Uncertainty analysis in systems biology

Uncertainty Analysis in Systems Biology

Cover:
The mechanisms underpinning biochemical systems are often inferred from measurements on the system in various conditions. By integrating various data sources, we hope to obtain one coherent picture of the underlying system. The cover shows a mountain range composed of several different pictures representing data sources. The underlying system, the mountain range, only becomes apparent after integrating the different parts.

A catalogue record is available from the Eindhoven University of Technology Library.

ISBN: 978-94-6191-995-3

Cover design: M. Vanlier

Printed by: Ipskamp Drukkers, Enschede, The Netherlands

This work was supported by the Netherlands Consortium for Systems Biology (NCSB)

Uncertainty Analysis in Systems Biology

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Technische Universiteit Eindhoven, op gezag van de
rector magnificus, prof.dr.ir. C.J. van Duijn, voor een
commissie aangewezen door het College voor
Promoties in het openbaar te verdedigen
op dinsdag 28 januari 2014 om 16:00 uur

door

Joep Vanlier

geboren te Helden

Dit proefschrift is goedgekeurd door de promotor:

prof.dr. P.A.J. Hilbers

Copromotor:
dr.ir. N.A.W. van Riel

Contents

1 Brief Introduction

Biology has made significant progress unraveling the complexity of life and its constituents. By studying phenotypical differences due to interventions and diseases, biology aims to infer causal relationships between various interacting components at a large scale. Classical biochemistry relied mostly on isolating parts of the system from the whole and studying the chemistry of the individual components. The field of genetics dealt with the structure and function of genes. Over time, these fields produced a large body of knowledge, from sequencing the entire genome and unraveling enzyme kinetics, to identifying the various molecules and pathways that play a role in biological systems. Nevertheless, it has also become increasingly apparent that for many diseases and syndromes a clear chain of causation is absent and that the large scale behavior emerges from the interactions between the different components (genes, proteins and small molecules). With the rise of molecular biology, the search for the mechanistic laws that drive these systems has been taken to the molecular level. Merely studying the individual components does not suffice and investigation of the interplay between the various modules of such systems is essential for understanding their emergent properties and behavior [1,2].

This is where systems biology (also known as integrative biology) comes into the picture. Systems biology is a scientific discipline and paradigm where the central theme is integration. Its main focus is the study of how emergent properties arise from the sum of their parts. This is a daunting task which requires integration in terms of system components, time scales as well as different data sources. Integrative approaches involve studying the system as a whole, while quantitatively observing multiple system components simultaneously. Subsequently, methods from computational biology, control theory and bio-informatics are applied to perform data integration and construct testable hypotheses. Though the aim of biology and biochemistry has always been to understand the mechanisms underlying biological processes and pathways, it is due to recent advances in computational abilities and measurement technologies that the systems biology approach becomes feasible for increasingly complex systems.

In many cases, it is possible to formulate conceptual understanding in the form of mathematical models. This process forces the investigator to concretely formalize the mechanisms that play a role in the system. This endeavor requires information on how the different model components are connected as well as knowledge on the kinetic behavior of the molecular species involved. Subsequently, these models can be implemented in computer software to be simulated and put to the test quantitatively. Tools from systems analysis and applied statistics can be used to predict behavior and formulate biological hypotheses in a quantifiable and testable manner. Any computer model is a simplification of reality, which manifests itself as some features being over-expressed, while others are absent. In this way, computer models provide caricatures of reality that aid in understanding the phenomena they were developed to describe. By repeatedly challenging such models with new informative data, our understanding can iteratively be improved. The ambition is that over time, such models become increasingly predictive and explanatory for the phenomena they were designed to describe, thereby providing answers to some of the difficult questions in the

life sciences. Which aspects of the real system should (or can) be incorporated in the model depends on the aims of the modeling endeavor and the availability of measurement data. The development of mathematical models and the generation of new data should therefore not be seen as two separate entities but a tightly interwoven process.

Dynamical models

Dynamical models are a way to describe the time evolution of a system by considering the time dependence of its state. Such models consist of state variables (or states), which are quantities that change over time and parameters (which are fixed with respect to time). These are then embedded in a system of equations which relate the different states of the model to time. To simulate the model and make predictions, model parameter values are required. Obtaining (or estimating) these parameter values constitutes a major challenge in the development of computational models. Measurements are typically used by simulating the experiments performed in the lab and calibrating the observable model outputs to their measured counterparts by adjusting model parameters. Initially, a model can be too simple and merely estimating the parameters does not achieve a sufficient agreement with the experimentally acquired data. When this occurs changes in the model are required. The iterative cycle of experimentation and model refinement often leads to relatively complex models with many parameters.

Though the molecular bio-sciences have made great advances in providing access to the various components of the system, not all states are accessible and not all data is equally suitable for dynamical modeling. Since several measurement technologies employed in the molecular bio-sciences are of a qualitative or semi-quantitative nature, one is often forced to incorporate ratios of molecular species rather than absolute quantities or adding unknown experimental scaling factors to the model parameters. Additionally, biological variability and noise result in large experimental variabilities for those components that are actually measured. These experimental limitations raise doubts on how reliably parameter values can be inferred for models of increasing complexity. In practical situations, this lack of data often results in several poorly constrained and therefore uncertain model parameters and predictions [3–7]. This makes devising a mechanistic dynamical model a difficult and often iterative task. Whereas the larger models are plagued by large uncertainties, lumped approximations are not always considered as useful for unraveling mechanistic detail as the original (albeit uncertain) model. No consensus has been reached on the scope and level of mechanistic detail on which inference can still reliably be performed. Being able to assess, report and effectively reduce both parameter and prediction uncertainties is crucial for the future of model development in systems biology.

Metabolic Syndrome

Metabolic syndrome is a combination of whole-body symptoms associated with overweight individuals at an increased morbidity risk. The actual mechanisms

behind metabolic syndrome are still poorly understood. Unraveling long-term adaptations in the underlying biochemical network is complicated by the multi-level aspects of the system and the time-scale on which the syndrome develops. One long-term goal is to unravel the changes taking place in the system by integrating various sources of data and mathematical modeling. Considering the complexity of the system and the ethical and financial cost of performing measurements, it is important that prediction uncertainties are identified, quantified and reduced as efficiently as possible. Classical methodologies are typically only suited to either work well on small models or in the limit of large amounts of data. Within this project, neither of these situations apply. Hence, there is a need to identify methodologies that perform well on non-linear models associated with large uncertainties. One of the requirements of the methodologies presented in this thesis is that they are applicable in practice and perform well on models of representative complexity.

The aim of this dissertation is to provide tools and strategies for performing uncertainty analysis and assess their applicability to computational models of biochemical pathways. Secondly, to provide methods for selecting experiments which reduce uncertainties in an optimal manner. Finally, to apply the developed methodologies to some relevant cases.

This work is divided into the following chapters:

Chapter 2 - Parameter Uncertainty in Biochemical Models

This chapter provides a brief introduction to parameter uncertainty analysis for dynamic models described by ordinary differential equations. The suitability of different techniques for addressing uncertainty in such models is discussed. Additionally, some state-of-the-art methods for uncertainty analysis are introduced.

Chapter 3 - A Strategy for Prediction Uncertainty Analysis

Here, a strategy for prediction uncertainty analysis is presented. The approach integrates Profile Likelihood analysis with Bayesian sampling and involves performing different analyses sequentially to detect and avoid problems associated with the individual techniques. The result is a sample from the posterior predictive distribution, which forms the basis for Chapters 4 and 5. Finally, the methodology is illustrated on a model of the JAK-STAT signaling pathway.

Chapter 4 - A Bayesian Approach to Experimental Design

The uncertainties of different model predictions are often related. Here, we present a method to select experiments that will reduce the uncertainty of specific predictions in an optimal manner. By applying importance sampling to the posterior predictive distribution, the efficacy of a new measurement is predicted. Because predictive distributions can be computed for a wide range of quantities, this approach is very flexible. The combinatorial effect of designing multiple experiments simultaneously can also be probed.

Chapter 5 - Optimal Experimental Design for Model Selection

Because of the limited experimental accessibility of the system, different hypotheses on how a biochemical pathway operates typically exist. Differences between these hypotheses can exist both in the kinetic equations as well as the model topologies. In some cases, certain models clearly outperform others in terms of providing a good description of the data. When this is not the case, more experiments are required. This chapter describes a method which can be used to determine which predictions should be measured to effectively discriminate between various models.

Chapter 6 - Computational Modeling of Diacylglycerol Transferase

Excessive accumulation of triglycerides (TG) or hypertriglyceridemia is one of the symptoms of Metabolic Syndrome and has been implicated as an important risk factor for various diseases. Production of TG occurs via two major pathways which converge into a final reaction where fatty acids (FA) and diacylglycerol (DAG) are bound into TG. Diacylglycerol acyltransferases (DGAT) are membrane bound enzymes which are primarily responsible for catalyzing this acylation of DAG. The first steps towards a computational model of the DGAT system are presented. Based on literature, a network topology is proposed and subsequently used to formulate a mathematical model. Though certain observations seemed contradictory at first, the model consolidates multiple datasets revealing that they are not inherently inconsistent. Lack of flux measurements at the system boundaries turned out to be detrimental for the predictive power of the model.

Chapter 7 - Optimal Experimental Design for Identifying Progressive Adaptations

This chapter describes a generally applicable method for designing experiments in order to improve predictions of progressive changes in biological systems. The proposed method captures the modulating effects on the metabolic level using time-dependent descriptions (or trajectories) of the model parameters, which is useful when mechanistic interactions are unknown. Subsequently, these parameter trajectories are used to obtain predictions for experiment design. Additionally a bias correction for the methodology presented in Chapter 4 is introduced for cases where the effective sample size is too low.

References

[1] Westerhoff H, Palsson B: **The evolution of molecular biology into systems biology**. *Nature biotechnology* 2004, **22**(10):1249–1252.

[2] Bruggeman F, Westerhoff H: **The nature of systems biology**. *TRENDS in Microbiology* 2007, **15**:45–50.

[3] Vanlier J, Tiemann C, Hilbers P, van Riel N: **An integrated strategy for prediction uncertainty analysis**. *Bioinformatics* 2012, **28**(8):1130–1135.

[4] Brännmark C, Palmér R, Glad S, Cedersund G, Strålfors P: **Mass and information feedbacks through receptor endocytosis govern insulin signaling as revealed using a parameter-free modeling framework**. *Journal of Biological Chemistry* 2010, **285**(26):20171.

[5] Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmüller U, Timmer J: **Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood**. *Bioinformatics* 2009, **25**(15):1923.

[6] Girolami M, Calderhead B: **Riemann manifold langevin and hamiltonian monte carlo methods**. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2011, **73**(2):123–214.

[7] Calderhead B, Girolami M: **Statistical analysis of nonlinear dynamical systems using differential geometric sampling methods**. *Interface Focus* 2011, **1**(6):821–835.

Parameter Uncertainty in Biochemical Models

2

Parts of this chapter are described in:

Vanlier, J. and Tiemann, C.A. and Hilbers, P.A.J. and van Riel, N.A.W. (in press), *Parameter Uncertainty in Biochemical Models described by Ordinary Differential Equations*, *Mathematical BioSciences*

Abstract

Improved mechanistic understanding of biochemical networks is one of the driving ambitions of Systems Biology. Computational modeling allows the integration of various sources of experimental data in order to put our conceptual understanding to the test in a quantitative manner. The aim of computational modeling is to obtain both predictive as well as explanatory models for complex phenomena, hereby providing useful approximations of reality with varying levels of detail. As the complexity required to model different systems increases, so does the need to determine with what accuracy model predictions can be made. Despite efforts to make tools for uncertainty analysis available to the field, these methods have not yet found widespread use in the field of Systems Biology. Additionally, the suitability of the different methods strongly depends on the problem and system under investigation. This chapter provides both an overview as well as a friendly introduction to parameter uncertainty analysis. A selection of techniques, including state-of-the-art methods for uncertainty analysis, will be discussed.

