

Optimal Design of Dynamic Experiments in Bioscience Engineering

Arno Strouwen

Supervisors:

Prof. dr. Peter Goos

Prof. dr. ir. Bart Nicolaï

Dissertation presented in partial
fulfillment of the requirements for the
degree of Doctor of Bioscience
Engineering (PhD)

September 2021

Optimal Design of Dynamic Experiments in Bioscience Engineering

Arno STROUWEN

Examination committee:

Prof. dr. ir. René De Mot, chair

Prof. dr. Peter Goos, supervisor

Prof. dr. ir. Bart Nicolaï, supervisor

Prof. dr. ir. Kristel Bernaerts

Prof. dr. ir. Wouter Saeys

Prof. dr. Julio Banga

(Instituto de Investigaciones Mariñas-
Consejo Superior de Investigaciones Científicas)

Dr. ir. Philippe Nimmemeers
(Universiteit Antwerpen)

Dissertation presented in partial
fulfillment of the requirements for
the degree of Doctor of Bioscience
Engineering (PhD)

September 2021

© 2021 KU Leuven – Faculty of Bioscience Engineering
Uitgegeven in eigen beheer, Arno Strouwen, Kasteelpark Arenberg 30, B-3001 Leuven (Belgium)

Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt worden door middel van druk, fotokopie, microfilm, elektronisch of op welke andere wijze ook zonder voorafgaande schriftelijke toestemming van de uitgever.

All rights reserved. No part of the publication may be reproduced in any form by print, photoprint, microfilm, electronic or any other means without written permission from the publisher.

Preface

During my bachelor studies as a bio-science engineer at KU Leuven, I followed a course on optimal experimental design by Professor Peter Goos. In this course, it became clear to me that statistical modeling is essential to properly describe living systems, due to the inherently high variability of such systems. It also became clear that carefully planned experiments are necessary to cope with this variability. The knowledge I gained from this course made me quite popular among my fellow bio-science engineering students, since the informative cost- and time-efficient experiments I planned for them saved them multiple days (and nights!) of lab work.

Thus, after completing my master studies, I knew immediately which direction I wanted to give to my further career. I e-mailed Professor Goos about a PhD in his research group. To my surprise, Professor Goos remembered me as the only student in his academic career who did not fail his exam, but nevertheless came to the ask feedback on their exam. Apparently, this left a good impression and I could soon start working on the thesis you are now reading. The first project Professor Goos suggested to me involved choice experiments with mixture constraints. I was already quite interested in this topic and was ready to agree to it, but, mostly out of curiosity, I asked whether there existed experimental design research about dynamic systems, since my master thesis focused heavily on dynamic systems and I had also considered doing a PhD in control theory. Here, I must say that the stars aligned as Professor Goos was just getting involved in a project about the metabolism of pear during hypoxia (together with Professors Bart Nicolaï and Wouter Saey), which required dynamic experiments. I was truly lucky to be able to combine my two research interests, control theory and experimental design, and I cannot imagine a research topic that would have interested me more. And thus it came to be that my thesis would be about optimal experimental design for dynamic systems, applied to the storage of pear fruit, with Professor Goos advising me on experimental design and Professor Nicolaï advising me on postharvest

modeling. Peter, I must thank you for teaching me to write a sound scientific manuscript. I am sitting besides a stack of all the drafts of our papers and they easily tower above me. Bart, your practical knowledge of postharvest models always rescued me when I got stuck in my PhD project.

Only a couple of months into my PhD, I was thrown to the lions of the Research Foundation - Flanders (FWO) strategic basic research panel. Thankfully, I made it out alive, and I am very grateful that the FWO believed in me and decided to fund my research for four years.

I want to thank the members of my supervisory comity, Professors Kristel Bernaerts and Wouter Saeys, for the constructive feedback throughout my PhD project. In particular, the PhD thesis of Professor Bernaerts and her many papers were a great starting point for my own research.

I also want to express my gratitude towards Professor Julio Banga and Doctor Philippe Nimmemeers for agreeing to be part of my jury.

Professor Banga is further thanked for agreeing to a research visit of mine to his research group in Vigo. I learned a lot about sensitivity analysis of dynamic systems and global optimization during this visit.

I also want to thank the first master thesis student I supervised, Karel Van Brantegem. Karel, you wrote an excellent master thesis and I am glad that you decided to join our research group. As you know by now, I have more research ideas than I could conceivable finish myself. So, it was really helpful to be able to collaborate with you to tackle some of them.

The last person I must thank on the academic side of things is somebody I have only communicated with in text, but was nevertheless instrumental in bringing this thesis to a successful end. This is Doctor Christopher Rackauckas, without whose scientific machine learning project, the text of this thesis would probably be twice as long, and the computer source code, upon which it is based, would be four times longer.

In the year 1993 my grandparents retired, but instead of enjoying their well-deserved free time, they got a new job. A baby was born in their family and the parents had very demanding careers. Bomma and Bompa without your love and care I would never have achieved half as much.

Arno,
Hasselt 2021

Abstract

Living systems are rife with variability. Well-chosen experimental designs are necessary to deal with this variability when modeling such systems. Models for living systems often rely on knowledge of physical, chemical and biological laws, such as mass balances, transport phenomena and reaction kinetics, and are often described by a system of non-linear differential equations. So, the structure of a model can often be determined from first principles. However, the model will generally also rely on parameters whose numerical values cannot be determined from physical laws. These parameters must then be inferred from experimental data before the model can be put to use. A well-chosen experiment can greatly improve the estimation of the model parameters. However, there exist several challenges for constructing such informative experiments for dynamic systems. One challenge is the dependence of the optimal experiment on the true model parameters, making it difficult to perform robust experiments that work well regardless of the specific model parameter values. Another challenge is the correlation of the observations due to the presence of process noise. The central research topic of this thesis revolves around solving these challenges by developing robust experimental design methodology for noisy dynamic systems. My novel methodology is developed to improve postharvest and other bio-science engineering applications, and is mainly applied to the estimation of respiration and fermentation parameters of pear fruit. Experiments that precisely estimate the model parameters of spring-mass-damper systems and compartment systems are also constructed.

After a brief introduction on dynamic design of experiments and its application areas, this PhD thesis deals with the estimation of respiration parameters of pear fruit inside a jar, modeled by Michaelis-Menten kinetics. Air flowing into the jar has to be controlled so that the parameters of the Michaelis-Menten model can be estimated as precisely as possible. The quality of this parameter estimation, and thus of the experimental design, is quantified using the determinant of the Fisher information matrix, which is inversely related to the area of the confidence

ellipse of the two Michaelis-Menten model parameters. The air flowing into the jar must thus be optimized so that the determinant of the Fisher information matrix is as large as possible. One major challenge here is the dependence of this Fisher information matrix on the true, but unknown, parameters of the system. The most commonly used method in the literature to deal with this issue is locally optimal design, where a single initial guess is used for the true parameter. One main conclusion from my work is that this optimal experimental design technique outperforms several commonly used heuristic experimental techniques.

Subsequently, also fermentation of the pear fruit is taken into account. The locally optimal design method only takes into account a single initial guess, and may not perform well if this guess deviates substantially from the true parameter values. Instead of a single initial guess, an entire distribution could be used to quantify the prior knowledge and uncertainty on the unknown parameters. Because of the use of a prior distribution, this method is called Bayesian experimental design. Most current techniques in the literature only allow for parametric prior distributions, such as normal distributions. The prior information about respiration and fermentation, coming from a previously gathered dataset, could not be summarized by any parametric distribution. For this reason, I developed a novel experimental design technique based on a Markov-chain Monte-Carlo (MCMC) analysis of this previously gathered data. This method is thus able to approximate arbitrary distributions. I found that this flexible experimental design technique is more robust than the commonly used locally optimal design method and other robust methods.

The final part of the thesis focuses on robust and adaptive experimental design techniques for dynamic systems with process noise. Current experimental design techniques for dynamic systems generally only incorporate measurement noise, but biological systems also often involve process noise. Calculating the Fisher information matrix for such systems requires estimating the uncertain dynamic states, using Bayesian filtering techniques. For linear dynamical systems, the optimal filter is the Kalman filter. However, deriving the Fisher information matrix for dynamic systems under process noise and then applying the methodology from the previous chapters is not sufficient to construct informative experiments. This is due to the difficulty in precisely predicting such systems far into the future, which causes those future measurements to contribute little to the Fisher information matrix. Adaptive experimental designs are able to deal with this issue. Adaptive designs use the already gathered data to re-optimize and thus adapt the remainder of the experiment. The already gathered measurements can help by reducing the uncertainty on the model parameters, which the optimal design depends upon. Adaptivity is thus always a good design strategy, even when no process noise is present.

But for systems with process noise, adaptivity is even more important, since the already gathered measurements help increase the prediction accuracy of future measurements, thus increasing the informativity of these measurements. I found that taking into account process noise in the experimental design greatly improves the quality of the experiment.

Beknopte Samenvatting

Levende systemen worden gekenmerkt door een grote maat aan variabiliteit. Goed gekozen experimentele ontwerpen zijn noodzakelijk om goed om te gaan met deze variabiliteit. Modellen van levende systemen steunen vaak op kennis van fysieke, chemische of biologische wetten, zoals massabalansen, transportverschijnselen en reactiekinetiek, en worden vaak beschreven door een systeem van niet-lineaire differentiaalvergelijkingen. Dus, de structuur van het model kan vaak bepaald worden door eerste principes. Echter, het model zal in het algemeen ook afhangen van parameters waarvan de numerieke waarden niet kunnen vastgesteld worden door fysieke wetten. Deze parameters moeten dan afgeleid worden uit experimentele data voordat het model toegepast kan worden. Een goed gekozen experiment kan de schatting van deze parameters sterk verbeteren. Er bestaan echter verschillende uitdagingen bij het opstellen van zo een informatief experiment voor dynamische systemen. Een eerste uitdaging is de afhankelijkheid van het optimale experiment van de waarden van de modelparameters, wat het moeilijk maakt om robuuste experimenten op te zetten die goed werken voor om het even welke modelparameterwaarden. Een andere uitdaging is de correlatie tussen de observaties door de aanwezigheid van procesruis. Het centrale onderzoeksthema van deze thesis draait rond het aanpakken van deze uitdagingen door het ontwerpen van een robuuste methodologie voor experimenteel ontwerp voor dynamische systemen met ruis. Mijn nieuwe methodologie is ontworpen om naoogsttoepassingen en andere toepassingen in het domein van de bio-ingenieurswetenschappen te verbeteren, en is voornamelijk gericht op de schatting van respiratie- en fermentatieparameters van peren. Experimenten die toelaten om de modelparameters van massa-veer-demper systemen en compartiment systemen precies te schatten worden evenwel ook opgesteld.

Na een korte inleiding over experimenteel ontwerp voor dynamische systemen en de toepassingsgebieden ervan, behandelt deze doctoraatsthesis het schatten van respiratieparameters van peer in een pot, gemodelleerd met behulp van

Michaelis-Menten kinetiek. Lucht die de pot instroomt kan aangestuurd worden zodat de parameters van het Michaelis-Menten model zo precies mogelijk kunnen geschat worden. De kwaliteit van deze parameterschatting, en dus ook het experimenteel ontwerp, wordt gekwantificeerd door middel van de determinant van de Fisher informatiematrix, die omgekeerd gerelateerd is aan de betrouwbaarheidsellips van de twee Michaelis-Menten modelparameters. Lucht die de pot instroomt moet geoptimaliseerd worden zodat de determinant van de Fisher informatiematrix zo groot mogelijk is. Een grote uitdaging is het feit dat deze Fisher informatiematrix afhangt van de (onbekende) modelparameters. De meest gebruikte techniek in de literatuur voor het omgaan met deze kwestie is lokaal optimaal ontwerp, waarbij een enkele initiële gok wordt gebruikt voor de echte parameters. De belangrijkste conclusie van mijn werk is dat dit optimaal experimenteel ontwerp verscheidene vaak gebruikte heuristische methodes van experimenteel ontwerp overtreft.

Vervolgens brengt de thesis ook de fermentatie van peerfruit in rekening. Het lokaal optimaal ontwerp houdt alleen rekening met een enkele initiële gok, en presteert doorgaans niet goed als deze gok substantieel afwijkt van de echte parameterwaarden. In de plaats van een enkele initiële gok kan een ganse kansverdeling gebruikt worden om de a priori kennis en onzekerheid omtrent de modelparameters te kwantificeren. Omwille van het gebruik van een a priori verdeling wordt deze methode Bayesiaans experimenteel ontwerp genoemd. De meeste huidige technieken in de literatuur staan enkel parametrische a priori verdelingen toe. De a priori informatie over respiratie en fermentatie, waarover ik beschikte, kon niet samengevat worden door een parametrische kansverdeling. Om deze reden ontwierp ik een nieuwe techniek gebaseerd op een Markov-chain Monte-Carlo analyse van deze voorheen verzamelde data. Deze methode kan arbitraire kansverdelingen benaderen. Ik ontdekte dat deze flexibele techniek robuuster is dan de vaak gebruikte techniek van lokaal optimaal ontwerp, alsook andere robuuste methodes.

Het laatste deel van de thesis richt zich op robuuste en adaptieve technieken voor experimenteel ontwerp voor dynamische systemen met procesruis. De huidige technieken voor dynamische systemen kunnen enkel omgaan met meetruis, maar procesruis komt ook vaak voor in biologische systemen. Het berekenen van de Fisher informatiematrix voor dergelijke systemen vereist het schatten van de onzekere dynamische toestanden met behulp van Bayesiaanse filtertechnieken. Voor lineaire dynamische systemen is de optimale filter de Kalman filter. Echter, het afleiden van de Fisher informatiematrix voor dynamische systemen onder procesruis en vervolgens de methodologie uit de voorgaande hoofdstukken toepassen volstaan niet om informatieve experimenten op te stellen. Dit is te wijten aan de moeilijkheid om dergelijke systemen ver in de toekomst nauwkeurig

te voorspellen, waardoor metingen ver in de toekomst weinig bijdragen aan de Fisher informatiematrix. Adaptieve experimentele ontwerpen zijn in staat om dit probleem te remediëren. Adaptieve ontwerpen gebruiken de reeds verzamelde gegevens om de rest van het experiment opnieuw te optimaliseren en dus aan te passen aan de vrijgekomen informatie. De reeds verzamelde metingen helpen de onzekerheid te verminderen omtrent de modelparameters, waarvan het optimale ontwerp afhankelijk is. Adaptiviteit is dus altijd een goede ontwerpstrategie, ook als er geen procesruis aanwezig is. Maar voor systemen met procesruis is adaptiviteit nog belangrijker dan voor systemen zonder procesruis, omdat de reeds verzamelde metingen de nauwkeurigheid van de voorspellingen van toekomstige metingen helpen te vergroten, waardoor de informativiteit van deze metingen toeneemt. In mijn derde hoofdstuk demonstreerde ik dat het beschouwen van procesruis in het experimentele ontwerp de kwaliteit van het experiment aanzienlijk verbetert.

List of Abbreviations

BLUE best linear unbiased estimator

DCA dynamic controlled atmosphere

FIM Fisher information matrix. When used without further specification, refers to the expected Fisher information matrix, as opposed to the observed Fisher information matrix.

KL-div Kullback-Leibler divergence

MAP modified atmosphere packaging

MCMC Markov-chain Monte-Carlo

SSE sum of squared errors

Contents

Contents	xi
List of Figures	xiii
List of Tables	xv
1 Introduction	1
1.1 Motivation	1
1.2 Goals	2
1.3 Chapter by Chapter Overview	3
2 Optimizing Oxygen Input Profiles for Efficient Estimation of Michaelis-Menten Respiration Models	5
2.1 Introduction	6
2.2 Methods	8
2.2.1 Dynamic Model for Respiration	8
2.2.2 Optimal Experimental Design Methodology	10
2.2.3 Information Optimization	12
2.3 Results	15
2.3.1 Proof of Concept Example	15
2.3.2 Refining the Design Space	20
2.4 Discussion	23
2.5 Conclusion	24
3 Robust Dynamic Experiments for the Precise Estimation of Respiration and Fermentation Parameters of Fruit and Vegetables	25
3.1 Introduction	26
3.2 Bayesian Optimal Experimental Design for Dynamical Systems	28
3.2.1 Dynamic Models	29
3.2.2 Information Content of an Experiment	29
3.2.3 Discretizing the Controls	30

3.2.4	Robustifying the Experiment	30
3.2.5	Numerical Details	31
3.3	Respiration and Fermentation Model of Pear Fruit	32
3.3.1	Model Description	33
3.3.2	Prior Information	35
3.3.3	Markov chain Details	38
3.4	Results	39
3.4.1	Locally Optimal Designs for the Respiration and Fermentation Model	39
3.4.2	Bayesian Optimal Designs for the Respiration and Fermentation Model	43
3.5	Discussion and Future Work	46
3.5.1	Alternative Information Criteria	46
3.5.2	State Discretization	47
3.5.3	Reverse Automatic Differentiation	47
3.6	Conclusion	48
4	Adaptive and Robust Experimental Design for Linear Dynamic Models using the Kalman Filter	49
4.1	Introduction	50
4.2	Modeling the Information Content of a Dynamic Experiment	52
4.2.1	The Model	52
4.2.2	Parameter Estimation	53
4.2.3	The Fisher Information Matrix	55
4.3	D-optimal Experimental Design	58
4.3.1	Locally Optimal Experimental Design	58
4.3.2	Robust Experiments	58
4.3.3	Adaptive Experiments	59
4.4	Numerical details	63
4.5	Case Studies	64
4.5.1	Mass-Spring-Damper System	64
4.6	Two Compartment System	71
4.7	Discussion and Conclusion	72
5	Conclusion and Outlook	74
5.1	Overarching Conclusion	74
5.2	Outlook	76
5.2.1	Model Selection	76
5.2.2	Process Noise	77
5.2.3	Kullback-Leibler Divergence	78
5.2.4	Koopman Expectation	78
5.2.5	Fully Risk Averse Experimental Design	78
5.2.6	Goal Driven Experimental Design	79
A	Markov chain details	81

List of Figures

2.1	Graphical representation of the respiration jar measurement setup.	9
2.2	Initial coordinate-exchange algorithm.	13
2.3	Modified coordinate-exchange algorithm.	14
2.4	Optimal design, assuming $l = 2$ and $n = 12$: pulse input after 6 h.	16
2.5	SSE contours for the optimal design in Figure 2.4a.	16
2.6	First benchmark: an always-on input profile.	18
2.7	Second benchmark: an always-off input profile.	19
2.8	Third benchmark: a constantly switching input profile.	19
2.9	Fourth benchmark: pulse input after 4 h.	19
2.10	Optimal design, assuming $l = 2$ and $n = 144$: two short pulses during the second half of the experiment.	21
2.11	Optimal design, assuming $l = 11$ and $n = 12$: two small pulses after 10 h and 16 h.	22
2.12	SSE contours for the optimal experimental designs in Figure 2.10a and 2.11a.	22
3.1	Schematic representation of the measurement setup.	33
3.2	Summary of the Bayesian data analysis of the data in Ho, Verboven, Verlinden, Lammertyn, et al. (2008) for the parameters of interest. This figure is stretched over two pages. The diagonal contains histograms of the Markov chain values of the six model parameters of interest. The orange dots denote the means. Above the diagonal, two-dimensional heatmaps are shown, visualizing the correlation between the Markov chain values for pairs of model parameters. Similarly, below the diagonal, scatter plots for 100 Markov chain points are shown.	37
3.3	Summary of our Markov-chain Monte-Carlo analysis of the data in Ho, Verboven, Verlinden, Lammertyn, et al. (2008) for the nuisance parameters E_{a,O_2} , E_{a,CO_2} and σ^2 .	38
3.4	Convergence of the local D-optimality criterion for finer discretizations of the controls.	40

3.5	Locally optimal experiment and simulated output, with control input switches every two hours.	41
3.6	Summary of the information gained from the locally optimal experiment in Figure 3.5, with local D-criterion equal to 8.67×10^8	42
3.7	95% confidence ellipsoids comparing the locally optimal experimental design (red) and the heuristic experimental design technique from Ho, Verboven, Verlinden, Lammertyn, et al. (2008) (blue).	42
3.8	Robust optimal experiment and simulated output, obtained using 100 parameter values from the Markov chain and control input switching every two hours.	44
3.9	Histogram of difference in determinants of the FIM of the robust and locally optimal designs for 6000 parameters from the prior distribution. Positive values mean the robust design performs better.	45
4.1	Schematic representation of mass spring damper system.	65
4.2	Comparison of optimal inputs compared to random inputs and the corresponding output behaviors.	66
4.3	Online maximum likelihood estimates for both the optimal experiment and the random experiment. The optimal experiment converges faster to the true parameters.	67
4.4	Likelihood grids at the end of the experiment. The likelihood decreases sharply away from the true parameters for the optimal experiment, unlike in the random experiment. The maximum likelihood estimate of the optimal experiment is much closer to the true value than the random experiment.	67
4.5	Mean and variance of online maximum likelihood estimates for different combinations of e , N and T in Algorithm 1.	70
4.6	Online maximum likelihood estimates for the two compartment model.	72

List of Tables

2.1	D-criterion values for both the optimal design as well as several benchmarks, assuming $l = 2$ and $n = 12$. These designs are shown in Figures 2.4 — 2.9.	17
2.2	D-criterion values for optimal experimental designs with different parametrizations in Sections 2.3.2 and 2.3.2. These designs are shown in Figures 2.4, 2.10 and 2.11.	20
3.1	Parameters used for the optimization of the construction of the locally optimal designs.	39
3.2	Parameters used for the optimization of the construction of the locally optimal designs	46
A.1	100 Markov-chain points used to generate robust designs in Chapter 3	82

Chapter 1

Introduction

1.1 Motivation

Experimentation is required in almost all areas of bio-science engineering. A well chosen experimental design can tremendously reduce this required experimental effort. One area of bio-science engineering that requires a lot of experimentation is the postharvest storage of fruit and vegetables.

Fresh fruit and vegetables are perishable and need to be stored at appropriate conditions after their harvest. The temperature is typically set as low as possible to reduce respiration, but above the freezing point of the product. Below this point, massive cell damage occurs and the product no longer can be considered as fresh. A further reduction of the respiration rate is possible by reducing the O₂ partial pressure of the storage atmosphere and increasing its CO₂ partial pressure.

The optimal storage conditions (temperature, O₂ and CO₂ partial pressures) of fruit and vegetables depend on the species, cultivar, ripeness stage and many other factors and must be experimentally determined. Traditionally, this has been achieved by storing the product at many combinations of temperature, O₂ partial pressure and CO₂ partial pressure, and monitoring the change in quality attributes during the storage period. This can be as much as one year for apple and pear fruit.

As an alternative, a mathematical model based approach could be used to optimize storage protocols. This approach relies on comprehensive mathematical

models that describe the behavior of the product as a dynamical system with inputs (temperature, O₂ partial pressure and CO₂ partial pressure) and outputs (respiration and fermentation rate, quality attributes), based on physical, chemical and biological laws. These mathematical models rely on parameters whose numerical values cannot be determined from first laws, and must be experimentally determined. The experimental effort then shifts from parameter estimation for black box response surface models to parameter estimation for mechanistically based differential equation models. We then no longer have to independently test the fruit and vegetables at different storage conditions via a factorial design. Instead, one single experiment involving time varying O₂ and CO₂ partial pressures would allow us to estimate the respiration and fermentation parameters of the differential equation models. The latter can subsequently be used to simulate the behavior of the fruit in arbitrary storage conditions, and the optimization of the storage process can then be done in silico (Ho, Verboven, Verlinden, Herremans, et al. 2011). This shift towards mathematical models also allows us to use model based experimental design methods (Franceschini and Macchietto 2008). The challenge is to construct optimal time-varying O₂ and CO₂ partial pressures such that the respiration and fermentation parameters can be precisely estimated in a single experiment.

Besides improving postharvest experimentation, the methods we derive in this thesis are much more broadly applicable and can be used to design experiments for the precise estimation of model parameters in many differential equation models in other fields of bio-science engineering and engineering.

1.2 Goals

The general goal of this thesis is to develop the statistical methodology required to create dynamic inputs that lead to precise parameter estimation for non-linear differential equation models. In Chapters 2 and 3, we apply this to the estimation of respiration and fermentation parameters of pear fruit. In Chapter 4, experiments to estimate the model parameters of spring-mass-damper systems as well as compartment systems are constructed. In doing so, several challenges must be overcome:

- **Challenge 1.** Constructing optimal experiments for non-linear dynamic systems is a circular problem, as the best experiment to learn about the model parameters depends on the specific values of those parameters. Some form of prior information about the model parameters is required to solve this issue.

- **Challenge 2.** If the prior information required in **Challenge 1** were perfect, then no further experiments would be needed. Generally, however, there is substantial uncertainty concerning the prior information. Therefore, a robust experimental design is desired that performs well over a range of possible parameter values. Current robust experimental design methods are only capable of achieving robustness when this prior information is summarized well with a parametric distribution. However, not all prior information can easily be summarized in such a way. To overcome this challenge, a non-parametric method should be developed.
- **Challenge 3.** Current experimental design methodology ignores process noise. Often, all errors are attributed to measurement noise. In the presence of process noise, the challenge lies in generalizing the definition of the Fisher information matrix to deal with observations that are correlated in time.
- **Challenge 4.** Adapting the experimental design using information from measurements as they are being gathered is another way to deal with the dependence of the optimal experiment on the true parameters. Current experimental design methodology, in the context of differential equation models in combination with process noise, is not suitable for online computation. This is because of the growing complexity of the optimization problem over time, particularly when also dealing with process noise. An alternative experimental design method with constant memory and time complexity therefore has to be developed.

1.3 Chapter by Chapter Overview

- **Chapter 2** uses the traditional locally optimal experimental design method to optimize oxygen input profiles, which lead to a precise estimation of respiration parameters of pear fruit. This method deals with the dependence of these input profiles on the true respiration parameters, described in **Challenge 1**, by requiring an initial guess for these parameters. This initial guess is provided by a parameter estimate from the literature.
- **Chapter 3** addresses **Challenge 2** by developing a robust experimental design method involving a quantification of the uncertainty about the model parameters by applying a Markov-chain Monte-Carlo technique to a prior data-set. This robust experimental design method is then used to jointly optimize both O₂ and CO₂ input profiles to precisely estimate both respiration and fermentation parameters of pear fruit.

- **Chapter 4** uses the Kalman filter to deal with **Challenge 3** of experimental design under process noise. This filter is used to track an estimate of the hidden states over time. This estimate is required to calculate the likelihood of the model parameters, which in turn is required to calculate the Fisher information matrix for these more complicated dynamic systems with process noise. The state estimate is also required for **Challenge 4** concerning adaptive designs, as we need information about the true state of the system when redesigning the remainder of the experiment. The issue of growing complexity of the optimization problem over time is dealt with by using a combination of the observed and expected Fisher information matrix.
- **Chapter 5** presents an overarching conclusion of the work done in my thesis. This chapter also gives an overview of alternative or related experimental design strategies for dynamic systems, as well as an overview of numerical methods that might improve the speed or precision with which an optimized experiment can be constructed. I also discuss future research prospects that might grow out of the techniques described in this thesis.

Chapter 2

Optimizing Oxygen Input Profiles for Efficient Estimation of Michaelis-Menten Respiration Models

This chapter has been published in Food and Bioprocess Technology.¹

Source code available upon request.

Abstract

Models based on mass balances and Michaelis-Menten respiration kinetics are increasingly used to determine optimal storage conditions of fresh fruits and vegetables. The model parameters are usually estimated from respiration experiments at different, but fixed, gas conditions according to a response surface design. This is a tedious procedure that requires a gas mixing facility or a series

¹Strouwen, A., Nicolai, B.M., Goos, P. (2019). Optimizing Oxygen Input Profiles for Efficient Estimation of Michaelis-Menten Respiration Models. Food and Bioprocess Technology, 12 (5), 769-780.

of gas cylinders with appropriate composition. In this chapter, we consider a simpler approach, in which the respiration kinetics of pear fruit are modeled using a single experiment with a time varying O₂ input profile. To optimize the information content produced by the O₂ profile, we apply optimal dynamic experimental design principles and present a modified coordinate-exchange algorithm to achieve this goal. Finally, we demonstrate the added value of our approach by comparing the optimal O₂ input profiles to several intuitive benchmark experiments.

