomes possible.

The results obtained after applying the two sequential procedures (with and without the use of a compromise experiment) to a case study showed that it is interesting to design and perform a compromise experiment before starting the model discrimination procedure. Although model discrimination was not achieved in less experiments compared to

the sequential procedure without a compromise experiment, the quality of the parameter estimates improved faster when a compromise experiment was performed first. The performance of the simultaneous procedure was not as good as the performance of the sequential procedures, because the wrong model was identified as the most appropriate more frequently than with the sequential procedures. However, this may be due to the way in which the adequacy of the models is evaluated in the simultaneous procedure and may not necessarily be associated with the rationale of the joint design criterion used in the procedure. The simultaneous procedure, which is interesting from a conceptual point of view, may still be further improved.

In addition, the results also confirmed the similarity between the anticipatory approach developed to design optimal discriminatory experiments on the one hand, and optimal experimental design for parameter estimation on the other hand. No matter which procedure was applied (sequential or simultaneous), the D-optimality design criterion values converged to more or less the same value after performing ten experiments, regardless of the number of discriminatory experiments that were needed to identify the most appropriate model. In other words, it did not really matter how much effort is spent on model discrimination, since the discriminatory experiments were informative with regard to the model parameters as they were designed using the anticipatory approach.

## 9.7 Suggestions for further research

In the following sections, some suggestions for further research are briefly discussed.

### 9.7.1 Application to other (real-world) model discrimination problems

In this dissertation, the general procedure for model discrimination and its related methods were evaluated by applying them to the working example described in Chapter 3. Applying these methods to *in silico* case studies has the big advantage that the characteristics of the generated experimental data are known as the characteristics of the random noise that is added to the simualation results in order to mimic experimental data (Section 3.6) are well-defined. This is important because in this context/dissertation, the focus lies on the methods and the presence of experimental abnormalities, that are unavoidable when performing actual experiments, may hamper the interpretability of the results with regard to the methods used.

Still, the ultimate test for the proposed procedure for model discrimination would be to apply it to a real-world case study. When doing so, some issues may arise that do not occur when performing *in silico* case studies. Some of them are briefly discussed below:

- In this dissertation, the uncertainty on the measurements was well defined through the choice of $\varsigma_y$ and $lb_y$, which respectively represent a constant minimal relative error and a lower accuracy bound on the measurement of $y$ (Section 3.6). However, in practice, it may be difficult to get an idea of the precision and/or accuracy of the measurements.

- Another issue that may arise in practical applications is the presence of outliers. As the latter may result in wrong parameter estimates, the evaluation of the experiments in the experimental design steps may be misleading. Therefore, it would be advisable if the presence of outliers in the experimental data could be detected before actually using the experimental data. Several techniques have been developed for so-called outlier detection. The work of Liang and Kvalheim (1996); Pell (2000); Rousseeuw and Leroy (1987); Vankeerberghen et al. (1995) deal with these techniques and may be good starting points.

### 9.7.2 Experimental identifiability

In Section 2.9, the notions of structural and practical identifiability were defined. The structural identifiability of a particular model structure is examined under the assumption that perfect or *error-free* measurements are available for the response variables. From the structural identifiability analysis, one may conclude that only certain combinations of the model parameters can be identified. Practical identifiability, on the other hand, determines whether the available data is sufficiently informative to identify the model parameters.

It would make sense to define *experimental identifiability*, where the possibilities of the experimental setup to collect information about the studied process are considered. Indeed, structural identifiability does not consider the practical aspects of the experimentation process because it assumes perfect and continuous data, whereas practical identifiability only considers the data from the already performed experiments. It would therefore be interesting to look at the potential of the experimental setup, and to detect an imbalance between the complexity of the model structure and the possibilities of the experimental setup. One could define experimental identifiability as follows: *Given the experimental*

*setup, is it possible to collect the necessary information from dedicated experiments allowing to give a unique value to the parameters?.*

### 9.7.3 Anticipatory approach to OED/MD in a sequential procedure

In the case studies described in Chapter 4, it was shown that model discrimination was achieved in the lowest number of experiments with the anticipatory approach and that this approach generally resulted in experiments with a larger information content compared to the other approaches to design optimal discriminatory experiments. The reason why the anticipatory approach performed better than the original approach of Buzzi-Ferraris et al. (1984) could be related to the uncertainty on the parameter estimates and by the similarity between the anticipatory approach and optimal experimental design for parameter estimation (both design criteria benefit from a larger information content of the designed experiment with regard to the model parameters). However, it would be interesting for future research to examine whether the following theory holds.

In optimal experimental design for parameter estimation, the expected uncertainty on the parameter estimates is taken as an objective function (for instance, using the D-optimality design criterion). When applying the anticipatory approach to design optimal discriminatory experiments, on the other hand, the optimal experiment is the one where the expected difference in the model predictions is high and where the uncertainty on the model predictions is low. However, although the latter is an important feature of the objective function, it is not considered as such when evaluating the adequacy of the rival models. Following the same reasoning as with optimal experimental design for parameter estimation, one could argue why this is done so.

However, the fact that the designed experiment is expected to result in a decreased uncertainty on the model predictions in potentially informative regions with regard to model discrimination, may be very important in a sequential procedure (as the one proposed in Section 2.7). Indeed, the latter may have as a consequence that the discriminatory potential of the next designed experiment is higher.

### 9.7.4 Design criteria proposed by Munack (1992)

In Sections 4.4.3 and 4.4.4, the design criteria proposed by Munack (1992) were described. However, it should be noted that, to our knowledge, these design criteria were never applied or further investigated (for dynamic models). Although it is not easy to predict the

performance of these design criteria, they seem appealing from a conceptual point of view. Indeed, because the model parameters are re-estimated based on the predictions obtained with the rival model(s), one can expect that these approaches to design discriminatory experiments are more robust, in the sense that they are less dependent on the currently available parameter estimates. However, two important remarks should be make with respect to these design criteria:

- As discussed in Sections 4.4.3 and 4.4.4, the experiment proposed by the optimization algorithm is simulated with both rival models, giving rise to two additional data sets. Each data set corresponds to a scenario in which one of the rival models is assumed to be the *true* model and for each scenario the design criterion is evaluated, that is, the difference between the model predictions (Section 4.4.3) or the expected change in the parameter estimates (Section 4.4.4). The smallest design criterion value is eventually used as a measure of the discriminatory potential of the proposed experiment and is maximized. However, in our opinion, it would be better to take the largest value as a measure of the discriminatory power of the proposed experiment, and not the smallest.

- As the results obtained in this dissertation showed that one preferably uses the anticipatory approach to evaluate the discriminatory potential of the proposed experiment, it is suggested to also use the anticipatory approach for the design criterion described in Section 4.4.3.

- The rationale behind the anticipatory approach (that is, take into account the information content of the newly designed experiment into account when evaluating its discriminatory potential) could also be applied for the design criterion described in Section 4.4.4 (as illustrated in Fig. 9.2). Indeed, the uncertainty on the currently available parameter estimates (dotted lines in Fig. 9.2) of the model that was not assumed to be the *true* model (or that is put under jeopardy, say model $m_j$) is compared with the uncertainty of its parameters obtained after re-estimating them based on the *in-silico* generated data set generated with the other model (which was assumed to be the true one, say model $m_i$). However, in our opinion, it would be better to compare this uncertainty with the uncertainty obtained when the information content of the proposed experiment into account, but without re-estimating the model parameters. The discriminatory potential of the proposed experiment should thus be evaluated as indicated by the full lines in Fig. 9.2, where, for clarity, the dotted lines represent the uncertainty on the currently available parameter estimates

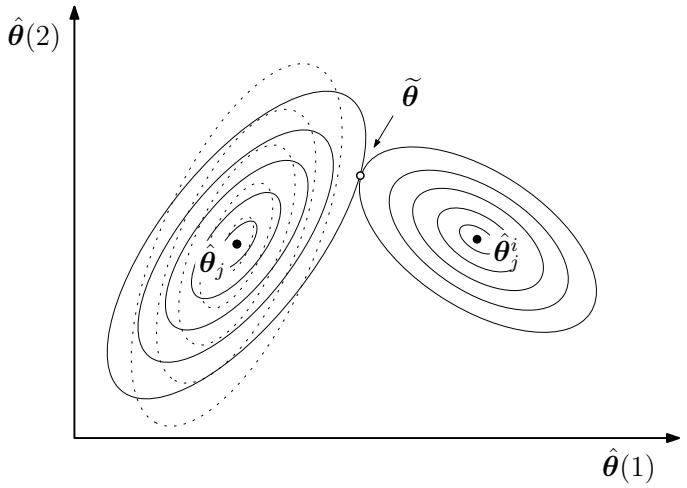



**Figure 9.2:** Illustration of the Mahalanobis distance function is used to quantify the expected change in parameter estimates while designing an optimal discriminatory experiment according to Munack (1992) for the scenario in which model $m_i$ is assumed to be the *true* model. The dotted lines represent the uncertainty on the currently available parameter estimates of model $m_j$ calculated without taking into account the information content of the designed experiment, whereas the full lines do take this into account.

of model $m_j$ calculated without taking into account the information content of the designed experiment.

### 9.7.5 Criterion to evaluate the discriminatory potential of an experiment

As advocated in Section 4.7 it would be very interesting if a criterion were available that could be used to indicate whether (further) model discrimination may be expected from a given experiment or not. In other words, if (further) model discrimination is impossible given the experiment(al setup), it would be very interesting if the experimenter could be informed about this. However, the currently available criterion to evaluate the discriminatory potential of an experiment is not appropriate for experiments where more than one response variable is measured and/or when measurements are taken at several time instants (as discussed in Section 4.7). One shortcoming of the currently available evaluation criterion is that it is based on the overall (or average) discriminatory potential of the different measurements, whereas the fact that, at least in theory, the most appropriate

model can be identified from a set of rival models when the other models fail to describe one single experimental data point is not considered.

However, the evaluation criterion does makes sense for situation with one response variable and one measurement thereof (as considered in Buzzi-Ferraris and Forzatti (1983), where this evaluation criterion was proposed for the first time). The evaluation criterion states that an experiment is expected to result in model discrimination if the corresponding $T_{ij}(\boldsymbol{\xi})$-value is larger than one. The rationale behind this criterion is that a $T_{ij}(\boldsymbol{\xi})$-value smaller than one indicates that the variance of the difference between the model predictions can be explained in terms of the measurement error ($\boldsymbol{\Sigma}$) and the uncertainty on the model predictions ($\boldsymbol{\Omega}_i + \boldsymbol{\Omega}_j$). In other words, one expects that no significant difference in the model predictions will be observed after performing the designed experiment. Instead of using the, in our opinion, incorrect extrapolations of the original evaluation criterion for more general applications (with more response variables and/or more sampling times), one could, for instance, determine the $T_{ij}$-values of the individual measurements.

For the special cases with only one measured response variable, the $T_{ij}$-values of the individual samples, denoted as $T_{ij}(\boldsymbol{\xi}(t_k))$ with $k = 1, \ldots, n_{sp}$, can be simply calculated as follows

$$T_{ij}(\boldsymbol{\xi}) = \sum_{l=1}^{n_{sp}} \Delta\hat{\mathrm{y}}_{ij}(\boldsymbol{\xi}, t_l)' \cdot \Psi_{\mathrm{ij}}(\boldsymbol{\xi}, t_l)^{-1} \cdot \Delta\hat{\mathrm{y}}_{ij}(\boldsymbol{\xi}, t_l) \ , \tag{9.1}$$

when the design criterion proposed by Buzzi-Ferraris et al. (1984) is used (see Section 4.4.7 for more information on this design criterion). As one can see in Eq. (9.1), $\hat{\mathrm{y}}_{ij}$ and $\Psi_{ij}$ are scalar values when only one response variable is measured. Each individual sampling time would then contribute to the discriminatory potential of the experiment if its corresponding $T_{ij}(\boldsymbol{\xi}(t_k))$-value is larger than one. One way of formulating an evaluation criterion would be to state that a given percentage of the $T_{ij}(\boldsymbol{\xi}(t_k))$-values is larger than one (as illustrated in Fig. 9.3).