2.1 Introduction

In the past decades, molecular biology has unraveled various pathways that play a role in biological phenomena. Many of the components and interactions involved have been identified by employing a reductionist approach, isolating parts of the system from the whole. Gradually, a complex network of various interacting subsystems is being revealed. Investigation of the interactions between various components of a system is essential for understanding its emergent behavior. Consequently, there has been a shift in paradigm from reductionism to integration [1–4]. By formalizing hypotheses on how a pathway operates in the form of computational models our conceptual understanding can be put to the test quantitatively [5–7]. By repeatedly challenging such models with new data, it is possible to iteratively obtain models that are decreasingly wrong. The aim of Systems Biology is to obtain both predictive as well as explanatory models for complex phenomena, hereby providing useful approximations of reality. This chapter focuses on dynamical models which consist of state variables, which are quantities that change over time, and parameters (fixed with respect to time). These state variables are embedded in a system of equations which relate the different state variables in the model.

To simulate the model and make predictions, parameter values are required. These typically have to be estimated from data. Due to the fact that data is measured with finite accuracy and only a subset of the state variables is accessible experimentally, uncertainty analysis is a highly relevant topic. Despite efforts to make tools for uncertainty analysis available [8–11] these methods have not yet found widespread use in the field. Nevertheless, some examples of successful applications in computational biology are listed in Table 2.1. This chapter provides an introduction to some of the techniques available as well as give an overview of the state-of-the-art methods for parameter uncertainty analysis [12–17]. Which methods are applicable to a specific problem strongly depends on the system under investigation and the assumptions one is willing to make. The suitability of different techniques for addressing uncertainty in computational modeling of dynamical systems is discussed.

Name	Requirements	Result	Papers
Sensitivity Analysis	None	Assessment of (local) sensitivity	[18–21]
Profile Likelihood	Likelihood function	Assessment of identifiability	[7,22]
Bootstrap	None	Distribution of parameters based on simulated replicates	[15,23,24]
Markov Chain Monte Carlo	Weak identifiability	Posterior distribution	[25–29]
Sequential Monte Carlo	Proper priors for all parameters	(Approximate) posterior distribution	[30–32]

Table 2.1: A list of the methods discussed in this paper and some relevant applications in the field of Systems Biology

2.2 Computational modeling of biochemical systems

The use of dynamical models has a long history within several disciplines of science. In the realm of classical physics, models often comprise of physical laws with well established and invariant physical constants acting as parameters. In engineering, parameters are often application specific, but the individual components are usually well characterized and their interactions known. In computational biology, the challenges are different from those in many other fields. Though methods for discovering interactions are well established [33–35], techniques for accurately determining biochemical parameters remain limited [12]. Moreover, relying on measurements that were performed *in vitro* can lead to inconsistencies [36] due to differences in the biochemical environment and regulatory mechanisms that were not accounted for [16]. Despite the wealth of information that kinetic assays provide, such issues require attention and warrant future research. Since such measurements are both expensive as well as time consuming, the amount of data is often relatively scarce. Considering that models are required to have a certain level of complexity to describe biological pathways, this leads to large uncertainties in the inferred values of these biochemical parameters. Consequently, the investigator is faced with several poorly constrained and therefore uncertain model predictions [10,24,37–39]. To deal with such uncertainties and to ascertain their implications on scientific conclusions, several methods have been developed. Some of these are useful for probing model properties, others for designing informative experiments.

2.3 Parameter estimation

The scope of this chapter is restricted to dynamical systems that can be described by ordinary differential equations (ODEs). Such a description is appropriate when the number of particles involved in the biochemical network is large enough to be able to consider concentrations and when spatial effects are negligible [40–45]. Typically, such models take the following form:

$$\begin{aligned} \dot{\vec{x}}(t) &= \mathbf{N}\vec{f}(\vec{x}(t), \vec{u}(t), \vec{p}) \\ \vec{y}(t) &= \vec{g}(\vec{x}(t), \vec{q}) \\ \vec{y}_{obs}(t) &= \vec{g}(\vec{x}(t), \vec{q}) + \vec{\xi}(t) \\ \vec{x}(0) &= \vec{h}(\vec{r}) \end{aligned} \tag{2.1}$$

Here $\mathbf{N}$ represents the reaction stoichiometry of the system or the quantitative relationships among substances as they participate in chemical reactions. Vector $\vec{f}$ is a vector of flux expressions and $\vec{g}$ is a vector of equations which map the internal state variables to an output. These equations contain parameters $\vec{p}$ (which are fixed constants with respect to time), inputs $\vec{u}(t)$ which depend on time, and state variables $\vec{x}(t)$. Given a set of parameters, inputs and initial conditions $\vec{x}(0)$, these equations can subsequently be simulated, and time courses of the model state variables can be obtained. Such systems are typically partially observed, which means that measurements $\vec{y}_{obs}(t)$ can only be performed on a subset or a

combination of the total number of state variables N in the model. Additionally, these measurements are hampered by measurement noise $\vec{\xi}$ and many of the employed techniques necessitate the use of scaling and offset parameters $\vec{q}$ [46]. For ease of notation, $\vec{\theta}$ is defined as $\vec{\theta} = \{\vec{p}, \vec{q}, \vec{r}\}$, which lists all the parameters that should be defined in order to simulate the model. After the model is postulated and data is acquired, parameter estimation can be performed. To do this, consider the probability density of observing data $\mathbf{y}$ consisting of data points y_{ij} given parameter values $\vec{\theta}$. For the sake of notation, we assume independent additive Gaussian noise with constant variance for each measurement, which results in a probability density function defined as:

$$\begin{aligned} p(\mathbf{y}|\vec{\theta}) &= \prod_{i=1}^{M}\prod_{j=1}^{N_i} p(y_{ij}, \vec{\theta}) \\ &= \left(\prod_{i=1}^{M}\left(\sqrt{2\pi}\sigma_i\right)^{N_i}\right)^{-1} \exp\left(-\sum_{i=1}^{M}\sum_{j=1}^{N_i}\left(\frac{y_{ij} - y_i(t_j, \vec{\theta})}{\sqrt{2}\sigma_i}\right)^2\right) \end{aligned} \tag{2.2}$$

Here p refers to a probability density. It is at this point where the so-called frequentist and Bayesian methodologies begin to diverge. The former opts for a purely data-based approach, aiming to find all the parameter sets which describe the observed data $\mathbf{y^D}$ to an acceptable degree. What is considered acceptable is evaluated by means of the likelihood function $L(\vec{\theta})$, whose right hand side is the same as (2.2). Although the likelihood function is a function of the parameters, it should not be confused with a parameter probability density as it makes no probabilistic assessment of the parameter values. The threshold L_{lim} for determining whether a parameter set is acceptable $L(\vec{\theta}) > L_{lim}$, is typically determined by choosing a significance level and computing the associated critical values of the uncertainty distribution associated with the noise on the data.

The Bayesian approach treats the parameters as random variables and uses Bayes' theorem (2.3) to perform inference probabilistically (2.4) [25, 47].

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{2.3}$$

$$p(\vec{\theta}|\mathbf{y}^D) = \frac{p(\mathbf{y}^D|\vec{\theta})p(\vec{\theta})}{p(\mathbf{y}^D)} \quad = \frac{p(\mathbf{y}^D|\vec{\theta})p(\vec{\theta})}{\int_\Omega p(\mathbf{y}^D|\vec{\theta})p(\vec{\theta})d\vec{\theta}} \tag{2.4}$$

Here, the probability density of the parameter values is given by $p(\vec{\theta}|\mathbf{y}^D)$ which can be computed from the probability density of the data $p(\mathbf{y}^D|\vec{\theta})$ given parameters $\vec{\theta}$, the prior probability density of the parameters $p(\vec{\theta})$ and the marginal likelihood or model evidence $p(\mathbf{y}^D)$. Since the marginal likelihood does not depend on the parameters, this merely acts as a normalizing constant.

Priors for the parameters are specified in the form of probability densities and are defined with respect to a specific parameterization of the model. They usually represent either current belief [48] or attempt to be non-informative. The latter is

usually accomplished either by choosing wide priors (such as gamma priors for parameters with only positive support [39]) or priors that exhibit invariance to parameter transformations [49]. In brief, the aim is to sample from the posterior parameter probability density and determine bounds enclosing $(100-\alpha)\%$ of the probability density.

In the frequentist paradigm, one usually proceeds by maximizing the likelihood function $L(\vec{\theta})$ (Maximum Likelihood Estimation or MLE). The result of maximizing the likelihood is finding those parameters for which the probability density of obtaining the observed data is maximal. In practice, one often minimizes the negative log-likelihood as it allows for efficient optimization algorithms due to its quadratic nature. This is permitted since the logarithm does not change the location of the optimum with respect to the parameters. When the data variances are known, the normalization constant is independent of the parameters and minimizing the Residual Sum of Squares (RSS) becomes equivalent to maximizing the likelihood (2.5):

$$\chi^2(\vec{\theta}) = \sum_{i=1}^{M}\sum_{j=1}^{N_i}\left(\frac{y_{ij}^D - y_i(t_j,\vec{\theta})}{\sigma_i}\right)^2 \tag{2.5}$$

When the variances σ_i have to be estimated from the data however, they should preferably be treated as additional parameters. Since σ_i appears in the normalizing constant of the additive noise distribution, such an approach leads to an additional term in the negative log likelihood resulting in:

$$-\ln\left(L(\vec{\theta})\right) = \frac{1}{2}\left(\chi^2(\vec{\theta}) + \sum_{i=1}^{M} N_i \ln\left(2\pi\sigma_i^2\right)\right) \tag{2.6}$$

When additional information regarding the parameters is available, one can include a prior distribution on the parameters and perform estimation based on the combined quantity. Maximizing this quantity results in the *Maximum A Posteriori* (MAP) estimator, which can be interpreted as a regularized form of MLE.

Determining which parameters maximize these quantities is a non-trivial problem due to the existence of locally optimal solutions, large discrepancies in the ranges of the different parameter values and various ridges present in these likelihood surfaces [13, 37, 38]. Estimation is commonly performed using an optimization algorithm. These can roughly be categorized into two classes: probabilistic methods, which propose and accept or reject steps in parameter space in a probabilistic manner, and deterministic methods. Whereas the stochastic methods usually take longer to converge, they run a smaller risk of converging to a local sub-optimal minimum [50, 51]. In the deterministic case, this is usually approached by starting the deterministic optimizer using widely dispersed starting values [16]. Note that both come in forms which either incorporate information about the local geometry or not and that including such information usually leads to faster convergence [11, 51]. Once an optimal parameter set is obtained, the residuals, or weighted errors can be inspected for any residual trend or autocorrelation to see whether the model describes the data to an acceptable degree [15].

2.4 Parameter uncertainty

Once the model describes the data, it is possible to proceed to the next question: with how much certainty can predictions be made? And what should be measured in order to put the model to the test? This section will briefly discuss different approaches for assessing parameter uncertainty. We will briefly discuss the topic of sensitivity and identifiability before introducing frequentist and Bayesian sampling based methods for uncertainty quantification.

2.4.1 Sensitivity

Common practice in the field of systems biology is to perform some sort of sensitivity analysis during this step, either by performing a local parameter sensitivity analysis (LPSA) or by varying multiple parameters simultaneously (global parameter sensitivity analysis, GPSA). The first technique is performed by perturbing one parameter at a time and comparing the simulation result to a reference output (2.7) [5]:

$$S_{\theta_i}^{M} = \frac{M(\theta_i + \Delta\theta_i) - M(\theta_i)}{M(\theta_i)} \frac{\theta_i}{\Delta\theta_i} \tag{2.7}$$

Here M refers to some prediction of interest, which is based on model simulations obtained with parameter value θ_i. Despite being intuitive, results from such an analysis should be considered with care since different reference parameters can lead to different sensitivities [52, 53]. To mitigate this issue, it is sensible to compute a weighted average of local sensitivities based on the probability of the chosen parameter values when the true parameters are highly uncertain [54]. Though informative, what this methodology provides is not an assessment of the prediction uncertainty, but rather an assessment of the consequences a parameter perturbation would have on specific predictions of interest.

The second class of techniques, the so-called Global Parameter Sensitivity Analyses [52, 55, 56] are meant to circumvent the issue of only perturbing one parameter at a time. Here large regions of parameter space are sampled and used to compute model sensitivities. One specific method for GPSA, known as Monte Carlo Filtering classifies simulations corresponding to the sampled parameter sets in two classes. Subsequently, the samples are sorted according to each parameter independently and cumulative distributions of both classifications are computed. The supremum of the difference between these distributions (Kolmogorov-Smirnov distance) now serves as a metric indicating how strongly the classification correlates to that specific parameter, *i.e.*, how sensitive the quantity the filter is based on is to the uncertainty in that specific parameter (see Figure 2.1). In this sense, the name sensitivity analysis might be a bit of a misnomer and uncertainty analysis might be a preferable term. The uncertainty distributions used when performing this sampling are crucial for the answer that one obtains. If the data implies that specific model parameters have to be correlated, it is important to incorporate these relations when performing the sampling.