2.1 Introduction

Fresh fruit needs to be stored at appropriate temperature as well as O₂ and CO₂ conditions, after its harvest. The ideal storage conditions depend on the species, cultivar, ripeness stage and many other factors, and must be experimentally determined. Traditionally, this is achieved by independently storing fruit at many different combinations of temperature as well as O₂ and CO₂ partial pressures, and by monitoring the change of quality attributes during, the sometimes year long, storage period (Fidler and North 1967). The optimal storage conditions are then inferred from response surface modeling. This method is black box in nature and labor intensive, due to the large number of time-consuming experimental combinations that have to be tested (Saltveit 2003).

Many modern storage applications, such as modified atmosphere packaging (MAP) (Jacxsens et al. 2000; Fonseca, Oliveira, and Brecht 2002a) and dynamic controlled atmosphere (DCA) (Bessemans et al. 2016), rely on knowledge of mass balances, transport phenomena and reaction kinetics, and use comprehensive mathematical models that describe the behavior of the product as a dynamical system with inputs (temperature, O₂ and CO₂ partial pressures) and outputs (respiration and fermentation rate, quality attributes). The respiration kinetics are a key feature of such dynamic models, and are generally described by a non-linear model of the Michaelis-Menten type (Hertog et al. 1998a). The standard Michaelis-Menten model contains two parameters: the maximum respiration rate and the affinity or Michaelis-Menten constant (Peppelenbos and Leven 1996).

Michaelis-Menten kinetics are generally used to describe the kinetics of a simple enzymatic pathway in which, first, a substrate-enzyme complex is formed in a reversible way. In a second rate limiting step, the complex dissociates into the reaction product and the enzyme (Michaelis and Menten 1913). Although respiration, and specifically O₂ consumption, is a far more

complex pathway, that contains both linear and circular routes and a multitude of intermediate compounds (Berg, Tymoczko, and Gatto Jr 2002), it can be described surprisingly well by Michaelis-Menten like kinetics at various spatial scales, from the fruit tissue level to the intact fruit level (Ho, Verboven, Verlinden, Herremans, et al. 2011). The maximum respiration rate and the Michaelis-Menten constant should thus be considered as apparent parameters, that summarize the behavior of the more complex pathway underneath. These parameters are typically estimated from respiration experiments at various combinations of temperature as well as O₂ and CO₂ partial pressure, which are then kept constant throughout the experiment. Appropriate experimental design techniques are then applied to reduce the number of experimental runs. Typically, response surface designs, such as central composite designs, are used to define the combinations to be tested, and response surface models are fitted to the resulting data. However, even then, the experimental effort remains considerable (Fidler and North 1967; Saltveit 2003).

Alternatively, dynamic experiments can be conceived in which the experimental factors may vary over time in such a way that they maximize the information content of the experimental data set. The shift from traditional response surface modeling towards dynamic models for estimating respiration kinetics entails unique challenges for designing experiments. As we explain below, this is due to the fact that the former modeling approach is based on linear algebraic equations and parameter estimation, while the latter requires differential equations and non-linear parameter estimation.

The optimal design of dynamic experiments is an unexplored research topic within postharvest research and fruit storage. It has, however, received a limited amount of attention in other biological fields. One example is the construction of an optimal temperature profile to aid the estimation of several kinetic growth parameters in predictive microbiology (Bernaerts, Versyck, and Van Impe 2000; Bernaerts, Servaes, et al. 2002; Balsa-Canto, Rodriguez-Fernandez, and Banga 2007). Another example can be found in food engineering, where temperature profiles have to be optimized to better determine thermo-physical properties of food products (Nahor, Scheerlinck, Verniest, et al. 2001a; Nahor, Scheerlinck, Van Impe, et al. 2003). However, applications are not limited to temperature profiles. For instance, during the identification of biochemical networks, the input profile of several (bio-)chemical substances has to be controlled (Balsa-Canto, Alonso, and Banga 2010). This is similar to our own problem, where an O₂ input profile has to be optimized.

In this chapter, we show how to design dynamic experiments to efficiently estimate the two Michaelis-Menten respiration parameters of the specific oxygen

consumption rate of pear fruit. First, in Section 2.2.1, we present a model for pear fruit respiring inside a jar, with O_2 as the sole time varying input. In Section 2.2.2, we introduce the main concepts of optimal dynamic experimental design and we subsequently apply them to our model to derive expressions for quantifying the information content of an experiment. In Section 2.2.3, We propose two methods to optimize the information content of the experiment, based on the coordinate-exchange algorithm. In Section 2.3, we construct three optimal designs for different admissible input profiles. The first proof of concept design in Section 2.3.1 is constructed assuming that the O_2 input level can only be modified every 2 h, and is either in an on or an off state. In Section 2.3.2, two optimal designs are constructed by relaxing the restrictions on the input profile in two different ways. First, by allowing the input to be modified every 10 min. And secondly, by allowing 11 possible input levels. Our decisions on how to quantify information, and on how to parametrize the input profile are discussed in Section 2.4, as well as some alternatives. We also discuss further research options in this section.

2.2 Methods

The methods section is divided into three parts. First, we propose a respiration model for pear fruit, in an experimental unit: the jar. Next, we discuss the main concepts of optimal dynamic experimental design and apply them to our model. Finally, we adapt the coordinate-exchange algorithm to work better for the dynamic experiments considered in this chapter.

2.2.1 Dynamic Model for Respiration

The respiration of pear fruit inside a jar was modeled using the following mass balance,

$$V_j \frac{dC_m}{dt}(t) = Q(t)(C_{in} - C_m(t)) - m_p r(t), \quad (2.1)$$

where $C_m(t)$ is the modeled O_2 concentration in the jar [mol m^{-3}], at time t [s]. The left-hand side of the equation shows the change of mass inside the jar, with a volume V_j of 5 dm^3 . The right-hand side consists of two terms, the first of which describes the air flowing in and out of the jar. The flow rate $Q(t)$ [$\text{m}^3 \text{s}^{-1}$] is the controllable input to the system. More specifically, we assumed that air is blown into the jar with concentration C_{in} equal to that of regular air (21 volume %), and that, air flows out of the system at the same rate $Q(t)$ and with the same O_2 concentration $C_m(t)$ as that inside the jar, due to the

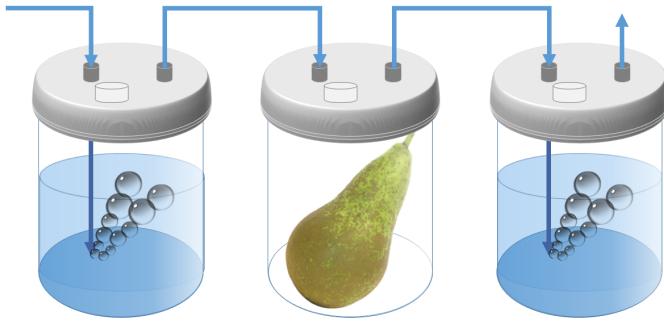


Figure 2.1: Graphical representation of the respiration jar measurement setup.

assumption of perfect mixing. The second term in the right-hand side models the consumption of O_2 inside the jar, which is proportional to the mass of the pears, m_p [kg]. Throughout the chapter, we assumed that all measurements are performed on 4 kg of pears. The reaction rate $r(t)$ [$\text{mol kg}^{-1}\text{s}^{-1}$], at time t is described by the Michaelis-Menten kinetics model,

$$r(t) = \frac{V_{\max}C_m(t)}{K_m + C_m(t)}, \quad (2.2)$$

in which the maximal respiration rate V_{\max} [$\text{mol kg}^{-1}\text{s}^{-1}$] and the Michaelis-Menten constant K_m [mol m^{-3}] are the two unknown parameters. In using this model, we ignored the effect of fermentation at low oxygen concentration. The jar was assumed to be filled with (regular) air at the start of every experiment. A representation of the model is given in Figure 2.1. Finally, we assumed that the measured concentration $C(t)$ only differs from the modeled concentration by an additive white noise term $\epsilon(t)$:

$$C(t) = C_m(t) + \epsilon(t). \quad (2.3)$$

We also assume that the composition of the gas concentration in the jar is measured after sampling by means of a syringe through a septum in the lid of the jar.

2.2.2 Optimal Experimental Design Methodology

Estimation of the parameter vector $\mathbf{p} = [V_{\max} \quad K_m]'$ was done by minimizing the squared error identification functional:

$$\mathbf{J} = \int_0^{t_f} (C(t) - C_m(\mathbf{p}, Q(t), t))^2 dt, \quad (2.4)$$

where t_f [s] is the end time of the experiment. The functional \mathbf{J} is often called the cost function. This estimation problem was solved using the non-linear least squares algorithm, which required a starting value \mathbf{p}^0 for the parameters. Around that starting value, the model and \mathbf{J} were linearized in the first iteration of the algorithm. When using that approach, the contours of \mathbf{J} are ellipses, whose principal axes and surface area depend on the Fisher information matrix,

$$\mathbf{F} = \int_0^{t_f} \left(\frac{dC_m}{d\mathbf{p}} \right)' \left(\frac{dC_m}{d\mathbf{p}} \right) dt, \quad (2.5)$$

(Fedorov and Leonov 2013). The derivatives in this expression are generally referred to as parameter sensitivities:

$$\frac{dC_m}{d\mathbf{p}} = \begin{bmatrix} \frac{\partial C_m}{\partial V_{\max}} & \frac{\partial C_m}{\partial K_m} \end{bmatrix}. \quad (2.6)$$

The larger the Fisher information matrix, the smaller the elliptic contours, and the steeper the cost functional \mathbf{J} around its minimum, and thus the better the least squares identification problem is defined. The Fisher information matrix also has a second interpretation due to the fact that it is the inverse of the asymptotic variance covariance matrix of the best linear unbiased estimator (BLUE) (Munack and Posten 1989). It summarizes the available information about the quality of the parameter estimates. The connection between the two interpretations of the Fisher information matrix is logical: the steeper the cost function, the better the least squares estimates fit the data compared to other parameter estimates. In that case, the quality of the estimates is large.

There exist more general definitions of the Fisher information matrix, which we use in subsequent chapters, in Sections 3.2.2 and 4.2.3. The other definitions would lead to the same optimal respiration experiments presented in this chapter. However, the converse does not hold, the definition of the FIM in this chapter leads to different results in the other chapters.

By modifying the input profile $Q(t)$, the magnitude of the Fisher information matrix, the steepness of the cost function and the surface areas of its contours can be optimized. To measure how large the Fisher information matrix is,

different scalar functions of the Fisher information matrix have been proposed in the literature. These functions can be used to compare and score different experimental designs (Atkinson, Donev, and Tobias 2007b). In this chapter, we used the determinant of the Fisher information matrix, $|\mathbf{F}|$, to compare different input profiles, which is referred to as the D-criterion. This criterion is inversely related to the area of the confidence ellipses of the parameter estimates, which should be minimal in an optimal experiment. Therefore, the optimized experimental design should maximize the determinant $|\mathbf{F}|$ over all admissible O_2 input profiles $Q(t)$:

$$\operatorname{argmax}_{\substack{\text{admissible} \\ Q(t)}} |\mathbf{F}|. \quad (2.7)$$

The parameter sensitivities $dC_m/d\mathbf{p}$ cannot be calculated analytically but have to be appended to the system of differential equations in equation (2.1), so they can be solved numerically. This is different from non-dynamical experiments, where analytical expressions exist for the sensitivities. The parameter sensitivity for the maximal respiration rate can be calculated from the following differential equation:

$$\begin{aligned} \frac{d}{dt} \left(\frac{\partial C_m}{\partial V_{\max}} \right) &= \frac{\partial}{\partial V_{\max}} \left(\frac{dC_m}{dt} \right) \\ &= -\frac{1}{V_j} Q(t) \frac{\partial C_m}{\partial V_{\max}} - \\ \frac{1}{V_j} m_p \left(\frac{\left(C_m(t) + V_{\max} \frac{\partial C_m}{\partial V_{\max}} \right) (K_m + C_m(t)) - \left(\frac{\partial C_m}{\partial V_{\max}} V_{\max} C_m(t) \right)}{(K_m + C_m(t))^2} \right). \end{aligned} \quad (2.8)$$

The parameter sensitivity for the Michaelis-Menten constant $\partial C_m / \partial K_m$ can be calculated in a similar manner:

$$\begin{aligned} \frac{d}{dt} \left(\frac{\partial C_m}{\partial K_m} \right) &= \frac{\partial}{\partial K_m} \left(\frac{dC_m}{dt} \right) \\ &= \frac{1}{V_j} \left(-Q(t) \frac{\partial C_m}{\partial K_m} - m_p V_{\max} \left(\frac{\frac{\partial C_m}{\partial K_m} (K_m + C_m(t)) - \left(1 + \frac{\partial C_m}{\partial K_m} \right) C_m(t)}{(K_m + C_m(t))^2} \right) \right). \end{aligned} \quad (2.9)$$

In our search for optimal input profiles, we solved these differential equations using the Matlab implementation of the Dormand-Prince 45 method (Shampine and Reichelt 1997). Whenever the solver reached a discontinuity in the input profile, the solver was terminated and restarted again with initial conditions equal to the previous final conditions. This prevented the adaptive time stepper from remaining stuck on the discontinuity. The integral in the expression

of the Fisher information matrix \mathbf{F} in equation (2.5) and the squared error identification functional \mathbf{J} in equation (2.4) were calculated using the trapezoid rule with the time points provided by the differential equation solver.

The sensitivities in equations (2.8) and (2.9) and thus also the information matrix in equation (2.5) depend on the unknown parameters V_{\max} and K_m . Therefore, we needed a starting parameter vector \mathbf{p}^0 , to be able to calculate D-optimal designs. This technique is called locally optimal design, as the optimal designs obtained, are only truly optimal for the specific starting parameters in \mathbf{p}^0 .

2.2.3 Information Optimization

To search for an optimal experimental design, we first had to characterize the input profile $Q(t)$, to form a set of admissible input profiles. This method is called control vector parametrization in the optimal control theory literature (Vassiliadis 1993). In that literature, methods for efficient numerical optimization of dynamic experiments were suggested as well, with a focus on continuous parameters (Telen, Logist, et al. 2012; Bauer et al. 2000). Our work is different in that we work with an integer parametrization of the input profile.

In this chapter, we focused on experiments lasting 24 h. The input $Q(t)$ could take l equidistant different levels between a maximal input level, with a flow rate of 10 l/h, and a minimal input level, with zero flow rate. In Section 2.3, we first consider examples with $l = 2$. In these examples, only an on or off state was allowed. Next, we also consider a scenario with $l = 11$, where the input could take 11 values. The total time duration of the experiment was divided into n equally long time-intervals. At the beginning of each of the n intervals, the input could be changed from one input level to another. Note that this does not mean that the input had to be changed at these time points. We first consider a scenario in which the input could be changed every 2 h ($n = 12$). Next, we also consider a case where changes were allowed every 10 min ($n = 144$).

Because a full enumeration of the design space is not feasible for larger values of n and j , a numerical routine had to be used to optimize the experiment. Popular choices in design of experiments are variants of the coordinate-exchange algorithm (Meyer and Nachtsheim 1995). Therefore, we have also used a coordinate-exchange algorithm. In its simplest variant, similar to the one described in Goos and Jones (2011), a random feasible starting design was generated first. Next, the most informative of the possible input levels during the first of the n time-intervals was identified by comparing all $l - 1$ alternatives

```

optimal profile = random feasible design;
for iteration = 1 to m do
    generate random feasible design;
    improvement = yes;
    while improvement == yes do
        improvement = no;
        for k = 1 to n do
            information = information of current profile;
            for j = 1 to l - 1 other levels in interval k do
                switch input of kth time-interval to jth other level;
                new information = information of the resulting profile;
                if new information > information then
                    accept change;
                    information = new information;
                    improvement = yes;
                end
            end
        end
    end
    if information mth input profile > information optimal profile then
        | optimal profile = mth input profile
    end
end

```

Figure 2.2: Initial coordinate-exchange algorithm.

to the original. If the best of the alternative levels led to a more informative experiment, the change was accepted. Otherwise, the previous input profile was reinstated. Subsequently, the same kind of change was evaluated for the next time-interval. When all n intervals had been considered, the algorithm evaluated the first time-interval again. This procedure was continued until no more improvements could be made to the design, in a whole pass through the entire set of n time points. This optimization was performed m times, where m is a specified number of iterations. The best of the m results was selected as the optimal input profile. Pseudo-code for this algorithm is shown in Figure 2.2.

When changing the input at a certain time t' , the output $C_m(t)$ for all $t \geq t'$ is affected. This distinguishes dynamic experiments from static experiments. Due to this feature, running through the n time-intervals in order might lead the coordinate-exchange algorithm to get stuck in a local minimum. Because our

```

optimal profile = random feasible design;
for iteration = 1 to m do
    generate random feasible design;
    improvement = yes;
    while improvement == yes do
        improvement = no;
        S = {1, 2, ..., n};
        for i = 1 to n do
            k = randomly selected element of S;
            information = information of current profile;
            for j = 1 to l - 1 other levels in interval k do
                switch input of kth time-interval to jth other level;
                new information = information of the new profile;
                if new information > information then
                    accept change;
                    information = new information;
                    improvement = yes;
                end
            end
        end
        remove k from S
    end
    if information mth input profile > information optimal profile then
        | optimal profile = mth input profile
    end
end

```

Figure 2.3: Modified coordinate-exchange algorithm.

initial coordinate-exchange algorithm indeed did get stuck in the same local optima quite often, as shown in the examples in Section 2.3, we propose a variant of the coordinate-exchange algorithm in which the optimization does not start with the first time-interval and subsequently moves to the next interval, but picks the n intervals in a random order, without repetition. This procedure is repeated until no more improvements can be made. Pseudo-code of this modified coordinate-exchange algorithm is shown in Figure 2.3.

2.3 Results

In this section, we report results concerning three different optimal experimental designs we constructed, each with their own parametrization of the input profile. We start with a proof of concept example, where the input could only take two levels (on/off) and the level can be changed every 2 h. The resulting optimal experimental design is compared to several intuitive benchmarks. Next, we explain how we refined the design space of our experiment in two different manners to study the resulting increase in information. We first subdivide every time-interval of 2 h into 12 intervals of 10 min each. Second, we allowed 11 different input levels.

As initial parameter vector \mathbf{p}^0 for all our computations, we used the respiration parameter estimates for Conference pear, as reported by Lammertyn et al. (2003). More specifically, we used a maximal respiration rate, V_{\max} , of $0.2324 \text{ mmol kg}^{-1}\text{s}^{-1}$, calculated at a temperature of 20°C , and the value 5.9 % for the Michaelis-Menten constant K_m . Concentrations in volume percentages are calculated by assuming atmospheric pressure in the jar, and applying the ideal gas law.

2.3.1 Proof of Concept Example

Optimal design

The optimal experimental design depicted in Figure 2.4a was found after an exhaustive search over all 2^{12} possible input profiles. The optimal input is a pulse and only pumps air into the jar during a single time-interval, from 6 until 8 h. The corresponding simulated output, the O_2 concentration $C(t)$ is depicted in Figure 2.4b. The D-criterion value for the optimal design equals 3.56×10^{15} . No alternative optima with the same information content were found. Why this input profile is optimal can be intuitively explained by the concentration in the jar near the end of the experiment. That concentration is roughly equal to the Michaelis-Menten constant K_m , which determines the switching behavior between the linear and saturated part of the Michaelis-Menten curve. Measuring in this concentration range provides information about this parameter. If the two-hour interval in which air is pumped into the jar would have started earlier, the concentration in the jar at the end of the experiment would have dropped even lower, and the information content of the experiment would decrease. This is due to the fact that, in our model, which neglects fermentation, a very low O_2 concentration means almost no respiration, and thus little information about respiration parameters. Conversely, if the two-hour interval would start later,

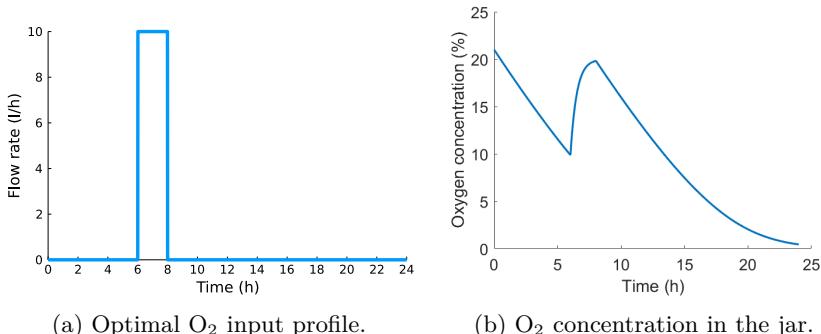


Figure 2.4: Optimal design, assuming $l = 2$ and $n = 12$: pulse input after 6 h.

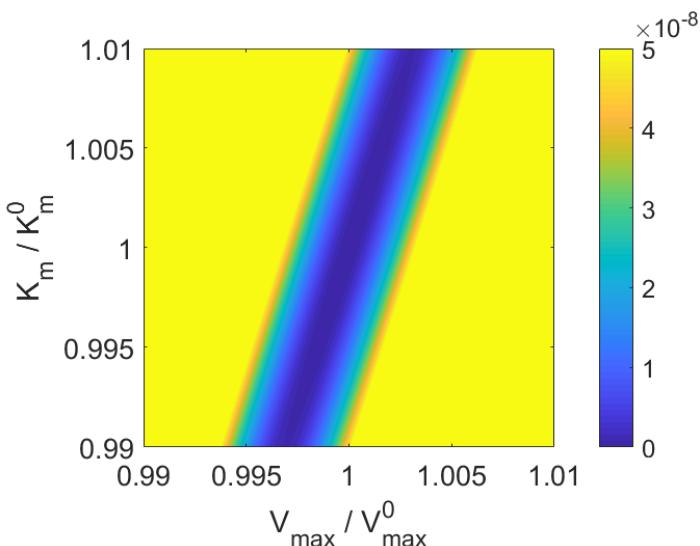


Figure 2.5: SSE contours for the optimal design in Figure 2.4a.

little information about the switching behavior would be gained.

While a full enumeration of the design space was feasible for our proof of concept example, we also tested the performance of the initial and the modified coordinate-exchange algorithm. In each case, we ran the algorithm 10,000 times. When applying the initial algorithm, the optimal experiment was found only ten times. So, while this simple algorithm performs well for response surface models, it is clearly not appropriate for our dynamical models, with our

Experimental design	D-criterion
Optimal design: pulse after 6 h	3.56×10^{15}
Alternative design: Always on	9.37×10^7
Alternative design: Always off	2.66×10^{15}
Alternative design: Switching on/off	2.64×10^{11}
Alternative design: pulse after 4 h	3.16×10^{15}

Table 2.1: D-criterion values for both the optimal design as well as several benchmarks, assuming $l = 2$ and $n = 12$. These designs are shown in Figures 2.4 — 2.9.

specific parametrization of the O_2 input profile. When applying the modified algorithm, we obtained the global optimum in 1,014 of the 10,000 iterations of the algorithm, a considerable improvement.

Comparison to benchmark designs

We now compare the first optimal experimental design of Figure 2.4a to four intuitive benchmark designs, also considered in the full enumerative search. More specifically, we provide a comparison with two constant input profiles, a maximally dynamic profile and an alternative pulse. Table 2.1 lists these benchmark designs and their D-criterion values.

We also performed a detailed graphical comparison by calculating the squared error identification functional in Equation (2.5) for different parameter vectors, in the neighborhood of the initial parameter vector \mathbf{p}^0 . If other parameters would fit the system almost equally well, the sum of squared errors (SSE) would not vary substantially in the neighborhood of the initial parameters values, suggesting imprecise parameter estimation. The SSE contours resulting from the first optimal experimental design are shown in Figure 2.5. All values larger than the maximum on the color bar are displayed in the same yellow color as the maximum itself. The values on the axes are relative to the initial parameter values. The function is steep in the vertical and horizontal direction, indicating precise parameter estimates. However, the principal axes of the ellipses are not parallel with the horizontal and vertical axes, and the contours are elongated diagonally. This indicates that even the optimal experimental design would result in correlated parameter estimates. In other words, the two parameters cannot be estimated independently. This is a well-known problem when estimating Michaelis-Menten models.

Figure 2.6b depicts the SSE contours resulting from the constant input profile

that continuously pumps air into the jar, for the full duration of the experiment in Figure 2.6a. The D-criterion corresponding to that experimental design is much lower, 9.37×10^7 . As a result, this benchmark experiment performs considerably worse in terms of the quality of the estimates than the optimal design in Figure 2.4: the SSE value is insensitive to the values of the parameters K_m and V_{max} .

The other constant input profile we considered as a benchmark and which is shown in Figure 2.7a, is the always off profile. This profile has a D-criterion value much closer to that of the optimal design, namely 2.66×10^{15} . The resulting contours for the SSE value are shown in Figure 2.7b, and resemble those in Figure 2.5. However, they have a slightly larger surface area.

Figure 2.8 demonstrates that the SSE contours produced by a maximally dynamic input profile are quite different from the contours produced by the optimal design. The maximally dynamic profile switches input every 2 h and the D-criterion value equals 2.64×10^{11} , for that profile. This demonstrates that optimal dynamic design of experiments is not equivalent to changing the input as frequently as possible.

In Figure 2.9, we show a pulse input function that occurs 2 h earlier than the pulse in the optimal experimental design, as well as the SSE contours corresponding to that input. This profile performs much better than the other benchmark experimental designs. This is also reflected in the D-criterion value of 3.16×10^{15} , which shows that designs that are close to the optimal design are also similar in information content.

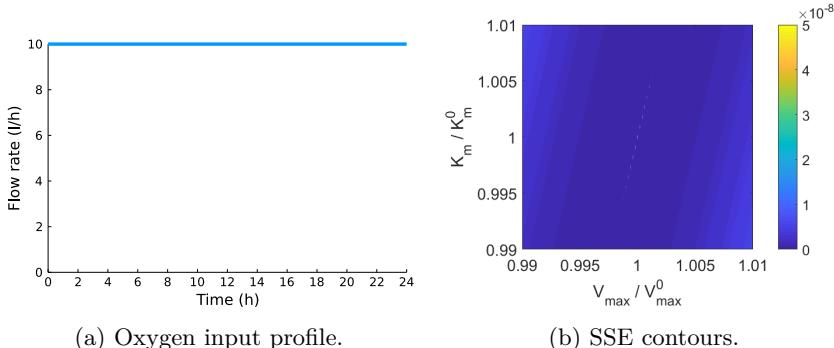
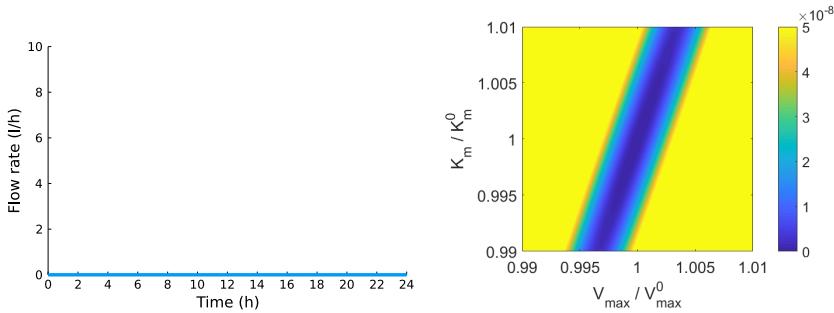


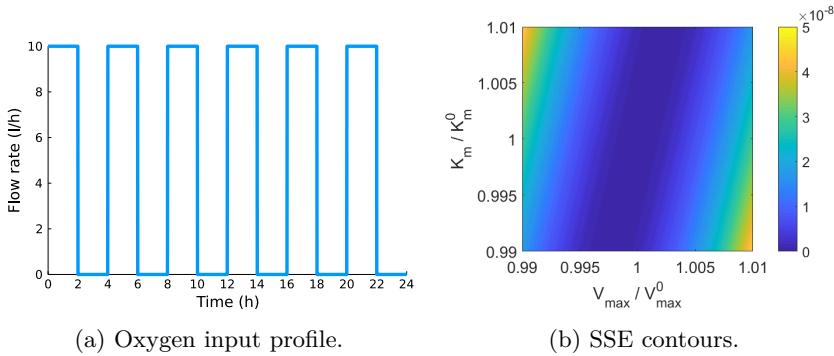
Figure 2.6: First benchmark: an always-on input profile.