However, the calculation of the $T_{ij}(\boldsymbol{\xi}(t_k))$-values is not straightforward for cases with more than one response variable because it is possible that correlations exist between them. Indeed, in such cases, one can see from Eq. (9.1) that the $T_{ij}(\boldsymbol{\xi}(t_k))$-values at the individual sampling times can only be calculated (independently) if $\boldsymbol{\Psi}_{ij}(\boldsymbol{\xi}, t_l)^{-1}$ is a diagonal matrix, which is generally not the case.

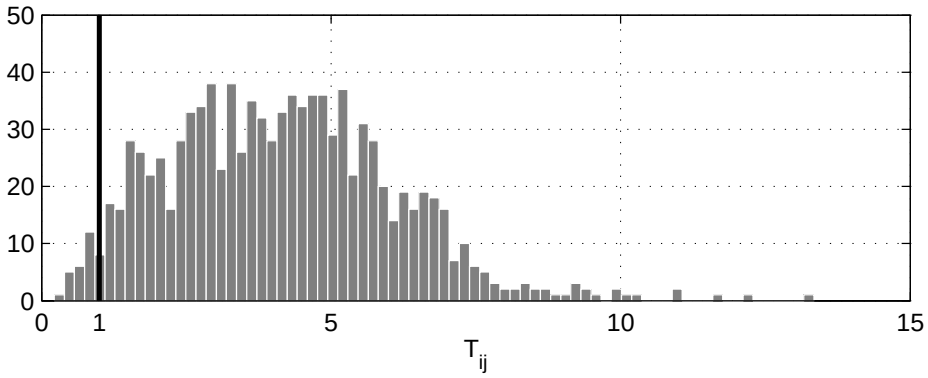



**Figure 9.3:** Histogram of hypothetical $T_{ij}\left(\boldsymbol{\xi}\left(t_k\right)\right)$-values.

### 9.7.6 Optimal experimental design for more accurate and precise parameter estimates

The established methods to perform optimal experimental design for parameter estimation (OED/PE) basically assume that the currently available parameter estimates are the *true* ones and design an experiment such that the corresponding uncertainty (assessed by means of the parameter estimation covariance matrix) is expected to reduce after it is performed. However, it would be interesting if a method were available that did not only consider the uncertainty on the parameter estimates, but also aims at finding the true parameter estimates. In other words, where the established methods for OED/PE were developed to improve the precision of the parameter estimates, it would be interesting to develop methods that also aim at an improved accuracy of the parameter estimates.

However, this is not only important for OED/PE, but also for optimal experimental design for model discrimination (OED/MD). Indeed, the discriminatory potential of the designed experiment may not be (as large) as expected because the currently available parameter estimates differ too much from the true ones. If the designed experiment is performed and the parameters re-estimated using the newly obtained experimental data, their values may change considerably and consequently also the discriminatory potential will change. If the true values of the model parameters are found faster, the performance of the overall model discrimination procedure may improve.

Suppose, for instance, that only a limited number of samples can be taken and that the time interval around $t_k$ is particularly informative with regard to the model parameters (if OED/PE) or the discriminatory potential (if OED/MD). When designing an experiment

using the methods described in this dissertation, one would expect that the samples will be scheduled in the time interval around $t_k$, where the information content is largest. However, if this time interval is rather small (compared with the duration of the entire experiment), one will not capture the overall dynamics of the process, which may result in wrong parameter estimates. The latter is obviously not desired.

One approach to include the accuracy of the parameter estimates in the experimental design, could be to make sure that the sampling times are spread over the whole duration of the experiment such that the dynamics of the process are captured. One could of course distribute the sampling times uniformly over the length of the experiment, but it would be better if the evolution of the information content over time could be taken into account.

For OED/MD problems, for instance, one could possibly achieve this as follows. In Chapter 4 it was explained that the trajectory of $T_{ij}(\boldsymbol{\xi})$ can be calculated for a given experiment $\boldsymbol{\xi}$, as the discriminatory potential at a given point in time can be calculated using the appropriate design criterion by assuming that one only samples at that time (for instance, as shown in Fig. 4.7). Suppose that such a trajectory is shown in the upper graph of Fig. 9.4. The cumulated sum of the trajectory of this design criterion is shown in the lower graph. Ten uniformly distributed sampling times are obtained as illustrated by the gray, dashed lines in the figure. Indeed, when the values of the design criterion are low, the cumulated sum will increase slowly, whereas the latter will increase rapidly in important/informative regions and these will eventually contain more sampling times.

Note that it is not possible to calculate a similar trajectory of the design criterion in the context of OED/PE as the latter is given by the determinant of $\mathbf{FIM}(\boldsymbol{\xi})$, which is calculated as

$$\mathbf{FIM}(\boldsymbol{\xi}) = \sum_{k=1}^{n_{sp}} \mathbf{FIM}(\boldsymbol{\xi}(t_k)) \ , \tag{9.2}$$

and the information content of all sampling times is thus required to evaluate the design criterion. In other words, one cannot simply evaluate the value of the design criterion at the individual sampling times. For OED/PE problems, however, the design criteria recently described by Buzzi-Ferraris and Manenti (2009) in order to obtain an optimal filling of the experimental design space for a sequence of (steady-state) experiments may be useful. Although these design criteria originated from another problem statement and

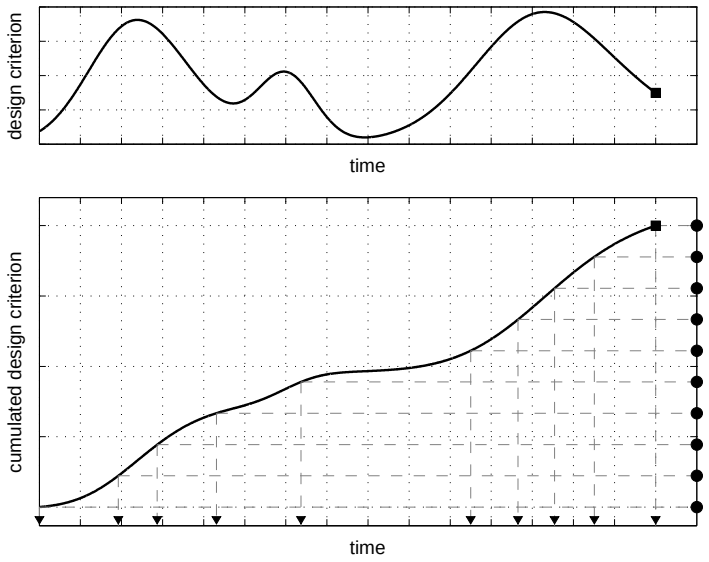



**Figure 9.4:** Illustration of an undeveloped method to distribute the sampling times uniformly over the length of the experiment (indicated by ■), thereby taking into account the information content with regard to the parameters (if OED/PE) or the discriminatory potential (if OED/MD). The upper graph shows the trajectory of the design criterion and its cumulative sum is shown in the lower graph. Ten uniformly distributed sampling times are obtained as illustrated by the gray, dashed lines.

they were not applied (yet) for dynamic models (or experiments), it would be interesting to investigate whether these design criteria could be used in this context as well.

### 9.7.7 Ds-optimality, combined with parameter subset selection

The nine rival models proposed in Chapter 3 to describe the kinetics glucokinase were rather simple models. In many applications, however, the models are much more complex and contain more parameters that obviously have to be estimated from experimental data. It is clear that the information content of the experiment has to be higher when more model parameters have to be estimated. In addition, one can expect that some model parameters will be more important than others, in the sense that the sensitivity of the state variables to the values of these parameters is higher. In literature, several approaches have been described to identify which parameters are the most important ones (Brun et al., 2002; De Pauw, 2005; Li et al., 2004). Sometimes, the least important model parameters are fixed and the original set of model parameters is thus reduced to the set of most important ones.

It could, however, be interesting to combine this with the so-called Ds-optimality design criterion. The Ds-optimality design criterion is used when one is especially interested in estimating a subset of the $n_p$ parameters as precisely as possible (Atkinson and Donev, 1992). As stated above, the most intuitive approach to do this would be to ignore the other parameters and design an experiment after removing their corresponding rows and columns from the **FIM**. In other words, by treating these model parameters as constants. However, this is a naive approach because in this way one ignores the fact that the uncertainty on the estimates of the parameter of special interest is also determined by the uncertainty of the other parameter estimates. A better approach is described below.

Without loss of generality, we may consider this subset to consist of the first $n_{p_s}$ elements of parameter vector $\boldsymbol{\theta}$ that can be partitioned as

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \end{bmatrix} \tag{9.3}$$

with $\boldsymbol{\theta}_1$ being the vector of the $n_{p_s}$ parameters of special interest and $\boldsymbol{\theta}_2$ the vector of the remaining $n_p - n_{p_s}$ parameters. The **FIM** can be partitioned accordingly as

$$\mathbf{FIM} = \begin{bmatrix} \mathbf{FIM}_{11} & \mathbf{FIM}_{12} \\ \mathbf{FIM}_{12} & \mathbf{FIM}_{22} \end{bmatrix} , \tag{9.4}$$

where $\mathbf{FIM}_{11}$ is the $n_{p_s} \times n_{p_s}$ matrix containing the rows and columns of the original **FIM** that are related to the parameters of special interest, $\mathbf{FIM}_{22}$ is the $(n_p - n_{p_s}) \times (n_p - n_{p_s})$ matrix with those of the other parameters, and $\mathbf{FIM}_{12}$ is the $n_{p_s} \times (n_p - n_{p_s})$ matrix containing the other elements of the **FIM** (those that represent the interaction between the parameters of special interest and the other ones).

The parameter estimation error covariance matrix corresponding with the $n_{p_s}$ parameters of special interest, denoted as $\boldsymbol{\Phi}_{11}$, can be calculated as

$$\boldsymbol{\Phi}_{11} = \left(\mathbf{FIM}_{11} - \mathbf{FIM}_{12} \cdot \mathbf{FIM}_{22}^{-1} \cdot \mathbf{FIM}_{12}'\right)^{-1} . \tag{9.5}$$

The Ds-optimality design criterion can thus be found by maximizing the determinant

$$\left|\mathbf{FIM}_{11} - \mathbf{FIM}_{12} \cdot \mathbf{FIM}_{22}^{-1} \cdot \mathbf{FIM}_{12}'\right| = \frac{|\mathbf{FIM}|}{|\mathbf{FIM}_{22}|} . \tag{9.6}$$

Note that, as with the D-optimality design criterion, problems arise when particular parameters are unidentifiable. Indeed, if this is the case, the **FIM** will be singular and it will not be possible to calculate $\mathbf{\Phi}_{11}$ (as seen from Eqs. (9.5) and (9.6)).

### 9.7.8 Extention of the kernel-based method to design compromise experiments

In Section 7.2.5, the difficulties encountered when extending the kernel-based method for more general experimental design problems were discussed. One of the problem occurs when one or more manipulatory variables (or process inputs) are optimized. For instance, when one of the manipulatory variables (the feed rate or the concentration of one of the process variables) is set at another value a certain time after the start of the experiment, the task of the experimental design then consists of finding the optimal new value for the manipulatory variable and the optimal time to change it. It is clear that both experimental degrees of freedom are dependent of eachother and the optimal value for the one is depends on the values given to the other. They can thus not simply be determined separately, as done when the kernel-based method is applied.

One possible approach to extend the kernel-based method as it was proposed in Section 7.2, is to use multivariate kernel density estimation instead of univariate kernel density estimation. In multivariate kernel density estimation (Wand and Jones, 1993), the general form of the kernel estimator of the multivariate probability density function, denoted as $\hat{p}(\boldsymbol{x})$, is given by

$$\hat{p}(\boldsymbol{x}) = \frac{1}{n} \cdot \sum_{i=1}^{n} \kappa_{\mathbf{H}}(\boldsymbol{x} - \boldsymbol{x}_i) \tag{9.7}$$

where $\boldsymbol{x}$ represents a $d$-dimensional (random) vector with density $p(\boldsymbol{x})$ and $n$ represents the number of vectors $\boldsymbol{x}_i$. Similar to univariate kernel density estimation, $\kappa_{\mathbf{H}}$ is taken to be a standard Gaussian function with zero mean and a variance equal to one, given by

$$\kappa_{\mathbf{H}}(\boldsymbol{x} - \boldsymbol{x}_i, \mathbf{H}) = \frac{1}{(2\pi)^{\frac{d}{2}} \cdot |\mathbf{H}|^{\frac{1}{2}}} \cdot \exp\left(-\frac{1}{2} \cdot (\boldsymbol{x} - \boldsymbol{x}_i)' \cdot \mathbf{H}^{-1} \cdot (\boldsymbol{x} - \boldsymbol{x}_i)\right). \tag{9.8}$$

The symmetric $d \times d$ matrix $\mathbf{H}$ is called the bandwidth matrix and has a similar function as the bandwidth or smoothing parameter ($h$) used in univariate kernel density estimation (and the kernel-based method). For more detailed information on multivariate kernel density estimation, the reader is referred to Wand and Jones (1993).