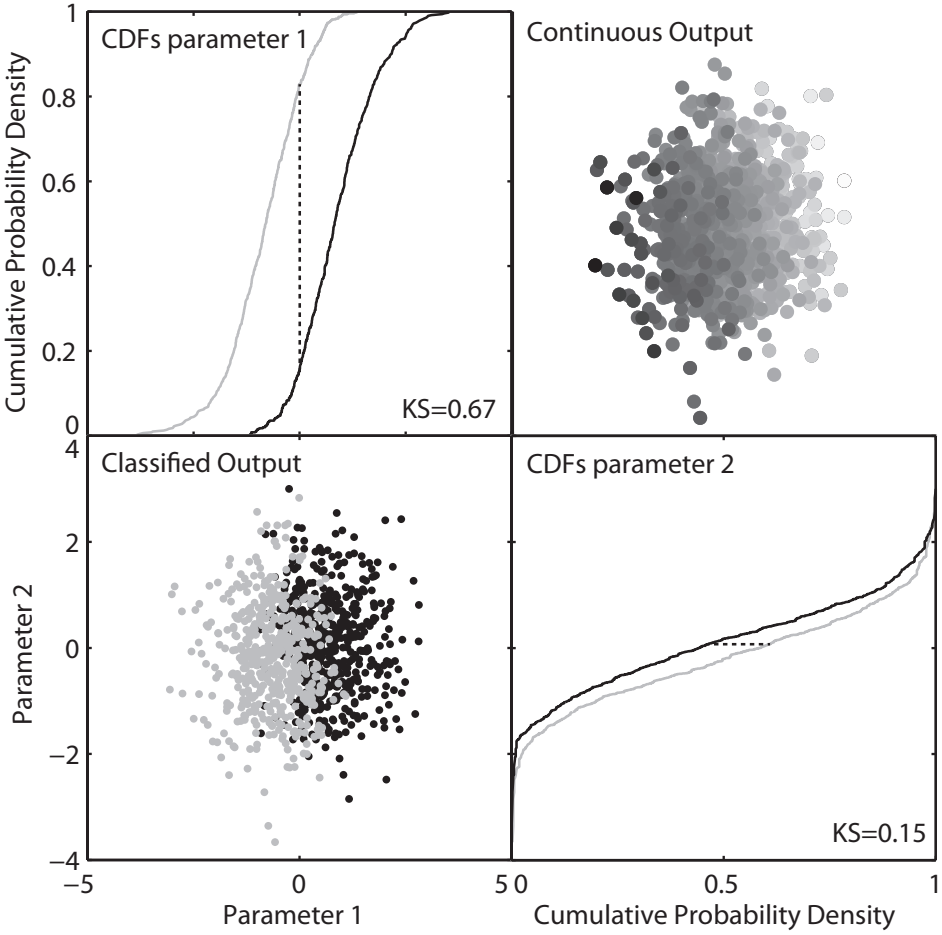

Figure 2.1: Graphical illustration of the Global Sensitivity Analysis based on Monte Carlo Filtering. Top right: Output quantity shown for various combinations of parameters. Lower left: Classification of the sampled parameter sets. The different classes are indicated with gray and black respectively. Top left and bottom right: Cumulative distribution functions of the two classes (solid lines) with the KS distance between the two (dashed lines). Note how the parameter responsible for the largest difference in the output quantity results in the largest KS distance.

2.4.2 Asymptotics

Although sensitivity analyses are a useful tool for analyzing the consequences of specific uncertainties, they give little information regarding how well these parameters can be estimated. If the model is identifiable and certain regularity conditions are met (see [57] for specifics), the MLE distribution tends to a Gaussian distribution for large amounts of data. The covariance matrix of this parameter estimate can be obtained as $C = H^{-1}$, where H refers to the Fisher Information Matrix (FIM) which is given by the negative Hessian of the log-likelihood. Calculating the true Hessian is costly and numerically challenging, which is why an approximation based on the first order model sensitivities is often used:

$$\begin{aligned} H_{kl} =& \frac{\delta^2 \ln L(\vec{\theta})}{\delta\theta_k \delta\theta_l} \\ =& \sum_{i=1}^{M} \sum_{j=1}^{N_i} \left(\frac{1}{2\sigma_{ij}^2} \left(\left(y_{ij}^D - y_i(t_j, \vec{\theta}) \right) \frac{\delta^2 y_i(t_j, \vec{\theta})}{\delta\theta_k \delta\theta_l} - \frac{\delta y_i(t_j, \vec{\theta})}{\delta\theta_l} \frac{\delta y_i(t_j, \vec{\theta})}{\delta\theta_k} \right) \right) \\ \approx& - \sum_{i=1}^{M} \sum_{j=1}^{N_i} \left(\frac{1}{2\sigma_{ij}^2} \left(\frac{\delta y_i(t_j, \vec{\theta})}{\delta\theta_l} \frac{\delta y_i(t_j, \vec{\theta})}{\delta\theta_k} \right) \right) \end{aligned} \tag{2.8}$$

This approximation is reasonable when the residuals are sufficiently small. Depending on the model, these first order sensitivities can either be computed by solving the sensitivity equations (preferred) [12, 58], or by means of finite differences (for which strict solver tolerances are required to ensure reliable derivatives). Approximate parameter confidence intervals can subsequently be determined using Student's t-distribution [59]:

$$\theta_i^{\pm} = \theta_i \pm t_{1-\alpha/2,df} \sqrt{cov(\vec{\theta})_{ii}} \tag{2.9}$$

Here t refers to Student's t-distribution with df, the degrees of freedom given by $df = N_{total} - N_{pars} + N_{bounds}$. In this equation, N_{total}, refers to the number of data points, N_{pars}, to the number of parameters, and, when parameter bounds are enforced, N_{bounds}, to the number of boundary constraints that the current estimate satisfies.

2.4.3 Identifiability

Basing conclusions on a single best fit parameter set (and its asymptotic confidence bounds) can be unreliable. Therefore, it is important to assess model behavior for all parameter values that correspond to simulations consistent with the data (and optionally prior knowledge). One concept that plays an important role here is parameter identifiability. An identifiable model is characterized by finite confidence intervals. Non-identifiability can roughly be divided in two classes. *Structural non-identifiability*, is a property of the model and observation equations and occurs when the consequences of a change in parameters can either not be observed from the measurements or when parameters are functionally related.

Several methods exist for detecting such non-identifiabilities [60, 61] and generally involve working with either Taylor expansions, generating series or assessments of the rank of the sensitivity matrix. For a review on these techniques see [61]. The second issue is *practical non-identifiability*, which results from lack of information in the data and manifests itself as a flattening out of the likelihood in a certain direction [10, 62]. In a Bayesian setting, choosing a sufficiently informative prior can prevent this flatness from propagating to the posterior parameter probability distribution.

2.4.4 Profile likelihood

Although attractive because of their low computational cost, asymptotic confidence intervals are often not appropriate. In many cases, identifiability is questionable and there is insufficient data to consider the asymptotic case. Parameters in likelihood functions can be related in a non-linear fashion and multiple optima may exist. In such cases, more reliable confidence intervals can be obtained using the Profile Likelihood method [37]. This method works by systematically tracing an optimal path over the likelihood, or, in the case of MAP estimation, the probability density function. Every profile is initiated at the best fit parameters, after which one parameter i is selected to be profiled. This parameter is subsequently changed, after which all the unchanged parameters are re-optimized. This process is subsequently repeated until the fit becomes unacceptably bad or some other criterion is met. In the Profile Likelihood case, the Residual Sum of Squares (RSS) along the path can be written as:

$$\chi^2_{PL,i}(x) = \min_{\theta_{j\neq i}} \left[\chi^2(\vec{\theta}|\theta_i = x)\right] \tag{2.10}$$

The parameter vectors associated with such a path shall be denoted as $\vec{\theta}_{PL}$. Nested models are models where one model can be transformed into the other by imposing linear constraints on the parameter values. When two models are nested ($\mathtt{M}(\vec{\theta}_{PL})$ and $\mathtt{M}(\vec{\theta}_{opt})$), their likelihood ratio is approximately distributed according to a χ^2_p distribution. Here p refers to the number of degrees of freedom which is defined as the difference in the number of parameters (which in this case is 1). Therefore, a bound based on the likelihood ratio can be used (2.11) [37]. Other equations or criteria can be employed for checking whether a fit is still acceptable. In this equation, α denotes a desired significance level.

$$-2\ln\left(\frac{L(\vec{\theta}_{PL})}{L(\vec{\theta}_{opt})}\right) \leq \chi^2_{1-\alpha,1} \tag{2.11}$$

Structural non-identifiability manifests itself as a completely flat profile, while practical non-identifiability involves flattening out of the likelihood, preventing it from reaching an appropriate bound. Note that in order to get a complete picture, profiles should be run for each separate acceptable optimum and subsequently be merged. An example is shown in Figure 2.2 where the profiles corresponding to different modes are shown in blue and red. Note how in the blue profile the second mode of parameter 4 can remain undetected otherwise.

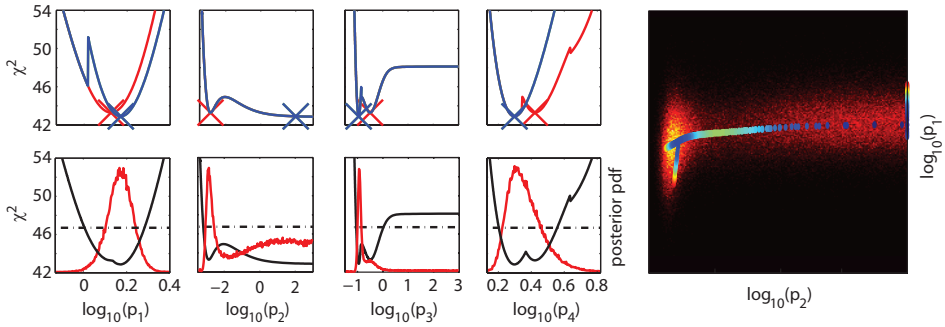

Figure 2.2: Top, Profile Likelihoods for different modes of a model of the JAK-STAT signaling pathway. Bottom, merged profile likelihoods (black, solid) with the associated likelihood threshold (black, dashed) contrasted with the Bayesian posterior distributions obtained using Markov Chain Monte Carlo (red). On the right is the joint posterior density for two model parameters along with a scatter plot of the profile likelihood (colored according to the RSS). Note how one mode clearly dominates in the Bayesian posterior density (best observed in parameter 1). This relatively large probability is caused by the flattening out of the likelihood with respect to parameter 2. As shown in the joint posterior density (again obtained using Markov Chain Monte Carlo) on the right, the mode obtains its mass from the large area corresponding to practical non-identifiability of parameter 2.

2.4.5 Sampling based methods: frequentist

A different approach to deal with uncertainty, known as bootstrapping, is to attempt to sample replicates $\{y^D_{*1}, y^D_{*2}, y^D_{*Q}\}$ of the observed data and repeat the estimation for each of these samples. The result is a distribution of parameter estimates, which gives insight into the parameter uncertainty. There are several mechanisms for obtaining bootstrap samples (which serve as simulated replicates in the estimation), which can roughly be divided in two classes: parametric and non-parametric bootstraps.

A parametric bootstrap is based on fitting a model to the data, and drawing samples from its parameterized error distribution [15, 16, 23]. In a parametric bootstrap it is important that the model used to obtain the samples closely matches the process that generated the data. Alternatively, one can make use of probabilistic models such as the Gaussian Process (GP) to obtain bootstrap samples [63]. A GP is a collection of random variables defined by a mean and a covariance function which relate the different measurements. Such covariance functions incorporate a description of the relationship between the various data points. Using the available data in combination with Bayes Rule, the parameters of such a GP are updated to a posterior distribution from which samples can be drawn. These samples can then be supplemented with noise for use as bootstrap samples [64]. The alternative (non-parametric bootstrap) is to sample with replacement from the available experimental replicates [15, 65, 66]. The general idea here is that the variability of the estimate around the true value is mimicked by the variability of estimates based on bootstrap replicates around the original estimate (see Figure 2.3). Note however that bootstrap methods of this sort require a sufficient number of replicates in order to adequately reflect variability.