(a) Oxygen input profile.

(b) SSE contours.

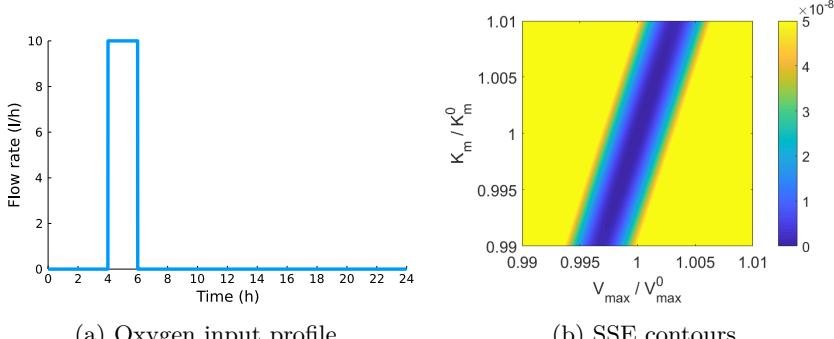
Figure 2.7: Second benchmark: an always-off input profile.



(a) Oxygen input profile.

(b) SSE contours.

Figure 2.8: Third benchmark: a constantly switching input profile.



(a) Oxygen input profile.

(b) SSE contours

Figure 2.9: Fourth benchmark: pulse input after 4 h.

Optimal experimental design	D-criterion
$n = 12$ and $l = 2$	3.56×10^{15}
$n = 144$ and $l = 2$	5.15×10^{17}
$n = 12$ and $l = 11$	5.13×10^{17}

Table 2.2: D-criterion values for optimal experimental designs with different parametrizations in Sections 2.3.2 and 2.3.2. These designs are shown in Figures 2.4, 2.10 and 2.11.

2.3.2 Refining the Design Space

We refined the parametrization of the proof of concept experimental designs in two different ways. The D-criterion values of the optimal input profiles are listed in Table 2.2.

Allowing more frequent input changes

For the construction of our second optimal experimental design, we allowed the inputs to be changed every 10 min, instead of every 2 h, while still only allowing an on or off input. As all admissible input profiles from the proof of concept example are also admissible under the new settings, the second optimal design problem generalizes the initial one, since 120 min is divisible by 10 min. The set of admissible input profiles in our second optimal design problem can thus be considered as a refinement of the first. To identify the optimal experimental design for the new problem, we used 10,000 iterations of the modified coordinate-exchange algorithm. The unique best design found by the modified coordinate-exchange algorithm is depicted in Figure 2.10a. It was found in 45 of the 10,000 iterations of the algorithm and involves two short 10 min pulses during the second half of the experiment. The resulting O₂ concentration in the jar is shown in Figure 2.10b. The D-criterion value of this design amounts to 5.15×10^{17} , which is much better than the value of the optimal design in the proof of concept example.

The new optimal profile differs substantially from the first in Figure 2.4a: instead of one longer pulse from 6 to 8 h, two shorter 10 min pulses after 12 and 17 h are preferred. The reason why this design is optimal is, however, very similar to that of the previous example. After an initial decrease, the O₂ concentration in the jar keeps hovering around the value of the Michaelis-Menten constant K_m, providing information about the switching behavior of the system between saturated and linear respiration. If the O₂ concentration would drop even lower, little respiration would occur and thus little information would be gained

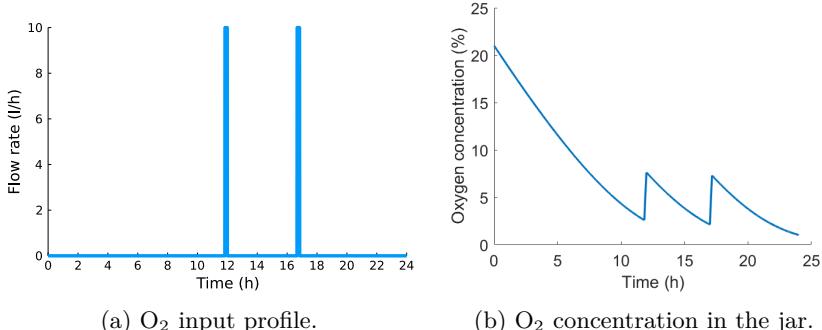


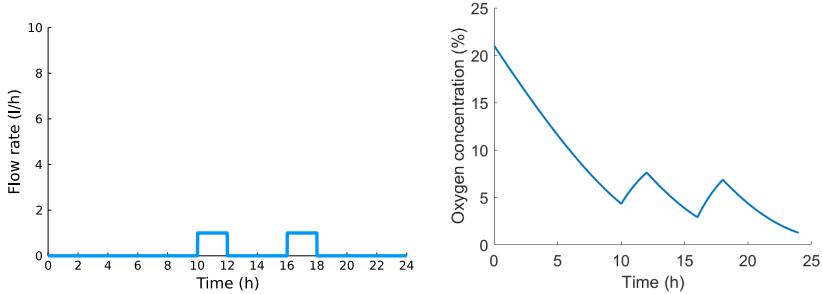
Figure 2.10: Optimal design, assuming $l = 2$ and $n = 144$: two short pulses during the second half of the experiment.

concerning the two model parameters. The preference for two short pulses over one longer pulse can be explained by comparing Figure 2.10 to Figure 2.4b. A longer pulse increases the O₂ concentration in the jar almost to the initial one. Since the system always starts at atmospheric conditions, information about maximal respiration is always present in the initial phase of the experiment. Returning to that high concentration does not yield new information. Instead, there is added value in studying the respiration at concentrations around K_m . Therefore, the optimal profile involving two shorter pulses focuses more on the switching behavior instead of revisiting the region of saturated respiration.

Allowing more input levels

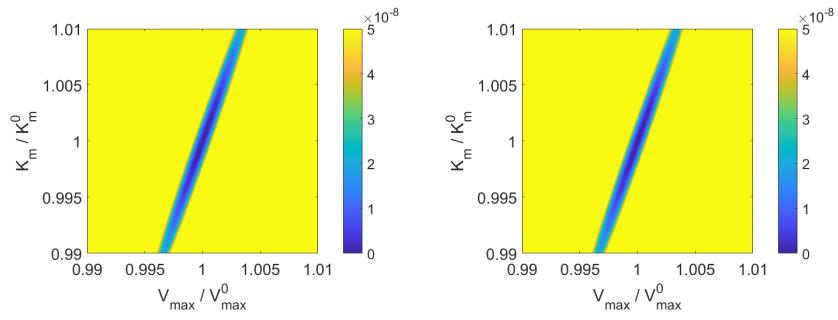
In our final optimization, we allowed the input to take 11 different levels, equally spaced between the maximum and minimum input level. The input was only allowed to be changed every 2 h, unlike in the previous example. Again, all possible designs encountered during the first optimization, were considered in this third optimization problem. So, this example also generalizes the proof of concept example. The unique best design found using 10,000 iterations of the modified coordinate-exchange algorithm, is depicted in Figure 2.11a. That input profile was found in 75 of the 10,000 iterations. The corresponding O₂ concentration in the jar is shown in Figure 2.11b. The D-criterion value of 5.13×10^{17} is very similar to that of the input profile in Figure 2.10a, as can be seen in Table 2.2.

A similar dynamic input, consisting of two pulses, seems to be preferred in the last two examples. In the third, experiment the small pulses take the minimum



(a) Optimal O₂ input profile. (b) O₂ concentration in the jar.

Figure 2.11: Optimal design, assuming $l = 11$ and $n = 12$: two small pulses after 10 h and 16 h.



(a) Optimal design in Figure 2.10a. (b) Optimal design in Figure 2.11a.

Figure 2.12: SSE contours for the optimal experimental designs in Figure 2.10a and 2.11a.

non-zero input value of 0.1, or 10 % of the maximum allowed flow rate. The simulated outlet O₂ concentrations shown on Figure 2.10b and Figure 2.11b, look remarkably similar, which shows that even with different parametrizations of the input curve, a similar dynamic behavior can be achieved.

The SSE contours in Figure 2.12 correspond to the input profiles of Figure 2.10a and Figure 2.11a. As already noted, the designs involve similar dynamic O₂ concentration trends, have similar D-criterion values and thus also have very similar contour plots. The contours in Figure 2.12 are also much steeper than those in Figure 2.5, but they are still elongated diagonally. So, even a refinement of the design space did not help to remove the correlation between the two parameter estimates.

2.4 Discussion

In Section 2.2.2, we mentioned that several design optimality criteria have been proposed in literature. We opted for the D-criterion because it does not depend on the chosen units to describe the model, unlike other criteria such as the A-optimality criterion, which minimizes the average variance of the parameter estimates. Another criterion that is independent from the chosen units is the I-optimality criterion, which seeks to minimize the average prediction variance over the family of all admissible input profiles (Goos and Jones 2011). However, this criterion is computationally more expensive than the D-criterion, which, combined with the already computationally intensive dynamic experimental design idea, would lead to prohibitively long computation times.

In this chapter, we calculated locally optimal designs for a given set of initial values \mathbf{p}^0 . Most of the literature on design of experiments for non-linear models focuses on locally optimal design (Fedorov and Leonov 2013). This is especially so in the dynamic case (Bernaerts, Versyck, and Van Impe 2000; Bernaerts, Servaes, et al. 2002; Balsa-Canto, Rodriguez-Fernandez, and Banga 2007; Balsa-Canto, Alonso, and Banga 2010; Nahor, Scheerlinck, Verniest, et al. 2001a; Nahor, Scheerlinck, Van Impe, et al. 2003). However, we often possess more a priori information concerning a parameter than just a point estimate. For instance, we might also have access to a confidence interval for each model parameter. We could also incorporate this information into our design, with a technique called Bayesian optimal design (Chaloner and Verdinelli 1995), at the expense of a substantial amount of computing time.

In this chapter, we considered input profiles that are practically feasible. More specifically, when determining the first optimal experimental design, we parametrized our input profile, as a function that could only be on or off. These kinds of input profiles can be easily implemented in practice. We improved on this first design in two different ways: by allowing more than two input levels, and by allowing the input to be switched more rapidly. Similar gains in information were achieved by both improvements, so that there is flexibility in how to perform an informative experiment.

In most of the existing literature, as well as in this chapter, only a single input profile has been optimized (Bernaerts, Versyck, and Van Impe 2000; Bernaerts, Servaes, et al. 2002; Balsa-Canto, Rodriguez-Fernandez, and Banga 2007; Balsa-Canto, Alonso, and Banga 2010; Nahor, Scheerlinck, Verniest, et al. 2001a; Nahor, Scheerlinck, Van Impe, et al. 2003). A useful extension of this work would be to optimize multiple input profiles at the same time. Another interesting avenue for future research would be to optimize input profiles for

estimating distributed parameter models, which describe the dynamical system by partial differential equations, instead of our lumped approach, which uses ordinary differential equations. The assumption of perfect mixing can then also be relaxed. Examples of such models are the reaction-diffusion models developed by Ho, Hertog, et al. (2018). To optimize input profiles for such models, however, we first need to develop measures for quantifying parameter uncertainty in these models and for quantifying the information content of the experiments used for collecting the data required for model estimation. (As is often the case in statistics and data science, model development is ahead of tools to verify model quality (Efron and Hastie 2016).)

2.5 Conclusion

This chapter presented a pioneering study about the usefulness of optimal dynamic experiments for non-linear modeling in postharvest research. Three D-optimal dynamic experimental designs for the estimation of the Michaelis-Menten respiration parameters of Conference pear were constructed. Each design was constructed using a different parametrization of the input profiles.

The design of protocols for controlled or modified atmosphere of fruit and vegetables is increasingly based on dynamical models of their respiration. Successful implementation depends on accurate knowledge of the model parameters, since every fruit cultivar has to be stored under slightly different conditions. Determining all these conditions is a labor-intensive activity, and requires efficient experimentation. Our three optimal designs show that optimal dynamic design of experiments increases the quality of the respiration parameters' estimates compared to some benchmarks. Therefore, optimal dynamic design of experiments has the potential to facilitate the successful implementation of controlled and modified atmosphere storage. As a matter of fact, dynamic experimentation is a generic method: while our experiments were optimized for Conference pear, a different choice of prior information can be used to obtain optimal input profiles for different cultivars. However, in this chapter, we only took respiration of pear fruit into account. We also used a locally optimal experimental design method which may not be robust. In the next chapter, we improve both these aspects.

Chapter 3

Robust Dynamic Experiments for the Precise Estimation of Respiration and Fermentation Parameters of Fruit and Vegetables

This chapter has been written by Arno Strouwen with feedback from Professors Goos and Nicolaï.

Source code available upon request.

Abstract

Dynamic models based on non-linear differential equations are increasingly being used in many biological applications. Highly informative dynamic experiments are valuable for the identification of these dynamic models. The storage of fresh fruit and vegetables is one such application where dynamic experimentation is gaining momentum. In this chapter, we construct optimal O₂ and CO₂ gas input profiles to estimate the respiration and fermentation kinetics of pear fruit. The optimal input profiles, however, depend on the true values of the respiration

and fermentation parameters. Locally optimal design of input profiles, which uses a single initial guess for the parameters, is the traditional method to deal with this issue. This method, however, is very sensitive to the initial values selected for the model parameters. Therefore, we suggest a Bayesian experimental design approach to robustify our input profiles to the unknown values of the parameters.

3.1 Introduction

Model-based approaches are commonly used in the analysis, control and optimization of biological systems. These models rely on knowledge of physical, chemical and biological laws, such as mass balances, transport phenomena and reaction kinetics, and are often described by a system of non-linear differential equations, with inputs and outputs. So, often, the structure of a model can be determined from first principles. However, the model will generally also rely on parameters whose numerical values cannot be determined from physical laws. These parameters must then be inferred from experimental data before the model can be put to use. The experiments required to estimate these parameters are often laborious and cost prohibitive. It is therefore important to determine experimental conditions that are rich in information and thus allow a precise estimation of the unknown model parameters.

At present, the model parameters are often estimated from data collected using multiple experiments at various combinations of input levels, which are kept constant throughout each individual experiment. Even if an appropriate experimental design technique is used to reduce the number of experiments that have to be performed, the experimental effort remains considerable. Alternatively, experiments in which the inputs vary in time can be conducted. This has been shown to be a cost-effective way to generate a highly informative data set (Versyck, Claes, and Van Impe 1997). Such experiments are called dynamic. In optimal design of dynamic experiments, time-varying input profiles are constructed to optimize the information content of a single experiment.

The major challenge for experimental design for any non-linear model is the dependence of the optimal input profiles on the true values of the unknown model parameters, whose estimation is the primary goal of the experiment. This dependence is due to the non-linearity of the model. Most of the experimental design literature uses a scalar function of the Fisher information matrix as the measure of information content in an experiment, as this matrix is inversely related to the covariance matrix of the model parameters. An informative experiment thus ensures a small covariance matrix of the model parameters.

Locally optimal design of input profiles uses initial guesses for the model parameters to calculate this information matrix (Fedorov and Leonov 2013). This method has already been used to construct informative time-varying inputs in chemical engineering (Franceschini and Macchietto 2008) and in biological fields such as systems biology (Balsa-Canto, Rodriguez-Fernandez, and Banga 2007), predictive microbiology (Bernaerts, Versyck, and Van Impe 2000; Bernaerts, Servaes, et al. 2002) and food engineering (Balsa-Canto, Alonso, and Banga 2010; Nahor, Scheerlinck, Verniest, et al. 2001b). However, an input profile that is highly informative for one set of parameter values may lack information for other parameter values. Thus, if the initial guesses for the parameters differ substantially from the true values, then the locally optimal design might not allow precise parameter estimates. So, locally optimal design is sensitive to the initial parameter guesses and is thus not robust (S. Asprey, Macchietto, and Pantelides 2000).

Much recent research in experimental design for non-linear models aims at robustifying the design to the true, but unknown, values of the model parameters. A robust design provides a large information content regardless of the true values of these parameters. For dynamic experiments, in particular, a min-max based approach has been used by Bauer et al. (2000) and Körkel et al. (2004). Here, the design is optimized for a worst case scenario. Fisher information matrices are calculated for a set of possible parameter values and the experiment is scored based on the least informative matrix in this set. In contrast, an expected value approach is taken by Schenkendorf, Kremling, and Mangold (2009), Telen, Logist, et al. (2012), and Nimmergeers et al. (2020). In this approach, the average information content of the experiments over all possible parameter values is optimized. The expected value approach tends to perform better for a large subset of probable parameter values than the min-max approach, but not for extreme sets of parameter values. The expected value approach is also called Bayesian experimental design, because the possible parameter values can be expressed using a prior distribution (Chaloner and Verdinelli 1995). The expected value approaches of Telen, Logist, et al. (2012) and Nimmergeers et al. (2020) rely on parametric distributions to describe the uncertainty on the model parameters before the experiment has taken place. However, a parametric distribution will often not be appropriate to fully quantify the uncertainty on the model parameters, as we show in the examples in this chapter. Therefore, in this chapter, we allow arbitrary distributions to quantify the model parameter uncertainty. More specifically, we show how the results of a Bayesian analysis of historical data using Markov-chain Monte-Carlo can be used as a prior distribution. This Markov chain can then be used to calculate the average Fisher information matrix, and has the advantage that it can represent arbitrary distributions (Gelman et al. 2013).

Postharvest storage of fresh fruit and vegetables is one biological application where optimal experimental design is useful. The ideal storage temperature as well as O₂ and CO₂ partial pressures depend on the respiration characteristics, which in turn depend on species, cultivar, ripeness and multiple other factors. Determining the ideal storage conditions therefore requires much experimentation. Traditionally, this was done by independently storing the product at many different combinations of temperature as well as O₂ and CO₂ partial pressures, and by monitoring the respiration and fermentation (Saltveit 2003). Many modern storage applications, such as modified atmosphere packaging (Fonseca, Oliveira, and Brecht 2002b) and dynamic controlled atmosphere (Bessemans et al. 2016), adopt a model-based approach, where the product is described as a dynamic system with inputs and outputs. The resulting models enable us to use the tools of optimal dynamic experimental design to construct informative experiments. The respiration and fermentation kinetics are generally described by a model of the Michaelis-Menten type (Hertog et al. 1998b). Robust experimental design is particularly needed for postharvest applications because of the large biological variability of fresh fruit and vegetables. As a consequence of that variability, the parameters of the aforementioned kinetic models vary substantially between seasons and origins (Ho, Verboven, Verlinden, Herremans, et al. 2011). In this chapter, we therefore focus on constructing robust experimental techniques to estimate the respiration and fermentation kinetics of pear fruit. This chapter describes the first use of robust optimal experimental design techniques in postharvest research.

This chapter is structured as follows. First, in Section 3.2, we present our robust experimental design methodology. Next, in Section 3.3, we present a state of the art dynamic respiration and fermentation model for pear fruit, and quantify the initial uncertainty on the model parameters using a Markov-chain Monte-Carlo analysis of a prior data-set from Ho, Verboven, Verlinden, Lammertyn, et al. (2008). In Section 3.4, we construct both locally and Bayesian optimal designs for this model. Finally, in Section 3.5, we end with a discussion of alternative techniques.

3.2 Bayesian Optimal Experimental Design for Dynamical Systems

In this section, we first present the type of dynamical models considered in this chapter. Then, we show how to quantify the information gained from measurements, using the Fisher information matrix. Next, we discuss how to

maximize this information content using appropriate control inputs. Finally, we explain how to make the optimal control inputs robust.

3.2.1 Dynamic Models

In this chapter, we consider experimental design for dynamic models of the form:

$$\begin{aligned} \frac{d\mathbf{x}}{dt} &= f(t, \mathbf{x}, \boldsymbol{\theta}, \mathbf{u}(t)), & \text{with } \mathbf{x}(t=0) = \mathbf{x}_0; \\ \mathbf{y}_k &= h(\mathbf{x}(t_k)) + \boldsymbol{\epsilon}_k, \end{aligned} \quad (3.1)$$

where t denotes the time ranging from 0 to t_e , the end time of the experiment. The column vector \mathbf{y}_k contains the measurements taken at time point t_k , with k ranging from 1 to N , the number of measurement times. The time between measurements is equally spread, so that $t_k = kt_e/N$. A measurement at the end of the experiment is thus included, but not at the start. The measurements are subject to independent zero mean Gaussian noise. More specifically, $\boldsymbol{\epsilon}_k$ is identically and independently multivariate normally distributed with zero mean and covariance matrix \mathbf{R} . The measurements depend on the dynamic state column vector $\mathbf{x}(t)$ through the measurement function h . The states $\mathbf{x}(t)$ have to be calculated from the system of ordinary differential equations f , with initial conditions \mathbf{x}_0 . This system depends on the unknown model parameter column vector $\boldsymbol{\theta}$, and the controllable input column vector $\mathbf{u}(t)$.

3.2.2 Information Content of an Experiment

Our goal is to optimize the controllable inputs $\mathbf{u}(t)$ so that the measurements \mathbf{y}_k contain as much information as possible about the unknown parameters $\boldsymbol{\theta}$. A popular way to quantify the information content is the Fisher information matrix (FIM) (Fedorov and Leonov 2013; Walter and Pronzato 1994):

$$\mathbb{F}(\boldsymbol{\theta}, \mathbf{u}(t)) = \sum_{k=1}^N \frac{\partial \mathbf{x}(t_k)}{\partial \boldsymbol{\theta}}^T \frac{\partial h}{\partial \mathbf{x}}^T \mathbf{R}^{-1} \frac{\partial h}{\partial \mathbf{x}} \frac{\partial \mathbf{x}(t_k)}{\partial \boldsymbol{\theta}}. \quad (3.2)$$

The sensitivities of the states to the unknown parameters, $\partial \mathbf{x}(t_k)/\partial \boldsymbol{\theta}$, cannot be computed directly, as the evolution over time of the states $\mathbf{x}(t)$ is described by the system of differential equations in Equation (3.1). However, these sensitivities can be calculated from the forward sensitivity differential equations:

$$\frac{d}{dt} \frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}} = \frac{\partial}{\partial \boldsymbol{\theta}} \frac{d \mathbf{x}}{dt} = \frac{\partial f}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}} + \frac{\partial f}{\partial \boldsymbol{\theta}}, \quad \text{with } \frac{\partial \mathbf{x}(t=0)}{\partial \boldsymbol{\theta}} = \mathbf{0}. \quad (3.3)$$

The FIM is the inverse of the Cramer-Rao lower bound of the variance of an unbiased estimator of θ . This lower bound can be interpreted geometrically as a hyper-ellipsoid defined by the eigenvectors and the inverse of the eigenvalues of the FIM. We want this lower bound to be as small as possible, and thus the eigenvalues of the FIM to be as large as possible, because this implies precise estimates for θ are possible. Several scalar functions have been proposed to quantify the size of the FIM (Atkinson, Donev, and Tobias 2007a). In our work, we use the determinant of the FIM, also called the D-optimality criterion. This criterion is inversely related to the volume of the hyper-ellipsoids, measuring the uncertainty about the parameter vector θ .

3.2.3 Discretizing the Controls

Optimal experimental design for dynamic systems is an infinite dimensional optimization problem since it requires finding optimal controls $\mathbf{u}(t)$ for every $t \in [0, t_e]$. To make this problem tractable, the controls have to be discretized in time. We utilize a bounded piecewise constant discretization allowing $\mathbf{u}(t)$ to switch values at M equally spaced time points,

$$\mathbf{u}_{\min} \leq \mathbf{u}(t) = \sum_{j=1}^M \mathbf{u}_j \chi_{[(j-1)t_e/M, jt_e/M[}(t) \leq \mathbf{u}_{\max}, \quad (3.4)$$

where \mathbf{u}_j is the constant control vector during the interval $[(j-1)t_e/M, jt_e/M[$, and χ_A is the indicator function,

$$\chi_A(t) = \begin{cases} 1 & t \in A \\ 0 & t \notin A, \end{cases} \quad (3.5)$$

and \mathbf{u}_{\min} and \mathbf{u}_{\max} are the minimum and maximum control values allowed. Piecewise constant input profiles do not only have the benefit of making our optimization problem tractable, but they are also easy to implement in practice.

3.2.4 Robustifying the Experiment

Another issue with experimental design for models that are non-linear in the parameter vector θ , such as the model we described in Equation (3.1), is the dependence of the FIM on θ . This presents us with a cyclic problem because we are performing the experiment to quantify those parameters. Locally optimal experimental design is the traditional method to deal with this issue. In this approach, the FIM is calculated using a single initial guess θ^* obtained from

any available prior knowledge. A locally optimal experimental design is thus given by:

$$\arg \max_{\boldsymbol{u}_1 \dots \boldsymbol{u}_M} |\mathbb{F}(\boldsymbol{\theta}^*, \boldsymbol{u}(t))| \quad \text{subject to } \boldsymbol{u}_{\min} \leq \boldsymbol{u}(t) \leq \boldsymbol{u}_{\max}. \quad (3.6)$$

This method might give poor results if the initial guess deviates from the true value, and is thus not robust. Bayesian optimal design offers a solution to this robustness problem by averaging the information content of the experiment over multiple possible values of the parameters, taking into account the likelihood of each parameter value. More specifically, a weighted average is used to quantify the information content, where the weights are given by a prior distribution of the parameters $p(\boldsymbol{\theta})$. This distribution represents the knowledge of uncertainty concerning these parameters, before the experiment has been performed. A robust experimental design is therefore given by:

$$\arg \max_{\boldsymbol{u}_1 \dots \boldsymbol{u}_M} \int |\mathbb{F}(\boldsymbol{\theta}, \boldsymbol{u}(t))| p(\boldsymbol{\theta}) d\boldsymbol{\theta} \quad \text{subject to } \boldsymbol{u}_{\min} \leq \boldsymbol{u}(t) \leq \boldsymbol{u}_{\max}. \quad (3.7)$$

This criterion is also called the Bayesian D-optimality criterion (Chaloner and Verdinelli 1995), because of the use of a prior distribution.

Generally, the integral in Equation (3.7) cannot be evaluated analytically. To approximate it numerically, we draw R random model parameter vectors, $\boldsymbol{\theta}_r$, from the prior $p(\boldsymbol{\theta})$ and average the D-criterion over these values:

$$\int |\mathbb{F}(\boldsymbol{\theta}, \boldsymbol{u}(t))| p(\boldsymbol{\theta}) d\boldsymbol{\theta} \approx \frac{1}{R} \sum_{r=1}^R |\mathbb{F}(\boldsymbol{\theta}_r, \boldsymbol{u}(t))| \quad \boldsymbol{\theta}_r \sim p(\boldsymbol{\theta}). \quad (3.8)$$

The robust criterion in Equation (3.7) is thus calculated by averaging the FIM over a sample drawn from the prior distribution. The details of the generation of this sample are explained in Section 3.3.3.

3.2.5 Numerical Details

The entire optimization problem was implemented in the Julia programming language (Bezanson et al. 2017). All differential equations were solved using the Tsitouras 5/4 Runge-Kutta method (Tsitouras 2011), as implemented in OrdinaryDiffEq.jl (Rackauckas and Nie 2017). The piecewise constant control switches were implemented using a periodic callback, provided by DiffEqCallbacks.jl. Note that $\boldsymbol{u}(t_e)$ in Equation 3.4 is undefined, but this value does not influence the FIM.