# PART V

## APPENDICES

# APPENDIX A

Supplementary figures for Chapter 4

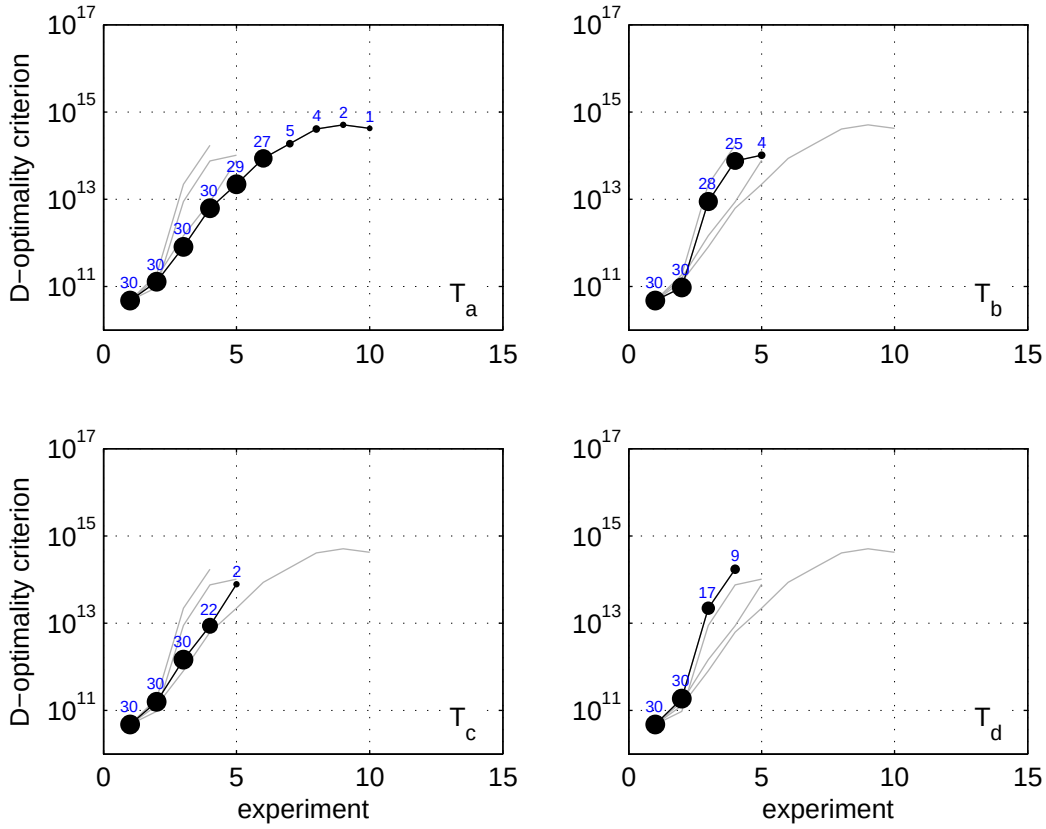



**Figure A.1:** Evolution of the median D-optimality criterion values of model $m_5$ for the (thirty) runs of the model discrimination procedure that was initiated with experiment $\boldsymbol{\xi}_1^2$, for each of the selected approaches for OED/MD. The evolution of the median criterion values of the other approaches are shown in gray to ease the comparison. The number of criterion values from which the median was determined will be indicated by the size of a bullet symbol and the corresponding integer.

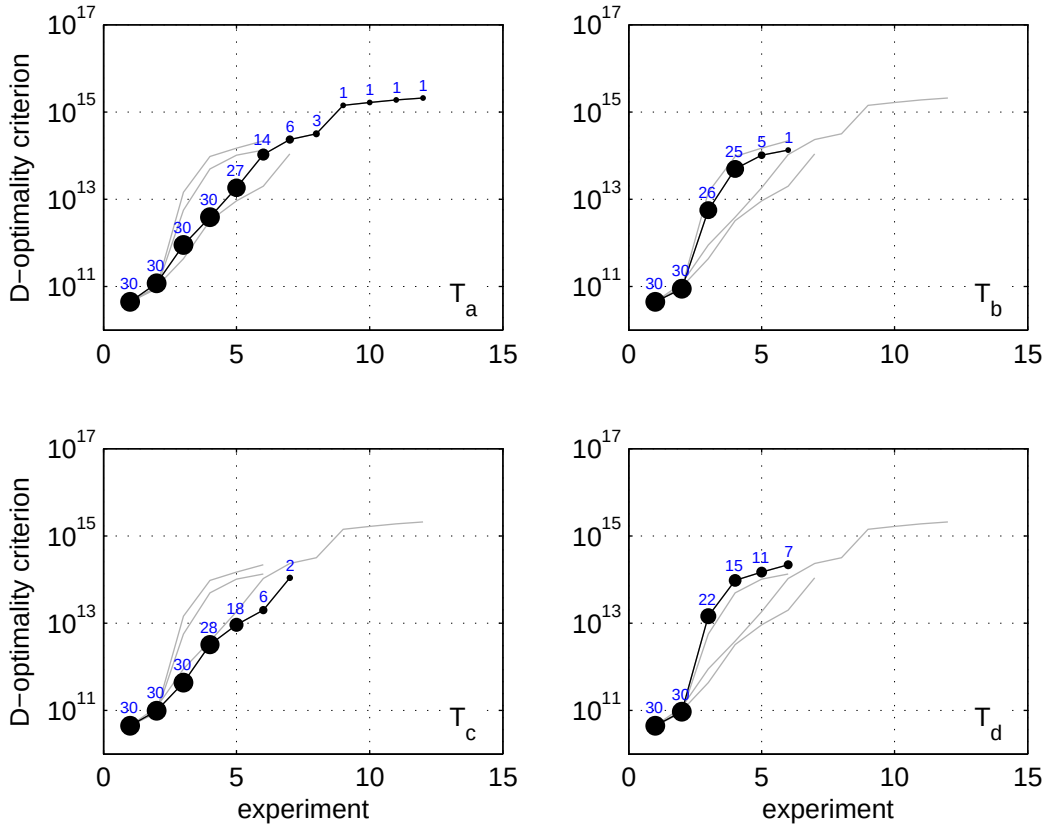



**Figure A.2:** Evolution of the median D-optimality criterion values of model $m_5$ for the (thirty) runs of the model discrimination procedure that was initiated with experiment $\boldsymbol{\xi}_1^3$, for each of the selected approaches for OED/MD. The evolution of the median criterion values of the other approaches are shown in gray to ease the comparison. The number of criterion values from which the median was determined will be indicated by the size of a bullet symbol and the corresponding integer.

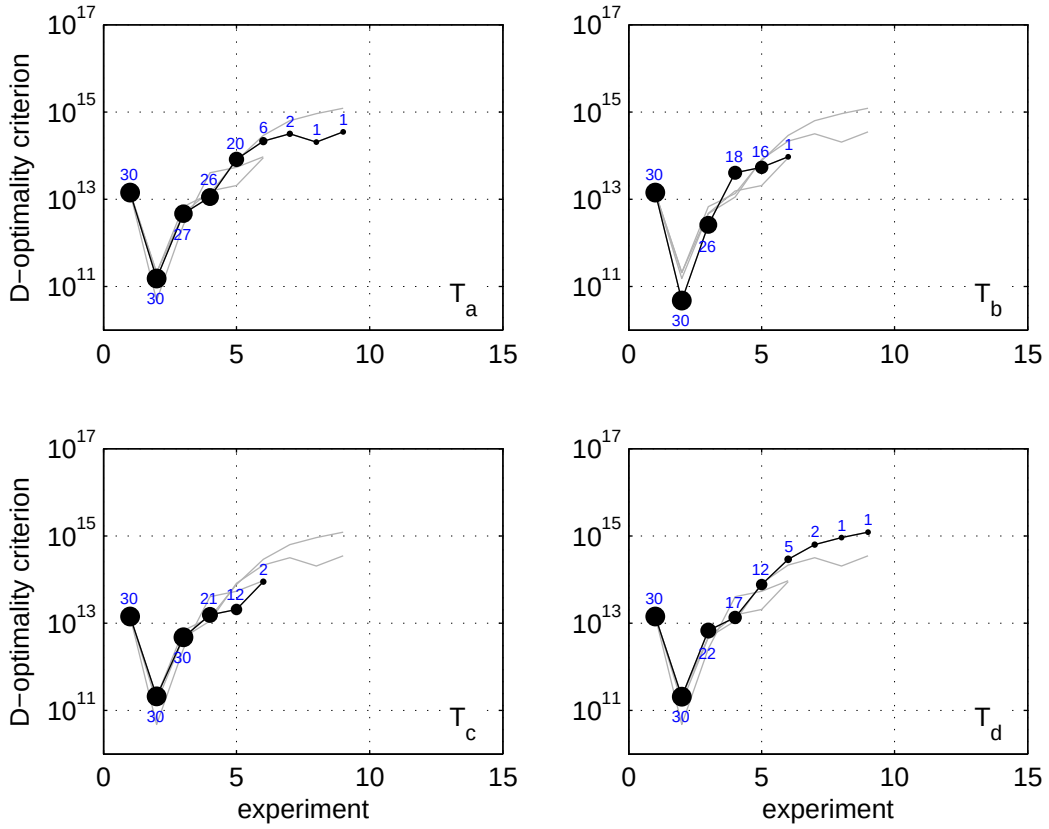
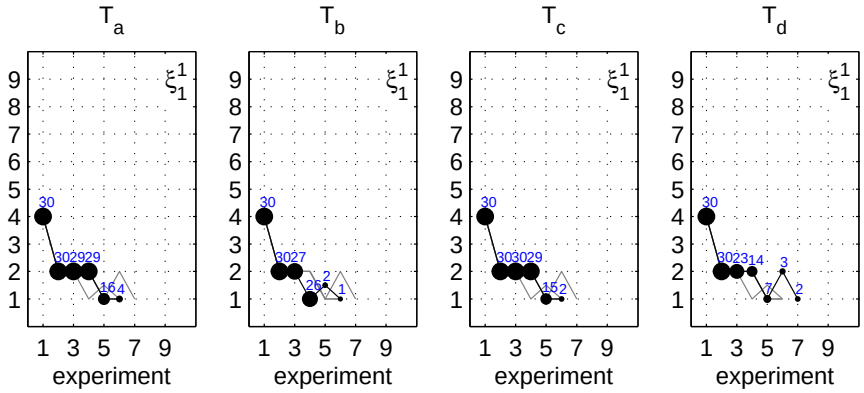
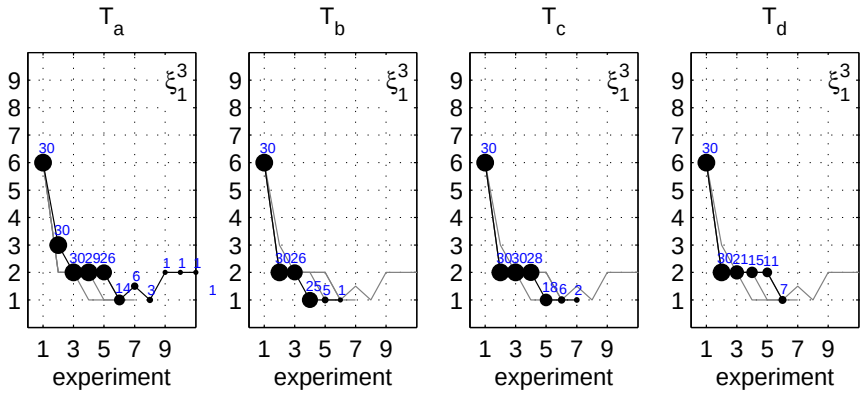



**Figure A.3:** Evolution of the median D-optimality criterion values of model $m_5$ for the (thirty) runs of the model discrimination procedure that was initiated with experiment $\boldsymbol{\xi}_1^5$, for each of the selected approaches for OED/MD. The evolution of the median criterion values of the other approaches are shown in gray to ease the comparison. The number of criterion values from which the median was determined will be indicated by the size of a bullet symbol and the corresponding integer.