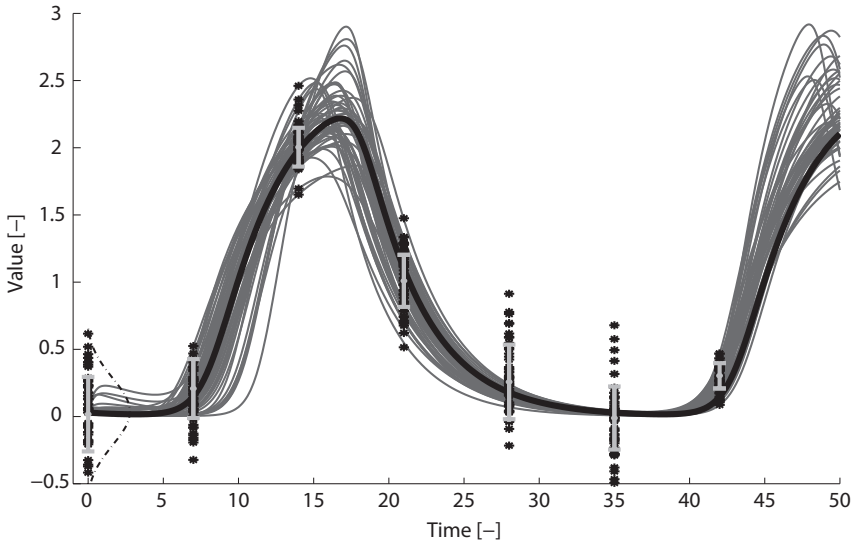

Figure 2.3: Graphical illustration of the bootstrap procedure. For each data point multiple realizations (denoted by the stars) are generated using a parametric normal distribution. Bootstraps were based on a model of a Goodwin oscillator [71]. 8 'data replicates' are generated for each time point, which were used to compute the mean and standard deviation required for the parametric bootstrap. Crosses denote the individual bootstrap samples. True model is shown in black. Grey lines indicate the different fits.

Note that combined with the data resampling, it is important that optimizations are either initiated from widely dispersed initial parameter values or that a global optimization method is used. Though no method can guarantee finding the global optimum in the non-linear case, global optimization methods have the ability to leave local modes and (initially) traverse the parameter space more freely. This is necessary since new modes can arise from the bootstrapping procedure. Based on the resulting parameter estimates, confidence intervals can be obtained by considering a significance level and determining percentile bounds [65, 67, 68]. These confidence intervals tend to underestimate the true confidence region however and need to be corrected for bias and skewness [69, 70]. Bootstraps have also been used to perform model selection [23] and testing [15].

2.4.6 Sampling based methods: Bayesian

Whereas most methods discussed earlier revolve around minimization or determining bounds, the Bayesian methods often require evaluation of probabilistic integrals. These approaches usually involve sampling probability densities to determine bounds which enclose $(100 - \alpha)\%$ of the desired density. An example of the difference between frequentist and Bayesian inference is shown in Figure 2.2. These problems often involve integration over many dimensions, which are typically evaluated using sampling based methods. An example of such an integral is shown in equation (2.12), where p in the integral refers to the probability

density of a specific parameter set $\vec{\theta}$ and x refers to the quantity whose expected value is being computed. In the sampled version, $\vec{\theta}_i$ refers to a vector of random parameter values independently drawn from a uniform distribution.

$$E[x] = \int_\Omega p(\vec{\theta})x(\vec{\theta})d\vec{\theta} \approx \sum_{i=1}^{M} p(\vec{\theta}_i)x(\vec{\theta}_i) \tag{2.12}$$

In practice, uniform random sampling is inefficient due to the low likelihood of most samples. One alternative is importance sampling. Here, samples are drawn from a different probability density function (g) than the target distribution which requires including a probability density ratio in the summation:

$$E[x] = \int_\Omega g(\vec{\theta})\frac{p(\vec{\theta})}{g(\vec{\theta})}x(\vec{\theta})d\vec{\theta} \approx \sum_{i=1}^{M} \frac{p(\vec{\theta}_i)}{g(\vec{\theta}_i)}x(\vec{\theta}_i) \tag{2.13}$$

This procedure is efficient when the importance sampling distribution overlaps well with the target distribution (see Figure 2.4). For non-linear models, it is often non-trivial to find a distribution that suits this need. Additionally, p and g refer to probability densities which should to integrate to one. Computing the required normalizing constants is difficult, since this involves integrating over the parameters. One alternative is self-normalized importance sampling but this adds the requirement that p may not have heavier tails than g:

$$E[x] \approx \frac{\sum_{i=1}^{M} w(\vec{\theta}_i)x(\vec{\theta}_i)}{\sum_{i=1} w(\vec{\theta}_i)} \text{ with } w(\vec{\theta}_i) = \frac{p(\vec{\theta}_i)}{g(\vec{\theta}_i)} \tag{2.14}$$

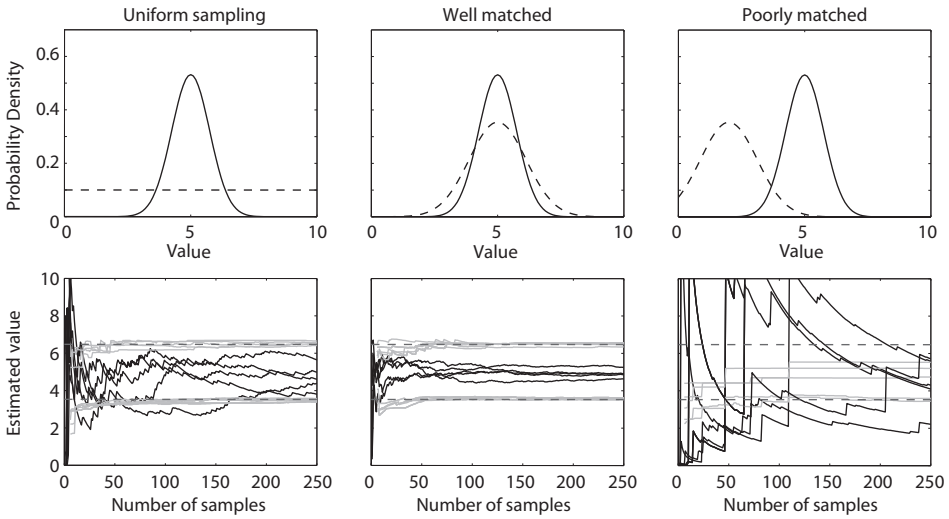

Figure 2.4: Importance sampling performance. Top row: Target probability density (solid) and importance sampling distribution (dashed lines). Bottom row: Estimate based on the importance sampling identity. Grey lines indicate the 95% interval based on the acquired samples. Dashed lines indicate true 95% interval.

Markov Chain Monte Carlo

The difficulty of finding suitable importance sampling densities make sampling methods such as Markov Chain Monte Carlo (MCMC) an attractive option. This class of methods can sample directly from any arbitrary probability density function which is known up to a normalizing constant [72]. Note that to ensure convergence, one needs to make sure that the posterior distribution is proper (integrable), which is easily checked by computing MAP profiles and making sure they do not flatten out [10, 62].

One MCMC algorithm, known as Metropolis Hastings, performs a random walk through parameter space, where each proposal step $\vec{\theta}^*$ is based on a local proposal distribution based on the current state $\vec{\theta}$. This step is either accepted or rejected based on the probability densities at the sampled points [72]. This is typically done by computing the likelihood $L(\vec{\theta}^*)$ and calculating the non-normalized density $\tilde{p}(\vec{\theta}^*|\mathbf{y^D}) = L(\vec{\theta}^*)p(\vec{\theta}^*)$, where $p(\vec{\theta}^*)$ refers to the prior probability density. This proposal is subsequently accepted with probability:

$$P_{accept} = \min\left(\frac{\tilde{p}(\vec{\theta}^*|\mathbf{y^D})q(\vec{\theta}^* \to \vec{\theta})}{\tilde{p}(\vec{\theta}|\mathbf{y^D})q(\vec{\theta} \to \vec{\theta}^*)}, 1\right) \tag{2.15}$$

The ratio of q is known as the Hastings correction and ensures detailed balance, a sufficient condition for the Markov Chain to converge to the equilibrium distribution. Typically, the first sequence of iterations are spent on moving towards the region of high probability density. This process is referred to as burn-in. Samples corresponding to the burn-in phase are usually discarded, since these are not very representative of the posterior probability distribution. How long it takes for a sampler to reach converge and obtain a representative sample of the posterior distribution strongly depends on the problem under investigation and the sampler that is used. See Cowles et al for an in-depth review on diagnostics for detecting non-convergence [73].

A critical component of MCMC samplers is the proposal distribution. When the proposal distribution matches the local shape and size of the probability density function poorly, then this will lead to either too many rejections (proposals in regions with much lower probability densities) or slow convergence due to slow traversal of the parameter space. Both these scenarios are characterized by large autocorrelations between different samples of the chain (see Figure 2.5). This is why samplers have been developed which exploit the local geometry of the posterior PDF [12, 38] either by using proposal distributions that adapt to the local geometry using an approximation of the FIM, based on linear sensitivities, or including higher order information as well. One example of such a method is MMALA, a method based on Langevin diffusion on a Riemannian manifold which scales well to large systems with widely disparate parameter scales [38]. Note that singularity of the FIM approximation must be avoided. One option is to use blocking strategies, which alternately run the MCMC for different subgroups of parameters [39], thus improving numerical stability by using smaller FIMs.

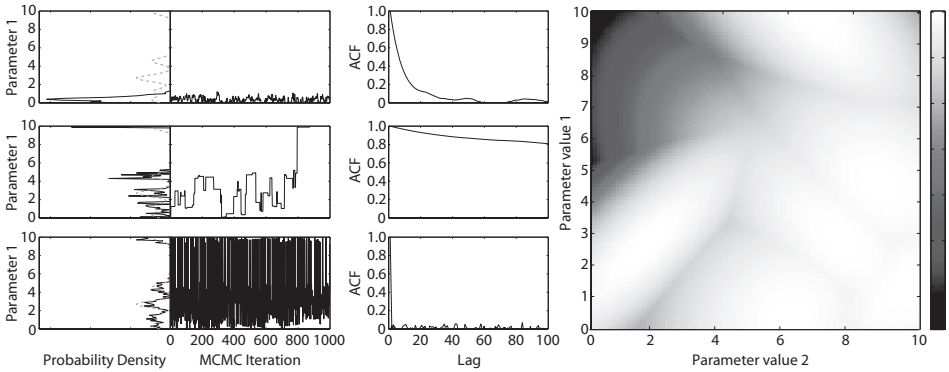

Figure 2.5: On the left: Graphical representation of various Markov Chain Monte Carlo runs. From top to bottom, a chain with a proposal kernel that is too small leading to poor traversal of the parameter space and inability to leave a local mode, a chain with a proposal kernel that is too large resulting in few accepted trial moves and a chain which incorporates exchange between chains using different proposal sizes, combining both large and small moves. Also note the differences in autocorrelation length as evidenced by their autocorrelation function (ACF). On the right the two dimensional objective function that was used in this problem. A mixture of 6 multivariate normal Gaussians with $\sigma = 0.1$ and correlation coefficients chosen according to $\rho = \frac{k}{7} sign(U[0,1] - 0.5)$ with k = [1...7]

Population based MCMC

In theory, MCMC methods should be able to reach all probable regions in parameter space when considering extremely large sample sizes. In practice however, the number of samples required to be able to reach this may be unreasonably high. In some cases, there can be distinct peaks (or modes) in the PDF. Moving between modes requires traversing a region of low probability density, which is improbable, and, does not occur often. This results in poor mixing between different modes in the probability density function. To mitigate these problems, researchers have proposed the idea of using multiple interacting MCMC chains [71, 74–76]. One example of such a sampler involves running multiple chains at different temperatures, where the different posterior densities are proportional to:

$$p_T(\vec{\theta}|\mathbf{y^D}) \propto p(\mathbf{y^D}|\vec{\theta})^{\frac{1}{T}} p(\vec{\theta}) \tag{2.16}$$

Since the likelihood function will flatten out for higher temperatures, chains at higher temperatures are able to traverse more freely. The highest temperature results in sampling directly from the prior. As these chains are updated, samples from the higher temperatures are exchanged with samples at lower temperatures using switch moves. These are performed by randomly selecting two temperatures and computing a Metropolis-Hastings step using the acceptance probability given by:

$$P_{switch} = \min\left(1, \frac{p(\mathbf{y^D}|\vec{\theta}_b)^{\frac{1}{T_a}} p(\mathbf{y^D}|\vec{\theta}_a)^{\frac{1}{T_b}}}{p(\mathbf{y^D}|\vec{\theta}_a)^{\frac{1}{T_a}} p(\mathbf{y^D}|\vec{\theta}_b)^{\frac{1}{T_b}}}\right) \tag{2.17}$$

Updates for each temperature are performed as usual. After convergence, the distribution at $T = 1$ is the desired posterior. One issue with this approach is that it requires all priors to be proper as the likelihood will completely flatten out for high temperatures. If one does not wish to assume such an informative prior, or the acceptance rate of switch moves between temperatures is unacceptably low, Parallel Hierarchical Sampling can be considered [77] as an alternative. Here multiple chains using different proposal kernels run at the same temperature (hence every switch move is accepted) and one chain is marked as 'mother' chain, which is involved in all exchanges. Since by design such samplers accept all exchange moves, autocorrelation is highly reduced (see Figure 2.5). For an application of such a sampler to a high dimensional system see [78]. Note that for both techniques, proposal kernels that exploit local geometry can still be used.