Because coding the sensitivity differential equations in (3.3) by hand is quite laborious and error prone, we calculated them exploiting the automatic differentiation capabilities of DiffEqSensitivity.jl (Rackauckas, Ma, Dixit, et al. 2018), more specifically its implementation of the discrete forward sensitivity analysis method.

To solve the non-linear optimization problems, we used the box constrained optimization capacities of Optim.jl (Mogensen and Riseth 2018), which utilizes a log barrier interior point algorithm (Nocedal and Wright 2006). This method requires the specification of an inner solver, for which we used the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm, as also implemented in Optim.jl. The required gradients for this method are calculated by the nested differentiation capabilities of ForwardDiff.jl (Revels, Lubin, and Papamarkou 2016). This interior point optimization method requires an initial experimental design to improve upon and is a local optimizer. Thus, it is not guaranteed to find the absolute best experimental design. To deal with this issue, we utilize multiple starts, each with a different initial design. The TikTak global optimization algorithm (Arnoud, Guvenen, and Kleineberg 2019) can carefully select these initial designs from the design space, using Sobol points. We use 1000 starts of the interior point optimization method in combination with the TikTak implementation in MultiStartOptimization.jl.

3.3 Respiration and Fermentation Model of Pear Fruit

In this chapter, we apply our robust experimental design methodology to precisely estimate the respiration and fermentation characteristics of pear fruit. First, we present a respiration and fermentation model of pear fruit inside a jar. We then quantify the initial uncertainty concerning the various parameters in this model using a published data-set.

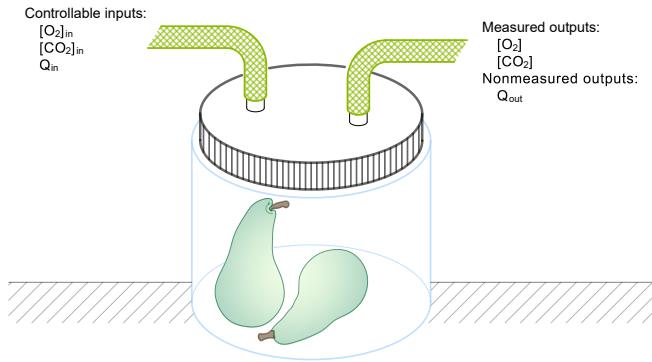


Figure 3.1: Schematic representation of the measurement setup.

3.3.1 Model Description

The respiration and fermentation of pear fruit inside a jar is modeled by two mass balances for O₂ and CO₂:

$$V_j \frac{d[O_2]}{dt} = Q_{in}(t)[O_2]_{in}(t) - Q_{out}(t)[O_2] - m_p r_{O_2}(t), \quad (3.9)$$

$$V_j \frac{d[CO_2]}{dt} = Q_{in}(t)[CO_2]_{in}(t) - Q_{out}(t)[CO_2] + m_p r_{CO_2}(t).$$

The square brackets in these expressions represent concentrations in mol/m³. These differential equations describe the change of O₂ and CO₂ concentrations inside a jar with volume V_j, which will equal 5 dm³, in our examples in Section 3.4. A time varying air mixture with an oxygen concentration [O₂]_{in}(t) and a carbon dioxide concentration [CO₂]_{in}(t) is blown into the jar with flow rate Q_{in}(t) (units: m³/h). These three time varying functions form the controllable inputs to our system. Our measurement set up is schematically shown in Figure 3.1.

Since the pressure inside the jar should remain equal to the atmospheric pressure, we can calculate the outflow Q_{out}(t) from the jar using the ideal gas law:

$$Q_{out}(t) \frac{P_{atm}}{\bar{R}T} = Q_{in} \frac{P_{atm}}{\bar{R}T} - m_p r_{O_2}(t) + m_p r_{CO_2}(t), \quad (3.10)$$

with \bar{R} the universal gas constant and P_{atm} the atmospheric pressure. Concentrations throughout the jar and at the outlet are considered to be similar, due to the assumption of well mixing. In the constructed experiments

in Section 3.4, the temperature equals 293.15 K, the amount of O₂ consumed and CO₂ produced is proportional to the mass of the pears m_p , taken to be 4 kg and the initial conditions for O₂ and CO₂ will be equal to regular air.

The respiration rates in Equation (3.10) are specified using models of the Michaelis-Menten type (Hertog et al. 1998b):

$$r_{O_2}(t) = \frac{V_{m,O_2}[O_2]}{(K_{m,O_2}/RT + [O_2]) \left(1 + \frac{[CO_2]}{K_{mn,CO_2}/RT}\right)}, \quad (3.11)$$

$$r_{CO_2}(t) = r_q r_{O_2}(t) + \frac{V_{m,f,CO_2}}{1 + \frac{[O_2]}{K_{m,f,O_2}/RT}}.$$

The models for these respiration rates contain six parameters that have to be identified. V_{m,O_2} and V_{m,f,CO_2} are the maximum respiration and fermentation reaction rates, respectively. The Michaelis-Menten constant K_{m,O_2} represents the saturation of respiration at high oxygen levels, whereas K_{mn,CO_2} models the inhibition of respiration by CO₂. The respiration quotient r_q represents the percentage of O₂ that is converted to CO₂ by respiration. Finally, K_{m,f,O_2} models the inhibition of fermentation by O₂. The three inhibition constants are expressed in units of pressure, which will be required for the analysis in the following section.

The measured gas concentrations [O₂]_{*m*} and [CO₂]_{*m*} at time point t_k are assumed to be equal to the true concentrations plus the additive Gaussian noise terms ζ_k and η_k respectively:

$$[O_2]_m(t_k) = [O_2](t_k) + \zeta_k, \quad (3.12)$$

$$[CO_2]_m(t_k) = [CO_2](t_k) + \eta_k.$$

Since both measurements are in the same range from 0 to 30 kPa, we assume that the error terms are identically and independently distributed for both outputs, and thus that variance, σ^2 , of the error terms ζ_k and η_k is equal. Since the two gasses are measured with different sensors, we assume that the two noise terms are independent. Several sensor principles are available for measuring O₂ concentrations in a practical setting, including gas chromatography, zirconium based sensors, paramagnetic sensors and fluorescence based optical sensors; CO₂ concentrations can be measured using gas chromatography, infrared absorption and chemical gas sensors.

Before considering experimental design for this model, we first checked whether the model parameters can be correctly identified at all. This is because it is

possible that two different model parameter values result in exactly the same output behavior, making it impossible to distinguish the true value of the model parameters from the data. We confirmed, using the STRIKE-GOLDD Matlab toolbox (Villaverde, Barreiro, and Papachristodoulou 2016), that our respiration and fermentation model is structurally identifiable even with constant input levels, that do not change in time. Our piecewise constant input profiles form a super-set of the set of constant input profiles, which thus ensures that our experimental designs will result in an identifiable model.

3.3.2 Prior Information

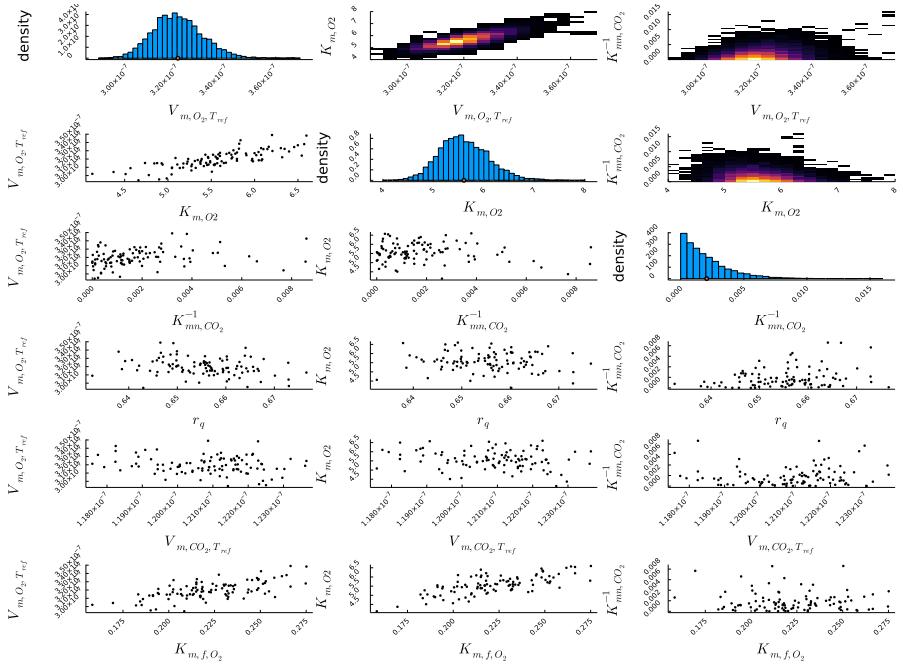
Optimal experimental design for the non-linear model introduced in Section 3.3.1 requires prior information, concerning the six respiration and fermentation parameters. Such prior information for pear respiration and fermentation can be found in Ho, Verboven, Verlinden, Lammertyn, et al. (2008). We cannot directly utilize the published results, as Ho, Verboven, Verlinden, Lammertyn, et al. (2008) only report confidence intervals for each individual parameter, but no correlations between estimates. To deal with this problem, we reanalyzed 50 time series data-sets, each containing O₂ and CO₂ measurements from a single jar, made available to us by the authors. We used a Bayesian data analysis technique to achieve this. More specifically, we utilized a Markov-chain Monte-Carlo method to re-estimate the parameters and quantify the uncertainty in the data (Betancourt 2017). The Markov chain stores values sampled from the posterior distribution of the parameters. The chain can thus be utilized for numerically approximating the expectation in the robust criterion in Equation (3.7). In other words, we use the MCMC chain as an input to the expression in Equation (3.8).

Because the data of Ho, Verboven, Verlinden, Lammertyn, et al. (2008) were collected at different temperatures, our analysis took into account the effect of temperature on the maximal respiration and fermentation rate by means of the Arrhenius equations:

$$V_{m,O_2} = V_{m,O_2,T_r} \exp \left(\frac{E_{a,O_2}}{\bar{R}} \left(\frac{1}{T_r} - \frac{1}{T} \right) \right),$$

$$V_{m,f,CO_2} = V_{m,f,CO_2,T_r} \exp \left(\frac{E_{a,CO_2}}{\bar{R}} \left(\frac{1}{T_r} - \frac{1}{T} \right) \right), \quad (3.13)$$

where V_{m,O_2,T_r} and V_{m,f,CO_2,T_r} are the maximal respiration rates at a reference temperature T_r of 293.15 K, and E_{a,O_2} and E_{a,CO_2} are activation energies that describe how the reaction rates increase with the temperature T . The different



This figure is continued on the next page.

temperatures in the data set influences the inhibition parameters in Equation (3.11). These activation energies are nuisance parameters in our computation of optimal input profiles in Section 3.4, as we only consider experiments at the reference temperature.

Another difference between the experiments of Ho, Verboven, Verlinden, Lammertyn, et al. (2008) and our experiments is that they used closed jars, instead of flow through experiments. Their system dynamics thus differ:

$$\begin{aligned} V_j \frac{d[O_2]}{dt} &= -m_p r_{O_2}(t), \\ V_j \frac{d[CO_2]}{dt} &= m_p r_{CO_2}(t). \end{aligned} \quad (3.14)$$

Figure 3.2 provides a summary of the results of our Bayesian analysis. On the diagonal of this figure, histograms of the Markov chain values of the six model parameters of interest (V_{m,O_2} , K_{m,O_2} , K_{mn,CO_2} , r_q , V_{m,f,CO_2} , K_{m,f,O_2}) are shown. Above the diagonal, two-dimensional heatmaps are shown, which visualize the correlation between the Markov chain values for pairs of model

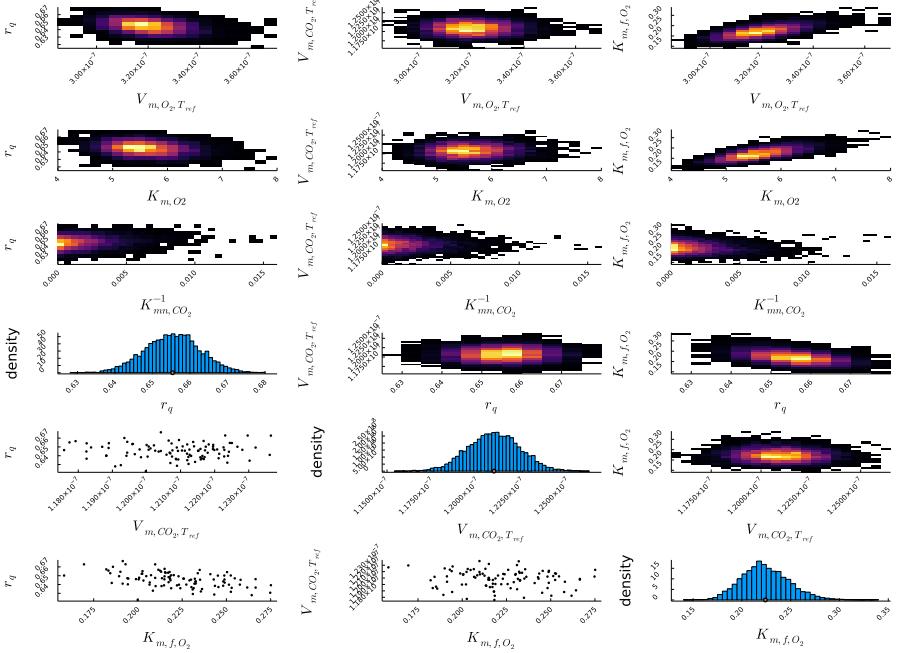


Figure 3.2: Summary of the Bayesian data analysis of the data in Ho, Verboven, Verlinden, Lammertyn, et al. (2008) for the parameters of interest. This figure is stretched over two pages. The diagonal contains histograms of the Markov chain values of the six model parameters of interest. The orange dots denote the means. Above the diagonal, two-dimensional heatmaps are shown, visualizing the correlation between the Markov chain values for pairs of model parameters. Similarly, below the diagonal, scatter plots for 100 Markov chain points are shown.

parameters. The figures below the diagonal provide similar information, but show only 100 pairs of values from the Markov chain. Histograms of the Markov chain values for the nuisance parameters E_{a,O_2} , E_{a,CO_2} , and σ are depicted in Figure 3.3. The estimation of the respiration inhibition parameter, K_{mn,CO_2} , from the available data was problematic. For this reason, we had to reparametrize the model using the inverse of K_{mn,CO_2} . As can be seen in the third histogram in Figure 3.2, K_{mn,CO_2}^{-1} takes values close to zero, which means K_{mn,CO_2} tends to infinity. The lack of information about K_{mn,CO_2}^{-1} can be explained as follows: to precisely estimate this parameter, both high O_2 and CO_2 concentrations are needed. A high O_2 concentration is required because otherwise there is no respiration that can be inhibited, and a high CO_2

concentration is required because otherwise the inhibition has a negligible effect. Data points in which both gasses possess a high concentration do not occur in the data of Ho, Verboven, Verlinden, Lammertyn, et al. (2008). The issues with the uncertainty about K_{mn,CO_2} were the catalyst for the development of our robust experimental design method. We found no other reliable way to quantify the uncertainty on this parameter, except working with a Markov chain. K_{m,f,O_2} is the second hardest parameter to identify, from the data of Ho, Verboven, Verlinden, Lammertyn, et al. (2008), because the O_2 concentration needs to be in a specific range for this parameter to have an effect on the outputs. An O_2 concentration much higher than K_{m,f,O_2} means no fermentation is happening at all, and an O_2 concentration much lower than K_{m,f,O_2} implies maximal fermentation. The third most difficult parameter to identify is K_{m,O_2} , because the model is only sensitive to this parameter when O_2 concentrations are close to the value of this parameter.

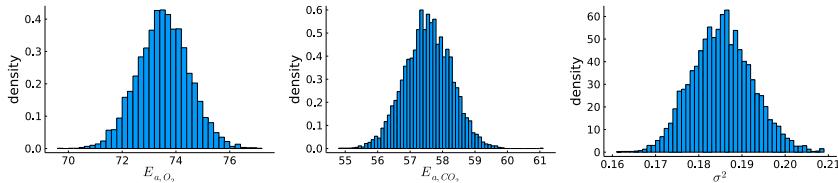


Figure 3.3: Summary of our Markov-chain Monte-Carlo analysis of the data in Ho, Verboven, Verlinden, Lammertyn, et al. (2008) for the nuisance parameters E_{a,O_2} , E_{a,CO_2} and σ^2 .

3.3.3 Markov chain Details

The data analysis was performed using the No U-Turn Sampling Markov-chain Monte-Carlo algorithm (Hoffman and Gelman 2014), as implemented in Turing.jl (Ge, Xu, and Ghahramani 2018). We ran 4 parallel Markov chains each starting from the maximum likelihood estimates of the parameters, and each chain comprises 1500 steps. We utilized a flat prior in the physically possible regions of the unknown parameters. For all parameters this means that only positive values are possible, with r_q being at most 1.

We used the posterior distribution resulting from this Bayesian analysis of the experiments in Ho, Verboven, Verlinden, Lammertyn, et al. (2008) as a prior distribution for the generation of the robust designs in this chapter. However, as calculating Bayesian optimal designs with the entire Markov chain is numerically quite intensive, we discarded the first 500 step, and thinned the remaining 1000 by a factor 40. This gave us 100 samples to calculate the Bayesian D-optimality

criterion in Equation (3.7). These are also the 100 values shown in the top right half of Figure 3.2.

3.4 Results

In our examples, we consider experiments lasting 24 h with measurements taken every 5 minutes, i.e. $t_e = 24$ h and $N = 288$. The maximum and minimum flow rates are equal to 1 l h^{-1} and 0.1 l h^{-1} , respectively. The input gas concentrations are allowed to vary between 0 kPa and 21 kPa. We did not use a minimum flow rate of 0 l h^{-1} , because at zero flow there is no difference in system response between a maximal or minimal gas input concentration. This implies that the design selection criteria from Equations (3.6) and (3.8) are flat in certain directions, causing numerical issues for gradient based optimizers. Working with a strictly positive minimum flow rate avoids this issue. Another reason to use strictly positive flows is to ensure that there is an outflow that can be measured.

3.4.1 Locally Optimal Designs for the Respiration and Fermentation Model

We take the average values of the Markov chains of the parameters of interest, as well as the average of the Markov chain values of the measurement variance σ^2 to evaluate the local D-criterion in Equation (3.7). These values are shown in Table 3.1, and indicated by a bullet in the histograms on the diagonal of Figure 3.2. We start by analyzing the effect of an increasing refinement of the discretization of the inputs $\mathbf{u}(t)$. In Figure 3.4, which shows the local D-criterion value as a function of the number of times the input signal is allowed to switch, we see that the D-criterion no longer improves noticeably after $M = 48$, which corresponds to a switch every half hour. The experimental design obtained with $M = 12$, which corresponds to a switch every two hours, already performs well. Therefore, in the remainder of this chapter, we consider O_2 , CO_2 and flow rate inputs that remain constant for two-hour time periods. The locally

V_{m,O_2}	K_{m,O_2}	K_{mn,CO_2}	r_q	V_{m,f,CO_2}	K_{m,f,O_2}	σ^2
$\mu\text{mol kg}^{-1}\text{ s}^{-1}$	kPa	kPa	-	$\mu\text{mol kg}^{-1}\text{ s}^{-1}$	kPa	$\text{mol}^2\text{ m}^{-6}$
0.320	5.61	484.4	0.655	0.121	0.224	0.186

Table 3.1: Parameters used for the optimization of the construction of the locally optimal designs.

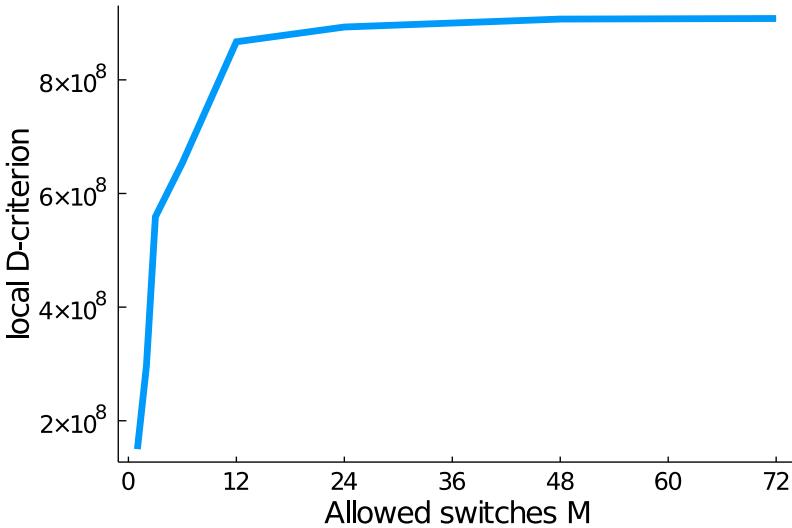


Figure 3.4: Convergence of the local D-optimality criterion for finer discretizations of the controls.

optimal design for the scenario in which the inputs are allowed to vary every two hours is shown in more detail in Figures 3.5a and 3.5b, together with the simulated outputs in Figure 3.5c, evaluated at the values of the model parameters used to optimize the design. The gas concentrations are converted to pressures to aid interpretation. During the interval from 2 h to 4 h, both O₂ and CO₂ are pumped into the jar, which causes high concentrations of the two gasses to be present at the same time. We noted before that, due to the lack of such conditions in Ho, Verboven, Verlinden, Lammertyn, et al. (2008), it was impossible to estimate K_{mn,CO_2} well from their data. However, the O₂ concentration cannot remain high throughout the entire experiment, as then there would be insufficient information about non-saturated respiration, i.e. to estimate K_{m,O_2} precisely. The O₂ concentration must decrease further to levels at which fermentation starts to occur for K_{m,f,O_2} to become estimable. This explains why air, with zero O₂ inlet concentrations is pumped into the jar in the interval between 10 h and 12 h. The CO₂ inlet concentration is also zero during that time interval due to the fact that the CO₂ concentration is not allowed to run up too high, as this compromises the ability to precisely determine the fermentation parameters. This can be intuitively understood by considering the system in an extreme scenario where the atmosphere in the jar consists entirely of CO₂. In this scenario, the outflow will also be pure CO₂ regardless of the values of the fermentation parameters. Information about the fermentation parameters is then only incorporated in the outflow rate, but this is

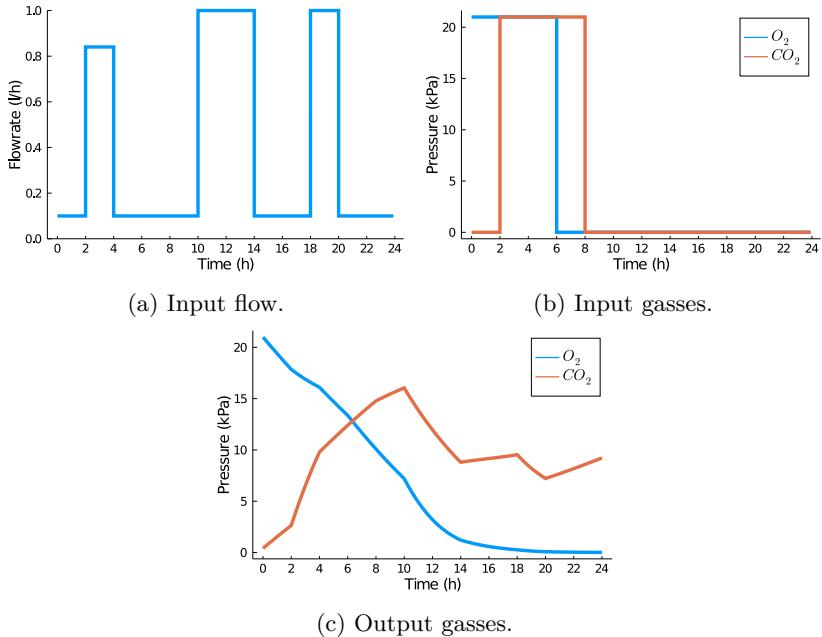
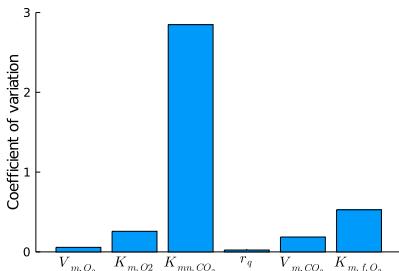
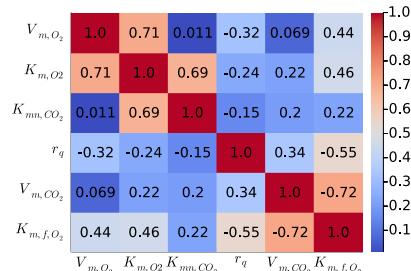


Figure 3.5: Locally optimal experiment and simulated output, with control input switches every two hours.

not a measured output. This illustrates how optimal experiments automatically take into account the specifics of the measurement setup. The pumping action at 18 h, again involving zero O_2 and CO_2 concentrations, is performed for similar reasons. Figure 3.6a shows the coefficients of variation of all six model parameters, defined as the ratio of the standard errors, as given by the diagonal elements of the inverse of the FIM, and the model parameter values used for the generation of the local optimal design. The three inhibition parameters (K_{m,O_2} , K_{mn,CO_2} and K_{m,f,O_2}) remain the most difficult ones to estimate. Now, however, in our experiment involving a single jar, instead of using the data from 50 jars, made available to us by Ho, Verboven, Verlinden, Lammertyn, et al. (2008), the model parameters can be estimated much more precisely. Figure 3.6b shows a map of the absolute values of the correlations of the estimates that would be obtained if the locally optimal experiment were to be used. Because there are no strong correlations between the first three parameters and the last three parameters, we can thus visualize the FIM in Equation (3.2) using two ellipsoids. The axes of the first ellipsoid are in the directions of the eigenvectors of the inverse of the top left quarter (a 3 by 3 matrix) of the FIM, and the lengths of the axes are the corresponding eigenvalues. The second ellipsoid



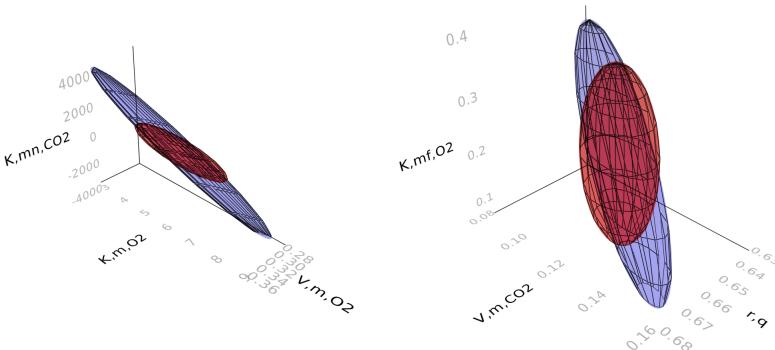
(a) coefficients of variation



(b) absolute value of correlations

Figure 3.6: Summary of the information gained from the locally optimal experiment in Figure 3.5, with local D-criterion equal to 8.67×10^8 .

is similarly based on the bottom right quarter. These ellipsoids are shown in Figures 3.7a and 3.7b for the experiment allowing changes in inputs every two hours, and are compared to the ellipsoids in blue resulting from the heuristic experimentation technique used by Ho, Verboven, Verlinden, Lammertyn, et al. (2008) with the volume of the jar, mass of pears, initial conditions, sampling times and total experimentation time equalized between the two methods. The volume of the ellipsoids of the locally optimal experiment are smaller.



(a) first three parameters
(V_{m,O_2} , K_{m,O_2} , K_{mn,CO_2}).

(b) last three parameters
(r_q , $V_{m,f}$, K_{m,f,O_2}).

Figure 3.7: 95% confidence ellipsoids comparing the locally optimal experimental design (red) and the heuristic experimental design technique from Ho, Verboven, Verlinden, Lammertyn, et al. (2008) (blue).