(a) starting from preliminary experiment $\boldsymbol{\xi}_1^1$




(b) starting from preliminary experiment $\boldsymbol{\xi}_1^3$

**Figure A.4:** The median values of the number of adequate models as a function of the number of experiments that have been performed, starting from preliminary experiment $\boldsymbol{\xi}_1^1$ (a) and $\boldsymbol{\xi}_1^3$ (b). The evolution of these median values shows the rate at which the number of adequate models decreases for the different approaches to design optimal discriminatory experiments ($T_a$, $T_b$, $T_c$ and $T_d$). The number of values (runs) from which the median was determined will be indicated by the size of a bullet symbol and the corresponding integer.

# APPENDIX B

Supplementary figures for Chapter 5

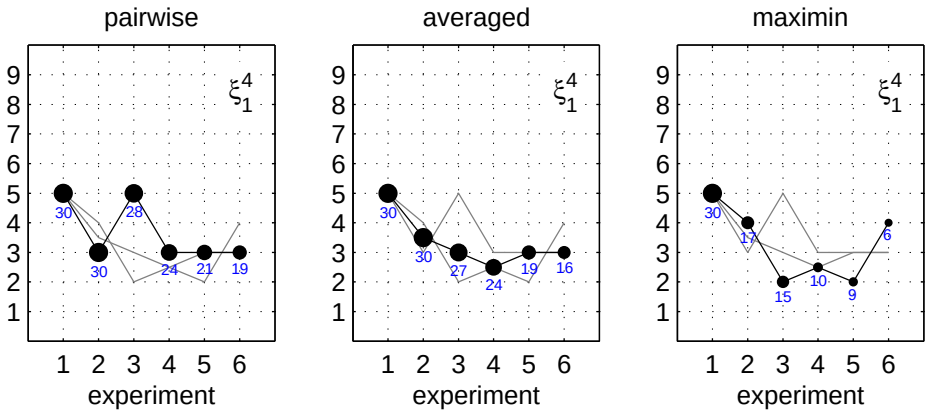



**Figure B.1:** The median of the number of adequate models, determined from the (thirty) runs of the model discrimination procedure that was initiated with experiment $\boldsymbol{\xi}_1^4$. In each of the subfigures, the results of one of the strategies are presented in black, while the evolution of the median values of the other strategies are shown in gray to ease their mutual comparison. The number of values (runs) from which the median was determined will be indicated by the size of a bullet symbol and the corresponding integer.

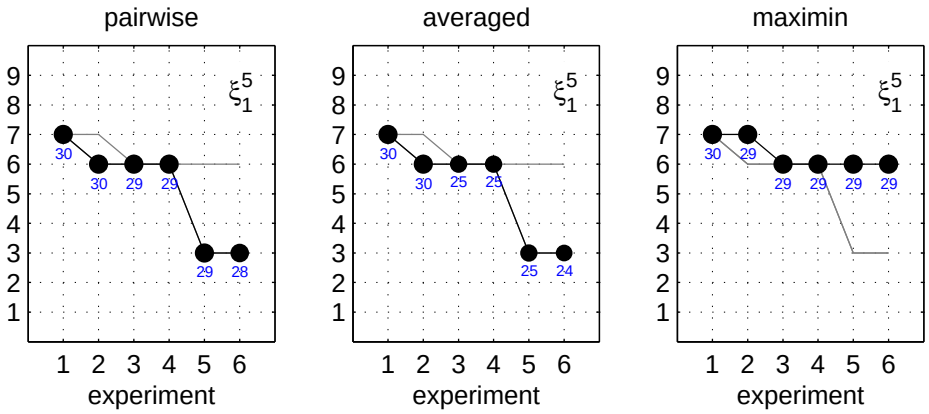



**Figure B.2:** The median of the number of adequate models, determined from the (thirty) runs of the model discrimination procedure that was initiated with experiment $\boldsymbol{\xi}_1^5$. In each of the subfigures, the results of one of the strategies are presented in black, while the evolution of the median values of the other strategies are shown in gray to ease their mutual comparison. The number of values (runs) from which the median was determined will be indicated by the size of a bullet symbol and the corresponding integer.

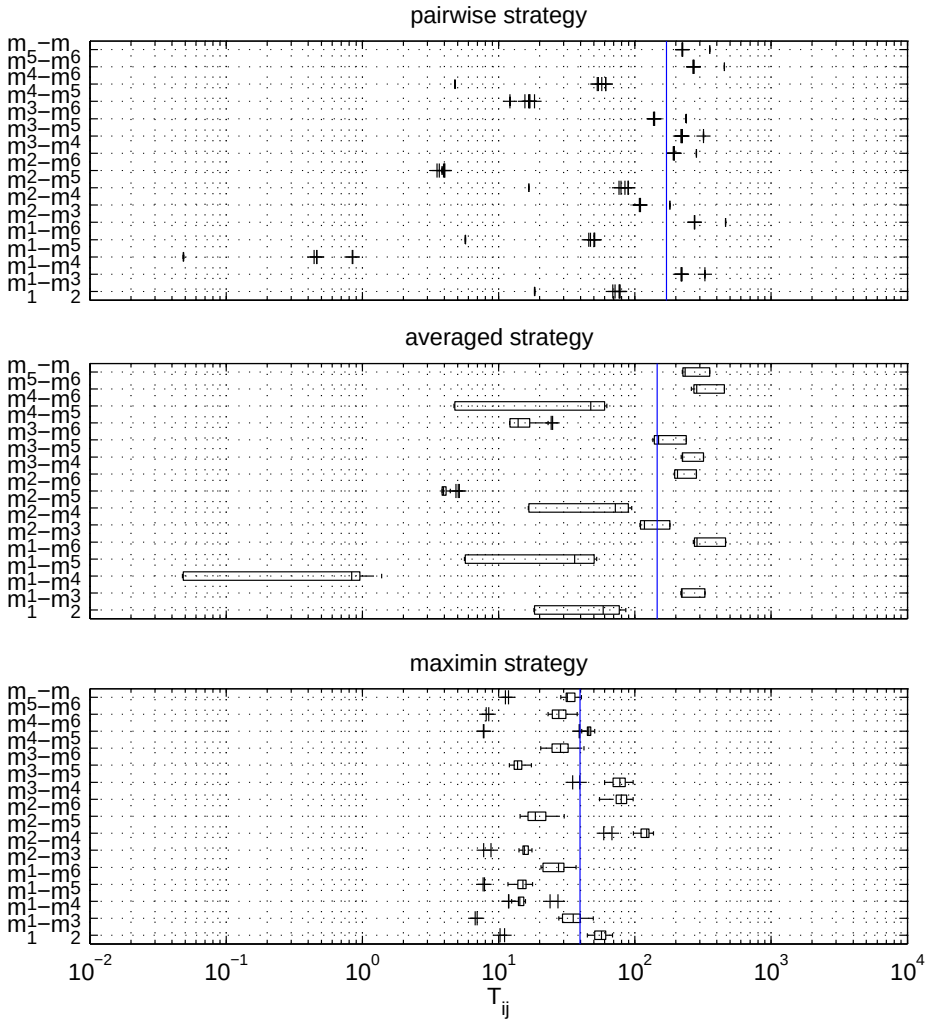



**Figure B.3:** Boxplots of the $T_{ij}$ ($\boldsymbol{\xi}_2^\star$)-values calculated for each of the model pairs, where $\boldsymbol{\xi}_2^\star$ refers to the first experiment that is designed when applying the model discrimination procedure, and is performed after preliminary experiment $\boldsymbol{\xi}_1^3$. The mean of these $T_{ij}$-values is indicated by the vertical (blue) line and gives an idea of the overall discriminatory potential of the designed experiment. Note that because the large number of model pairs, their labels were omitted for brevity.

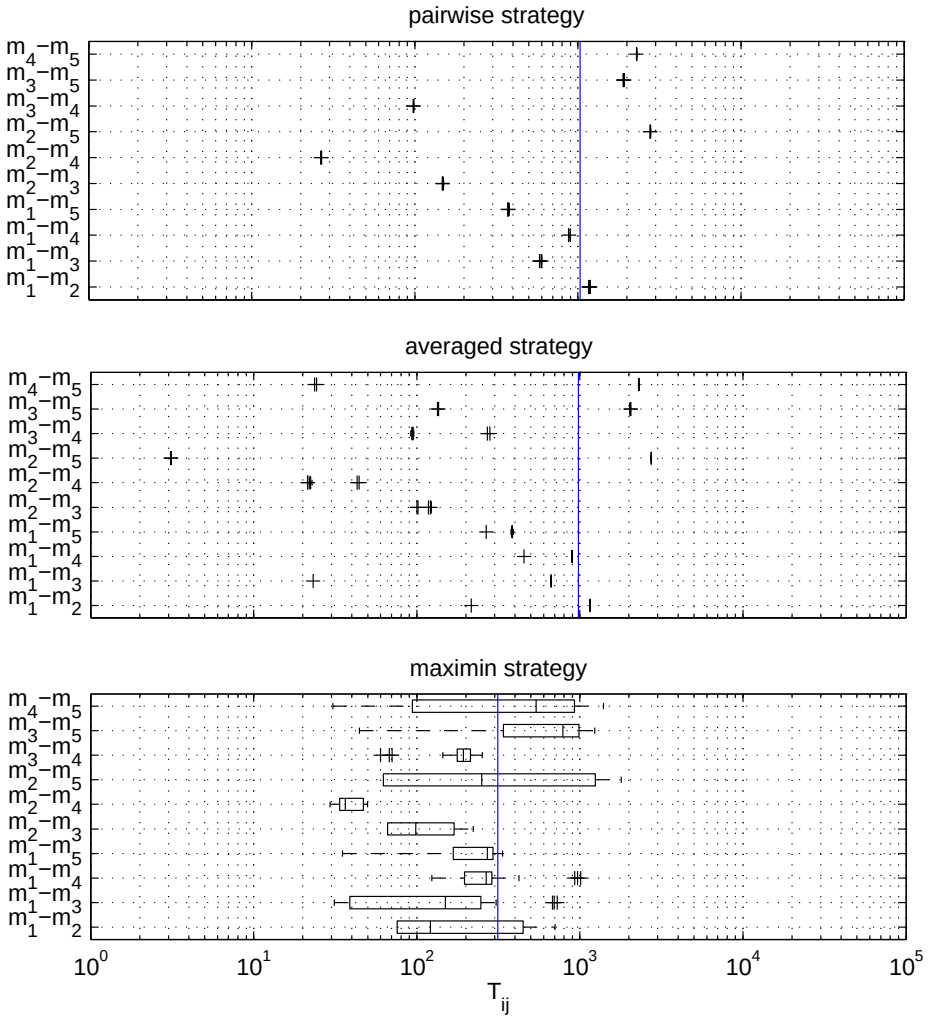



**Figure B.4:** Boxplots of the $T_{ij}$ ($\boldsymbol{\xi}_2^\star$)-values calculated for each of the model pairs, where $\boldsymbol{\xi}_2^\star$ refers to the first experiment that is designed when applying the model discrimination procedure, and is performed after preliminary experiment $\boldsymbol{\xi}_1^4$. The mean of these $T_{ij}$-values is indicated by the vertical (blue) line and gives an idea of the overall discriminatory potential of the designed experiment. Note that because the large number of model pairs, their labels were omitted for brevity.