Sequential Monte Carlo

Another option to obtain samples from the posterior distribution is provided by Sequential Monte Carlo algorithms. Rather than having multiple interacting chains, these samplers propagate multiple parameter sets, referred to as particles, through a series of intermediate distributions. These intermediate distributions incrementally become harder to sample from. One such approach is Sequential Importance Resampling, which is a recursive version of Importance Sampling. At each iteration, each particle is assigned a weight according to the ratio between the distribution in the sequence currently targeted $p_T(\vec{\theta}^t|\mathbf{y^D})$, and, the current proposal distribution v^t:

$$w_i^t = \frac{p_T(\vec{\theta}_i^t|\mathbf{y^D})}{v_t(\vec{\theta}_i^t)} \tag{2.18}$$

Here the superscript t indicates the iteration number (and thereby also the intermediate distribution number). Subscript i indicates the particle number. Subsequently, these weights are normalized to one and the particles are then resampled proportionally to these weights. These samples then form the basis for the next intermediate distribution. One issue with this approach is that the sampling quickly begins to show degeneracy. This is when most weights approach zero, resulting in only a few samples effectively contributing to the estimate. This degeneracy is typically avoided by perturbing the particles after resampling using a proposal kernel.

This process of weighting, resampling and perturbing is repeated until the posterior distribution is reached. This process is schematically illustrated in Figure 2.6. Two merits of these methods are the fact that these particles can easily be evaluated in parallel and that multi-modal probability density functions are less problematic. The performance of the algorithms strongly depends on the choice of intermediate distributions, number of particles and perturbation kernels. Currently, this requires some practical experience with the methods. Which perturbation kernels to use is still an open problem, but it could be beneficial to consider the local geometry of the problem as outlined before in the section on MCMC. Additionally, these methods require a prior distribution on all parameters, which should be wide enough to ensure a good coverage of the parameter

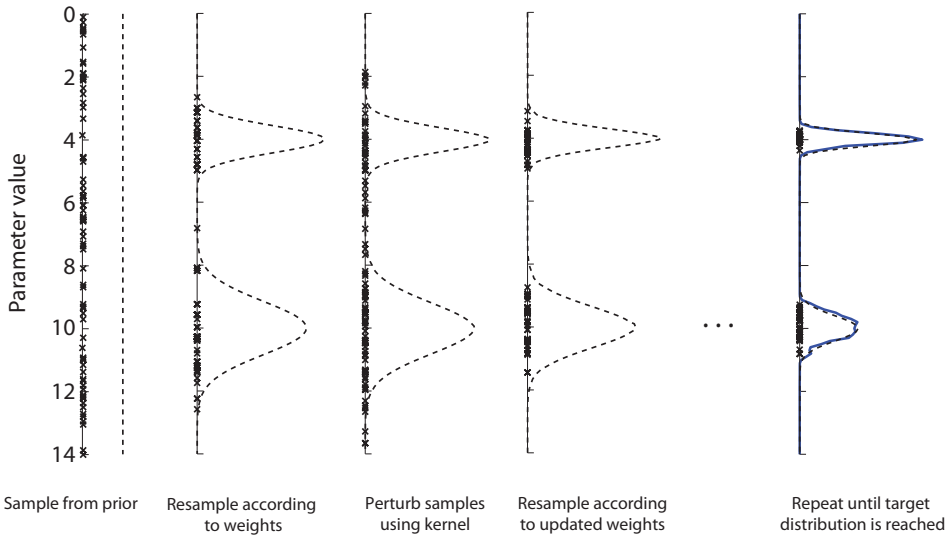

Figure 2.6: Graphic illustration of Sequential Importance Sampling. From left to right: Sampling from the prior distribution. Resampling based using the current target distribution (dashed lines) as weights. Perturbation of the samples (crosses) by a proposal kernel. Resampling based on the ratio between the target distribution and previous distribution. Until convergence to the desired target distribution. Histogram of the final distribution shown in gray.

space. One particular application of such a sampler in the setting of Approximate Bayesian Computation can be found in [79]. In Approximate Bayesian Computation, one defines a distance function between real and simulated data. Rather than having a likelihood function which flattens out at different temperatures as before, the distance threshold at which a parameter set is deemed acceptable is sequentially decreased. For sufficiently small distances, this posterior distribution approximates the true posterior distribution. One major advantage of such methods is that they are still applicable, even when the likelihood becomes prohibitively expensive to compute.

2.4.7 Simulation free MCMC

One issue to consider is that the probability density functions of larger models used in systems biology can be expensive to sample from. Profiling the execution times of any of these methods often reveals that the computational effort is usually dominated by the time required to solve the model and/or sensitivity equations. Consequently, simulation free methods are another promising avenue of research. Rather than solving the system of ordinary differential equations, derivatives of the state variables are estimated, enabling the modeler to sidestep the problem of model simulation and working with the right hand side of the differential equations directly. Initial approaches in this direction involved using spline based interpolation of the data [80] followed by optimization. However, these early approaches involved a requirement for all state variables to be mea-

sured and required choosing smoothing parameters. These methods were further generalized by [81], where both the spline fitting and data fidelity were combined in a single estimation step.

Approaches for absolving the need of an ODE integrator have also been suggested in a Bayesian setting [82]. Here, a GP was used to specify a distribution over the state variables and its derivatives. Subsequently, a correlation between the different state derivatives and the right hand side of the model equations is imposed. Values for the state variables are obtained from the GP. MCMC can then be performed on the joint posterior of the GP coefficients and model parameters, or sequentially by determining MAP estimates for the GP in an initial step. Note however, the priors for the GP also control the smoothness (and therefore dynamic behavior) of the state variable trajectories.

2.5 Predictions

Computational models are developed in order to make predictions. Often there is a need to predict system responses to a specific stimulus or to predict some internal (hidden) state variable that cannot be measured directly. In these cases it is of particular interest how the uncertainty in the data is propagated to the predictions. Depending on the formalism of choice, there are several ways to estimate prediction uncertainties.

2.5.1 Asymptotics and linearizations

The most straightforward approach is to assume the parameter uncertainty distribution to be Gaussian and to propagate the parameter uncertainty to the predictions by using a linearized version of the model. Such an approach can be realized by computing the first order sensitivities of the model [83]. These can subsequently be used to approximate the covariance matrix of the parameter estimates (2.8) and project the associated uncertainties onto the first order sensitivities of the predictions of interest:

$$Var(y) = \left[(\nabla y)^T \left(\nabla^T \nabla \ln(L(\vec{\theta})) \right)^{-1} (\nabla y) \right]_{\vec{\theta}=\vec{\theta}_{opt}} \tag{2.19}$$

where

$$\nabla = \left(\frac{d}{d\theta_1}, \frac{d}{d\theta_2}, \ldots, \frac{d}{d\theta_n} \right)^T \tag{2.20}$$

Here ∇y refers to the sensitivity of y with respect to each of the parameters. This method is applicable when there is little uncertainty in the parameters and the local model behavior can reasonably be approximated by a linear approximation. A Bayesian version of this linearized approach which takes into account thermodynamic constraints and assumes multiplicative errors and log-normal parameter distributions is proposed in [84]. When multiple modes exist, this approach can and should be repeated for each of the modes. Issues arise when parameters are non-identifiable or highly uncertain (see Figure 2.7). In these cases

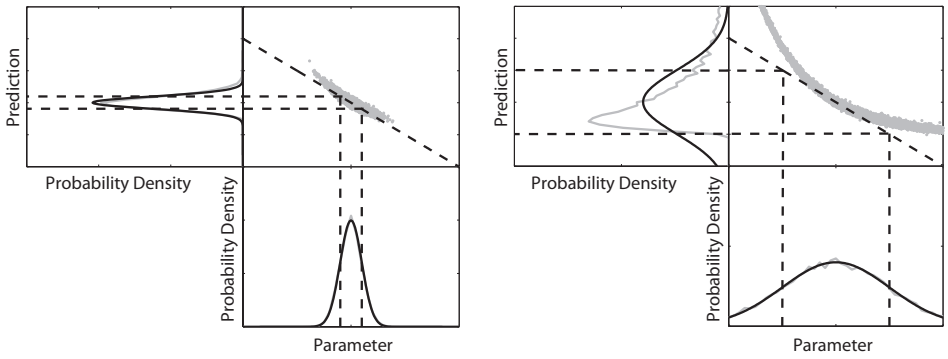

Figure 2.7: Estimating prediction variances using linearization. On the left the case with small uncertainty, where the linear parameter sensitivities provide an adequate description for projecting the parameter uncertainty onto the predictions. On the right, the case with large parameter uncertainty where the non-linearity of the model results in a poor estimate of the predictive distribution when it is estimated via linear projection.

such an approach would no longer be appropriate and one needs to resort to more advanced methods such as the ones mentioned below.

2.5.2 Prediction profiles

An approach similar to the Profile Likelihood method can also been applied in prediction space [13]. Prediction Profiles are obtained by augmenting the experimental data with an additional point which has to be satisfied by the simulation. Initially, this point is based on the simulation belonging to the optimal parameter values. Similar to the Profile Likelihood, this new constraint is then shifted followed by subsequent parameter optimization. This process is continued until the model simulations corresponding to the current parameter set no longer describe the measurement data (or some other suitable error bound is reached). Mathematically this can be expressed as follows:

$$\chi^2_{PPL}(z) = \min_{\vec{\theta},\vec{\theta}\in\{\vec{\theta}|Q(\vec{\theta})=z\}} \left[\chi^2(\vec{\theta})\right] \tag{2.21}$$

Where Q refers to a function which takes a parameter and produces the simulation point of the prediction that is being profiled. By repeating this procedure for different predictions, reliable confidence intervals can be obtained, considering only one-dimensional scans within prediction space. Predictions that are well constrained are useful for performing validation experiments while poorly constrained predictions are useful for performing experiments which optimally reduce the uncertainty in some sense.

2.5.3 Bayesian predictive simulation

In the Bayesian setting, samples from the posterior distribution can directly be used for prediction uncertainty analysis. The uncertainty in the predictions $\mathbf{y}$ can

be obtained by integrating the predictions over the posterior distribution of parameter values. Marginalizing the predictions over the converged MCMC chain provides a predictive distribution (note that this includes simulation from the noise model):

$$p(\mathbf{y_{obs}}|\mathbf{y^D}) = \int_\Omega p(\mathbf{y_{obs}}|\vec{\theta})p(\vec{\theta}|\mathbf{y^D})d\vec{\theta} \tag{2.22}$$

In practical situations, this amounts to simulating the model using parameter sets from the posterior distribution. For visualization, the state space can subsequently be divided into a discrete number of levels after which histograms of the posterior predictive distribution can be computed for each prediction. Alternatively, credible intervals can be computed by selecting a desired probability and determining bounds that enclose this fraction of the posterior density. Usually, these bounds are chosen in such a way that the posterior density between the bounds is maximal (Highest Posterior Density). Since the posterior predictive distribution forms a link between the various predictions of interest, while being constrained by the available data and prior knowledge, the relations within this distribution can be highly informative and useful for both system analysis and optimal experiment design [85].