3.4.2 Bayesian Optimal Designs for the Respiration and Fermentation Model

We now continue by searching a more robust design than the locally optimal design in Figure 3.5. Instead of optimizing the determinant of the FIM in Equation (3.2) for the means of the Markov chain values, we optimize the mean determinant for a thinned version of the Markov chain. These values are graphically shown by the blue bullets in the lower left hand part of Figure 3.2, these values can also be found in Appendix A. The design found by maximizing the robust optimality criterion in Equation (3.8) is shown in Figures 3.8a and 3.8b. The simulated output for all values of the thinned Markov chain is shown in Figure 3.8c. The robust design exhibits several similarities to the locally optimal design, but also some key differences. The occurrence of a pumping action in the time interval between 2 h and 4 h as well as the interval between 10 h and 14 h is similar to that in the locally optimal design in Figure 3.5a. Furthermore, there are no differences in O_2 and CO_2 input concentrations between the robust and locally optimal design. One main difference between the robust experimental design and the locally optimal one is that, in the former, the first pumping action is less intense, i.e. the flow rate only amounts to 0.5 lh^{-1} , as opposed to 0.8 lh^{-1} , in the locally optimal design. A second main difference is the absence of the third pumping action in the robust design. We hypothesize this is because less pumping actions must occur to ensure a high CO_2 concentration when inhibition of respiration by CO_2 only happens at high CO_2 concentrations, which is possible, since we have little prior knowledge about K_{mn,CO_2} .

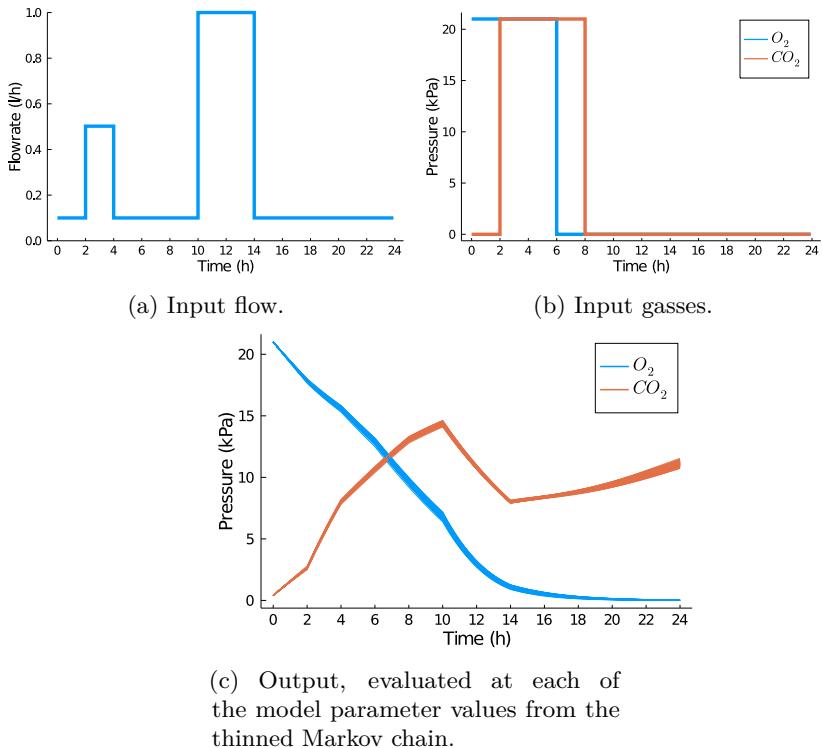


Figure 3.8: Robust optimal experiment and simulated output, obtained using 100 parameter values from the Markov chain and control input switching every two hours.

Figure 3.9 shows the difference in performance between the robust and the locally optimal experiment. More specifically, the histogram shows the difference between the determinant of the FIM in Equation (3.2) for both the robust and locally optimal design at all 6000 model parameter values from the Markov chain. For many possible model parameter values, the locally optimal and robust design perform almost equally well. However, the histogram clearly has a heavy right tail, showing that, for some parameter values, the Bayesian design performs significantly better. As the histogram does not possess a heavy left tail, the reverse is not true. The robust design is thus substantially less sensitive to the exact values of the model parameters. It therefore provides better guarantee for a highly informative experiment than the locally optimal design. Since this result holds for the entire Markov chain and not just the thinned version, we can be confident that the thinned version is sufficient to

summarize the prior uncertainty, and that the resulting robust design only works well for the specific model parameter values used to evaluate the optimality criterion in Equation (3.8). Finally, the positive mean value for the difference in determinants, indicated by the orange dot in Figure 3.9, proves that Bayesian design is expected to do better for the region of prior uncertainty.

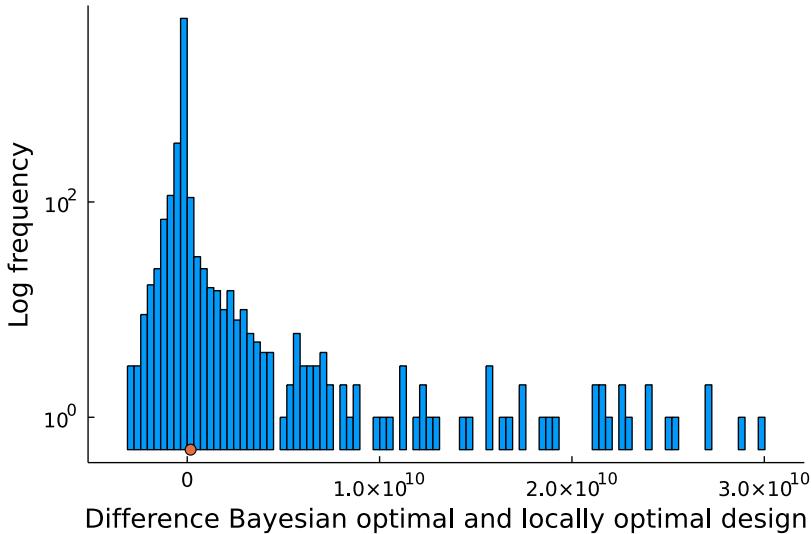


Figure 3.9: Histogram of difference in determinants of the FIM of the robust and locally optimal designs for 6000 parameters from the prior distribution. Positive values mean the robust design performs better.

In Table 3.2, we also compare our robust experimental design method based on a Markov-chain Monte-Carlo chain, which can approximate arbitrary distributions, to the robust experimental design method of Telen, Logist, et al. (2012), which uses 13 sigma points to summarize the uncertainty on the model parameters, and is thus computationally less intensive. The latter results in a design that performs worse than the locally optimal design, showing that sigma points do not always provide good summary of the uncertainty in the model parameters.

	Markov-chain Monte-Carlo optimal design	locally optimal design	sigma point optimal design
Robust D-criterion	1.26×10^{10}	1.22×10^{10}	9.83×10^9

Table 3.2: Parameters used for the optimization of the construction of the locally optimal designs

3.5 Discussion and Future Work

3.5.1 Alternative Information Criteria

In this chapter, we presented a robust experimental design methodology for dynamic models and applied that methodology to precisely estimate the respiration and fermentation parameters of pear fruit. We achieved this by quantifying the information content of the experiments using the determinant of the Fisher information matrix averaged over a prior distribution, represented by a Markov-chain. However, the Fisher information matrix is only an approximation of the inverse of the model parameter covariance matrix. Schenkendorf, Kremling, and Mangold (2009) suggest another approach to approximate the covariance matrix based on simulating multiple possible data-sets, and estimating the model parameters from each data-set. This approach is numerically much more intensive than an approach based on the FIM, and is therefore infeasible in the context of optimal experimental design. Another promising approach would be to quantify information based on the expected Kullback-Leibler divergence (KL-div) between the prior and posterior distributions (Lindley 1956). For experiments with a large number of observations, the expected KL-div is asymptotically equal to our robust D-optimality criterion. For a small number of measurements, the KL-div approach might be superior as it does not utilize normal approximations. One major downside of using the KL-div is its computational complexity, as it involves calculating a high dimensional integral over all possible outcomes \mathbf{y} of the experiment. This is generally done using a double loop Monte-Carlo integration method (Ryan et al. 2016). Some research has been done using this method for determining optimal sampling times of dynamic systems (Overstall, Woods, and Parker 2019), but this work does not consider optimal control. Therefore, the dynamic system in Overstall, Woods, and Parker (2019) does not need to be solved repeatedly for different control actions. Instead, only a single dynamic system needs to be solved that can be evaluated at different possible sampling times. Selection of input profiles based on the KL-div is considered in Liepe et al. (2013), but only for a small discrete set of possible input profiles. In contrast,

in our paper, we optimize the experimental design over a large continuous space of possible experimental designs. Recently, advances have been made in lowering the computational burden of the KL-div based approach by considering surrogate functions that approximate the KL-div based on polynomial chaos expansions (Paulson, Martin-Casas, and Mesbah 2019). Another approach to lower the computational burden is based on variational Bayesian techniques (Foster, Jankowiak, Bingham, et al. 2019), where the inner Monte Carlo loop is replaced by optimizing a variational distribution. This technique then allows for jointly optimizing the design and variational parameters (Foster, Jankowiak, O’Meara, et al. 2020). In future work, we will employ these variational Bayesian techniques for adaptive dynamic experiments, where the design is modified online as data is collected.

3.5.2 State Discretization

In this chapter, we discretized only the controls \mathbf{u} , but not the states \mathbf{x} . The discretization of states is applied using multiple shooting and collocation based dynamic optimization approaches (Biegler 2010). Generally, these methods lead to optimization problems that are faster to solve, as they do not require repeatedly solving differential equals, but which might lead to less accurate results. This is because we found that solving the dynamic system with model parameters values equal to those of the heavy right tail in Figure 3.9 requires many more steps than the other parameter values form the Markov chain. Thus unless the discretization is chosen fine enough, the approximation for the troublesome parameters in the tail might not suffice. Besides the faster computing time, one additional reason to consider discretization of the states would be the presence of additional constraints on the states. This is because violations of such constraints are difficult to check without a discretization. Our optimization problem did not involve any such additional state constraints. Nimmermeier et al. (2020) provide a more detailed discussion on the incorporation of constraints into optimal dynamic experiments.

3.5.3 Reverse Automatic Differentiation

We used forward mode automatic differentiation to calculate the sensitivities to the unknown parameters $\boldsymbol{\theta}$, required to obtain the FIM in Equation (3.2), and this was further nested to calculate the gradient of the D-criteria in Equations (3.6) and (3.8) with respect to the control parameters \mathbf{u} . Forward mode automatic differentiation generally performs well for functions with a small number of inputs, relative to the number of outputs. For our respiration model,

it thus was a good choice for calculating the unknown parameter sensitivities present in the FIM. Reverse mode automatic differentiation performs better for functions with many more outputs than inputs. It thus seems like a natural choice to calculate the gradients necessary for the optimization of the control parameters. However, there is currently not yet a mature implementation of reverse over forward mode automatic differentiation in the Julia ecosystem (Rackauckas, Ma, Dixit, et al. 2018).

3.6 Conclusion

This chapter presented a pioneering study about the usefulness of robust optimal dynamic experiments for non-linear modeling in postharvest research. Robust designs work well for a range of possible model parameter values. Robust design methods require the specification of a prior distribution for the model parameters. Current methods in the literature only work for parametric prior distributions. The available prior information about respiration and fermentation of pear fruit could not be adequately summarized using a parametric distribution, thus we developed a novel experimental design methodology based on a Markov-chain Monte-Carlo analysis of a prior data-set. This Markov chain can approximate arbitrary distributions. We used this methodology to construct robust D-optimal dynamic experimental designs for the estimation of the Michaelis-Menten respiration and fermentation parameters of Conference pear. These robust designs perform better than locally D-optimal designs, which are optimized using a single initial guess for the model parameters, and also outperform other robust methods.

Chapter 4

Adaptive and Robust Experimental Design for Linear Dynamic Models using the Kalman Filter

This chapter has been written by Arno Strouwen with feedback from Professors Goos and Nicolaï.

Source code available upon request.

Abstract

Current experimental design techniques for dynamic systems often only incorporate measurement noise, while dynamic systems also involve process noise. To construct experimental designs we need to quantify their information content. The Fisher information matrix is a popular tool to do so. Calculating the Fisher information matrix for linear dynamic systems with both process and measurement noise involves estimating the uncertain dynamic states using a Kalman filter. The Fisher information matrix, however, depends on the true but unknown model parameters. In this paper we combine two methods to solve this issue and develop a robust experimental design methodology. First, Bayesian experimental design averages the Fisher information matrix over a prior

distribution of possible model parameter values. Second, adaptive experimental design allows for this information to be updated as measurements are being gathered. This updated information is then used to adapt the remainder of the design.

4.1 Introduction

Control, optimization and analysis of dynamic systems are increasingly being performed using parametric models (Findeisen and Allgöwer 2002). High-quality data are needed to precisely identify these models. Optimal input design for dynamic systems deals with the cost-effective collection of these data (Goodwin and Payne 1977).

Most experimental design literature for precisely estimating model parameters of dynamic systems focuses on models with only measurement noise (Franceschini and Macchietto 2008), or on models with only process noise, when dealing with autoregressive models for time-series modeling (Hjalmarsson 2005; Pintelon and Schoukens 2012). The two previous chapters also only took into account measurement noise. Relatively little literature exists about designing informative experiments when both measurement and process noise are present. One approach that does combine process and measurement noise for experimental design is that of Telen, Houska, et al. (2013). These authors use a heuristic extension of the Fisher information matrix used by Franceschini and Macchietto (2008) to deal with process noise. Our approach differs as we use the formal definition of the Fisher information matrix, based on the variance of the score, which is the gradient of the log-likelihood function. The main challenge that arises in this approach is that estimating the unknown model parameters also requires the hidden dynamic states to be estimated.

Estimating such hidden states for continuous-time non-linear stochastic differential equations generally has no analytical solution (Särkkä and Solin 2019). In this chapter, we focus on linear discrete-time dynamic systems with both measurement and process noise. For these models analytical results exist. Particularly, the Kalman filter is used to estimate the dynamic state. The Kalman filter has hardly been used in the context of optimal experiments. Titterington (1980) uses the steady state Kalman filter to construct continuous optimal designs, which are asymptotically optimal when a large amount of data is gathered. This is in contrast to exact designs, which are optimized for a finite number of measurements, and which we use in this chapter. Because of our focus on a finite number of measurements, our chapter also does not rely on the steady state prediction error covariance. Stojanovic et al. (2016) use

a robust Kalman filter to generate optimal inputs for autoregressive models with non-Gaussian noise. Instead of autoregressive models, we work with linear state space models, where the matrices describing such a state space model may depend on model parameters that must be estimated as precisely as possible.

The Fisher information matrix (FIM) is a popular tool to quantify the quality of an experiment, as it is related to the inverse of the covariance matrix of the model parameter estimates (Fedorov and Leonov 2013). An informative experiment makes a scalar measure of the FIM as large as possible. The major issue with optimal experimental design is the dependence of the FIM, and thus also the optimal inputs for the experiment, on the true, but unknown, model parameters. This presents us with a circular problem as the experiment is needed to precisely estimate the parameters. Locally optimal design, where inputs are optimized for a single initial guess for the parameters, is the traditional method to deal with this issue (Atkinson, Donev, and Tobias 2007a). However, this method can be very sensitive to the single initial guess. Generally, there exist two directions to improve on the locally optimal design method, namely robustifying the experiment against the uncertainty in the model parameters and making the experiment adaptive (Walter and Pronzato 1994).

Robustifying the experiment can be achieved in various ways. One popular approach is min-max experimental design (Körkel et al. 2004). In this method, the experiment is optimized under the assumption of a worst case scenario. This means that Fisher information matrices are calculated for all elements of a set of possible parameter values and the quality of the experiment is judged based on the least informative matrix in this set. This guarantees that, regardless the true parameter values, the experiment will always have a minimal information content. Another popular approach is Bayesian optimal design, which uses an expected value approach (Chaloner and Verdinelli 1995). A prior distribution for the model parameters is then used, and the experiment is designed to perform well on average for this prior.

In the second approach to improve on the locally optimal design method, the measurements obtained during the execution of the experiment are used to improve on the initial guess of the model parameters. This is called adaptive or sequential experimental design. A new locally optimal design is then based on the updated model parameters and this process is repeated.

In our work, we combine the Bayesian robustification approach with adaptive experimental design and construct optimal adaptive Bayesian experiments. After every measurement, we update our knowledge of the unknown parameters using Bayesian statistics. A new Bayesian optimal design for the remainder

of the experiment is then constructed based on the updated prior distribution. Adaptive designs have an additional benefit for dynamic systems in the presence of process noise. This is because it is difficult to predict the dynamic state of such systems far into the future, because of the process noise. This causes these future measurements to be uninformative and contributing little to the Fisher information matrix. When adaptively designing an experiment, the estimate of the dynamic state based on the already gathered data will also reduce the prediction variance of future observations, meaning these future observations become more informative.

4.2 Modeling the Information Content of a Dynamic Experiment

4.2.1 The Model

Our goal is to find dynamic inputs $\mathbf{u}_{1:T}$ which lead to a precise estimation of the static model parameters $\boldsymbol{\theta}$ of a linear time-invariant discrete-time dynamic system with Gaussian noise,

$$\begin{aligned} \mathbf{x}_k &= F(\boldsymbol{\theta})\mathbf{x}_{k-1} + B(\boldsymbol{\theta})\mathbf{u}_k + \mathbf{w}_k, & 0 < k \leq T \\ \mathbf{y}_k &= H(\boldsymbol{\theta})\mathbf{x}_k + \mathbf{v}_k. \end{aligned} \tag{4.1}$$

In these equations, \mathbf{y}_k represents the measured outputs at time-step k . We assume that the experiment ends after T time-steps, and thus k can range from 1 to T . The measured output is dependent on the dynamic states \mathbf{x}_k through the output matrix $H(\boldsymbol{\theta})$. These states completely determine the stochastic evolution of the system over time. The transition of the states from one time-step to the next is impacted by the state matrix $F(\boldsymbol{\theta})$, as well as the inputs at that time-step, \mathbf{u}_k , through the input matrix $B(\boldsymbol{\theta})$. All three of these matrices $F(\boldsymbol{\theta})$, $B(\boldsymbol{\theta})$ and $H(\boldsymbol{\theta})$ can depend on the model parameters $\boldsymbol{\theta}$.

Noise is present in both the measurements and the state transitions. We make the following assumptions about the measurement noise \mathbf{v}_k and the uncontrollable and unobserved process noise \mathbf{w}_k at time-step k :

$$\begin{aligned} \mathbf{w}_k &\sim \mathcal{N}(\mathbf{0}, Q(\boldsymbol{\theta})), \\ \mathbf{v}_k &\sim \mathcal{N}(\mathbf{0}, R(\boldsymbol{\theta})), \\ \text{Covar}(\mathbf{w}_k, \mathbf{v}_l) &= \mathbf{0}, \\ \text{Covar}(\mathbf{w}_k, \mathbf{w}_l) = \text{Covar}(\mathbf{v}_k, \mathbf{v}_l) &= \mathbf{0}, \quad k \neq l. \end{aligned} \tag{4.2}$$

We thus assume that \mathbf{v}_k and \mathbf{w}_k both follow a multivariate normal distribution with zero mean and covariance matrices equal to $Q(\boldsymbol{\theta})$ and $R(\boldsymbol{\theta})$, respectively. These covariance matrices may also depend on the unknown static parameters $\boldsymbol{\theta}$. The measurement and process noise are independent of each other, and there is also no correlation over time, neither for measurement nor process noise.

The initial state of the system is also assumed to be multivariate normally distributed, with mean \mathbf{m}_0 and covariance matrix P_0 . This state is independent of all later noise. So,

$$\begin{aligned} \mathbf{x}_0 &\sim \mathcal{N}(\mathbf{m}_0, P_0), \\ \text{Covar}(\mathbf{x}_0, \mathbf{v}_k) &= \text{Covar}(\mathbf{x}_0, \mathbf{w}_k) = \mathbf{0}. \end{aligned} \tag{4.3}$$

4.2.2 Parameter Estimation

Before presenting our experimental design methodology, we first discuss how to estimate the model parameters $\boldsymbol{\theta}$ of the model in Equation (4.1). One popular approach for parameter estimation is based on the likelihood of the unknown parameters given the observations $\mathbf{y}_{1:T}$. Here, $\mathbf{y}_{k:l}$ denotes the measured outputs from time-step k to l , with both endpoints included and $k \leq l$. We use a similar notation for other vectors. The log-likelihood after k observations have been collected can be computed with the recursive factorization

$$L_k(\boldsymbol{\theta}, \mathbf{u}_{1:k}, \mathbf{y}_{1:k}) = \log p(\mathbf{y}_{1:k} | \boldsymbol{\theta}, \mathbf{u}_{1:k}) \tag{4.4}$$

$$= \log p(\mathbf{y}_k | \boldsymbol{\theta}, \mathbf{u}_{1:k}, \mathbf{y}_{1:k-1}) + \log p(\mathbf{y}_{1:k-1} | \boldsymbol{\theta}, \mathbf{u}_{1:k-1}). \tag{4.5}$$

The first term in this expression can further be computed as

$$p(\mathbf{y}_k | \boldsymbol{\theta}, \mathbf{u}_{1:k}, \mathbf{y}_{1:k-1}) = \int p(\mathbf{y}_k | \mathbf{x}_k, \boldsymbol{\theta}) p(\mathbf{x}_k | \boldsymbol{\theta}, \mathbf{u}_{1:k}, \mathbf{y}_{1:k-1}) d\mathbf{x}_k. \tag{4.6}$$

In this equation, the first factor of the integrand $p(\mathbf{y}_k | \mathbf{x}_k, \boldsymbol{\theta})$ corresponds to $\mathcal{N}(\mathbf{H}(\boldsymbol{\theta})\mathbf{x}_k, R(\boldsymbol{\theta}))$, while the second factor, $p(\mathbf{x}_k | \boldsymbol{\theta}, \mathbf{u}_{1:k}, \mathbf{y}_{1:k-1})$, called the state predictive distribution, can be calculated as

$$p(\mathbf{x}_k | \boldsymbol{\theta}, \mathbf{u}_{1:k}, \mathbf{y}_{1:k-1}) = \int p(\mathbf{x}_k | \mathbf{x}_{k-1}, \boldsymbol{\theta}, \mathbf{u}_k) p(\mathbf{x}_{k-1} | \boldsymbol{\theta}, \mathbf{u}_{1:k-1}, \mathbf{y}_{1:k-1}) d\mathbf{x}_{k-1}. \tag{4.7}$$

In this equation, the first factor of the integrand, $p(\mathbf{x}_k | \mathbf{x}_{k-1}, \boldsymbol{\theta}, \mathbf{u}_k)$, can be calculated as $\mathcal{N}(\mathbf{F}(\boldsymbol{\theta})\mathbf{x}_{k-1} + \mathbf{B}(\boldsymbol{\theta})\mathbf{u}_k, Q(\boldsymbol{\theta}))$, while the second factor, $p(\mathbf{x}_{k-1} | \boldsymbol{\theta}, \mathbf{u}_{1:k-1}, \mathbf{y}_{1:k-1})$, is called the state filtering distribution. This

distribution can be computed by Bayes' law,

$$p(\mathbf{x}_k | \boldsymbol{\theta}, \mathbf{u}_{1:k}, \mathbf{y}_{1:k}) = \frac{p(\mathbf{y}_k | \mathbf{x}_k, \boldsymbol{\theta}) p(\mathbf{x}_k | \boldsymbol{\theta}, \mathbf{u}_{1:k}, \mathbf{y}_{1:k-1})}{p(\mathbf{y}_k | \boldsymbol{\theta}, \mathbf{u}_{1:k}, \mathbf{y}_{1:k-1})}. \quad (4.8)$$

The denominator of this fraction can be thought of as a normalization factor, ensuring that the state filtering distribution for time-step k integrates to one. The second factor in the numerator is the state predictive distribution. The state filtering distribution thus depends on the state predictive distribution, which in turn depends on the state predictive distribution at the previous time-step. These two recurring equations are known as the Bayesian filtering equations. If the state filtering distribution at time-step $k - 1$ is normally distributed as $p(\mathbf{x}_{k-1} | \boldsymbol{\theta}, \mathbf{u}_{1:k-1}, \mathbf{y}_{1:k-1}) = \mathcal{N}(\mathbf{m}_{k-1}, P_{k-1})$, then the state predictive distribution is also normally distributed $p(\mathbf{x}_k | \boldsymbol{\theta}, \mathbf{u}_{1:k}, \mathbf{y}_{1:k-1}) = \mathcal{N}(F(\boldsymbol{\theta})\mathbf{m}_{k-1} + B(\boldsymbol{\theta})\mathbf{u}_k, F(\boldsymbol{\theta})P_{k-1}F'(\boldsymbol{\theta}) + Q(\boldsymbol{\theta}))$. If the state predictive distribution is normally distributed as $p(\mathbf{x}_k | \boldsymbol{\theta}, \mathbf{u}_{1:k}, \mathbf{y}_{1:k-1}) = \mathcal{N}(\mathbf{m}_k^-, P_k^-)$, then the joint state and measurement prediction distribution is also normally distributed,

$$p(\mathbf{x}_k, \mathbf{y}_k | \boldsymbol{\theta}, \mathbf{u}_{1:k}, \mathbf{y}_{1:k-1}) = \mathcal{N}\left(\begin{bmatrix} \mathbf{m}_k^- \\ H(\boldsymbol{\theta})\mathbf{m}_k^- \end{bmatrix}, \begin{bmatrix} P_k^- & P_k^- H(\boldsymbol{\theta})' \\ H(\boldsymbol{\theta})P_k^- & H(\boldsymbol{\theta})P_k^- H(\boldsymbol{\theta})' + R \end{bmatrix}\right). \quad (4.9)$$

The state filtering distribution is then also normal and can be calculated from the conditional distribution of a partitioned multivariate normal distribution (Von Mises 2014):

$$p(\mathbf{x}_k | \boldsymbol{\theta}, \mathbf{u}_{1:k}, \mathbf{y}_{1:k}) = \mathcal{N}(\mathbf{m}_k, P_k),$$

$$\begin{aligned} \mathbf{m}_k &= \mathbf{m}_k^- + P_k^- H(\boldsymbol{\theta})'(H(\boldsymbol{\theta})P_k^- H'(\boldsymbol{\theta}) + R(\boldsymbol{\theta}))^{-1}(\mathbf{y}_k - H(\boldsymbol{\theta})\mathbf{m}_k^-), \\ P_k &= P_k^- - P_k^- H(\boldsymbol{\theta})'(H(\boldsymbol{\theta})P_k^- H'(\boldsymbol{\theta}) + R(\boldsymbol{\theta}))^{-1}H(\boldsymbol{\theta})P_k^-. \end{aligned} \quad (4.10)$$

The first state predictive distribution $p(\mathbf{x}_1 | \boldsymbol{\theta}, \mathbf{u}_1)$ is normally distributed since the initial state distribution $p(\mathbf{x}_0)$ is normally distributed, and thus all subsequent state predictive and filtering distributions are normally distributed as well. The above derivation for state predictive and filtering distributions is equivalent to the Kalman filter, with an explicit dependence on the model

parameters $\boldsymbol{\theta}$ (Särkkä 2013). The recursion can thus also be written as:

$$\begin{aligned}
\mathbf{m}_k^- &= F(\boldsymbol{\theta})\mathbf{m}_{k-1} + B(\boldsymbol{\theta})\mathbf{u}_k, \\
P_k^- &= F(\boldsymbol{\theta})P_{k-1}F(\boldsymbol{\theta})' + Q(\boldsymbol{\theta}), \\
p(\mathbf{x}_k|\boldsymbol{\theta}, \mathbf{u}_{1:k}, \mathbf{y}_{1:k-1}) &= \mathcal{N}(\mathbf{m}_k^-, P_k^-), \\
\mathbf{v}_k &= \mathbf{y}_k - H(\boldsymbol{\theta})\mathbf{m}_k^-, \\
S_k &= H(\boldsymbol{\theta})P_k^-H(\boldsymbol{\theta})' + R(\boldsymbol{\theta}), \\
K_k &= P_k^-H(\boldsymbol{\theta})'S_k^{-1}, \\
\mathbf{m}_k &= \mathbf{m}_k^- + K_k\mathbf{v}_k, \\
P_k &= P_k^- - K_kS_kK', \\
p(\mathbf{x}_k|\boldsymbol{\theta}, \mathbf{u}_{1:k}, \mathbf{y}_{1:k}) &= \mathcal{N}(\mathbf{m}_k, P_k).
\end{aligned} \tag{4.11}$$

In these equations \mathbf{v}_k , S_k and K_k are called the innovation gain residual, innovation gain covariance and optimal Kalman gain, respectively. The Kalman filter recurses back to the initial state distribution in Equation (4.3). Since we know that $p(\mathbf{y}_k|\mathbf{x}_k, \boldsymbol{\theta}) = \mathcal{N}(H(\boldsymbol{\theta})\mathbf{x}_k, R(\boldsymbol{\theta}))$ and $p(\mathbf{x}_k|\boldsymbol{\theta}, \mathbf{u}_{1:k}, \mathbf{y}_{1:k-1}) = \mathcal{N}(\mathbf{m}_k^-, P_k^-)$, it is easy to see that $p(\mathbf{y}_k|\boldsymbol{\theta}, \mathbf{u}_{1:k}, \mathbf{y}_{1:k-1}) = \mathcal{N}(H(\boldsymbol{\theta})\mathbf{m}_k^-, H(\boldsymbol{\theta})P_k^-H'(\boldsymbol{\theta}) + R(\boldsymbol{\theta}))$. This leads to the following expression for the log-likelihood of the model parameters:

$$L_k(\boldsymbol{\theta}, \mathbf{u}_{1:k}, \mathbf{y}_{1:k}) = L_{k-1}(\boldsymbol{\theta}, \mathbf{u}_{1:k-1}, \mathbf{y}_{1:k-1}) - \frac{1}{2} \log |2\pi S_k| - \frac{1}{2} \mathbf{v}_k' S_k^{-1} \mathbf{v}_k, \tag{4.12}$$

where S_k and \mathbf{v}_k come from the Kalman filter recursion. The likelihood at a certain model parameter value $\boldsymbol{\theta}$ can thus be updated at every time-step by running a Kalman filter, with that particular value of $\boldsymbol{\theta}$.