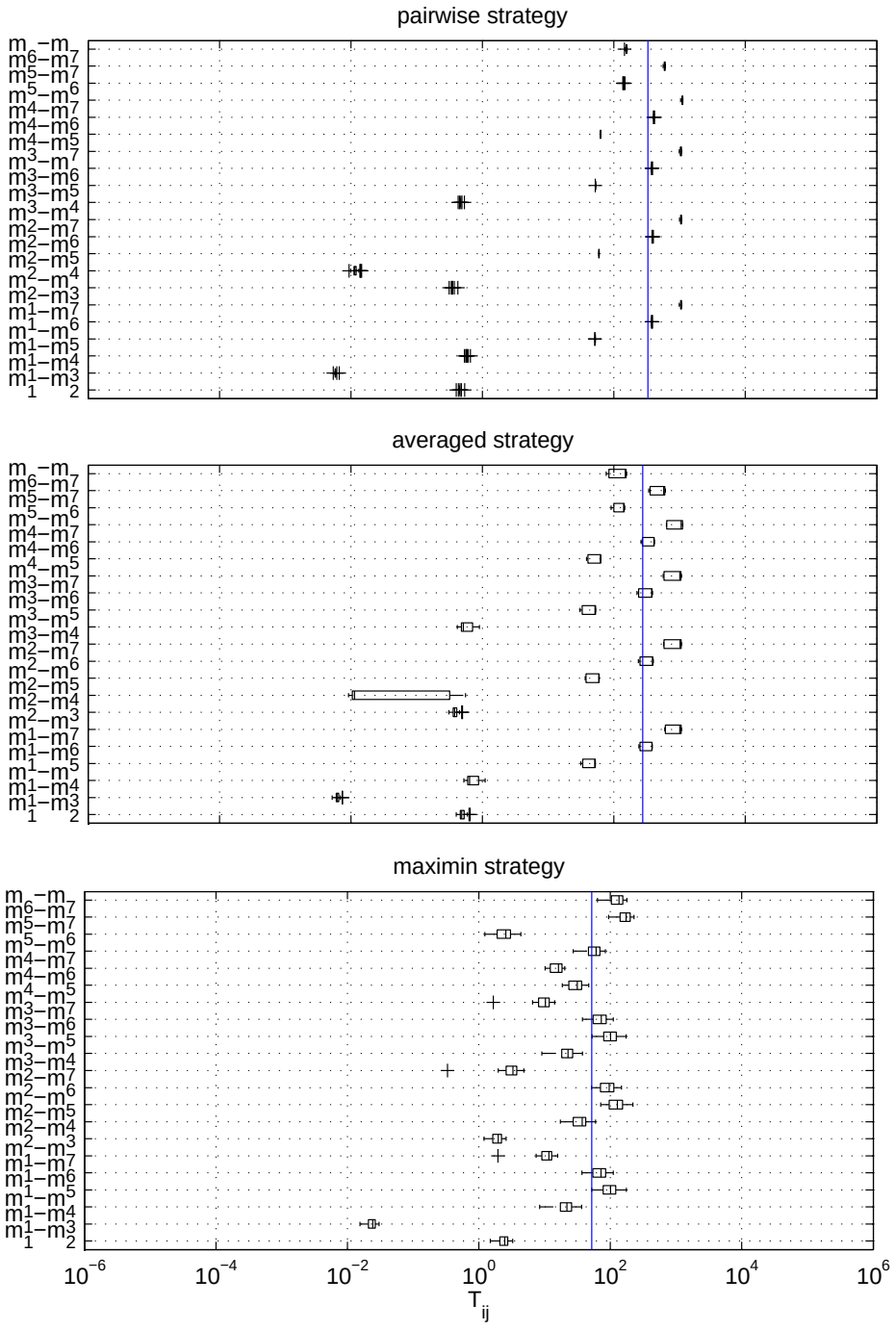
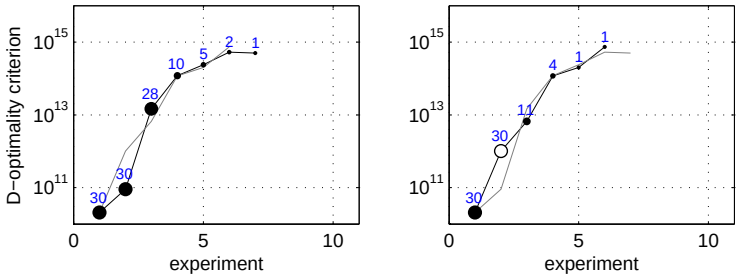


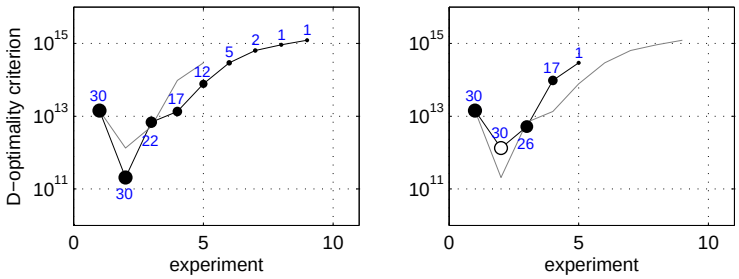

**Figure B.5:** Boxplots of the $T_{ij}$ $(\boldsymbol{\xi}_2^\star)$-values calculated for each of the model pairs, where $\boldsymbol{\xi}_2^\star$ refers to the first experiment that is designed when applying the model discrimination procedure, and is performed after preliminary experiment $\boldsymbol{\xi}_1^5$. The mean of these $T_{ij}$-values is indicated by the vertical (blue) line and gives an idea of the overall discriminatory potential of the designed experiment. Note that because the large number of model pairs, their labels were omitted for brevity.

# APPENDIX C

Supplementary figures for Chapter 8

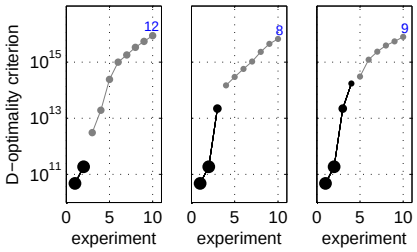



(a) starting from preliminary experiment $\boldsymbol{\xi}_1^4$

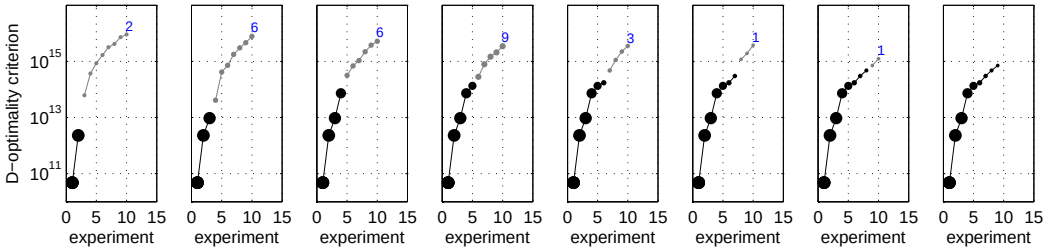



(b) starting from preliminary experiment $\boldsymbol{\xi}_1^5$

**Figure C.1:** Evolution of the median of the D-optimality criterion values of model $m_5$ for the (thirty) runs of the model discrimination procedure initiated with (a) experiment $\boldsymbol{\xi}_1^4$ and (b) experiment $\boldsymbol{\xi}_1^5$. The results obtained without the use of a compromise experiment are shown in the left subfigure and the results obtained with a compromise experiment are shown in the right subfigure. The evolution of the median criterion values of the other strategy is shown in gray to ease the comparison. The number of criterion values from which the median was determined is mentioned and also indicated by the size of the bullet symbol.

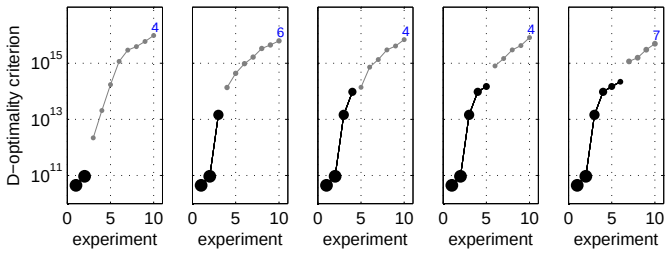



(a) starting from preliminary experiment $\boldsymbol{\xi}_1^2$

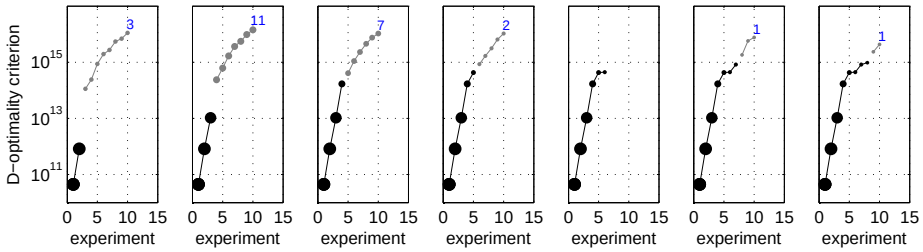



(b) starting from preliminary experiment $\boldsymbol{\xi}_1^2$ with compromise experiment

**Figure C.2:** Evolution of the median D-optimality criterion values of model $m_5$ for the (thirty) runs of the model discrimination procedure that was initiated with experiment $\boldsymbol{\xi}_1^2$ for the procedures with (a) and without the use of a compromise experiment (b). The black lines and dots indicate the D-optimality criterion values obtained in the model discrimination procedure, whereas the gray lines and dots shows how the D-optimality criterion values evolves after model $m_5$ has been identified as the most appropriate model. The latter occurs after a varying number of experiments, and the results obtained for the different occurrences are presented in the different subfigures.

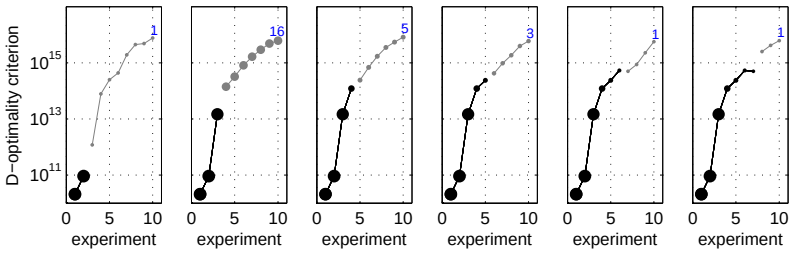



(a) starting from preliminary experiment $\boldsymbol{\xi}_1^3$

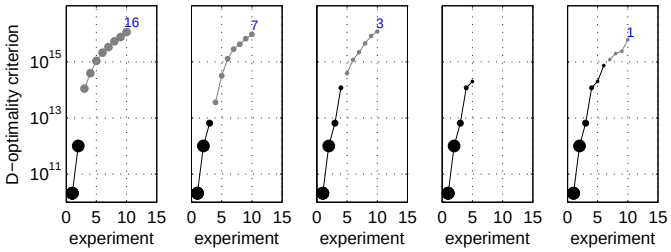



(b) starting from preliminary experiment $\boldsymbol{\xi}_1^3$ with compromise experiment

**Figure C.3:** Evolution of the median D-optimality criterion values of model $m_5$ for the (thirty) runs of the model discrimination procedure that was initiated with experiment $\boldsymbol{\xi}_1^3$ for the procedures with (a) and without the use of a compromise experiment (b). The black lines and dots indicate the D-optimality criterion values obtained in the model discrimination procedure, whereas the gray lines and dots shows how the D-optimality criterion values evolves after model $m_5$ has been identified as the most appropriate model. The latter occurs after a varying number of experiments, and the results obtained for the different occurrences are presented in the different subfigures.