2.6 Concluding remarks

Uncertainty analysis is an important tool to assess the reliability of various model predictions. Moreover, its results can also be used for model analysis and improvement. Predictions that are well constrained (core predictions [15, 86]) are useful for testing whether our current assumptions and model provide a sufficient description of reality, while highly uncertain predictions point to opportunities for gathering additional data which reduces uncertainty in some optimal sense.

Relations between these different parameter uncertainties and their predictive counterparts can be determined. This helps uncovering how different predictions in the model relate to each other dynamically and therefore gives insights in how the system behaves [87, 88]. Additionally, correlations and covariances between different state variables of interest can be exploited in order to determine which state variables would be most beneficial to measure. Techniques for this exist in both a frequentist [13, 37, 89] as well as a Bayesian setting [85, 88]. This makes it possible to tailor experiments specifically to predictions that are of particular interest, thereby saving time and resources.

Note that many biological systems exhibit log-normally distributed errors arising from multiplicative errors [46, 90]. In order to convert this to additive error, data and model simulations can be log-transformed and inferences performed in logarithmic space [46]. This chapter provided an overview of state-of-the-art methods for uncertainty analysis applicable to dynamical systems comprised of ODEs. The suitability of each of the aforementioned methods strongly depends on the problem and system under investigation, as well as the assumptions the investigator is willing to make. Some of the issues that often arise in systems biology models have been mentioned. As shown in Table 2.1, some of

these techniques have already been applied to systems biology models. Nevertheless, it seems that there is still a gap between groups working on methodology and groups involved in the development of new models. This gap is particularly apparent in the case of the sampling based methods which are more involved computationally. Hopefully in the future, and with the advent of new tutorials and tools [8–11] this gap will be bridged, by providing the means to effectively reduce uncertainty and to pave the way for conclusions despite uncertainties.

References

[1] Westerhoff H, Palsson B: **The evolution of molecular biology into systems biology**. *Nature biotechnology* 2004, **22**(10):1249–1252.

[2] Bruggeman F, Westerhoff H: **The nature of systems biology**. *TRENDS in Microbiology* 2007, **15**:45–50.

[3] Chou I, Voit E: **Recent developments in parameter estimation and structure identification of biochemical and genomic systems**. *Mathematical biosciences* 2009, **219**(2):57.

[4] Voit E, Neves A, Santos H: **The intricate side of systems biology**. *Proceedings of the National Academy of Sciences* 2006, **103**(25):9452–9457.

[5] van Riel NAW: **Dynamic modelling and analysis of biochemical networks: mechanism-based models and model-based experiments**. *Briefings in Bioinformatics* 2006, **7**(4):364–374.

[6] Swameye I, Müller T, Timmer J, Sandra O, Klingmüller U: **Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling**. *Proceedings of the National Academy of Sciences* 2003, **100**(3):1028.

[7] Becker V, Schilling M, Bachmann J, Baumann U, Raue A, Maiwald T, Timmer J, Klingmuller U: **Covering a broad dynamic range: information processing at the erythropoietin receptor**. *Science Signalling* 2010, **328**(5984):1404.

[8] Liepe J, Barnes C, Cule E, Erguler K, Kirk P, Toni T, Stumpf M: **ABC-SysBio approximate Bayesian computation in Python with GPU support**. *Bioinformatics* 2010, **26**(14):1797.

[9] Vyshemirsky V, Girolami M: **BioBayes: a software package for Bayesian inference in systems biology**. *Bioinformatics* 2008, **24**(17):1933–1934.

[10] Vanlier J, Tiemann C, Hilbers P, van Riel N: **An integrated strategy for prediction uncertainty analysis**. *Bioinformatics* 2012, **28**(8):1130–1135.

[11] Maiwald T, Timmer J: **Dynamical modeling and multi-experiment fitting with PottersWheel**. *Bioinformatics* 2008, **24**(18):2037–2043.

[12] Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP: **Universally Sloppy Parameter Sensitivities in Systems Biology Models**. *PLoS Comput Biol* 2007, **3**(10):e189.

[13] Kreutz C, Raue A, Timmer J: **Likelihood based observability analysis and confidence intervals for predictions of dynamic models**. *BMC Systems Biology* 2012, **6**:120.

[14] Hasenauer J, Waldherr S, Wagner K, Allgower F: **Parameter identification, experimental design and model falsification for biological network models using semidefinite programming**. *Systems Biology, IET* 2010, **4**(2):119–130.

[15] Cedersund G, Roll J: **Systems biology: model based evaluation and comparison of potential explanations for given biological data**. *FEBS Journal* 2009, **276**(4):903–922.

[16] Tiemann C, Vanlier J, Hilbers P, van Riel N: **Parameter adaptations during phenotype transitions in progressive diseases**. *BMC Systems Biology* 2011, **5**:174.

[17] Lillacci G, Khammash M: **Parameter estimation and model selection in computational biology**. *PLoS computational biology* 2010, **6**(3):e1000696.

[18] Jiang N, Cox R, Hancock J: **A kinetic core model of the glucose-stimulated insulin secretion network of pancreatic β cells**. *Mammalian Genome* 2007, **18**(6):508–520.

[19] Jeneson J, Westerhoff H, Kushmerick M: **A metabolic control analysis of kinetic controls in ATP free energy metabolism in contracting skeletal muscle**. *American Journal of Physiology-Cell Physiology* 2000, **279**(3):C813–C832.

[20] Wu F, Jeneson J, Beard D: **Oxidative ATP synthesis in skeletal muscle is controlled by substrate feedback**. *American Journal of Physiology-Cell Physiology* 2007, **292**:C115–C124.

[21] Groenendaal W, Schmidt K, von Basum G, van Riel N, Hilbers P: **Modeling glucose and water dynamics in human skin**. *Diabetes technology & therapeutics* 2008, **10**(4):283–293.

[22] Bachmann J, Raue A, Schilling M, Becker V, Timmer J, Klingmüller U: **Predictive mathematical models of cancer signalling pathways**. *Journal of internal medicine* 2012, **271**(2):155–165.

[23] Müller T, Faller D, Timmer J, Swameye I, Sandra O, Klingmüller U: **Tests for cycling in a signalling pathway**. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 2004, **53**(4):557–568.

[24] Brännmark C, Palmér R, Glad S, Cedersund G, Strålfors P: **Mass and information feedbacks through receptor endocytosis govern insulin signaling as revealed using a parameter-free modeling framework**. *Journal of Biological Chemistry* 2010, **285**(26):20171.

[25] Klinke D: **An empirical Bayesian approach for model-based inference of cellular signaling networks**. *BMC bioinformatics* 2009, **10**:371.

[26] Finley S, Gupta D, Cheng N, Klinke D: **Inferring relevant control mechanisms for interleukin-12 signaling in naïve CD4+ T cells**. *Immunology and cell biology* 2010, **89**:100–110.

[27] Konukoglu E, Relan J, Cilingir U, Menze B, Chinchapatnam P, Jadidi A, Cochet H, Hocini M, Delingette H, Jaïs P, et al.: **Efficient probabilistic model personalization integrating uncertainty on data and parameters: Application to eikonal-diffusion models in cardiac electrophysiology**. *Progress in Biophysics and Molecular Biology* 2011, **107**:134–146.

[28] Xu T, Vyshemirsky V, Gormand A, von Kriegsheim A, Girolami M, Baillie G, Ketley D, Dunlop A, Milligan G, Houslay M, et al.: **Inferring signaling pathway topologies from multiple perturbation measurements of specific biochemical species**. *Science Signalling* 2010, **3**(113).

[29] Kalita MK, Sargsyan K, Tian B, Paulucci-Holthauzen A, Najm HN, Debusschere BJ, Brasier AR: **Sources of cell-to-cell variability in canonical nuclear factor-κB (NF-κB) signaling pathway inferred from single cell dynamic images**. *Journal of Biological Chemistry* 2011, **286**(43):37741–37757.

[30] Holmes G, Anderson S, Dixon G, Robertson A, Reyes-Aldasoro C, Billings S, Renshaw S, Kadirkamanathan V: **Repelled from the wound, or randomly dispersed? Reverse migration behaviour of neutrophils characterized by dynamic modelling**. *Journal of The Royal Society Interface* 2012.

[31] Taylor H, Barnes C, Huvet M, Bugeon L, Thorne T, Lamb J, Dallman M, Stumpf M, et al.: **Calibrating spatio-temporal models of leukocyte dynamics against in vivo live-imaging data using approximate Bayesian computation**. *Integrative Biology* 2012, **4**(3):335–345.

[32] Silk D, Kirk P, Barnes C, Toni T, Rose A, Moon S, Dallman M, Stumpf M: **Designing attractive models via automated identification of chaotic and oscillatory dynamical regimes**. *Nature communications* 2011, **2**:489.

[33] Maerkl S, Quake S: **A systems approach to measuring the binding energy landscapes of transcription factors**. *Science* 2007, **315**(5809):233.

[34] Barabási A, Oltvai Z: **Network biology: understanding the cell's functional organization**. *Nature Reviews Genetics* 2004, **5**(2):101–113.

[35] Voit E, Almeida J: **Decoupling dynamical systems for pathway identification from metabolic profiles**. *Bioinformatics* 2004, **20**(11):1670–1681.

[36] Teusink B, Passarge J, Reijenga C, Esgalhado E, van der Weijden C, Schepper M, Walsh M, Bakker B, van Dam K, Westerhoff H, et al.: **Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry**. *European Journal of Biochemistry* 2000, **267**(17):5313–5329.

[37] Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmüller U, Timmer J: **Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood**. *Bioinformatics* 2009, **25**(15):1923.

[38] Girolami M, Calderhead B: **Riemann manifold langevin and hamiltonian monte carlo methods**. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2011, **73**(2):123–214.

[39] Calderhead B, Girolami M: **Statistical analysis of nonlinear dynamical systems using differential geometric sampling methods**. *Interface Focus* 2011, **1**(6):821–835.

[40] Schmitz J, Van Riel N, Nicolay K, Hilbers P, Jeneson J: **Silencing of glycolysis in muscle: experimental observation and numerical analysis**. *Experimental physiology* 2010, **95**(2):380–397.

[41] Schilling M, Maiwald T, Hengl S, Winter D, Kreutz C, Kolch W, Lehmann W, Timmer J, Klingmüller U: **Theoretical and experimental analysis links isoform-specific ERK signalling to cell fate decisions**. *Mol Syst Biol* 2009, **5**:334.

[42] Borisov N, Aksamitiene E, Kiyatkin A, Legewie S, Berkhout J, Maiwald T, Kaimachnikov N, Timmer J, Hoek J, Kholodenko B: **Systems-level interactions between insulin–EGF networks amplify mitogenic signaling**. *Mol Syst Biol* 2009, **5**:256.

[43] Cedersund G, Roll J, Ulfhielm E, Danielsson A, Tidefelt H, Strålfors P: **Model-based hypothesis testing of key mechanisms in initial phase of insulin signaling**. *PLoS Comput Biol* 2008, **4**(6):799–806.

[44] Koschorreck M, Gilles E: **Mathematical modeling and analysis of insulin clearance in vivo**. *BMC Syst Biol* 2008, **2**:43.

[45] Schoeberl B, Eichler-Jonsson C, Gilles E, Müller G: **Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors**. *Nat Biotechnol* 2002, **20**(4):370–375.

[46] Kreutz C, Rodriguez M, Maiwald T, Seidl M, Blum H, Mohr L, Timmer J: **An error model for protein quantification**. *Bioinformatics* 2007, **23**(20):2747.

[47] Vyshemirsky V, Girolami M: **Bayesian ranking of biochemical system models**. *Bioinformatics* 2008, **24**(6):833–839.

[48] Gelman A, Bois F, Jiang J: **Physiological pharmacokinetic analysis using population modeling and informative prior distributions**. *Journal of the American Statistical Association* 1996, **91**(436):1400–1412.

[49] Jeffreys H: **An Invariant Form for the Prior Probability in Estimation Problems**. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* 1946, **186**(1007):pp. 453–461.