4.2.3 The Fisher Information Matrix

One common approach to quantify the quality of the inputs $\mathbf{u}_{1:T}$ for precisely estimating the static parameters $\boldsymbol{\theta}$ is the expected Fisher information matrix

(FIM):

$$\begin{aligned}\mathcal{I}(\boldsymbol{\theta}, \mathbf{u}_{1:T}) &= E_{\mathbf{y}_{1:T}|\boldsymbol{\theta}, \mathbf{u}_{1:T}} \left(\frac{\partial \log p(\mathbf{y}_{1:T}|\boldsymbol{\theta}, \mathbf{u}_{1:T})}{\partial \boldsymbol{\theta}} \frac{\partial \log p(\mathbf{y}_{1:T}|\boldsymbol{\theta}, \mathbf{u}_{1:T})}{\partial \boldsymbol{\theta}}' \right), \\ &= -E_{\mathbf{y}_{1:T}|\boldsymbol{\theta}, \mathbf{u}_{1:T}} \left(\frac{\partial^2 \log p(\mathbf{y}_{1:T}|\boldsymbol{\theta}, \mathbf{u}_{1:T})}{\partial \boldsymbol{\theta}^2} \right).\end{aligned}\tag{4.13}$$

In this equation, $\mathbf{y}_{1:T}|\boldsymbol{\theta}, \mathbf{u}_{1:T}$ denotes the joint distribution of all the measurements, given the parameters $\boldsymbol{\theta}$ and the inputs $\mathbf{u}_{1:T}$. In addition, $\frac{\partial \log p(\mathbf{y}_{1:T}|\boldsymbol{\theta}, \mathbf{u}_{1:T})}{\partial \boldsymbol{\theta}}$ is the gradient of the log-likelihood, and $\frac{\partial^2 \log p(\mathbf{y}_{1:T}|\boldsymbol{\theta}, \mathbf{u}_{1:T})}{\partial \boldsymbol{\theta}^2}$ is the Hessian matrix of the log-likelihood. Formally, the Cramér-Rao bound states that the inverse of the expected FIM is a lower bound, by Loewner ordering, of the covariance matrix of an unbiased estimator of $\boldsymbol{\theta}$. This lower bound can be visualized by an hyperellipsoid. The directions and lengths of the principal axes are determined by the eigenvectors and eigenvalues of the inverse of the expected FIM, respectively. The inputs $\mathbf{u}_{1:T}$ should thus be chosen such that this hyperellipsoid is as small as possible.

We can also explain the expected FIM in the following way: we want to chose inputs $\mathbf{u}_{1:T}$ such that the true parameter $\boldsymbol{\theta}$ will fit the obtained data well, while other parameters should fit poorly. This is equivalent to saying that we want the log-likelihood function $\log p(\mathbf{y}_{1:T}|\boldsymbol{\theta}, \mathbf{u}_{1:T})$ to exhibit a sharp peak around the true model parameter value $\boldsymbol{\theta}$. When the Hessian matrix in the FIM, has large negative eigenvalues, the peak is sharp. Of course, when planning the experiment we do not know the outcomes $\mathbf{y}_{1:T}$ of the experiment yet. Therefore, the expectation should be taken with regard to all possible outcomes. So, it requires solving a high-dimensional integral.

Calculating the expected FIM for arbitrary non-linear models is often intractable, because generally no analytical results are available for the high-dimensional integral that is involved in this calculation. These integrals are then often numerically approximated using Monte Carlo methods (Ryan et al. 2016). However, our model in Equation (4.1) only contains linear transformations, and multivariate normal distributions remain normal under such transformations. As a result, the measurements $\mathbf{y}_{1:T}|\boldsymbol{\theta}, \mathbf{u}_{1:T}$ also follow a multivariate normal distribution. More specifically, the expected value and covariance of the measurements can be calculated by using following recursion relations, as adapted from Cavanaugh and Shumway (1996) to allow for models with control

inputs $\mathbf{u}_{1:T}$:

$$\begin{aligned}
E(\mathbf{y}_r|\boldsymbol{\theta}, \mathbf{u}_{1:T}) &= H(\boldsymbol{\theta})E(\mathbf{x}_r|\boldsymbol{\theta}, \mathbf{u}_{1:T}), \\
E(\mathbf{x}_r|\boldsymbol{\theta}, \mathbf{u}_{1:T}) &= F(\boldsymbol{\theta})E(\mathbf{x}_{r-1}|\boldsymbol{\theta}, \mathbf{u}_{1:T}) + B(\boldsymbol{\theta})\mathbf{u}_r, \\
E(\mathbf{x}_0|\boldsymbol{\theta}, \mathbf{u}_{1:T}) &= \mathbf{m}_0, \\
\text{Var}(\mathbf{y}_r|\boldsymbol{\theta}, \mathbf{u}_{1:T}) &= H(\boldsymbol{\theta}) \text{Var}(\mathbf{x}_r|\boldsymbol{\theta}, \mathbf{u}_{1:T})H(\boldsymbol{\theta})' + R(\boldsymbol{\theta}), \\
\text{Covar}(\mathbf{y}_r|\boldsymbol{\theta}, \mathbf{u}_{1:T}; \mathbf{y}_s|\boldsymbol{\theta}, \mathbf{u}_{1:T}) &= H(\boldsymbol{\theta})F^{r-s} \text{Var}(\mathbf{x}_s|\boldsymbol{\theta}, \mathbf{u}_{1:T})H(\boldsymbol{\theta})', \quad \forall r > s, \\
\text{Covar}(\mathbf{y}_r|\boldsymbol{\theta}, \mathbf{u}_{1:T}; \mathbf{y}_s|\boldsymbol{\theta}, \mathbf{u}_{1:T}) &= \text{Covar}(\mathbf{y}_s|\boldsymbol{\theta}, \mathbf{u}_{1:T}; \mathbf{y}_r|\boldsymbol{\theta}, \mathbf{u}_{1:T})', \quad \forall r < s, \\
\text{Var}(\mathbf{x}_r|\boldsymbol{\theta}, \mathbf{u}_{1:T}) &= F(\boldsymbol{\theta}) \text{Var}(\mathbf{x}_{r-1}|\boldsymbol{\theta}, \mathbf{u}_{1:T})F(\boldsymbol{\theta})' + Q(\boldsymbol{\theta}), \\
\text{Var}(\mathbf{x}_0|\boldsymbol{\theta}, \mathbf{u}_{1:T}) &= P_0.
\end{aligned} \tag{4.14}$$

The names Var and Covar are somewhat arbitrary in these equations. For example, if \mathbf{y}_k is bivariate, then the matrix $\text{Var } \mathbf{y}_k$ is a two by two matrix. We make the distinction between Var and Covar to stress the correlation of measurements over time.

The expected FIM for multivariate normal data is well known (Fedorov and Leonov 2013), its $[i, j]$ th element is

$$\begin{aligned}
\mathcal{I}(\boldsymbol{\theta}, \mathbf{u}_{1:T})[i, j] &= \frac{\partial E(\mathbf{y}_{1:T}|\boldsymbol{\theta}, \mathbf{u}_{1:T})'}{\partial \boldsymbol{\theta}[i]} \text{Covar}(\mathbf{y}_{1:T}|\boldsymbol{\theta}, \mathbf{u}_{1:T})^{-1} \frac{\partial E(\mathbf{y}_{1:T}|\boldsymbol{\theta}, \mathbf{u}_{1:T})}{\partial \boldsymbol{\theta}[j]} \\
&+ \\
&\frac{1}{2} \text{tr} \left(\text{Covar}(\mathbf{y}_{1:T}|\boldsymbol{\theta}, \mathbf{u}_{1:T})^{-1} \frac{\partial \text{Covar}(\mathbf{y}_{1:T}|\boldsymbol{\theta}, \mathbf{u}_{1:T})}{\partial \boldsymbol{\theta}[i]} \right. \\
&\left. \text{Covar}(\mathbf{y}_{1:T}|\boldsymbol{\theta}, \mathbf{u}_{1:T})^{-1} \frac{\partial \text{Covar}(\mathbf{y}_{1:T}|\boldsymbol{\theta}, \mathbf{u}_{1:T})}{\partial \boldsymbol{\theta}[j]} \right).
\end{aligned} \tag{4.15}$$

In this equation, square brackets are used to select an element of a vector or matrix. This equation does not only involve the expectation and covariance of all the observations, but also the derivative of these quantities with respect to the parameters $\boldsymbol{\theta}$. Similar recursions as in Equation (4.14) exist for these parameter sensitivities. We do not explicitly state these recursions as we calculate them by applying forward mode automatic differentiation on this recursion; see the numerical details in Section 4.4.

4.3 D-optimal Experimental Design

To find an optimal experimental design, we thus have to optimize the inputs $\mathbf{u}_{1:T}$ such that the expected FIM is as large as possible, as this leads to precise model parameter estimates. But generally, the expected FIM is not a scalar, and we thus need to define what constitutes a large matrix. Since the definition of the expected FIM involves Loewner ordering, it seems natural to also use this ordering to compare the quality of different inputs. However, Fedorov and Leonov (2013) demonstrate why it is impossible to directly use this partial ordering of positive semi-definite matrices, and that instead a scalar function of the expected FIM is needed. A popular choice is the determinant of the expected FIM, $|\mathcal{I}(\boldsymbol{\theta}, \mathbf{u}_{1:T})|$. This criterion is known as D-optimality and it is related to the inverse of the volume of the confidence hyperellipsoid.

4.3.1 Locally Optimal Experimental Design

One of the main difficulties in experimental design is the dependence of the expected FIM in Equation (4.13) on the true parameters of the system, which are exactly those parameters that the experiment should inform us about. This leads to a circular problem. The simplest way to deal with this issue is by using a single initial guess $\boldsymbol{\theta}^*$ for the parameters at the start of the experiment, and to compute D-optimal designs as

$$\arg \max_{\mathbf{u}_{\min} \leq \mathbf{u}_{1:T} \leq \mathbf{u}_{\max}} |\mathcal{I}(\boldsymbol{\theta}^*, \mathbf{u}_{1:T})|. \quad (4.16)$$

This method is called locally optimal design as the design only performs well if this single initial guess for the model parameters is close to the true value. In Equation (4.16), \mathbf{u}_{\min} and \mathbf{u}_{\max} are the minimal and maximal allowed control values, respectively.

4.3.2 Robust Experiments

To make the design more robust, so that it provides much information if the initial guess is not very close to the true value of the model parameters, we can replace the single initial guess $\boldsymbol{\theta}^*$ with a prior probability distribution $p(\boldsymbol{\theta})$, which represents our knowledge of possible values of $\boldsymbol{\theta}$ before the experiment has started. We want the experiment to perform well over the parameter values in the domain of $p(\boldsymbol{\theta})$, where the most likely parameter values have the largest weight. Averaging the determinant of the expected FIM over this prior

distribution and then optimizing this average achieves this:

$$\arg \max_{\mathbf{u}_{\min} \leq \mathbf{u}_{1:T} \leq \mathbf{u}_{\max}} \int |\mathcal{I}(\boldsymbol{\theta}, \mathbf{u}_{1:T})| p(\boldsymbol{\theta}) d\boldsymbol{\theta} \approx \arg \max_{\mathbf{u}_{\min} \leq \mathbf{u}_{1:T} \leq \mathbf{u}_{\max}} \frac{1}{N} \sum_{i=1}^N |\mathcal{I}(\boldsymbol{\theta}^i, \mathbf{u}_{1:T})|,$$

draw $\boldsymbol{\theta}^i$ from $p(\boldsymbol{\theta})$. (4.17)

So, the expectation is approximated by Monte Carlo integration with N draws from the prior distribution $p(\boldsymbol{\theta})$, each draw having the same weight, $\frac{1}{N}$. Besides Monte Carlo integration, other methods to numerically calculate this robust D-criterion exist, such as the sigma-point based method of Telen, Logist, et al. (2012). However, these methods complicate the updating of the weights for adaptive experimental design in the following section, and might not be compatible with the jittering described in the discussion section.

Due to the use of a prior distribution, this technique is also called pseudo-Bayesian optimal design (Chaloner and Verdinelli 1995). We want to stress that the optimality criterion is only pseudo-Bayesian, and not fully Bayesian. This is because, while the prior information is used to construct the inputs, it is not directly used to influence the estimation of the parameters. Only the information coming from the measurements acquired from the experiment is incorporated in the information criterion.

4.3.3 Adaptive Experiments

Concept

In the local optimal design criterion in Equation (4.16) and the pseudo-Bayesian optimal design criterion in Equation (4.17), prior information was only incorporated at the start of the experiment. But as soon as the experiment has started, knowledge is accumulating. That additional information can be exploited to optimize the remainder of the experiment. To formalize this, we now assume that we have already performed an experiment with inputs $\mathbf{u}_{1:k}$ and measured outputs $\mathbf{y}_{1:k}$. These measurements are used to form an updated prior distribution after k measurements, $p(\boldsymbol{\theta}|\mathbf{u}_{1:k}, \mathbf{y}_{1:k})$, which represents our belief in the possible values of the model parameters $\boldsymbol{\theta}$ given the additional information the first k measurements contain. We use this updated prior to optimize the remaining inputs $\mathbf{u}_{k+1:T}$ of the experiment:

$$\arg \max_{\mathbf{u}_{\min} \leq \mathbf{u}_{k+1:T} \leq \mathbf{u}_{\max}} \int |\mathcal{I}(\boldsymbol{\theta}, \mathbf{u}_{1:T})| p(\boldsymbol{\theta}|\mathbf{u}_{1:k}, \mathbf{y}_{1:k}) d\boldsymbol{\theta}.$$

Computational Challenges

This approach, however, has the drawback that optimizing the remaining $N - k$ input vectors for every time-step is computationally much too expensive, especially at the beginning of the experiment, as the behavior of the system then has to be predicted far into the future, using the recursions in equation (4.14). To reduce the computational burden, we only optimize the expected FIM for the next e measurements:

$$\arg \max_{\mathbf{u}_{\min} \leq \mathbf{u}_{k+1:k+e} \leq \mathbf{u}_{\max}} \int |\mathcal{I}(\boldsymbol{\theta}, \mathbf{u}_{1:k+e})| p(\boldsymbol{\theta} | \mathbf{u}_{1:k}, \mathbf{y}_{1:k}) d\boldsymbol{\theta}.$$

While this criterion does not require predicting very far into the future, due to the moving horizon $k + 1 : k + e$, it is still problematic for online computation, as the time and memory required in calculating the expected FIM increases at every time-step, as the dimension of the covariance matrix of $\mathbf{y}_{1:k+e}$ keeps growing (Cavanaugh and Shumway 1996). This is because, the expected FIM in Equation (4.13) involves an expectation over $\mathbf{y}_{1:k+e} | \boldsymbol{\theta}, \mathbf{u}_{1:k+e}$, even when the outputs $\mathbf{y}_{1:k}$ have already been measured.

Computational Shortcut

A computationally feasible alternative approach uses the observed Fisher information matrix, rather than the expected Fisher information matrix. This observed FIM does not require averaging over all possible measurements. Instead, it uses the actually observed values $\mathbf{y}_{1:k}$,

$$\mathcal{J}(\boldsymbol{\theta}, \mathbf{u}_{1:k}, \mathbf{y}_{1:k}) = -\frac{\partial^2 \log p(\mathbf{y}_{1:k} | \boldsymbol{\theta}, \mathbf{u}_{1:k})}{\partial \boldsymbol{\theta}^2}. \quad (4.18)$$

The observed FIM has been argued to be a superior tool to quantify the variance of the model parameter estimates (Efron and Hinkley 1978), and has already been used in sequential experimental design by Lane (2017) to produce more precise parameter estimates than a method purely based on the expected FIM. A straightforward solution to keep the optimization cost constant at every time-step would be to combine both the observed FIM, to quantify the information of the k already performed measurements, and the expected FIM, to quantify the information of the e future observations. This leads to the following optimal design criterion at time-step k :

$$\arg \max_{\mathbf{u}_{\min} \leq \mathbf{u}_{k+1:k+e} \leq \mathbf{u}_{\max}} \int |\mathcal{J}(\boldsymbol{\theta}, \mathbf{u}_{1:k}, \mathbf{y}_{1:k}) + \mathcal{I}(\boldsymbol{\theta}, \mathbf{u}_{k+1:k+e})| p(\boldsymbol{\theta} | \mathbf{y}_{1:k}, \mathbf{u}_{1:k}) d\boldsymbol{\theta}. \quad (4.19)$$

To calculate the expected FIM $\mathcal{I}(\boldsymbol{\theta}, \mathbf{u}_{k+1:k+e})$ only an expectation over $\mathbf{y}_{k+1:k+e}$ is needed. So,

$$\mathcal{I}(\boldsymbol{\theta}, \mathbf{u}_{k+1:k+e})[i, j] =$$

$$\begin{aligned} & \frac{\partial \mathbf{E}(\mathbf{y}_{k+1:k+e} | \boldsymbol{\theta}, \mathbf{y}_{1:k}, \mathbf{u}_{1:k+e})'}{\partial \boldsymbol{\theta}[i]} \text{Covar}(\mathbf{y}_{k+1:k+e} | \boldsymbol{\theta}, \mathbf{y}_{1:k}, \mathbf{u}_{1:k+e})^{-1} \frac{\partial \mathbf{E}(\mathbf{y}_{k+1:k+e} | \boldsymbol{\theta}, \mathbf{y}_{1:k}, \mathbf{u}_{1:k+e})}{\partial \boldsymbol{\theta}[j]} \\ & + \\ & \frac{1}{2} \text{tr} \left(\text{Covar}(\mathbf{y}_{k+1:k+e} | \boldsymbol{\theta}, \mathbf{y}_{1:k}, \mathbf{u}_{1:k+e})^{-1} \frac{\partial \text{Covar}(\mathbf{y}_{k+1:k+e} | \boldsymbol{\theta}, \mathbf{y}_{1:k}, \mathbf{u}_{1:k+e})}{\partial \boldsymbol{\theta}[i]} \right. \\ & \quad \left. \text{Covar}(\mathbf{y}_{k+1:k+e} | \boldsymbol{\theta}, \mathbf{y}_{1:k}, \mathbf{u}_{1:k+e})^{-1} \frac{\partial \text{Covar}(\mathbf{y}_{k+1:k+e} | \boldsymbol{\theta}, \mathbf{y}_{1:k}, \mathbf{u}_{1:k+e})}{\partial \boldsymbol{\theta}[j]} \right). \end{aligned} \quad (4.20)$$

The recursion formulas in Equation (4.14) must thus also be changed to not recurse all the way back to time point 0. The recursion instead should end at time k , with the state mean \mathbf{m}_k and covariance estimate P_k coming from the Kalman filter in Equation (4.11):

$$\begin{aligned} E(\mathbf{y}_r | \boldsymbol{\theta}, \mathbf{u}_{1:k+e}, \mathbf{y}_{1:k}) &= H(\boldsymbol{\theta}) E(\mathbf{x}_r | \boldsymbol{\theta}, \mathbf{u}_{1:k+e}, \mathbf{y}_{1:k}), \\ E(\mathbf{x}_r | \boldsymbol{\theta}, \mathbf{u}_{1:k+e}, \mathbf{y}_{1:k}) &= F(\boldsymbol{\theta}) E(\mathbf{x}_{r-1} | \boldsymbol{\theta}, \mathbf{u}_{1:k+e}, \mathbf{y}_{1:k}) + B(\boldsymbol{\theta}) \mathbf{u}_r, \\ E(\mathbf{x}_k | \boldsymbol{\theta}, \mathbf{u}_{1:k+e}, \mathbf{y}_{1:k}) &= \mathbf{m}_k, \\ \text{Var}(\mathbf{y}_r | \boldsymbol{\theta}, \mathbf{u}_{1:k+e}, \mathbf{y}_{1:k}) &= H(\boldsymbol{\theta}) \text{Var}(\mathbf{x}_r | \boldsymbol{\theta}, \mathbf{u}_{1:k+e}, \mathbf{y}_{1:k}) H(\boldsymbol{\theta})' + R(\boldsymbol{\theta}), \\ \text{Covar}(\mathbf{y}_r | \boldsymbol{\theta}, \mathbf{u}_{1:k+e}, \mathbf{y}_{1:k}; \mathbf{y}_s | \boldsymbol{\theta}, \mathbf{u}_{1:k+e}, \mathbf{y}_{1:k}) &= \\ H(\boldsymbol{\theta}) F^{r-s} \text{Var}(\mathbf{x}_s | \boldsymbol{\theta}, \mathbf{u}_{1:k+e}, \mathbf{y}_{1:k}) H(\boldsymbol{\theta})', \quad \forall r > s, \\ \text{Covar}(\mathbf{y}_r | \boldsymbol{\theta}, \mathbf{u}_{1:k+e}, \mathbf{y}_{1:k}; \mathbf{y}_s | \boldsymbol{\theta}, \mathbf{u}_{1:k+e}, \mathbf{y}_{1:k}) &= \\ \text{Covar}(\mathbf{y}_s | \boldsymbol{\theta}, \mathbf{u}_{1:k+e}, \mathbf{y}_{1:k}; \mathbf{y}_r | \boldsymbol{\theta}, \mathbf{u}_{1:k+e}, \mathbf{y}_{1:k})', \quad \forall r < s, \\ \text{Var}(\mathbf{x}_r | \boldsymbol{\theta}, \mathbf{u}_{1:k+e}, \mathbf{y}_{1:k}) &= F(\boldsymbol{\theta}) \text{Var}(\mathbf{x}_{r-1} | \boldsymbol{\theta}, \mathbf{u}_{1:k+e}, \mathbf{y}_{1:k}) F(\boldsymbol{\theta})' + Q(\boldsymbol{\theta}), \\ \text{Var}(\mathbf{x}_k | \boldsymbol{\theta}, \mathbf{u}_{1:k+e}, \mathbf{y}_{1:k}) &= P_k. \end{aligned} \quad (4.21)$$

The integral in Equation (4.19) can again be approximated by Monte Carlo integration:

$$\arg \max_{\mathbf{u}_{\min} \leq \mathbf{u}_{k+1:k+e} \leq \mathbf{u}_{\max}} \sum_{i=1}^N |\mathcal{J}(\boldsymbol{\theta}_k^i, \mathbf{u}_{1:k}, \mathbf{y}_{1:k}) + \mathcal{I}(\boldsymbol{\theta}_k^i, \mathbf{u}_{k+1:k+e})| w_k^i. \quad (4.22)$$

In this equation, the weights w_k^i and the model parameters $\boldsymbol{\theta}_k^i$ come from the recursion:

$$\begin{aligned} w_k^i &= \frac{p(\mathbf{y}_k | \boldsymbol{\theta}_{k-1}^i, \mathbf{u}_{1:k}, \mathbf{y}_{1:k-1}) w_{k-1}^i}{\sum_{j=1}^N p(\mathbf{y}_k | \boldsymbol{\theta}_{k-1}^j, \mathbf{u}_{1:k}, \mathbf{y}_{1:k-1}) w_{k-1}^j}, \\ \boldsymbol{\theta}_k^i &= \boldsymbol{\theta}_{k-1}^i, \\ w_0^i &= \frac{1}{N}, \\ \boldsymbol{\theta}_0^i &\text{ drawn from } p(\boldsymbol{\theta}). \end{aligned} \tag{4.23}$$

These weights are thus updated to give higher importance to model parameters according to their likelihood. This adaptive experimental design routine only works when using the same Monte Carlo draws $\boldsymbol{\theta}_k^i$ at each time-step. If different values were used at every time-step the likelihoods would have to be calculated again from the beginning, instead of relying on the recursive Equation (4.12).

Final Algorithm

Putting together all these computations leads to the following Algorithm 1 which summarizes all the steps of the algorithm for adaptively generating a robust experiment to estimate the model parameters of a linear dynamic system in the presence of both process noise and measurement noise.

Algorithm 1: Robust and adaptive experimental design algorithm for dynamic systems in the presence of both process and measurement noise

initialize at step zero

for $i = 1$ through N **do**

 draw θ_0^i from $p(\theta)$

 set initial state distribution $\mathbf{m}_0^i = \mathbf{m}_0$ and $P_0^i = P_0$

 set initial state distribution sensitivities $\frac{\partial \mathbf{m}_0^i}{\partial \theta} \Big|_{\theta=\theta_0^i} = 0$, $\frac{\partial^2 \mathbf{m}_0^i}{\partial \theta^2} \Big|_{\theta=\theta_0^i} = 0$,

$\frac{\partial P_0^i}{\partial \theta} \Big|_{\theta=\theta_0^i} = 0$, $\frac{\partial^2 P_0^i}{\partial \theta^2} \Big|_{\theta=\theta_0^i} = 0$

 set log-likelihood $L(\theta_0^i) = 0$

 set observed FIM $\mathcal{J}(\theta_0^i) = 0$

 set weight $w_0^i = \frac{1}{N}$

end

run the experiment

for each time-step $k = 0$ through T **do**

 optimize next e controls $\mathbf{u}_{k+1:k+e}$ using Eq (4.22), with expected FIM from Eq (4.20)

 use control \mathbf{u}_{k+1}

 perform measurement \mathbf{y}_{k+1}

for $i = 1$ through N **do**

 Push forward Kalman filter (using hyper-dual numbers) in Eq (4.11)
 with θ_{k-1}^i

 Use results (and intermediate results) from Kalman filter to:

 1) update state distribution mean \mathbf{m}_k^i and covariance P_k^i

 2) update state distribution sensitivities $\frac{\partial \mathbf{m}_0^i}{\partial \theta} \Big|_{\theta=\theta_{k-1}^i}$, $\frac{\partial^2 \mathbf{m}_0^i}{\partial \theta^2} \Big|_{\theta=\theta_{k-1}^i}$,

$\frac{\partial P_0^i}{\partial \theta} \Big|_{\theta=\theta_{k-1}^i}$ and $\frac{\partial^2 P_0^i}{\partial \theta^2} \Big|_{\theta=\theta_{k-1}^i}$

 3) update log-likelihood using Eq (4.12)

 4) update observed FIM using Eq (4.18)

 5) update weights w_k^i using Eq (4.23)

 6) update model parameters θ_k^i using Eq (4.23)

end

end

4.4 Numerical details

The entire Algorithm 1 was implemented in the Julia programming language (Bezanson et al. 2017). The recursion in Equation (4.21) can be efficiently calculated in batch form. This batch form can easily be adapted from the

results in Cavanaugh and Shumway (1996). These authors, however, do not give a batch form of the expected FIM in Equation (4.20). For batch form of the Fisher information matrix for normally distributed data see Fedorov and Leonov (2013).