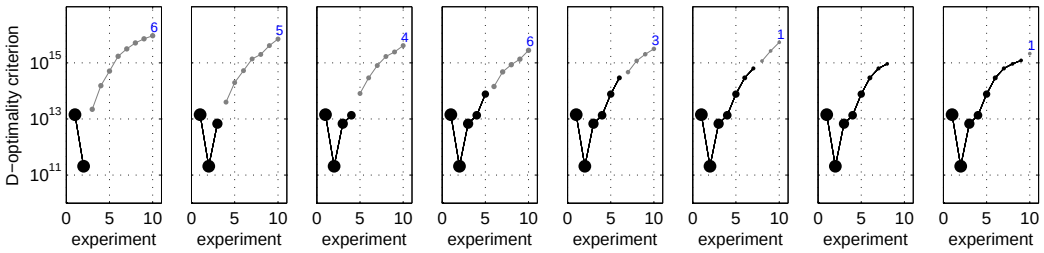



(a) starting from preliminary experiment $\boldsymbol{\xi}_1^4$

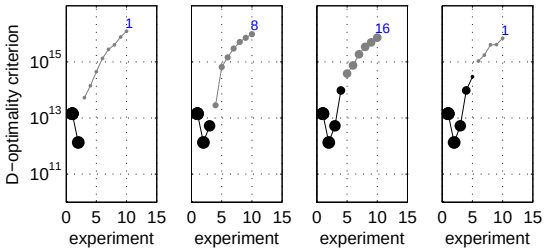



(b) starting from preliminary experiment $\boldsymbol{\xi}_1^4$ with compromise experiment

**Figure C.4:** Evolution of the median D-optimality criterion values of model $m_5$ for the (thirty) runs of the model discrimination procedure that was initiated with experiment $\boldsymbol{\xi}_1^4$ for the procedures with (a) and without the use of a compromise experiment (b). The black lines and dots indicate the D-optimality criterion values obtained in the model discrimination procedure, whereas the gray lines and dots shows how the D-optimality criterion values evolves after model $m_5$ has been identified as the most appropriate model. The latter occurs after a varying number of experiments, and the results obtained for the different occurrences are presented in the different subfigures.




(a) starting from preliminary experiment $\boldsymbol{\xi}_1^5$




(b) starting from preliminary experiment $\boldsymbol{\xi}_1^5$ with compromise experiment

**Figure C.5:** Evolution of the median D-optimality criterion values of model $m_5$ for the (thirty) runs of the model discrimination procedure that was initiated with experiment $\boldsymbol{\xi}_1^5$ for the procedures with (a) and without the use of a compromise experiment (b). The black lines and dots indicate the D-optimality criterion values obtained in the model discrimination procedure, whereas the gray lines and dots shows how the D-optimality criterion values evolves after model $m_5$ has been identified as the most appropriate model. The latter occurs after a varying number of experiments, and the results obtained for the different occurrences are presented in the different subfigures.

# Bibliography


Asprey, S. and Macchietto, S. (2000). Statistical tools for optimal dynamic model building. *Computers and Chemical Engineering*, 24:1261–1267.

Asprey, S. and Macchietto, S. (2002). Designing robust optimal dynamic experiments. *Journal of Process Control*, 12:545–556.

Atherton, R., Schainker, R., and Ducot, E. (1975). Statistical sensitivity analysis of models for chemical-kinetics. *AIChE Journal*, 21(3):441–448.

Atkinson, A. and Donev, A. (1992). *Optimum experimental designs*. Oxford University Press.

Atkinson, A. C. (2003). Horwitzs rule, transforming both sides and the design of experiments for mechanistic models. *Journal of the Royal Statistical Society Series C – Applied Statistics*, 52(3):261–278.

Atkinson, A. C. (2008). Dt-optimum designs for model discrimination and parameter estimation. *Journal of Statistical Planning and Inference*, 138:56–64.

Atkinson, A. C. and Fedorov, V. (1975). The design of experiments for discriminating between two rival models. *Biometrika*, 62:57–70.

Balsa-Canto, E., Rodriguez-Fernandez, M., and Banga, J. R. (2007). Optimal design of dynamic experiments for improved estimation of kinetic parameters of thermal degradation. *Journal of Food Engineering*, 82(2):178–188.

Banga, J., Balsa-Canto, E., Moles, C., and Alonso, A. (2005). Dynamic optimization of bioprocesses: efficient and robust numerical strategies. *Journal of Biotechnology*, 117:407–419.

Banga, J., Moles, C., and Alonso, A. (2003). Global optimization of bioprocesses using stochastic and hybrid methods. In Floudas, C. and Pardalos, P., editors, *Nonconvex Optimization and its Applications*, volume 74, pages 45–70. Kluwer Academic Publishers.

Banga, J., Versyck, K., and Impe, J. V. (2002). Computation of optimal identification experiments for nonlinear dynamic process models: a stochastic global optimization approach. *Industrial and Engineering Chemistry Research*, 41:2425–2430.

Bellman, R. and Aström, K. (1970). On structural identifiability. *Mathematical Biosciences*, 7:329–339.

Benabbas, L., Asprey, S. P., and Macchietto, S. (2005). Curvature-based methods for designing optimally informative experiments in multiresponse nonlinear dynamic situations. *Industrial and Engineering Chemistry Research*, 44:7120–7131.

Bernaerts, K., Versyck, K., and Impe, J. V. (2000). On the design of optimal dynamic experiments for parameter estimation of a Ratkowsky-type growth kinetics at suboptimal temperatures. *International Journal of Food Microbiology*, 54:27–38.

Box, G. and Hill, W. (1967). Discrimination among mechanistic models. *Technometrics*, 9(1):57–71.

Box, G. and Hunter, W. (1967). A useful method for model-building. *Technometrics*, 4(3):301–318.

Box, G. and Lucas, H. (1959). Design of experiments in non-linear situations. *Biometrika*, 46:77–90.

Brun, R., Kühni, M., Siegrist, H., Gujer, W., and Reichert, P. (2002). Practical identifiability of ASM2d parameters–systematic selection and tuning of parameter subsets. *Water Research*, 36(16):4113–4127.

Burke, A. L., Duever, T. A., and Penlidis, A. (1995). Model discrimination via designed experiments: discrimination between the terminal and penultimate models based on rate data. *Chemical Engineering Science*, 50(10):1619–1634.

Burke, A. L., Duever, T. A., and Penlidis, A. (1996). An experimental verification of statistical discrimination between the terminal and penultimate polymerization models. *Journal of Polymer Science: Part A: Polymer Chemistry*, 34:2665–2678.

Burke, A. L., Duever, T. A., and Penlidis, A. (1997). Discrimination between the terminal and penultimate models using designed experiments: An overview. *Industrial and Engineering Chemistry Research*, 36:1016–1035.

Buzzi-Ferraris, G. (1999). Planning of experiments and kinetic analysis. *Catalysis Today*, 52:125–132.

Buzzi-Ferraris, G. and Forzatti, P. (1983). A new sequential experimental design procedure for discriminating among rival models. *Chemical Engineering Science*, 38(2):225–232.

Buzzi-Ferraris, G., Forzatti, P., and Canu, P. (1990). An improved version of a sequential design criterion for discriminating among rival multiresponse models. *Chemical Engineering Science*, 45(2):477–481.

Buzzi-Ferraris, G., Forzatti, P., Emig, G., and Hofmann, H. (1984). Sequential experimental design procedure for model discrimination in the case of multiple responses. *Chemical Engineering Science*, 39(1):81–85.

Buzzi-Ferraris, G. and Manenti, F. (2009). Kinetic models analysis. *Chemical Engineering Science*, 64:1061–1074.

Cardoso, M. F., Salcedo, R., and Feyo de Azevedo, S. (1996). The simplex-simulated annealing approach to continuous non-linear optimization. *Computers and Chemical Engineering*, 20(9):1065–1080.

Checchi, N. and Marsili-Libelli, S. (2005). Reliability of parameter estimation in respirometric models. *Water Research*, 39:3686–3696.

Chen, B. H. and Asprey, S. P. (2003). On the design of optimally informative dynamic experiments for model discrimination in multiresponse nonlinear situations. *Industrial and Engineering Chemistry Research*, 42:1379–1390.

de Brauwere, A., De Ridder, F., Pintelon, R., Elskens, M., Schoukens, J., and Baeyens, W. (2005). Model selection through a statistical analysis of the minimum of a weighted least squares cost function. *Chemometrics and Intelligent Laboratory Systems*, 76:163–173.

De Pauw, D. (2005). *Optimal experimental design for calibration of bioprocess models: a validated software toolbox*. PhD thesis, Ghent University.

De Pauw, D. J. and Vanrolleghem, P. A. (2006a). Designing and performing experiments for model calibration using an automated iterative procedure. *Water Science and Technology*, 53(1):117–127.

De Pauw, D. J. and Vanrolleghem, P. A. (2006b). Practical aspects of sensitivity function approximation for dynamic models. *Mathematical and Computer Modelling of Dynamical Systems*, 12(5):395–414.

Deb, K. (2001). *Multi-Objective Optimization using Evolutionary Algorithms*. John Wiley & Sons.

Dochain, D. and Vanrolleghem, P. A. (2001). *Dynamical Modelling and Estimation in Wastewater Treatment Processes*. IWA Publishing.

Duda, R. O., Hart, P. E., and Stork, D. G. (2001). *Pattern Classification*. John Wiley & Sons.

Dumez, F. (1977). The use of sequential discrimination in the study of 1-butane dehydrogenation. *Industrial and Engineering Chemistry Fundamentals*, 16(2):298–301.

Ellwein, L. M., Tran, H. T., Zapata, C., Novak, V., and Olufsen, M. S. (2008). Sensitivity analysis and model assessment: mathematical models for arterial blood flow and blood pressure. *Cardiovascular Engineering*, 8:94–108.

Espie, D. and Macchietto, S. (1989). The optimal design of dynamic experiments. *AIChE Journal*, 35(2):223–229.

Fersht, A. (1999). *Structure and Mechanism in Protein Science – A Guide to Enzyme Catalysis and Protein Folding*. W.H. Freeman and Company.

Ford, I., Titterington, D., and Kitsos, C. (1989). Recent advances in non-linear experimental design. *Technometrics*, 31(1):49–60.

Franceschini, G. and Macchietto, S. (2008). Model-based design of experiments for parameter precision: state of the art. *Chemical Engineering Science*, 63:4846–4872.

Froment, G. and Mezaki, R. (1970). Sequential discrimination and estimation procedures for rate modelling in heterogeneous catalysis. *Chemical Engineering Science*, 25:293–300.

Goel, T., Vaidyanathan, R., Haftka, R., Shyy, W., Queipo, N., and Tucker, K. (2007). Response surface approximation of Pareto optimal front in multi-objective optimization. *Computer Methods in Applied Mechanics and Engineering*, 196:879–893.

Goodwin, G. and Payne, R. (1977). *Dynamic system identification; Experiment design and data analysis*. Acadamic Press. New York.

Hill, P. D. (1978). A review of experimental design procedures for regression model discrimination. *Technometrics*, 20(1):15–21.

Hill, W. J., Hunter, W. G., and Wichern, D. W. (1968). A joint criterion for the dual problem of model discrimination and parameter estimation. *Technometrics*, 10(1):145–160.

Hsiang, T. and Reilly, P. (1971). A practical method for discriminating among mechanistic models. *Canadian Journal of Chemical Engineering*, 49:865–871.

Hunter, W., Hill, J., and Wichern, D. (1968). A joint design criterion for the dual problem of model discrimination and parameter estimation. *Technometrics*, 10(1):145–160.

Hunter, W. and Reiner, A. (1965). Designs for discriminating between two rival models. *Technometrics*, 7(3):307–323.

Janssen, P. and Heuberger, P. (1995). Calibration of process-oriented models. *Ecological Modelling*, 83(1–2):55–66.

Jenssen, R., Principe, J. C., Erdogmus, D., and Eltoft, T. (2006). The Cauchy–Schwarz divergence and parzen windowing: Connections to graph theory and mercer kernels. *Journal of the Franklin Institute*, 343:614–629.

Jones, M., Marron, J., and Sheather, S. (1996). A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association*, 91(433):401–407.

Joshi, M., Seidel–Morgenstern, A., and Kremling, A. (2006). Exploiting the bootstrap method for quantifying parameter confidence intervals in dynamic systems. *Metabolic Engineering*, 8:447–455.

Kennedy, J. B. and Neville, A. M. (1985). *Basic Statistical Methods for Engineers and Scientists*. Harper & Row, Publishers, Inc.

Kirkpatrick, S., Gelatt, C., and Vecchi, M. (1983). Optimization by simulated annealing. *Science*, 220(4598):671–680.

Krause, P., Boyle, D., and Bäse, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences*, 5:89–97.

Kremling, A., Fischer, S., Gadkar, K., Doyle, F. J., Sauter, T., Bullinger, E., Allgöwer, F., and Gilles, E. D. (2004). A benchmark for methods in reverse engineering and model discrimination: Problem formulation and solution. *Genome Research*, 14:1773–1785.