[50] Moles CG, Mendes P, Banga JR: **Parameter estimation in biochemical pathways: a comparison of global optimization methods**. *Genome research* 2003, **13**(11):2467–2474.

[51] Ashyraliyev M, Fomekong-Nanfack Y, Kaandorp J, Blom J: **Systems biology: parameter estimation for biochemical models**. *FEBS Journal* 2009, **276**(4):886–902.

[52] Zhang Y, Rundell A, et al.: **Comparative study of parameter sensitivity analyses of the TCR-activated Erk-MAPK signalling pathway**. *Systems biology* 2006, **153**(4):201.

[53] Hafner M, Koeppl H, Hasler M, Wagner A: **Glocalrobustness analysis and model discrimination for circadian oscillators**. *PLoS computational biology* 2009, **5**(10).

[54] Bentele M, Lavrik I, Ulrich M, Stösser S, Heermann D, Kalthoff H, Krammer P, Eils R: **Mathematical modeling reveals threshold mechanism in CD95-induced apoptosis**. *The Journal of cell biology* 2004, **166**(6):839–851.

[55] Cho K, Shin S, Kolch W, Wolkenhauer O: **Experimental Design in Systems Biology, Based on Parameter Sensitivity Analysis Using a Monte Carlo Method: A Case Study for the TNFα-Mediated NF-κ B Signal Transduction Pathway**. *Simulation* 2003, **79**(12):726–739.

[56] Zi Z, Cho K, Sung M, Xia X, Zheng J, Sun Z: **In silico identification of the key components and steps in IFN-γ induced JAK-STAT signaling pathway**. *FEBS letters* 2005, **579**(5):1101–1108.

[57] Philippou A, Roussas G: **Asymptotic normality of the maximum likelihood estimate in the independent not identically distributed case**. *Annals of the Institute of Statistical Mathematics* 1975, **27**:45–55.

[58] Hindmarsh A, Brown P, Grant K, Lee S, Serban R, Shumaker D, Woodward C: **SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers**. *ACM T Math Software* 2005, **31**(3):363–396.

[59] Schaber J, Klipp E: **Model-based inference of biochemical parameters and dynamic properties of microbial signal transduction networks**. *Current opinion in biotechnology* 2011, **22**:109–116.

[60] Anguelova M, Cedersund G, Johansson M, Franzén C, Wennberg B: **Conservation laws and unidentifiability of rate expressions in biochemical models**. *IET Systems Biology* 2007, **1**(4):230–237.

[61] Chis OT, Banga JR, Balsa-Canto E: **Structural Identifiability of Systems Biology Models: A Critical Comparison of Methods**. *PLoS ONE* 2011, **6**(11):e27755.

[62] Raue A, Kreutz C, Theis F, Timmer J: **Joining forces of Bayesian and frequentist methodology: A study for inference in the presence of non-identifiability**. *Phil. Trans. Roy. Soc. A* 2012.

[63] Williams C, Rasmussen C: **Gaussian processes for machine learning** 2006.

[64] Kirk PDW, Stumpf MPH: **Gaussian process regression bootstrapping: exploring the effects of uncertainty in time course data**. *Bioinformatics* 2009, **25**(10):1300–1306.

[65] Joshi M, Seidel-Morgenstern A, Kremling A: **Exploiting the bootstrap method for quantifying parameter confidence intervals in dynamical systems**. *Metabolic Engineering* 2006, **8**(5):447–455.

[66] DiCiccio T, Efron B: **Bootstrap confidence intervals**. *Statistical Science* 1996, :189–212.

[67] Efron B: **Nonparametric standard errors and confidence intervals**. *Canadian Journal of Statistics* 1981, **9**(2):139–158.

[68] Efron B, Efron B: *The jackknife, the bootstrap, and other resampling plans*, *Volume 38*. SIAM 1982.

[69] DiCiccio T, Tibshirani R: **Bootstrap confidence intervals and bootstrap approximations**. *Journal of the American Statistical Association* 1987, **82**(397):163–170.

[70] Diciccio T, Romano J: **A review of bootstrap confidence intervals**. *Journal of the Royal Statistical Society. Series B (Methodological)* 1988, :338–354.

[71] Calderhead B, Girolami M: **Estimating Bayes factors via thermodynamic integration and population MCMC**. *Computational Statistics & Data Analysis* 2009, **53**(12):4028–4045.

[72] Geyer C: **Practical markov chain monte carlo**. *Statistical Science* 1992, :473–483.

[73] Cowles M, Carlin B: **Markov Chain Monte Carlo convergence diagnostics: a comparative review.** *Journal of the American Statistical Association* 1996, **91**(434).

[74] Jasra A, Stephens D, Holmes C: **On population-based simulation for static inference**. *Statistics and Computing* 2007, **17**(3):263–279.

[75] Neal R: **Sampling from multimodal distributions using tempered transitions**. *Statistics and Computing* 1996, **6**(4):353–366.

[76] Altekar G, Dwarkadas S, Huelsenbeck J, Ronquist F: **Parallel metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference**. *Bioinformatics* 2004, **20**(3):407–415.

[77] Rigat F, Mira A: **Parallel hierarchical sampling: A general-purpose interacting Markov chains Monte Carlo algorithm**. *Computational Statistics & Data Analysis* 2011.

[78] Hug S, Raue A, Hasenauer J, Bachmann J, Klingmüller U, Timmer J, Theis F: **High-Dimensional Bayesian Parameter Estimation: Case Study for a Model of JAK2/STAT5 Signaling**. *Mathematical biosciences* 2013.

[79] Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf M: **Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems**. *Journal of the Royal Society Interface* 2009, **6**(31):187–202.

[80] Varah J: **A spline least squares method for numerical parameter estimation in differential equations**. *SIAM Journal on Scientific and Statistical Computing* 1982, **3**:28.

[81] Ramsay J, Hooker G, Campbell D, Cao J: **Parameter estimation for differential equations: a generalized smoothing approach**. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2007, **69**(5):741–796.

[82] Calderhead B, Girolami M, Lawrence N: **Accelerating Bayesian inference over nonlinear differential equations with Gaussian processes**. *Advances in neural information processing systems* 2009, **21**:217–224.

[83] Casey F, Baird D, Feng Q, Gutenkunst R, Waterfall J, Myers C, Brown K, Cerione R, Sethna J: **Optimal experimental design in an epidermal growth factor receptor signalling and down-regulation model**. *Systems Biology, IET* 2007, **1**(3):190–202.

[84] Liebermeister W, Klipp E, et al.: **Biochemical networks with uncertain parameters**. *Syst Biol (Stevenage)* 2005, **152**(3):97–107.

[85] Vanlier J, Tiemann C, Hilbers P, van Riel N: **A Bayesian approach to targeted experiment design**. *Bioinformatics* 2012, **28**(8):1136–1142.

[86] Nyman E, Brannmark C, Palmer R, Brugard J, Nystrom F, Stralfors P, Cedersund G: **A hierarchical whole body modeling approach elucidates the link between in vitro insulin signaling and in vivo glucose homeostasis**. *Journal of Biological Chemistry* 2011.

[87] Gomez-Cabrero D, Compte A, Tegner J: **Workflow for generating competing hypothesis from models with parameter uncertainty**. *Interface Focus* 2011, **1**(3):438.

[88] Weber P, Kramer A, Dingler C, Radde N: **Trajectory-oriented Bayesian experiment design versus Fisher A-optimal design: an in depth comparison study**. *Bioinformatics* 2012, **28**(18):i535–i541.

[89] Transtrum M, Qiu P: **Optimal experiment selection for parameter estimation in biological differential equation models**. *BMC bioinformatics* 2012, **13**:181.

[90] Limpert E, Stahel W, Abbt M: **Log-normal distributions across the sciences: keys and clues**. *BioScience* 2001, **51**(5):341–352.

A Strategy for Prediction Uncertainty Analysis

3

Parts of this chapter are described in:

Vanlier, J. and Tiemann, C.A. and Hilbers, P.A.J. and van Riel, N.A.W. (2012), *An integrated strategy for prediction uncertainty analysis*, *Bioinformatics*, 28(8), 1130-1135.

Abstract

To further our understanding of the mechanisms underlying biochemical pathways mathematical modelling is used. Since many parameter values are unknown, they need to be estimated using experimental observations. The complexity of models necessary to describe biological pathways in combination with the limited amount of quantitative data results in large parameter uncertainty, which propagates into model predictions. Therefore, prediction uncertainty analysis is an important topic that needs to be addressed in Systems Biology modelling. A strategy for prediction uncertainty analysis is proposed. This approach integrates Profile Likelihood analysis with Bayesian estimation. Our method is illustrated with an application to a model of the JAK-STAT signalling pathway.

3.1 Introduction

One of the driving ambitions of mathematical modeling is to formalize hypotheses about a biochemical network in such a manner that these can be tested. Computational models can also be used to learn more about the system under investigation, as well as predict unmeasured behavior or responses. Once the model describes the system to an acceptable degree, it is time to make predictions. How well these predictions can actually be made depends strongly on the constraints the data manage to impose on the dynamics as well as the required complexity of the model. In the previous chapter, some of the different techniques available in literature were discussed. This chapter introduces a strategy particularly suitable for uncertainty analysis of dynamical models. This strategy involves performing different analyses sequentially to detect and avoid problems associated with the individual techniques. The chapter is composed of two parts: First the strategy will be outlined after which it is applied to a test case.

We consider models of biochemical networks based on Ordinary Differential Equations (ODEs). Measurements are performed on a subset or a combination of the total number of state variables in the model $\vec{y}_{obs}(t)$, which results in a collection of data points $\mathbf{y}$ consisting of elements y_{ij}. Considering M time series of length N_i with additive independent Gaussian noise, the probability density function of the data is given by:

$$\begin{aligned} p(\mathbf{y}|\vec{\theta}) &= \prod_{i=1}^{M}\prod_{j=1}^{N_i} p(y_{ij},\vec{\theta}) \\ &= K\exp\left(-\sum_{i=1}^{M}\sum_{j=1}^{N_i}\left(\frac{y_{ij}-y_i(t_j,\vec{\theta})}{\sqrt{2}\sigma_{ij}}\right)^2\right) \end{aligned} \tag{3.1}$$

The first step in the approach is to employ Maximum Likelihood Estimation (MLE) to find model parameters for which the probability density function most likely produced the data. Here the likelihood function $L(\vec{\theta})$, whose right hand side is identical to $p(\mathbf{y}|\vec{\theta})$, is maximized over the parameters $\vec{\theta}$. Subsequently, likelihood based methods are used to determine how constrained the model parameters are when considering only the observed data. After ensuring that the posterior distribution is integrable, Bayesian inference is used to sample from the posterior probability distribution of parameters. This sample is then used to produce predictive distributions with their associated credible intervals. By performing these analyses sequentially, it is possible to increase our confidence that the results are reliable and reproducible.

3.2 Methods

The proposed approach is based on four steps which shall be discussed below.

Step 1. Obtaining parameters (MLE and MAP)

Maximum likelihood estimation corresponds to finding the maximum of $L(\vec{\theta})$ by minimizing the negative log of the likelihood. Assuming the data variances known, the normalizing constant in the likelihood function is independent of the parameters. Furthermore, since constant values have no effect on the location of the minimum, the quantity to be minimized during parameter estimation becomes:

$$\chi^2(\vec{\theta}) = \sum_{i=1}^{M} \sum_{j=1}^{N_i} \left(\frac{y_{ij}^D - y_i(t_j, \vec{\theta})}{\sigma_{ij}} \right)^2 \tag{3.2}$$

This quantity can be recognized as a weighted sum of squared differences between model and data. A regularized version of this method is the *Maximum A Posteriori* (MAP) estimator, which minimizes the negative log-likelihood multiplied by the log-prior. Finding the optimum can be challenging due to the existence of multiple local minima and wide range of model sensitivities [1]. The first step in the approach employs Monte Carlo Multiple Minimization (MCMM), which entails performing parameter estimation for a large number of widely dispersed initial values. Such an initial run enables the modeler to probe the parameter space for the existence of multiple minima at reasonable computational cost.