For the sensitivities required to calculate the expected FIM, forward mode automatic differentiation is used. In forward mode automatic differentiation, every variable is replaced with a dual number containing both the value of that number and the partial derivatives of that variable with respect to θ . Operators are then overloaded to correctly propagate the partial derivatives (Griewank and Walther 2008). For example, if in the original code there is an expression $c = a * b$, and we know the partial derivatives of a and b , these numbers are replaced by the dual numbers $(a, \frac{\partial a}{\partial \theta})$ and $(b, \frac{\partial b}{\partial \theta})$ and multiplication is overloaded as $(a, \frac{\partial a}{\partial \theta}) * (b, \frac{\partial b}{\partial \theta}) = (a * b, a * \frac{\partial b}{\partial \theta} + b * \frac{\partial a}{\partial \theta}) = (c, \frac{\partial c}{\partial \theta})$. Variables that do not depend on θ have zero partial derivatives, and the i th element of θ is initialized with one for the i th partial derivative and zero for the other partial derivatives. For the observed FIM, second order derivatives are needed. These can also be calculated using forward mode automatic differentiation using hyper-dual numbers, see Revels, Lubin, and Papamarkou (2016) for details how these are implemented in Julia.

Sequential quadratic programming, as implemented in NLOpt (Johnson 2014), is used to solve the optimization problem (Kraft et al. 1988; Kraft 1994) in Equation (4.22). The optimal controls found at the previous time-step, are reused as a hot starting point. A random value between u_{\min} and u_{\max} is selected for the controls at the end of the optimization horizon. The optimization algorithm is allowed a maximum of 20 function evaluations before termination, except for the first time-step where 120 evaluations are allowed. Gradients of the control objective are again calculated using hyper-dual numbers.

4.5 Case Studies

4.5.1 Mass-Spring-Damper System

Problem description

In the first case study, we consider experimental design for the mass-spring-damper system depicted in Figure 4.1. Mass-spring-damper systems are commonly used in bio-mechanics (Pieczywek and Zdunek 2017). The discretized

linear dynamics of this system are:

$$\begin{aligned} \boldsymbol{x}_k &= \begin{bmatrix} 1 & \Delta t \\ \frac{-\Delta t K}{M} & \frac{-\Delta t C}{M} + 1 \end{bmatrix} \boldsymbol{x}_{k-1} + \begin{bmatrix} 0 \\ \frac{\Delta t}{M} \end{bmatrix} F_k + \boldsymbol{w}_k, \\ \boldsymbol{w}_k &\sim \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{q\Delta t^3}{3M^2} & \frac{q\Delta t^2}{2M^2} \\ \frac{q\Delta t^2}{2M^2} & \frac{q\Delta t}{M^2} \end{bmatrix} \right), \quad (4.24) \\ y_k &= [1 \ 0] \boldsymbol{x} + v_k, \\ v_k &\sim \mathcal{N}(0, 0.1). \end{aligned}$$

In these equations, K and C are the spring and damper constant, respectively. These are the two unknown model parameters that must be estimated. Their true values are assumed equal to 1 and 2, respectively. The prior distributions we use for them are independent normal distributions centered around 1.4 and 4, with variances assumed equal to 0.2 and 2. The parameters q and M are the spectral density of the process noise and mass respectively, which we assume to be known and equal to 0.05 and 1. Finally, Δt is the time between measurements, assumed equal to 0.1. The position and velocity are the two states, but only the position is measured, with measurement noise on top of it. The initial state distributions are assumed to be independent normal distributions with means equal to zero and variances equal to 0.1. The controllable input at the k th time-step is a force F_k , which must be optimized such that K and C can be estimated as precisely as possible from the position measurements. The maximum absolute value of the force that can be applied is 1.

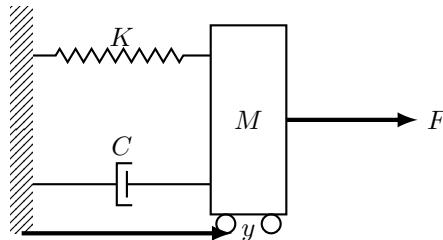


Figure 4.1: Schematic representation of mass spring damper system.

Optimal Versus Random Design

We start by comparing the performance of our optimal experimental design strategy to a random input signal. Both experiments last $T = 100$ time-steps.

The optimal experiment is generated with $N = 100$ draws from the prior distribution of the model parameters, and looks $e = 3$ steps ahead for optimizing the controls. In Figure 4.2a, the inputs for both experiments are shown, and the corresponding measurements are shown in Figure 4.2b. An always maximal or minimal control action (bang-bang control) seems to be preferred, since the optimal experimental design switches twice between the maximum and minimum allowed force. The influence of these optimal controls is clearly visible on the measurements, where the position is clearly lower after a negative force has been applied, and clearly higher after a positive force has been applied. The controls seem to switch from positive to negative and vice versa after the position stagnates around position values of 1 and -1 . This is logical since, once the position stagnates, nothing can be learned anymore about the damping constant.

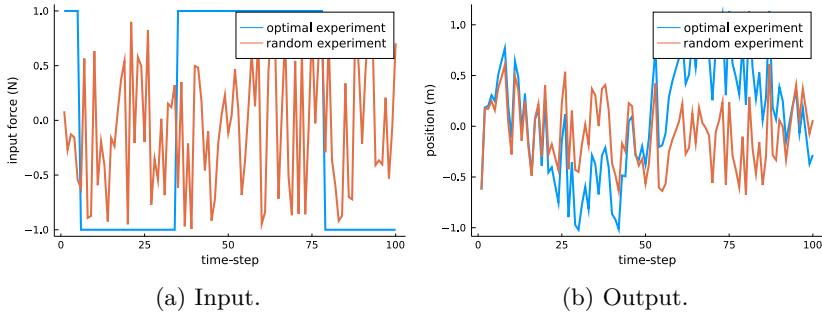


Figure 4.2: Comparison of optimal inputs compared to random inputs and the corresponding output behaviors.

In Figure 4.3, we show the evolution of the online maximum likelihood estimates as the experiment progresses. The optimal experiment hovers around the true model parameters after only 50 time-steps, while the random experiment can not even correctly estimate these parameters after 100 time-steps. The likelihood at the end of the experiment for the 100 parameters drawn from the prior distribution for the model parameters is shown in Figure 4.4. We see that the maximum likelihood estimate for the optimal experiment is one of the closest grid points to the true model parameter values, while this is not the case for the random experiment. Furthermore, for the optimal experiment the relative likelihood of other model parameters compared to the maximal likelihood estimate decreases rapidly when moving away from this estimate. This means that for the optimal experiment only model parameter values close to the true values fit the data well. For the random experiment the likelihood does not decrease rapidly when moving away from the maximum likelihood

estimate, which means almost all values fit the data almost equally well and we can not discern the true model parameter values from the data.

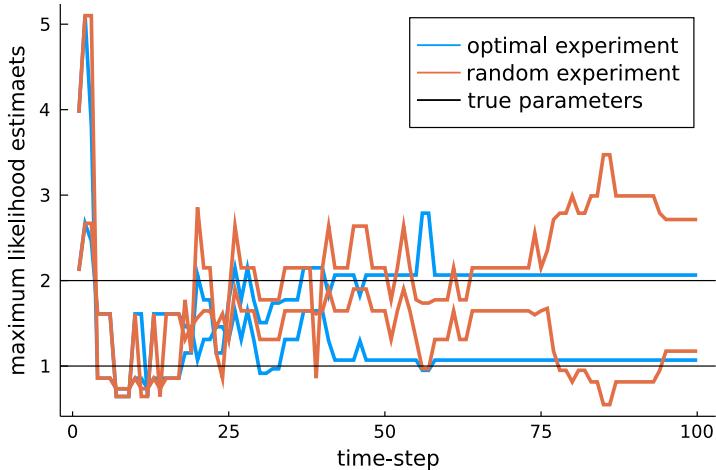


Figure 4.3: Online maximum likelihood estimates for both the optimal experiment and the random experiment. The optimal experiment converges faster to the true parameters.

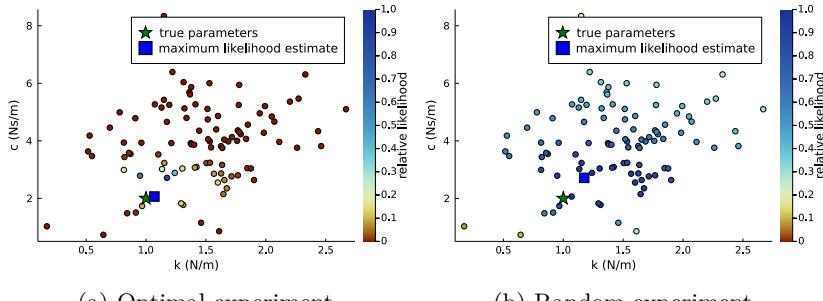


Figure 4.4: Likelihood grids at the end of the experiment. The likelihood decreases sharply away from the true parameters for the optimal experiment, unlike in the random experiment. The maximum likelihood estimate of the optimal experiment is much closer to the true value than the random experiment.

Added Value of the Robustness and Adaptivity

The above discussion already shows the value of experimental design methodology compared to random inputs. We now continue by showing the combined added value of robustness and adaptivity. In Figure 4.5, we study the behavior of Algorithm 1 for a variety of combinations of control horizon length e , number of model parameters drawn from the prior N , and number of time-steps, T . For each combination of e , N and T the experiment is repeated 100 times and the mean and variance of the online maximum likelihood estimates are plotted. Figures 4.5a and 4.5b show the same combination as was used before, in Figure 4.4, for the optimal experiment and random experiment, respectively. This allows us to confirm that the optimal experiment performs much better than the random experiment over an ensemble of 100 experiments, and, thus, that the better estimates of the experimental design methodology were not by chance.

In Figure 4.5c, the control horizon length, e , is reduced from 3 to 1. This causes this experiment to perform almost as bad as the random experiment. Increasing the control horizon length to 6, however, does not greatly increase the performance of the experiment, as shown in Figure 4.5d. Increasing the control horizon thus has diminishing marginal returns, which is good to know since a short horizon keeps the size of the matrices involved in the calculation of expected FIM in Equation (4.20) manageable. The lack of additional information when increasing the control horizon is probably due to the increased variance of the predictions further into the future.

The effect of the number of parameters drawn from the prior distribution, N , is shown in Figures 4.5e and 4.5f. For the non-robust experiment, the mean of the prior distribution of the model parameters was used in the calculation of the Fisher information matrices in Equations (4.20) and (4.18), instead of a single random value from this distribution. The effect of reducing robustness is much less pronounced than the effect of reducing the control horizon. Only the convergence of the estimate of the damper constant is slower. Increasing the number of draws from the prior distribution from 100 to 500 also seems to have little added value past a certain point. This is because the Monte Carlo integration involved in Bayesian optimal design is already adequate.

We also compare our adaptive experimental design technique to the non-adaptive strategy from Equation (4.17), in Figure 4.5g. In the beginning of the experiment, the non-adaptive experiment performs equally well as our adaptive strategy. However, later on in the experiment, between time-steps 50 and 75, the estimation of the damper constant is clearly performing worse than the adaptive experiment. An additional shortcoming of the non-adaptive

design is the large computational time required to generate this design. This is because the non-adaptive experiment requires predicting 100 steps into the future. The optimization of this design was much slower than the adaptive designs with a short control window, due to the presence of large matrices in the expected FIM in Equation (4.20).

Finally, the effect of a larger number of time-steps is shown in Figure 4.5h. The Figure shows that, also here, there are diminishing marginal returns for longer experiments.

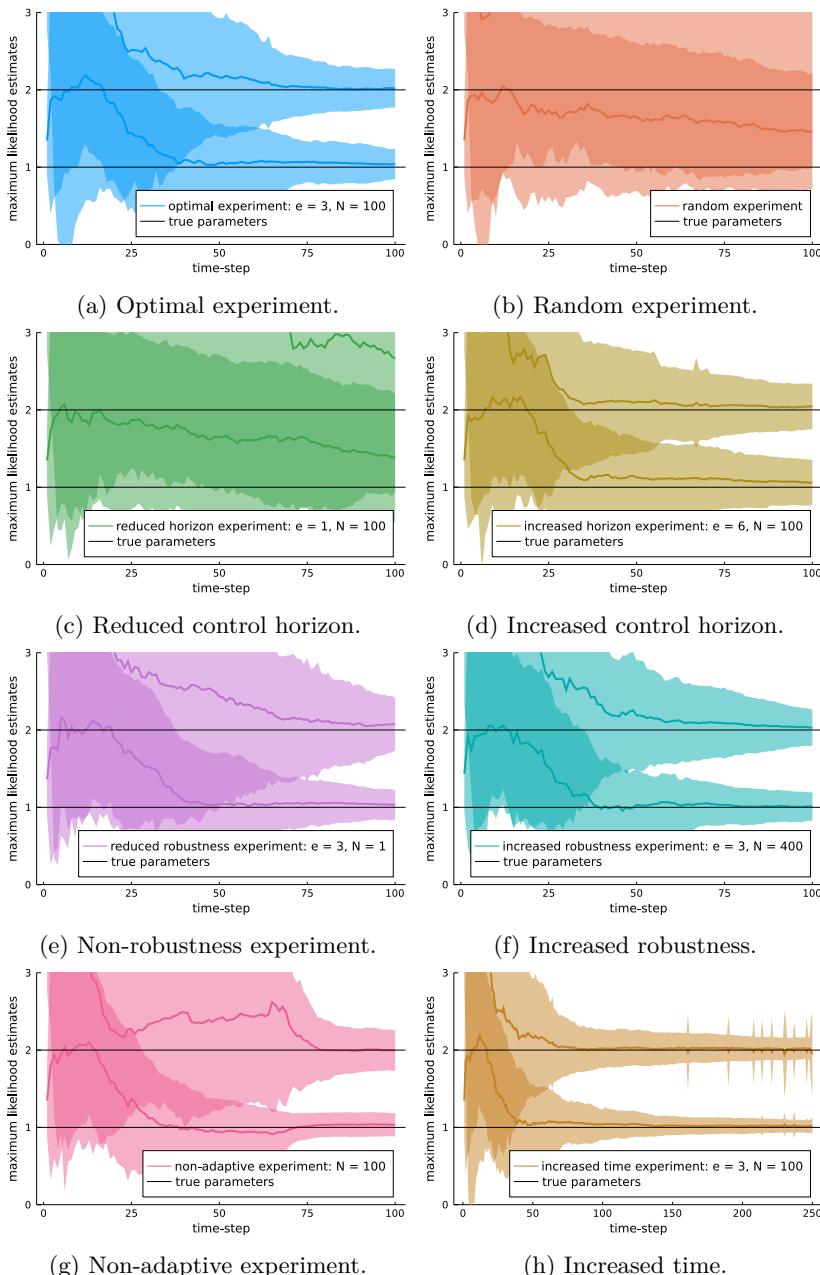


Figure 4.5: Mean and variance of online maximum likelihood estimates for different combinations of e , N and T in Algorithm 1.

4.6 Two Compartment System

In the second case study, we consider experimental design for a two compartment system. Compartment systems are common in postharvest modeling (Léchaudel and Joas 2007; Yoneyama et al. 2016), and pharmacokinetics (Fedorov and Leonov 2013). The discretized linear dynamics of this system are:

$$\begin{aligned} \mathbf{x}_k &= \begin{bmatrix} 1 - \Delta t(K_{1,0} + K_{1,2}) & \Delta t K_{2,1} \\ \Delta t K_{1,2} & 1 - \Delta t K_{2,1} \end{bmatrix} \mathbf{x}_{k-1} + \begin{bmatrix} \Delta t \\ \frac{\Delta t^2}{2} \end{bmatrix} u_k + \mathbf{w}_k, \\ \mathbf{w}_k &\sim \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} q\Delta t & \frac{q\Delta t^2}{2} \\ \frac{q\Delta t^2}{2} & \frac{q\Delta t^3}{3} \end{bmatrix} \right), \\ y_k &= [\Delta t K_{1,0} \quad 0] \mathbf{x} + v_k, \\ v_k &\sim \mathcal{N}(0, 0.0001). \end{aligned} \tag{4.25}$$

In these equations, $K_{1,2}$, $K_{2,1}$ and $K_{1,0}$ are the unknown model parameters that determine the flows between the two compartments and the flow from the first compartment to the environment. Their true values are assumed to all equal 0.2. The prior distributions we use for them are independent normal distributions centered around 0.22, with variances assumed equal to 0.0016. The time between measurements, Δt , is assumed equal to 0.1. The outflow of the first compartment to the environment is the measured output, with measurement noise on top of it. The initial distributions for the two compartments are assumed to be independent normal distributions with means equal to 10 and 1, and variances equal to 0.01 and 0.00001, respectively. The controllable input at time-step k , u_k , is a flow towards the first compartment. This input is constrained between 0 and 10. There is also an unknown stochastic input \mathbf{w}_k to the first compartment, represented by a discretization of Brownian motion with spectral density q equal to 0.001. Brownian motion is a continuous time stochastic process whose increments are independent, stationary and normally distributed. The variance of the increments is determined by the spectral density.

This example shows the added value of working with arbitrarily parametrized state space models, instead of linear autoregressive models. Some parameters, such as $K_{1,0}$, occur multiple times in the system dynamics. This is contrary to autoregressive models, where the output is assumed to be a linear combination of previous measurements and inputs, and each parameter of this linear combination is allowed to vary freely in the parameter estimation.

Figure 4.6 depicts the progression of the online maximum likelihood estimate as time goes on. The optimal experiment was generated with $N = 1000$ draws

from the prior distribution, and looks $e = 3$ steps ahead. The parameters are estimated precisely after roughly 50 time steps, while the random experiment does not correctly estimate the model parameters even after 200 steps.

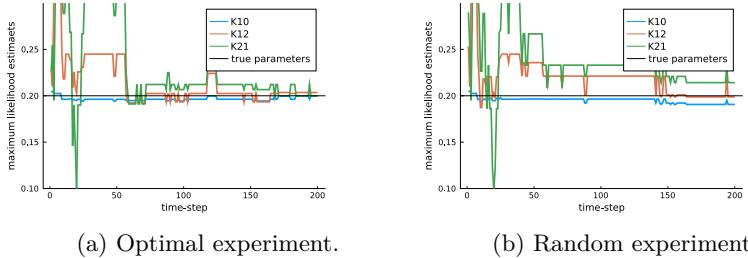


Figure 4.6: Online maximum likelihood estimates for the two compartment model.

4.7 Discussion and Conclusion

In this chapter, we presented a novel robust and adaptive experimental design method to estimate the model parameters of discrete-time linear state space models. We achieved this by quantifying the information content of an experiment using a combination of the expected and observed Fisher information matrix. In future research, we want to extend these results to non-linear dynamics. The Kalman filter must then be replaced with another Bayesian filter, such as the extended Kalman filter, sigma-point filter or particle filter. Continuous-time dynamics is another interesting direction for future research, as almost no literature exists on experimental design for stochastic differential equation models.

To be able to optimize our experiments adaptively, we were forced to evaluate the likelihood of the model parameters at the same location (in the model parameters space) at every time-step. Most locations quickly become very unlikely, as seen in Figure 4.4a. It would be better if this sample could slightly move towards regions of higher probability at every time-step. Kantas et al. (2009) give an overview of methods that allow for such jittering of the location of the model parameters, but none of the methods discussed are completely online. Since these authors are only interested in online parameter estimation, this is not a major issue for them. But when also considering adaptive experimental design, where an optimization over the input space has to be ran at every time-step, the jittering of the model parameters must be very efficient. Very

recently He, Khedher, and Spreij (2021), published a promising method for online parameter estimation for linear dynamic systems based on the Kalman filter, that could be used for this purpose, and which we plan to incorporate in future research.

Chapter 5

Conclusion and Outlook

5.1 Overarching Conclusion

The general goal of this thesis is to show the usefulness of dynamic experiments for bio-engineering applications. The case studies in Chapters 2 and 3 focus on postharvest modeling and show that dynamic experiments indeed lead to a much more precise and robust estimation of respiration and fermentation parameters of pear fruit with reduced experimental effort. Chapter 4 focuses on experimental design methodology for linear dynamical systems in the presence of process noise. This methodology leads to improved experiments for models commonly used in bio-mechanics and pharmacokinetics.

To reach this general goal, multiple challenges had to be overcome:

- **Challenge 1:** The optimal experiment depends on the true values of the model parameters which the experiment aims to learn about. Some form of prior information about these parameters is thus always required to begin optimizing the experiment. In **Chapter 2**, we solved this issue by taking parameter estimates from the literature and used them to construct locally optimal designs. Using this methodology, experiments to estimate Michaelis-Menten respiration parameters for pear fruit were constructed. These experiments perform better than heuristic techniques used in postharvest experimentation.
- **Challenge 2:** Robust experimental designs perform well over a wide range of possible parameter values. Bayesian experimental design is a popular technique to achieve robustness, where the prior information is

quantified using a probability distribution. The design is then optimized to perform well on average, where the information content is averaged over this prior distribution. Available Bayesian experimental design techniques from the literature only allow for parametric distributions to quantify this prior information, and these techniques were found to be insufficient to adequately describe the prior information we had available for the respiration and fermentation of pear fruit. In **Chapter 3**, an experimental design methodology was developed that can deal with arbitrary prior distributions obtained by using a Markov-Chain Monte-Carlo integration technique. This methodology was then found to perform better than the traditional locally optimal design method used in **Chapter 2**. It was also demonstrated to perform better than other robust experimental design methods, that use parametric distributions as prior information. This methodology was then used to construct robust experiments to estimate the respiration and fermentation of pear fruit.

- **Challenge 3:** Little experimental design literature exists on how to deal with dynamic systems with both process and measurement noise. In **Chapter 4**, the Fisher information matrix (FIM) is derived for linear dynamical systems with both kinds of noise. The use of this FIM in an experimental design context is new to the literature. The key discovery here is that experimental design for dynamic systems with hidden dynamic states requires tracking these hidden states using an appropriate Bayesian filter, which for linear systems equals the Kalman filter.
- **Challenge 4:** Adaptive experimental design techniques are needed when the prior information about model parameters is poor. These adaptive techniques re-optimize the remainder of the experimental design after every measurement. However, the optimization problem that occurs in the traditional experimental design technique, which is based on the expected Fisher information matrix, keeps growing in complexity at every time-step. In **Chapter 4**, this issue is solved by developing a novel experimental design criterion based on a combination of the observed and expected FIM. The observed FIM is computationally less expensive as it only utilizes the actually observed measurements, thus making it well suited for quantifying the information from the already gathered measurements. The expected FIM requires an expectation over all possible observations, making it well suited to quantify the information of future observations. A combination of these two types of Fisher information matrices results in an experimental design methodology that requires the same amount of computation at every time-step, which is a necessity for adaptive experimental design. Such a combination of the observed and expected FIM has not yet been used for dynamic experimentation.

5.2 Outlook

In this thesis, I developed novel experimental design methodology for dynamic systems. I believe that more interesting and important research can still be done in this area, including research on experimental design methodology for more complicated dynamic models, novel ways to quantify uncertainty and more efficient numerical techniques. In this section, I discuss some possible perspectives on future research.

5.2.1 Model Selection

It is commonly assumed in the experimental design literature that the entire model structure is known. However, it is often more realistic that:

- Competing model structures exist to explain a certain phenomenon.
- Only part of the model structure is known, and there are missing components in the model.

Appropriate experimental design techniques should be developed that can deal with both model parameter and model structure uncertainty.

Model Discrimination

Some experimental design techniques for model discrimination in dynamic systems exist (Chen and S. P. Asprey 2003), mostly based on the Hunter and Reiner criterion (Hunter and Reiner 1965). A weakness of this approach is that it is not robust, since it does not account for parameter uncertainty. Instead, the method only compares models at point estimates of the parameters. If there is much parameter uncertainty, the Hunter and Reiner method will be overly optimistic in its ability to discriminate models. Instead of this method, a Bayesian optimal design method, similar to the one used in Chapter 3, could be developed, where an information criterion, averaged out over a joint prior distribution for the model structure and the model parameters, is optimized. The challenge here is to generalize the Markov-Chain Monte-Carlo integration technique to deal with both the continuous model parameters and the discrete model structures.

Missing Model Components

Another form of model uncertainty occurs when only part of the model structure is known. One method to deal with missing model components is to replace them with a black box model, such as a Gaussian process or neural network. Replacing parts of a differential equation with a surrogate model is currently a hot topic within machine learning (Rackauckas, Ma, Martensen, et al. 2020; Dandekar et al. 2020). Optimal experimental design techniques have not been used by machine learning communities, as they often deal with large observational datasets. However, when combining the black box modeling approach from machine learning and mechanistic models for biological systems, in scenarios where acquiring experimental data is cumbersome and time-consuming, there is a need for carefully crafted experiments. Instead of quantifying the uncertainty of a finite dimensional parameter space, here the challenge lies in quantifying the uncertainty of an infinite dimensional function space. To achieve this, methodological concepts from optimal design of computer experiments can be borrowed, where the Fisher information and entropy based criteria have already been combined with Gaussian processes (Santner et al. 2003). Of course, computer simulations are often deterministic, while biology is rife with variability.

Throughout this thesis we assumed that Michaelis-Menten reaction kinetics perfectly summarize the respiration and fermentation behavior of pear fruit. However, in reality this is just a summary of a much larger biological pathway, and diffusion also plays a role, leading to a high dimensional partial differential equation model. Instead of working directly with this more complicated model, the misspecification on the lumped model could be taken into account through a Gaussian process (Kennedy and O'Hagan 2001), but experimental design for such misspecified systems has not yet been considered in literature.

5.2.2 Process Noise

In chapter 4, I developed an experimental design method to deal with process noise for discrete-time linear dynamical systems. For such systems, the Fisher information matrix has a tractable form, involving the Kalman filter to estimate the hidden state. The next step here is replacing the linear dynamics with non-linear dynamics. The Kalman filter then no longer suffices to estimate the hidden state. A more complicated filter, such as an extended Kalman, unscented or particle filter must then be used, greatly increasing the computational cost. If we then also consider continuous time models, we arrive at the unexplored research topic of experimental design for stochastic differential equation models.

5.2.3 Kullback-Leibler Divergence

In this thesis, I used the Fisher information matrix (FIM) to quantify the information content of an experiment. The Kullback-Leibler divergence (KL-div) is another method to quantify the information content. Unlike the FIM, the KL-div does not involve normal approximations, and might be a better criterion for small sample sizes. The major drawback is that the KL-div is numerically much more intensive to calculate. This is generally done using a double loop Monte-Carlo integration method (Ryan et al. 2016). Recently, advances have been made in lowering the computational burden of the KL-div based approach by considering variational Bayesian techniques (Foster, Jankowiak, Bingham, et al. 2019), where the inner Monte-Carlo loop is replaced by optimizing a variational distribution. This technique then allows for jointly optimizing the design and variational parameters (Foster, Jankowiak, O’Meara, et al. 2020).