Kullback, S. (1959). *Information Theory and Statistics*. John Wiley & Sons, Inc.

Legates, D. R. and McCabe Jr., G. J. (1999). Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, 35(1):233–241.

Leis, J. and Kramer, M. (1988). The simultaneous solution and sensitivity analysis of systems described by ordinary differential equations. *ACM Transactions on Mathematical Software*, 14(1):45–60.

Li, R., Henson, M., and Kurtz, M. (2004). Selection of model parameters for off-line parameter estimation. *IEEE on Control Systems Technology*, 12(3):402–412.

Li, S., Petzold, L., and Zhu, W. (2000). Sensitivity analysis of differential-algebraic equations: a comparison of methods on a special problem. *Applied Numerical Mathematics*, 32:161–174.

Liang, Y.-Z. and Kvalheim, O. M. (1996). Robust methods for multivariate analysis – a tutorial review. *Chemometrics and Intelligent Laboratory Systems*, 32(1):1–10.

Ljung, L. (1999). *System Identification: Theory for the user*. Prentice–Hall.

Lunin, V. V., Li, Y., Schrag, J. D., Iannuzzi, P., Cygler, M., and Matte, A. (2004). Crystal structures of *Escherichia coli* ATP-dependent glucokinase and its complex with glucose. *Journal of Bacteriology*, 186(20):6915–6927.

Madigan, M. T., Martinko, J. M., and Parker, J. (2000). *Brock Biology of Microorganisms, 9th edition*. Prentice–Hall.

Marler, R. and Arora, J. (2004). Survey of multi-objective optimization methods for engineering. *Structural and Multidisciplinary Optimization*, 26:369–395.

Marsili–Libelli, S., Guerrizio, S., and Checchi, N. (2003). Confidence regions of estimated parameters for ecological systems. *Ecological Modelling*, 165:127–146.

Mathews, C. K., van Holde, K., and Ahern, K. G. (2000). *Biochemistry*. Addison Wesley Longman, Inc.

Mehra, R. (1974). Optimal input signals for parameter estimation in dynamic systems – survey and new results. *IEEE Transactions on Automatic Control*, 19(6):753–768.

Mendes, P. and Kell, D. (1998). Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics*, 14(10):869–883.

Mitchell, P. (1997). Misuse of regression for emperical validation of models. *Agricultural Systems*, 54(3):313–326.

Moles, C., Mendes, P., and Banga, J. (2003). Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Research*, 13:2467–2475.

Monasterio, O. and Cárdenas, M. (2003). Kinetic studies of rat liver hexokinase D ('glucokinase') in non-co-operative conditions showing an ordered mechanism with MgADP as the last product to be released. *Biochemical Journal*, 371:29–38.

Moriasi, D., Arnold, J., Liew, M. V., Bingner, R., Harmel, R., and Veith, T. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the American Society of Agricultural and Biological Engineers*, 50(3):885–900.

Munack, A. (1989). Optimal feeding strategy for identification of monod-type models by fed-batch experiments. In Fish, N., Fox, R., and Thornhill, N., editors, *Computer applications in fermentation technology: Modelling and control of biotechnological processes (4th edition)*, pages 195–204. SCI Elsevier Applied Science Publishing, Amsterdam.

Munack, A. (1991). Optimization of sampling. In Schugerl, K., editor, *Biotechnology, a Multi-volume Comprehensive Treatise, Volume 4. Measuring, Modelling and Control*, pages 251–264. VCH, Weinheim.

Munack, A. (1992). Some improvements in the identification of bioprocesses. In Karim, M. and Stephanoloulos, G., editors, *Modelling and Control of Biotechnical Processes*, pages 89–94. Pergamon Press, Oxford.

Nash, J. and Sutcliffe, J. (1970). River flow forecasting through conceptual models: Part I – a discussion of principles. *Journal of Hydrology*, 10:282–290.

Nelder, J. and Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, 7:308–313.

Ogawa, T., Mori, H., Tomita, M., and Yoshino, M. (2007). Inhibitory effect of phosphoenolpyruvate on glycolytic enzymes in *Escherichia coli*. *Research in Microbiology*, 158:159–163.

Ogunnaike, B. A. and Ray, W. H. (1994). *Process Dynamics, Modeling and Control*. Oxford University Press.

Omlin, M. and Reichert, P. (1999). A comparison of techniques for the estimation of model prediction uncertainty. *Ecological Modelling*, 115:45–59.

Pell, R. J. (2000). Multiple outlier detection for multivariate calibration using robust statistical techniques. *Chemometrics and Intelligent Laboratory Systems*, 52(1):87–104.

Petersen, B. (2000). *Calibration, identifiability and optimal experimental design of activated sludge models*. PhD thesis, Ghent University.

Pinto, J., Lobao, M., and Monteiro, J. (1990). Sequential experimental design for parameter estimation: a different approach. *Chemical Engineering Science*, 45:883–892.

Power, M. (1993). The predictive validation of ecological and environmental models. *Ecological Modelling*, 68:33–50.

Pronzato, L. and Walter, E. (1988). Robust experiment design via maximin optimization. *Mathematical Biosciences*, 89(2):161–176.

Rodriguez-Fernandez, M., Mendes, P., and Banga, J. R. (2006). A hybrid approach for efficient and robust parameter estimation in biochemical pathways. *Biosystems*, 83(2–3):248–265.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, 27:832–837.

Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust regression and Outlier Detection*. John Wiley & Sons, New York.

Rykiel, E. J. J. (1996). Testing ecological models: the meaning of validation. *Ecological modeling*, 90:229–244.

Schwaab, M., Monteiro, J. L., and Pinto, J. C. (2008). Sequential experimental design for model discrimination. Taking into account the posterior covariance matrix of differences between model predictions. *Chemical Engineering Science*, 63:2408–2419.

Schwaab, M., Silva, F. M., Queipo, C. A., Jr., A. G. B., Nele, M., and Pinto, J. C. (2006). A new approach for sequential experimental design for model discrimination. *Chemical Engineering Science*, 61:5791–5806.

Seber, G. and Wild, C. (1989). *Nonlinear Regression*. John Wiley & Sons, Inc.

Segel, I. H. (1975). *Enzyme Kinetics – Behaviour and Analysis of Rapid Equilibrium and Steady–State Enzyme Systems*. John Wiley & Sons, Inc.

Shannon, C. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(4):623–656.

Shirt, R., Harris, T., and Bacon, D. (1994). Experimental design considerations for dynamic systems. *Industrial and Engineering Chemistry Research*, 33:2656–2667.

Sidoli, F. R., Mantalaris, A., and Asprey, S. P. (2004). Modelling of mammalian cells and cell culture processes. *Cytotechnology*, 44:27–46.

Smith, K. I., Everson, R. M., and Fieldsend, J. E. (2004). Dominance measures for multi-objective simulated annealing. In *Proceedings of Congress on Evolutionary Computation, CEC04*, pages 23–30.

Spriet, J. (1985). Structure characterization: an overview. In Baker, H. and Young, P., editors, *Identification and System Parameter Estimation 1985 – Proceedings of the 7th IFAC/IFORS Symposium*, pages 749–756. Pergamon Press.

Takors, R., Wiechert, W., and Weuster–Botz, D. (1997). Experimental design for the identification of macrokinetic models and model discrimination. *Biotechnology and Bioengineering*, 56(5):564–576.

Tedeschi, L. O. (2006). Assessment of the adequacy of mathematical models. *Agricultural Systems*, 89:225–247.

Ternbach, M. A. B. (2005). *Modeling based process development of fed-batch bioprocesses: L-valine production by Corynebacterium glutamicum*. PhD thesis, RWTH Aachen University.

Ternbach, M. B., Bollman, C., Wandrey, C., and Takors, R. (2005). Application of model discriminating experimental design for modeling and development of a fermentative fed-batch L-valine production process. *Biotechnology and Bioengineering*, 91(3):356–368.

Turányi, T. and Rabitz, H. (2000). Local methods. In Saltelli, A., Chan, K., and Scott, E. M., editors, *Mathematical and Statistical Methods: Sensitivity Analysis*. John Wiley & Sons, Chichester.

Turlach, B. (1993). Bandwidth selection in kernel density estimation: A review. *Discussion Paper 9307, Institut für Statistik und Ökonometrie, Humboldt-Universität zu Berlin.*

Vankeerberghen, P., Smeyers-Verbeke, J., Leardi, R., Karr, C. L., and Massart, D. L. (1995). Robust regression and outlier detection for non-linear models using genetic algorithms. *Chemometrics and Intelligent Laboratory Systems*, 28(1):73–87.

Vanrolleghem, P. A. and Dochain, D. (1998). Bioprocess model identification. In Van Impe, J. F., Vanrolleghem, P. A., and Iserentant, D. M., editors, *Advanced Instrumentation, Data Interpretation, and Control of Biotechnological Processes*, pages 251–318. Kluwer Academic Publishers.

Vanrolleghem, P. A. and Van Daele, M. (1994). Optimal experimental design for structure characterization of biodegradation models: on-line implementation in a respirographic biosensor. *Water Science and Technology*, 30(4):243–253.

Versyck, K. J., Bernaerts, K., Geeraerd, A. H., and Van Impe, J. F. (1999). Introducing optimal experimental design in predictive modeling: A motivating example. *International Journal of Food Microbiology*, 51(1):39–51.

Walter, E. and Pronzato, L. (1990). Qualitative and quantitative experiment design for phenomenological models - a survey. *Automatica*, 26(2):195–213.

Walter, E. and Pronzato, L. (1997). *Identification of Parametric Models from Experimental Data*. Springer–Verlag, Berlin, Heidelber, New York.

Wand, M. and Jones, M. (1993). *Kernel Smoothing*. Chapman & Hall, London.

Waterhouse, T. H., Redmann, S., Duffull, S. B., and Eccleston, J. A. (2005). Optimal design for model discrimination and parameter estimation for itraconazole population pharmacokinetics in cystic fibrosis patients. *Journal of Pharmacokinetics and Pharmacodynamics*, 32(3–4):521–545.

Wiechert, W. (2002). Modeling and simulation: tools for metabolic engineering. *Journal of Biotechnology*, 94:37–63.

Willmott, C. (1981). On the evaluation of model performance in physical geography. *Physical Geography*, 2:184–194.

Willmott, C. J., Ackleson, S. G., Davis, R. E., Feddema, J. J., Klink, K. M., Legates, D. R., O'Donnell, J., and Rowe, C. M. (1985). Statistics for the evaluation and comparison of models. *Journal of Geographical Research*, 90(C5):8995–9005.

# Summary

A mathematical model can be defined as a mathematical representation of the mechanisms that govern the behavior of a process being studied, and the aim of a modeling exercise is to obtain a mathematical model that adequately describes and even predicts the process behavior. Once a proper model has been identified, it becomes a powerful tool for both scientists and engineers. However, it is important to realize that the lack of insight in the modeled process may result in the proposal of several so-called rival models, each of which represents a certain hypothesis of how the process works. Obviously, one is especially interested in the model that describes the process behavior in the best way. The problem of identifying the best model from a set of rival models, often referred to as the *problem of model discrimination*, was dealt with in this dissertation.

A general procedure to deal with the problem of model discrimination consists of four steps that are performed in an iterative manner until a stopping criterion is met. In the first step, the parameters of the rival models are estimated from all the data that is available. A second step involves an evaluation of the model's adequacy to describe the available data and is thus performed in order to find out which models are able to describe the data in a reasonable manner and which ones do not. Models that pass this test are used in a third step, where an optimal discriminatory experiment is designed. This experiment is then performed in a fourth and last step, after which the loop is closed by re-estimating the parameters of all rival models using all data available at that time.

To identify the best model from the set of rival models, it is thus necessary to collect new information about the process, and thus additional experiments have to be performed. Since performing experiments can be time and money consuming, carefully designing them can significantly reduce the required experimental effort. To achieve model discrimination in a minimal number of experiments, a number of design criteria have been developed. A detailed study of these design criteria revealed that the experimental design basically

comes down to finding that experiment that maximizes the difference between the model predictions, preferably taking into account the uncertainty associated with this difference. Indeed, when designing the experiment, one assumes that one the rival models is the *true* model and that the outcome of the designed experiment can be predicted by this model. Under this assumption, one expects that it will be possible to identify the most appropriate model when the other models predict this experiment in a totally different way.