5.2.4 Koopman Expectation

In Chapter 3 of this thesis, we calculated the expectation required for Bayesian experimental design by pushing forward the uncertain extended state distribution in time, and then calculating some quantity of interest at each time point. Gerlach et al. (2020) recently proposed a new method to calculate expectations for dynamic systems that instead pulls back this quantity of interest in time, using the Koopman operator which relies on the change of variable formula. This idea is analogous to the back propagation (adjoint sensitivity analysis) of gradients for optimization. Calculating the expectation using the Koopman operator is thus generally a good choice when a low number of expectations have to be calculated, such as the expectation of a scalar function as opposed to the expectation of a vector function. In Bayesian experimental design, such an expectation of a scalar function of the FIM is calculated. This means calculating the expectation using the Koopman operator should be favorable from a computational cost perspective.

5.2.5 Fully Risk Averse Experimental Design

In this thesis, we took an expected value approach to deal with uncertainty, in the sense that uncertainty on the model parameters was quantified using a probability distribution, and experiments were optimized for the average of an information criterion over this distribution. Experimental design techniques that instead assume a worst case scenario for the model parameter uncertainty also exist, such as the techniques developed by Bauer et al. (2000). Here, the

experimental design is optimized for the minimum of an information criterion over the set of possible model parameters. This method, however, only uses the worst case scenario to deal with parameter uncertainty, but there is, of course, also uncertainty in the measurements, and other model disturbances such as process noise. If these errors are bounded, we can also assume a worst case scenario here. There exist parameter estimation techniques that produce the entire set of model parameters that are consistent with the measurements, while eliminating those that are not possible (Jaulin et al. 2001). However, these set based techniques have only recently been used for experimental design (Jauberthie et al. 2018). It would in particular be interesting to combine these methods with reachability analysis for dynamic systems (Bogomolov et al. 2019), which are techniques to numerically solve differential equations, which return a tube in which the true solution is guaranteed to exist. Fully risk averse experimental design techniques would particularly be of interest for designing experiments that guarantee the true model parameters lie within a set of a chosen size.

5.2.6 Goal Driven Experimental Design

Most experimental design literature for dynamic systems focuses on a precise estimation of all model parameters, putting equal importance on each parameter. Often, the reason for estimating the model parameters is to use the resulting model in the control and optimization of the bioprocess. To achieve an optimal control or to find the optimal input settings for a bioprocess, not all model parameters are equally important. Instead, it is typically a non-linear function of a subset of the model parameters that matters. More effort should therefore be spent on precisely estimating the relevant functions of the influential parameters. The challenge here is transforming the variability of the parameters into a variability of optimal operation levels. Experiments then must be constructed to minimize this latter uncertainty.

This topic of precisely estimating a non-linear function has only been briefly explored in the dynamic experimental design literature, more specifically by Houska et al. (2015). These authors started by using the Fisher information matrix to quantify the uncertainty about the model parameters and then linearize the effect that the model parameters have on the optimal conditions and transform the parameter uncertainty into an uncertainty on the optimal conditions, using the delta method. However, they did not use this technique to find an optimal point in the design region, which would be a particularly interesting application of this technique. Li, Ng, and Tan (2020) suggested a method to achieve this, but only for generalized linear models. Their method

splits the Kullback-Leibler divergence into two parts: (i) the Kullback-Leibler divergence between the observations and the unknown parameters, and (ii) an information loss term between the unknown parameters and the optimal conditions quantified by the data processing inequality theorem (Cover and Thomas 2012). However, this technique has not yet been used in combination with dynamic systems.

Appendix A

Markov chain details

Table A.1: 100 Markov-chain points used to generate robust designs in Chapter 3

V_{m,O_2} $\mu\text{mol kg}^{-1} \text{s}^{-1}$	K_{m,O_2} kPa	K_{mn,CO_2} kPa	r_q -	V_{m,f,CO_2} $\mu\text{mol kg}^{-1} \text{s}^{-1}$	K_{m,f,O_2} kPa
0.319	6.048	1 509.965	0.660	0.120	0.219
0.337	6.499	360.728	0.647	0.118	0.262
0.302	4.127	381.998	0.660	0.123	0.158
0.314	5.376	1 879.107	0.665	0.122	0.213
0.318	6.080	705.206	0.660	0.119	0.245
0.320	5.973	547.186	0.654	0.119	0.248
0.325	6.075	613.318	0.657	0.122	0.226
0.333	6.303	3 993.319	0.652	0.122	0.249
0.315	4.735	151.390	0.657	0.123	0.195
0.335	6.457	1 880.454	0.649	0.121	0.248
0.318	5.968	486.791	0.657	0.121	0.255
0.315	4.927	116.835	0.667	0.122	0.212
0.314	5.219	668.050	0.656	0.122	0.199
0.301	4.779	2 834.253	0.655	0.122	0.203
0.315	4.875	524.641	0.652	0.121	0.207
0.325	5.432	206.583	0.656	0.122	0.200
0.318	5.796	4 876.479	0.664	0.119	0.218
0.321	5.465	866.089	0.650	0.119	0.194
0.316	5.487	626.561	0.651	0.121	0.233
0.302	4.511	1 211.476	0.652	0.120	0.185
0.313	5.468	1 396.073	0.661	0.121	0.207
0.331	5.756	361.899	0.646	0.122	0.253
0.332	6.102	519.659	0.665	0.118	0.203
0.306	5.271	2 050.857	0.663	0.121	0.184
0.326	5.582	198.514	0.663	0.122	0.234
0.326	5.777	351.813	0.655	0.122	0.230
0.322	5.462	803.832	0.648	0.120	0.213
0.318	5.383	271.911	0.656	0.121	0.215
0.326	5.675	399.561	0.645	0.121	0.236

0.332	6.148	372.446	0.647	0.122	0.245
0.318	5.310	2 123.003	0.649	0.120	0.222
0.309	5.137	30 722.240	0.646	0.121	0.219
0.328	5.600	498.991	0.646	0.122	0.224
0.317	5.981	672.748	0.667	0.118	0.235
0.320	5.796	1 353.348	0.657	0.120	0.226
0.335	6.163	4 266.786	0.650	0.119	0.258
0.329	6.398	566.697	0.668	0.121	0.215
0.328	5.797	1 096.313	0.641	0.121	0.275
0.309	5.046	270.633	0.670	0.119	0.186
0.317	5.619	1 594.171	0.659	0.121	0.223
0.315	5.645	669.299	0.657	0.120	0.214
0.321	5.087	158.819	0.656	0.118	0.218
0.313	4.938	663.970	0.659	0.118	0.189
0.328	6.045	247.834	0.660	0.122	0.252
0.349	6.201	311.906	0.647	0.121	0.266
0.297	4.986	898.951	0.673	0.122	0.193
0.311	5.414	8 614.486	0.655	0.120	0.258
0.320	5.571	378.320	0.648	0.122	0.244
0.305	5.507	18 510.410	0.661	0.121	0.220
0.325	5.906	799.309	0.655	0.121	0.235
0.327	5.999	1 192.227	0.659	0.123	0.215
0.319	5.343	256.175	0.651	0.124	0.232
0.315	5.440	1 012.412	0.659	0.122	0.205
0.309	4.567	1 305.077	0.632	0.120	0.200
0.313	5.530	866.167	0.653	0.121	0.217
0.309	5.321	1 812.792	0.657	0.122	0.232
0.320	5.389	448.156	0.663	0.120	0.216
0.311	5.271	782.166	0.658	0.120	0.217
0.348	6.598	263.406	0.649	0.122	0.275
0.313	5.537	1 784.753	0.663	0.120	0.217
0.321	5.734	300.298	0.662	0.122	0.202
0.310	5.047	187.941	0.658	0.122	0.210
0.330	6.206	664.203	0.659	0.119	0.218
0.339	5.822	194.900	0.650	0.119	0.227
0.327	5.900	5 201.394	0.639	0.119	0.231
0.315	5.178	254.046	0.658	0.122	0.201
0.321	5.563	954.128	0.642	0.122	0.236
0.313	5.277	863.780	0.646	0.121	0.231
0.294	4.660	10 601.054	0.643	0.121	0.208
0.323	5.549	47 663.943	0.649	0.121	0.215
0.316	5.374	826.334	0.657	0.122	0.194
0.332	6.346	590.861	0.660	0.122	0.249
0.324	6.544	2 189.372	0.656	0.121	0.266

0.321	5.571	390.299	0.650	0.122	0.207
0.328	5.707	530.657	0.650	0.121	0.251
0.320	5.234	898.434	0.650	0.122	0.223
0.333	6.075	280.156	0.664	0.121	0.215
0.322	6.318	1 230.360	0.656	0.120	0.249
0.325	5.732	525.615	0.654	0.121	0.233
0.314	5.543	1 315.741	0.651	0.121	0.191
0.324	5.600	885.672	0.655	0.118	0.218
0.295	4.898	4 308.873	0.664	0.123	0.190
0.325	6.019	463.202	0.673	0.124	0.210
0.320	6.088	589.510	0.666	0.121	0.213
0.322	5.743	3 817.629	0.655	0.120	0.229
0.316	5.669	2 696.075	0.642	0.121	0.253
0.333	5.819	464.960	0.656	0.121	0.243
0.315	4.842	366.880	0.666	0.119	0.183
0.317	5.487	8 669.237	0.662	0.120	0.202
0.304	4.372	130.306	0.673	0.123	0.170
0.317	5.227	857.621	0.639	0.121	0.243
0.343	5.837	116.378	0.664	0.118	0.250
0.310	5.175	501.003	0.662	0.120	0.241
0.325	5.520	432.656	0.647	0.122	0.230
0.313	5.108	308 870.950	0.648	0.122	0.210
0.312	5.604	829.501	0.653	0.122	0.232
0.338	6.572	3 040.241	0.638	0.119	0.264
0.315	5.569	132 026.342	0.644	0.123	0.231
0.301	5.042	2 842.465	0.667	0.121	0.182
0.313	5.452	8 565.545	0.677	0.120	0.191

Bibliography

- Arnoud, Antoine, Fatih Guvenen, and Tatjana Kleineberg (2019). *Benchmarking Global Optimizers*. Tech. rep. National Bureau of Economic Research.
- Asprey, S, S Macchietto, and CC Pantelides (2000). “Robust optimal designs for dynamic experiments”. In: *IFAC Proceedings Volumes* 33.10, pp. 845–850.
- Atkinson, Anthony, Alexander Donev, and Randall Tobias (2007a). *Optimum experimental designs, with SAS*. Oxford University Press.
- (2007b). *Optimum experimental designs, with SAS*. Vol. 34. Oxford University Press.
- Balsa-Canto, Eva, Antonio A Alonso, and Julio R Banga (2010). “An iterative identification procedure for dynamic modeling of biochemical networks”. In: *BMC Systems Biology* 4.1, p. 11.
- Balsa-Canto, Eva, Maria Rodriguez-Fernandez, and Julio R Banga (2007). “Optimal design of dynamic experiments for improved estimation of kinetic parameters of thermal degradation”. In: *Journal of Food Engineering* 82.2, pp. 178–188.
- Bauer, Irene et al. (2000). “Numerical methods for optimum experimental design in DAE systems”. In: *Journal of Computational and Applied Mathematics* 120.1-2, pp. 1–25.
- Berg, Jeremy M, John L Tymoczko, and GJ Gatto Jr (2002). *Biochemistry*. WH Freeman.
- Bernaerts, Kristel, Roos D Servaes, et al. (2002). “Optimal temperature input design for estimation of the square root model parameters: parameter accuracy and model validity restrictions”. In: *International Journal of Food Microbiology* 73.2, pp. 145–157.
- Bernaerts, Kristel, Karina J Versyck, and Jan F Van Impe (2000). “On the design of optimal dynamic experiments for parameter estimation of a Ratkowsky-type growth kinetics at suboptimal temperatures”. In: *International Journal of Food Microbiology* 54.1, pp. 27–38.
- Bessemans, Niels et al. (2016). “A novel type of dynamic controlled atmosphere storage based on the respiratory quotient (RQ-DCA)”. In: *Postharvest Biology and Technology* 115, pp. 91–102.

- Betancourt, Michael (2017). “A conceptual introduction to Hamiltonian Monte Carlo”. In: *arXiv preprint arXiv:1701.02434*.
- Bezanson, Jeff et al. (2017). “Julia: A fresh approach to numerical computing”. In: *SIAM review* 59.1, pp. 65–98.
- Biegler, Lorenz T (2010). *Nonlinear programming: concepts, algorithms, and applications to chemical processes*. SIAM.
- Bogomolov, Sergiy et al. (2019). “JuliaReach: a toolbox for set-based reachability”. In: *Proceedings of the 22nd ACM International Conference on Hybrid Systems: Computation and Control*, pp. 39–44.
- Cavanaugh, Joseph E and Robert H Shumway (1996). “On computing the expected Fisher information matrix for state-space model parameters”. In: *Statistics & probability letters* 26.4, pp. 347–355.
- Chaloner, Kathryn and Isabella Verdinelli (1995). “Bayesian experimental design: A review”. In: *Statistical Science* 10.3, pp. 273–304.
- Chen, Bing H and Steven P Asprey (2003). “On the design of optimally informative dynamic experiments for model discrimination in multiresponse nonlinear situations”. In: *Industrial & engineering chemistry research* 42.7, pp. 1379–1390.
- Cover, Thomas M and Joy A Thomas (2012). *Elements of Information Theory*. John Wiley & Sons.
- Dandekar, Raj et al. (2020). “Bayesian Neural Ordinary Differential Equations”. In: *arXiv preprint arXiv:2012.07244*.
- Efron, Bradley and Trevor Hastie (2016). *Computer age statistical inference*. Cambridge University Press.
- Efron, Bradley and David V Hinkley (1978). “Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information”. In: *Biometrika* 65.3, pp. 457–483.
- Fedorov, Valerii V and Sergei L Leonov (2013). *Optimal design for nonlinear response models*. CRC Press.
- Fidler, JC and CJ North (1967). “The Effect of Conditions of Storage on the Respiration of Apples: I. The Effects of Temperature and Concentrations of Carbon Dioxide and Oxygen on the Production of Carbon Dioxide and Uptake of Oxygen”. In: *Journal of Horticultural Science* 42.2, pp. 189–206.
- Findeisen, Rolf and Frank Allgöwer (2002). “An introduction to nonlinear model predictive control”. In: *21st Benelux meeting on systems and control*. Vol. 11. Technische Universiteit Eindhoven Eindhoven, The Netherlands, pp. 119–141.
- Fonseca, Susana C, Fernanda AR Oliveira, and Jeffrey K Brecht (2002a). “Modelling respiration rate of fresh fruits and vegetables for modified atmosphere packages: a review”. In: *Journal of Food Engineering* 52.2, pp. 99–119.

- (2002b). “Modelling respiration rate of fresh fruits and vegetables for modified atmosphere packages: a review”. In: *Journal of Food Engineering* 52.2, pp. 99–119.
- Foster, Adam, Martin Jankowiak, Elias Bingham, et al. (2019). “Variational Bayesian optimal experimental design”. In: *Advances in Neural Information Processing Systems*, pp. 14036–14047.
- Foster, Adam, Martin Jankowiak, Matthew O’Meara, et al. (2020). “A Unified Stochastic Gradient Approach to Designing Bayesian-Optimal Experiments”. In: *International Conference on Artificial Intelligence and Statistics*, pp. 2959–2969.
- Franceschini, Gaia and S Macchietto (2008). “Model-based design of experiments for parameter precision: State of the art”. In: *Chemical Engineering Science* 63.19, pp. 4846–4872.
- Ge, Hong, Kai Xu, and Zoubin Ghahramani (2018). “Turing: A language for flexible probabilistic inference”. In: *Proceedings of Machine Learning Research*. 84, pp. 1682–1690.
- Gelman, Andrew et al. (2013). *Bayesian data analysis*. CRC press.
- Gerlach, Adam R et al. (2020). “The Koopman Expectation: An Operator Theoretic Method for Efficient Analysis and Optimization of Uncertain Hybrid Dynamical Systems”. In: *arXiv preprint arXiv:2008.08737*.
- Goodwin, Graham Clifford and RL Payne (1977). “Dynamic system identification, experiment design and data analysis.” In.
- Goos, Peter and Bradley Jones (2011). *Optimal design of experiments: a case study approach*. John Wiley & Sons.
- Griewank, Andreas and Andrea Walther (2008). *Evaluating derivatives: principles and techniques of algorithmic differentiation*. SIAM.
- He, Jian, Asma Khedher, and Peter Spreij (2021). “A Kalman particle filter for online parameter estimation with applications to affine models”. In: *Statistical Inference for Stochastic Processes*, pp. 1–51.
- Hertog, Maarten LATM et al. (1998a). “A dynamic and generic model of gas exchange of respiring produce: the effects of oxygen, carbon dioxide and temperature”. In: *Postharvest Biology and Technology* 14.3, pp. 335–349.
- (1998b). “A dynamic and generic model of gas exchange of respiring produce: the effects of oxygen, carbon dioxide and temperature”. In: *Postharvest Biology and Technology* 14.3, pp. 335–349.
- Hjalmarsson, Håkan (2005). “From experiment design to closed-loop control”. In: *Automatica* 41.3, pp. 393–438.
- Ho, Maarten LATM Hertog, et al. (2018). “Down-regulation of respiration in pear fruit depends on temperature”. In: *Journal of Experimental Botany* 69.8, pp. 2049–2060.

- Ho, Pieter Verboven, Bert E Verlinden, Els Herremans, et al. (2011). "A Three-Dimensional Multiscale Model for Gas Exchange in Fruit". In: *Plant Physiology* 155.3, pp. 1158–1168.
- Ho, Pieter Verboven, Bert E Verlinden, Jeroen Lammertyn, et al. (2008). "A continuum model for metabolic gas exchange in pear fruit". In: *PLoS Computational Biology* 4.3, e1000023.
- Hoffman, Matthew D and Andrew Gelman (2014). "The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo". In: *Journal of Machine Learning Research* 15.1, pp. 1593–1623.
- Houska, Boris et al. (2015). "An economic objective for the optimal experiment design of nonlinear dynamic processes". In: *Automatica* 51, pp. 98–103.
- Hunter, William G and Albey M Reiner (1965). "Designs for discriminating between two rival models". In: *Technometrics* 7.3, pp. 307–323.
- Jacxsens, Liesbeth et al. (2000). "Designing equilibrium modified atmosphere packages for fresh-cut vegetables subjected to changes in temperature". In: *LWT-Food Science and Technology* 33.3, pp. 178–187.
- Jauberthie, Carine et al. (2018). "Optimal input design for parameter estimation in a bounded-error context for nonlinear dynamical systems". In: *Automatica* 92, pp. 86–91.
- Jaulin, Luc et al. (2001). "Interval analysis". In: *Applied interval analysis*. Springer, pp. 11–43.
- Johnson, Steven G (2014). *The NLOpt nonlinear-optimization package*.
- Kantas, Nicholas et al. (2009). "An overview of sequential Monte Carlo methods for parameter estimation in general state-space models". In: *IFAC Proceedings Volumes* 42.10, pp. 774–785.
- Kennedy, Marc C and Anthony O'Hagan (2001). "Bayesian calibration of computer models". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63.3, pp. 425–464.
- Körkel, Stefan et al. (2004). "Numerical methods for optimal control problems in design of robust optimal experiments for nonlinear dynamic processes". In: *Optimization Methods and Software* 19.3-4, pp. 327–338.
- Kraft, Dieter et al. (1988). "A software package for sequential quadratic programming". In.
- Kraft, Dieter (1994). "Algorithm 733: TOMP–Fortran modules for optimal control calculations". In: *ACM Transactions on Mathematical Software (TOMS)* 20.3, pp. 262–281.
- Lammertyn, Jeroen et al. (2003). "A respiration–diffusion model for Conference pears: II. Simulations and relation to core breakdown". In: *Postharvest Biology and Technology* 30.1, pp. 43–55.
- Lane, Adam (2017). "Adaptive Designs for Optimal Observed Fisher Information". In: *arXiv preprint arXiv:1712.08499*.

- Léchaudel, Mathieu and Jacques Joas (2007). "An overview of preharvest factors influencing mango fruit growth, quality and postharvest behaviour". In: *Brazilian Journal of Plant Physiology* 19.4, pp. 287–298.
- Li, Guilin, Szu Hui Ng, and Matthias Hwai-yong Tan (2020). "Bayesian optimal designs for efficient estimation of the optimum point with generalised linear models". In: *Quality Technology & Quantitative Management* 17.1, pp. 89–107.
- Liepe, Juliane et al. (2013). "Maximizing the information content of experiments in systems biology". In: *PLoS Computational Biology* 9.1.
- Lindley, Dennis V (1956). "On a measure of the information provided by an experiment". In: *The Annals of Mathematical Statistics*, pp. 986–1005.
- Meyer, Ruth K and Christopher J Nachtsheim (1995). "The coordinate-exchange algorithm for constructing exact optimal experimental designs". In: *Technometrics* 37.1, pp. 60–69.
- Michaelis, Leonor and Maud Leonora Menten (1913). *Die Kinetik der Invertinwirkung*. Vol. 49. Universitätsbibliothek Johann Christian Senckenberg, pp. 333–369.
- Mogensen, Patrick Kofod and Asbjørn Nilsen Riseth (2018). "Optim: A mathematical optimization package for Julia". In: *Journal of Open Source Software* 3.24.
- Munack, Axel and Clemens Posten (1989). "Design of optimal dynamical experiments for parameter estimation". In: *American Control Conference, 1989*. IEEE, pp. 2010–2016.
- Nahor, H B, Nico Scheerlinck, Jan F Van Impe, et al. (2003). "Optimization of the temperature sensor position in a hot wire probe set up for estimation of the thermal properties of foods using optimal experimental design". In: *Journal of Food Engineering* 57.1, pp. 103–110.
- Nahor, H B, Nico Scheerlinck, R Verriest, et al. (2001a). "Optimal experimental design for the parameter estimation of conduction heated foods". In: *Journal of Food Engineering* 48.2, pp. 109–119.
- (2001b). "Optimal experimental design for the parameter estimation of conduction heated foods". In: *Journal of Food Engineering* 48.2, pp. 109–119.
- Nimmegeers, Philippe et al. (2020). "Optimal experiment design under parametric uncertainty: a comparison of a sensitivities based approach versus a polynomial chaos based stochastic approach". In: *Chemical Engineering Science*, p. 115651.
- Nocedal, Jorge and Stephen Wright (2006). *Numerical optimization*. Springer Science & Business Media.
- Overstall, Antony M, David C Woods, and Ben M Parker (2019). "Bayesian optimal design for ordinary differential equation models with application in biological science". In: *Journal of the American Statistical Association*, pp. 1–36.

- Paulson, Joel A, Marc Martin-Casas, and Ali Mesbah (2019). “Optimal Bayesian experiment design for nonlinear dynamic systems with chance constraints”. In: *Journal of Process Control* 77, pp. 155–171.
- Peppelenbos, Herman W and Johan van’t Leven (1996). “Evaluation of four types of inhibition for modelling the influence of carbon dioxide on oxygen consumption of fruits and vegetables”. In: *Postharvest Biology and Technology* 7.1-2, pp. 27–40.
- Pieczywek, Piotr M and Artur Zdunek (2017). “Compression simulations of plant tissue in 3D using a mass-spring system approach and discrete element method”. In: *Soft matter* 13.40, pp. 7318–7331.
- Pintelon, Rik and Johan Schoukens (2012). *System identification: a frequency domain approach*. John Wiley & Sons.
- Rackauckas, Christopher, Yingbo Ma, Vaibhav Dixit, et al. (2018). “A comparison of automatic differentiation and continuous sensitivity analysis for derivatives of differential equation solutions”. In: *arXiv preprint arXiv:1812.01892*.
- Rackauckas, Christopher, Yingbo Ma, Julius Martensen, et al. (2020). “Universal differential equations for scientific machine learning”. In: *arXiv preprint arXiv:2001.04385*.
- Rackauckas, Christopher and Qing Nie (2017). “Differentialequations. jl—a performant and feature-rich ecosystem for solving differential equations in Julia”. In: *Journal of Open Research Software* 5.1.
- Revels, Jarrett, Miles Lubin, and Theodore Papamarkou (2016). “Forward-mode automatic differentiation in Julia”. In: *arXiv preprint arXiv:1607.07892*.
- Ryan, Elizabeth G et al. (2016). “A review of modern computational algorithms for Bayesian optimal design”. In: *International Statistical Review* 84.1, pp. 128–154.
- Saltveit, Mikal E (2003). “Is it possible to find an optimal controlled atmosphere?” In: *Postharvest Biology and Technology* 27.1, pp. 3–13.
- Santner, Thomas J et al. (2003). *The design and analysis of computer experiments*. Vol. 1. Springer.
- Särkkä, Simo (2013). *Bayesian filtering and smoothing*. Vol. 3. Cambridge University Press.
- Särkkä, Simo and Arno Solin (2019). *Applied stochastic differential equations*. Vol. 10. Cambridge University Press.
- Schenkendorf, R, A Kremling, and M Mangold (2009). “Optimal experimental design with the sigma point method”. In: *IET Systems Biology* 3.1, pp. 10–23.
- Shampine, Lawrence F and Mark W Reichelt (1997). “The Matlab ODE suite”. In: *SIAM Journal on Scientific Computing* 18.1, pp. 1–22.
- Stojanovic, Vladimir et al. (2016). “Optimal experiment design for identification of ARX models with constrained output in non-Gaussian noise”. In: *Applied Mathematical Modelling* 40.13-14, pp. 6676–6689.

- Telen, Dries, Boris Houska, et al. (2013). “Optimal experiment design under process noise using Riccati differential equations”. In: *Journal of Process Control* 23.4, pp. 613–629.
- Telen, Dries, Filip Logist, et al. (2012). “Optimal experiment design for dynamic bioprocesses: a multi-objective approach”. In: *Chemical Engineering Science* 78, pp. 82–97.
- Titterington, DM (1980). “Aspects of optimal design in dynamic systems”. In: *Technometrics* 22.3, pp. 287–299.
- Tsitouras, Ch (2011). “Runge–Kutta pairs of order 5 (4) satisfying only the first column simplifying assumption”. In: *Computers & Mathematics with Applications* 62.2, pp. 770–775.
- Vassiliadis, Vassilios S (1993). “Computational Solution of Dynamic Optimization Problems with General Differential-Algebraic Constraints”. PhD thesis. Imperial College, University of London.
- Versyck, Karina J, Johan E Claes, and Jan F Van Impe (1997). “Practical identification of unstructured growth kinetics by application of optimal experimental design”. In: *Biotechnology progress* 13.5, pp. 524–531.
- Villaverde, Alejandro F, Antonio Barreiro, and Antonis Papachristodoulou (2016). “Structural identifiability of dynamic systems biology models”. In: *PLoS computational biology* 12.10, e1005153.
- Von Mises, Richard (2014). *Mathematical theory of probability and statistics*. Academic Press.
- Walter, Éric and Luc Pronzato (1994). *Identification de modèles paramétriques à partir de données expérimentales*. Masson.
- Yoneyama, Tadakatsu et al. (2016). “Whole-plant dynamic system of nitrogen use for vegetative growth and grain filling in rice plants (*Oryza sativa* L.) as revealed through the production of 350 grains from a germinated seed over 150 days: a review and synthesis”. In: *Frontiers in plant science* 7, p. 1151.

Publications

- **Strouwen, A.**, Nicolaï, B.M., Goos, P. (2019). Optimizing Oxygen Input Profiles for Efficient Estimation of Michaelis-Menten Respiration Models. *Food and Bioprocess Technology*, 12 (5), 769-780.
- **Strouwen, A.**, Goos, P. (2019). A Note on the Output of a Coordinate-Exchange Algorithm for Optimal Experimental Design. *Chemometrics and Intelligent Laboratory Systems*.

FACULTY OF BIOSCIENCE ENGINEERING
DEPARTMENT OF BIOSYSTEMS
MEBIOS
Kasteelpark Arenberg 30
B-3001 Leuven
arno.strouwen@kuleuven.be