The uncertainty associated with the difference in the model predictions originates from two sources: the uncertainty on the measurements and the uncertainty on the model predictions. Indeed, the uncertainty on the measurements, which can be seen as a measure of the reproducibility of the experiment(al) data, has to be taken into account when designing a discriminatory experiment, as well as the uncertainty on the model predictions since the evaluation of the discriminatory potential of an experiment is based on how it is predicted by the rival models.

These aspects are incorporated in the design criterion proposed by Buzzi-Ferraris et al. (1984), which was taken as a starting point because it is appealing from a conceptual point of view. However, this design criterion only uses the information content of the already performed experiments to evaluate the discriminatory potential of the designed experiment and, therefore, a modified design criterion was proposed in this dissertation, where the expected information content of the newly designed experiment is considered, even before the experiment is performed. Hence, this approach was called the anticipatory approach to optimal experimental design for model discrimination.

After applying the approaches to design optimal discriminatory experiments to a case study, one could conclude that the anticipatory approach performed better than the original approach of Buzzi-Ferraris et al. (1984). The reason why the anticipatory approach performs better than the original approach of Buzzi-Ferraris et al. (1984) is related to the uncertainty on the parameter estimates, and can be explained by the similarity between the anticipatory approach and optimal experimental design for parameter estimation (both design criteria benefit from a larger information content of the designed experiment with regard to the model parameters).

It is clear that high model prediction uncertainties hamper the efficacy and efficiency of the model discrimination procedure. These model prediction uncertainties are determined by the quality of the available data, since low quality data will result in poorly estimated parameters, which in turn result in uncertain model predictions. The discrimination among

several rival models may thus become more efficient and effective if this uncertainty could be reduced prior to the start of the model discrimination procedure. Reducing the uncertainty on the model predictions can be achieved by designing and performing experiments dedicated to reducing the uncertainty on the parameter estimates. However, performing an additional experiment for each rival model may undermine the overall goal of optimal experimental design, since this would require at least as many experiments as the number of rival models. Therefore, the possibility to design a so-called compromise experiment was investigated. Such a compromise experiment is defined as an experiment that is not optimal for one or more of the individual rival models, but is sufficiently informative to improve the overall precision of the parameters of all rival models.

To design compromise experiments, two methods were developed: the kernel-based method and the ideal point method. The kernel-based method was developed to design a compromise experiment for experimental design problems where only the sampling times are to be optimized. Because the use of this method is limited to such experimental design problems, the idea of designing a compromise experiment was further explored by treating it as a multi-objective problem. As a result, the so-called ideal point method was proposed, which can be used for experimental design problems with experimental degrees of freedom of any type (manipulations, initial conditions and sampling times).

Finally, three approaches to integrate optimal experimental design for parameter estimation and model discrimination were investigated. In a first procedure, both aspects are dealt with sequentially, that is, the model discrimination procedure is performed first, and then the parameters of the selected model are further refined through the design of optimally informative experiments. The second procedure, is similar, except that a compromise experiment is designed and performed prior to the start of the model discrimination procedure to improve the quality of the parameter estimates. In the third procedure, both issues are dealt with simultaneously. For this purpose, the joint design criterion proposed by Hill et al. (1968) is modified such that the anticipatory approach can be used to quantify the discriminatory potential of the proposed experiments. The results obtained after applying the three procedures to a case study showed that, although model discrimination was not achieved in less experiments compared to the sequential procedure without a compromise experiment, the quality of the parameter estimates improved faster when a compromise experiment was performed first. The performance of the simultaneous procedure appeared to be worse than the performance of the sequential procedures, because the wrong model was identified as the most appropriate more frequently than with the sequential procedures.

In addition, the results confirmed the similarity between the anticipatory approach to design optimal discriminatory experiments and optimal experimental design for parameter estimation.

# Samenvatting

Een wiskundig model kan gedefinieerd worden als een mathematische beschrijving van de mechanismen die het gedrag van een bepaald proces bepalen, en het doel van een modelleeroefening is het identificeren van een wiskundig model dat in staat is om het bestudeerde proces adequaat te beschrijven en zelfs te voorspellen. Eens dergelijk model is geïdentificeerd, vormt het een krachtig hulpmiddel voor wetenschappers en ingenieurs. Het is echter van groot belang te beseffen dat het inzicht in het gemodelleerde proces vaak onvolledig is en dat dit leidt tot het voorstellen van verschillende, rivaliserende modellen voor eenzelfde proces. Elk van deze modellen is de wiskundige vertaling van een bepaalde hypothese over hoe het proces werkt. Het spreekt voor zich dat men vooral geïnteresseerd is in dat model dat het verloop van het proces het beste beschrijft. De zoektocht naar het beste model uit een set van rivaliserende modellen wordt dikwijls het modeldiscriminatieprobleem genoemd en werd in dit doctoraat bestudeerd.

Een algemene procedure om dit modeldiscriminatieprobleem aan te pakken bestaat uit vier stappen die op iteratieve wijze worden uitgevoerd tot aan een stopcriterium is voldaan. In de eerste stap worden de parameters van de rivaliserende modellen geschat op basis van alle experimentele data die op dat ogenblik voor handen zijn. In een tweede stap wordt nagegaan welke van de rivaliserende modellen (nog) in staat zijn om de beschikbare data voldoende adequaat te beschrijven. De modellen die adequaat werden bevonden, worden gebruikt in de derde stap waarin een optimaal discriminerend experiment wordt ontworpen. Vervolgens, wordt dit experiment uitgevoerd (vierde en laatste stap) en wordt de lus gesloten door de parameters van alle rivaliserende modellen opnieuw te schatten.

Om het beste model te identificeren uit een set van rivaliserende modellen is het dus nodig om nieuwe informatie te verzamelen over het bestudeerde proces en moeten nieuwe experimenten worden uitgevoerd. Aangezien dit dikwijls veel tijd en geld kost, is het belangrijk deze experimenten zodanig te ontwerpen dat het aantal bijkomende experimenten tot

het minimum beperkt wordt. Om modeldiscriminatie te verwezenlijken werden reeds een aantal ontwerpcriteria ontwikkeld. In wezen komt het ontwerp van een optimaal discriminerend experiment neer op het zoeken naar dat experiment dat zo verschillend mogelijk wordt voorspeld door de rivaliserende modellen, rekening houdende met de onzekerheid die hiermee gepaard gaat. Inderdaad, bij het ontwerpen van het experiment wordt verondersteld dat een van de rivaliserende modellen het *echte* model is en dat de uitkomst van het experiment dus kan voorspeld worden door dit model. Onder deze veronderstelling, verwacht men dat dit model zal kunnen worden geïdentificeerd als het experiment anders wordt voorspeld door de andere, rivaliserende modellen.

De onzekerheid die gerelateeerd is aan het verschil tussen de modelvoorspellingen wordt bepaald door de onzekerheid op de experimentele data enerzijds, en door de onzekerheid op de modelvoorspellingen anderzijds. Inderdaad, de onzekerheid op de experimentele data moet in rekening gebracht worden bij het ontwerpen van een optimaal discriminerend experiment, odat dit kan gezien worden als een maat voor de reproduceerbaarheid ervan. Maar ook de onzekerheid op de modelvoorspellingen is van belang aangezien het evalueren van het discriminerend vermogen van een experiment gebeurt op basis van deze modelvoorspellingen.

Deze aspecten zijn vervat in het ontwerpcriterium dat werd ontwikkeld door Buzzi-Ferraris et al. (1984) en dit ontwerpcriterium werd als vertrekpunt genomen in dit onderzoek. Bij het evalueren van het discriminerend vermogen van een bepaald experiment gebruikt dit criterium echter enkel de informatie van de op dat ogenblik reeds uitgevoerde experimenten. Vandaar dat dit ontwerpcriterium werd aangepast zodat de verwachte informatie-inhoud van het ontworpen experiment reeds in rekening kon worden gebracht voordat het effectief wordt uitgevoerd. Vandaar dat deze aanpak de anticiperende aanpak voor het ontwerpen van optimaal discriminerende experimenten werd genoemd.

Na het toepassen van de methoden voor het ontwerpen van optimaal discriminerende experimenten, kon geconcludeerd worden dat de anticiperende aanpak beter werkt dan de originele aanpak van Buzzi-Ferraris et al. (1984). Een verklaring hiervoor is gerelateeerd aan de onzekerheid op de parameterschattingen, en kan worden verklaard door de analogie tussen de anticiperende aanpak en de methoden om experimenten te ontwerpen waarvan verwacht wordt dat die een maximum aan informatie over de parameters zullen opleveren. Beide methoden hebben er namelijk voordeel bij dat het ontworpen experiment informatierijk is met betrekking tot de modelparameters.

Het is duidelijk dat een grote onzekerheid op de modelvoorspellingen niet ten goede komt aan de effectiviteit en de efficiëntie van de modeldiscriminatieprocedure. Nu is het zo dat de onzekerheid op de modelvoorspellingen wordt bepaald door de kwaliteit van de beschikbare data, aangezien data van lage kwaliteit zullen resulteren in slecht geschatte parameters die op hun beurt zullen leiden tot onbetrouwbare modelvoorspellingen. Het lijkt dus mogelijk om de modeldiscriminatie tussen verschillende rivaliserende modellen efficienter te maken door deze onzekerheid te reduceren alvorens de modeldiscriminatieprocedure te beginnen. Een reductie van de onzekerheid op de modelvoorspellingen kan worden bekomen door experimenten uit te voeren die ontworpen zijn met als doel de onzekerheid op de parameterschattingen te verminderen. Het uitvoeren van dergelijk experiment voor elk van de rivaliserende modellen kan echter het ultieme doel van optimale proefopzet ondermijnen aangezien minstens zoveel experimenten nodig zijn als er rivaliserende modellen zijn. Vandaar dat in dit doctoraat de mogelijkheid werd onderzocht om een zogenaamd compromis-experiment te ontwerpen. Dergelijk compromis-experiment werd gedefinieerd als een experiment dat niet optimaal is voor een of meer individuele modellen, maar voldoende informatierijk om de globale precisie van de parameterschattingen van alle rivaliserende modellen te verbeteren.

Om dergelijke compromis-experimenten te ontwerpen werden twee methoden ontwikkeld: de *kernel*-gebaseerde methode en de ideaalpuntmethode. De *kernel*-gebaseerde methode werd ontwikkeld om compromis-experimenten te ontwerpen voor problemen waarbij enkel de meettijdstippen worden geoptimaliseerd. Aangezien het toepassingsgebied van deze methode beperkt is, werd het idee van het compromis-experiment verder verkend door het te benaderen als een multi-objectief probleem. De ideaalpuntmethode werd voorgesteld en kan worden gebruikt voor problemen met experimentele vrijheidsgraden van alle types (manipulaties, initiële condities en meettijdstippen).

Tot slot werden drie verschillende procedures onderzocht om het ontwerp van experimenten voor precieze parameterschattingen en modeldiscriminatie te integreren. In een eerst procedure worden beide aspecten sequentieel behandeld. De modeldiscriminatieprocedure wordt eerst uitgevoerd, en vervolgens wordt de precisie van de parameterschattingen verbeterd door het uitvoeren van gerichtte experimenten. Een tweede procedure is gelijkaardig, behalve dat een compromis-experiment wordt uitgevoerd alvorens de modeldiscriminatieprocedure te starten. In een derde procedure worden beide aspecten simultaan behandeld. Hiervoor werd het ontwerpcriterium ontwikkeld door Hill et al. (1968) aangepast zodat de anticiperende aanpak kon worden gebruikt voor het kwantificeren van het discriminerend

potentieel van de voorgestelde experimenten. Verder werden deze procedures toegepast in een gevalstudie waarin duidelijk werd dat het gebruik van een compromis-experiment nuttig is. Hoewel er evenveel experimenten nodig waren om modeldiscriminatie te bekomen, verbeterde de kwaliteit van de parameterschattingen sneller dan met de procedure zonder compromis-experiment. De performantie van de simultane procedure was minder goed dan de performantie van de twee sequentiele procedures omdat het verkeerde model frequenter werd geselecteerd als beste model.