Step 2. Parameter bounds and identifiability

When model predictions sufficiently describe the experimental data, confidence intervals are obtained using the Profile Likelihood method [2]. Each profile likelihood is initiated at the best fit parameters after which one parameter (denoted with i) is changed incrementally while optimizing all the other parameters. The Weighted Residual Sum of Squares (WRSS) along the path can be written as:

$$\chi^2_{PL,i}(x) = \min_{\theta_{j \neq i}} \left[\chi^2(\vec{\theta} | \theta_i = x) \right] \tag{3.3}$$

We obtain confidence bounds by determining the value of x for which the WRSS reaches a threshold. This threshold is based on the likelihood ratio test given by:

$$-2 \ln \left(\frac{L(\vec{\theta}_{PL})}{L(\vec{\theta}_{opt})} \right) = \chi^2(\vec{\theta}_{PL}) - \chi^2(\vec{\theta}_{opt}) \leq \chi^2_{1-\alpha,1} \tag{3.4}$$

The fact that each parameter is treated independently (in the sense that only a one dimensional traversal is required for each parameter) makes this method

efficient to compute. In some cases, model parameters can be *structurally non-identifiable*. Structural identifiability is a property of the model, observations and stimuli and does not depend on the actual precision of the data. Structurally non-identifiable parameters manifest themselves through a constant $\chi^2_{PL,i}$ for the involved parameters. Consequently, no parameter confidence bound can be computed for such a parameter. In other cases, parameter bounds cannot be determined due to insufficiently precise data. Such parameters are typically associated with profiles which flatten out in certain directions and are named *practically non-identifiable* [2].

After performing a profile likelihood analysis, it is important to verify that the confidence intervals cover all acceptable solutions obtained in the MCMM from step 1. If this is not the case, then another local optimum exists and the profile likelihood method should be repeated starting from this optimum. After profiles for the different modes have been estimated, these profiles are merged to obtain an overall profile (see Appendix 3.7.1 for further details).

Step 3. Assessing prediction uncertainty

The aim in the third step of the analysis is to use the measurement data to infer a posterior distribution over the parameters rather than a single parameter set with confidence intervals. Applying Bayes' rule, we obtain the following expression for the posterior parameter probability:

$$p(\vec{\theta}|\mathbf{y^D}) \propto p(\mathbf{y^D}|\vec{\theta})p(\vec{\theta}) \tag{3.5}$$

where $p(\mathbf{y^D}|\vec{\theta})$ represents the conditional probability density of data given a parameter set. Here, the data has already been observed, therefore this is replaced by the likelihood function. Finally, $p(\vec{\theta})$ denotes the prior probability density of the parameters. This prior is specified before the data has been observed and usually represents either the current state of knowledge or attempts to be non-informative (or objective). Note that most priors depend on the chosen parameterization, which demonstrates that uniform priors do not necessarily (and typically do not) reflect complete objectivity [3, 4] in the Bayesian setting.

To sample from the posterior, Markov Chain Monte Carlo (MCMC) is employed. MCMC can generate samples from probability densities which are known up to a normalizing factor [5]. The Metropolis-Hastings algorithm used in this strategy performs a correlated random walk through parameter space, where each step is based on a local proposal distribution and an acceptance criterion. This acceptance criterion is based on the posterior probability densities at the sampled points. Rather than sampling purely at random (where most of the samples would be from regions of low likelihood), such a chain draws samples proportionally to the posterior probability density. The histogram of such a parameter walk with respect to a specific parameter corresponds to the marginalized (integrated over all other variables) posterior parameter distribution for that specific parameter.

The algorithm proceeds by iteratively performing the following steps:

1. Generate a sample $\vec{\theta}_{n+1}$ by sampling from a proposal distribution based on the current state $\vec{\theta}_n$

2. Compute the likelihood of the proposed parameter set $L(\vec{\theta}_{n+1})$ and calculate $\tilde{p}(\vec{\theta}_{n+1}|\mathbf{y^D}) = L(\vec{\theta}_{n+1})p(\vec{\theta}_{n+1})$, where $p(\vec{\theta}_{n+1})$ refers to the prior probability density.

3. Draw a random number γ from a uniform distribution between 0 and 1 and accept the new step if $\gamma < \min\left(\frac{\tilde{p}(\vec{\theta}_{n+1}|\mathbf{y^D})q(\vec{\theta}_{n+1}\rightarrow\vec{\theta}_n)}{\tilde{p}(\vec{\theta}_n|\mathbf{y^D})q(\vec{\theta}_n\rightarrow\vec{\theta}_{n+1})}, 1\right)$.

Here $\tilde{p}$ denotes a density which is not normalized. The ratio of q is known as the Hastings correction and ensures detailed balance, a sufficient condition for the Markov Chain to converge to the equilibrium distribution. It corrects for the fact that the proposal density going from parameter set $\vec{\theta}_n$ to $\vec{\theta}_{n+1}$ and $\vec{\theta}_{n+1}$ to $\vec{\theta}_n$ is unequal when the proposal distribution depends on the current parameter set, and, is given by the inverse of this ratio. The simplicity of the algorithm makes it conceptually attractive. Note however that simple approaches can lead to MCMC samplers that converge slowly or only explore a single mode [6].

Proposals

Each iteration requires a new proposal, which is taken from a proposal distribution. To generate samples which allow the sampler to efficiently traverse parameter space, the proposal distribution should adapt to the local geometry of the probability density function. This becomes particularly important in the non-identifiable case since the parameters will be strongly correlated [7]. Ensuring that the sampler takes large steps along the parameter correlations helps accelerate convergence. To generate these proposals an adaptive multivariate Gaussian distribution is used. Its covariance matrix is based on a quadratic approximation to the negative logarithm of the probability density function at the current parameter set [1]. This matrix is computed by taking the inverse of an approximation to the Hessian matrix of the model under consideration. This proposal distribution is subsequently scaled by a problem specific proposal scaling factor that is tuned using short exploratory runs of the sampler.

Sometimes certain directions in parameter space can be so poorly constrained that this leads to a near singular Hessian (some singular values near zero). Consequently, the proposal distribution will become extremely elongated in these directions, leading to proposals with extremely large or small parameter values resulting in a decline of the acceptance ratio. To avoid such numerical difficulties, singular values below a certain threshold are set to a specified minimal cutoff (prior to inversion). Additionally, second derivatives of the non-uniform priors (when available) are included directly in the Hessian approximation.

Parameter representation and non-identifiability

The posterior distribution is required to be 'proper' (finitely integrable) otherwise the sampler will not converge and inference is not possible [8]. Without a finitely integrable prior to constrain the posterior distribution, improper likelihoods can lead to improper posteriors. This becomes an issue when parameters are non-identifiable and the data contains insufficient information to ensure a proper likelihood for those parameters. In such cases, one pragmatic approach involves incorporating empirical priors to ensure that parameters which are non-identifiable from the data, do not drift off to extreme values [8] hampering ODE integration and resulting in numerical instabilities. Although such priors make the following analyses feasible, they *artificially* reduce the parameter uncertainty. It is therefore prudent to study the effects of the assumed priors by repeating the analysis for different prior distributions.

Priors are typically not re-parameterization invariant. Although seemingly non-informative, a uniform prior in linear parameter space implies that extremely large rates have an equal *a priori* probability of occurring as slow rates. For positively defined parameters, a uniform distribution in logarithmic space corresponds to an non-informative prior [9]. Such a prior gives equal probability to different orders of magnitude (scales). An approximate scale invariance of kinetic parameters has been observed in biological models [10]. The transformation between parameters can be described by the matrix of partial derivatives with respect to the equations which transform the parameters from one parameterization to another (the Jacobian of the transformation). To calculate the prior density that is equivalent under a different parameterization and correct for the stretching and compression of the distribution, one needs to multiply the prior probability density by the absolute value of the determinant of the Jacobian of the transformation:

$$p(f(\vec{\theta})) = p(\vec{\theta}) \left| \begin{bmatrix} \frac{df(\theta_1)}{d\theta_1} & \cdots & \frac{df(\theta_1)}{d\theta_n} \\ \vdots & \ddots & \vdots \\ \frac{df(\theta_n)}{d\theta_1} & \cdots & \frac{df(\theta_n)}{d\theta_n} \end{bmatrix} \right| \tag{3.6}$$

Convergence

Burn in refers to the time it takes the chain to get to a region of high probability density and samples obtained during this period are discarded to avoid assigning too much weight to highly improbable samples. One approach to avoid a long burn-in period is by using a deterministic minimization algorithm to obtain a best fit parameter set [1] which is likely to be a reasonable sample within the posterior probability distribution. Determining whether a sufficient number of samples has been acquired is hard to assess, and in practical situations only non-convergence can be diagnosed [11]. In order to try and detect possible non-convergence, a single long chain was divided into several batches to determine whether systematic differences could be observed.

Step 4. Analysis of the posterior parameter and predictive distribution

The final stage of the strategy involves analyzing the obtained results to probe model properties. Having determined the posterior distribution of model predictions, there is now a direct link between different predictions and parameters, which can be exploited by determining how predictions relate to each other and to the model parameters. The uncertainty in predictions $\mathbf{y_{obs}}$ can be obtained by integrating the output over the posterior distribution of parameter values. Note that $\mathbf{y_{obs}}$ can be any prediction obtained using the model. In other words, marginalizing the predictions over the converged MCMC chain provides the prediction uncertainty as shown in:

$$p(\mathbf{y_{obs}}|\mathbf{y^D}) = \int_{-\infty}^{\infty} p(\mathbf{y_{obs}}|\vec{\theta}) p(\vec{\theta}|\mathbf{y^D}) d\vec{\theta} \tag{3.7}$$

The posterior predictive distribution of simulations can be visualized by computing predictive histograms per time point. Alternatively, credible intervals can be computed by selecting a desired probability and determining bounds that enclose this fraction of the posterior density. These bounds are chosen in such a way that the posterior density between the bounds is maximal. By examining correlations between different state variables of interest, it is possible to determine which states would be good measurement candidates if a prediction we happen to be interested in is not measurable. Similarly, such correlations can be explored between state variables and parameters in order to determine which measurement could be used to avoid the necessity of having to use an overly informative prior. In summary, the entire strategy consists of the following steps:

1. Detection of (multiple) acceptable parameter modes using an exploratory large scale search.
2. Detection of structural and practical non-identifiabilies using the profile likelihood method.
3. Perform a Bayesian analysis considering detected non-identifiabilities from the PL method.
4. Perform an analysis of the posterior parameter and predictive distributions.

3.3 Implementation

Algorithms were implemented in MATLAB (Natick, MA). Numerical integration of the differential equations was performed using compiled MEX files using numerical integrators from the SUNDIALS CVode package (Lawrence Livermore National Laboratory, Livermore, CA). Absolute and relative tolerances were set to 10^{-8} and 10^{-9} respectively. Integration time for a single integration was allowed to be 10 seconds at most after which an integration is assumed to fail and a large error is returned. Throughout the experiments, integration failures were carefully monitored.

For the MCMM, random sampling was performed using a log-uniform hypercube to obtain initial parameter values. These were subsequently optimized using the large scale algorithm from the MATLAB Optimization Toolbox. The best fit was subsequently selected and used for determining profile likelihoods. When profiling, the following heuristic was used to determine stepsize: Each time the parameter of interest is increased or decreased, the increase in the WRSS is evaluated. Subsequently the step size was increased when the increase in WRSS was below a certain threshold, while decreasing the step size and rejecting the step when the WRSS increased too much. These thresholds were defined relative to the difference in WRSS of the optimum, and, the threshold based on the likelihood ratio $[10^{-3}\chi_{range}, 10^{-2}\chi_{range}]$. Step sizes were forced to stay within the range of $[10^{-5}, 10^{-1}]$.

To attain an adequate acceptance rate and good mixing in the MCMC, the proposal scaling was determined during an initial tuning stage. This tuning was performed by running several short chains (100 iterations each), targeting an acceptance rate between 0.2 and 0.4 [12]. If the acceptance rate was higher, 10% was added to the scale (which scales the entire covariance matrix), while 10% was substracted in the case where the acceptance was too low. Interestingly, the chains at higher temperatures had very similar acceptance rates once started. The cutoff for the Hessian approximation was set to 10^{-6}. It was observed that this greatly affected the acceptance rate indicating (near) singularity of the Hessian approximation.

3.4 JAK-STAT model

In this section, the approach shall be demonstrated using a model of the STAT signaling pathway [2,13] depicted in Figure 3.1. The model is based on a number of hypothesized steps. First erythropoietin (EpoR) activates the EpoR receptor which phosphorylates cytoplasmic STAT (x_1). This phosphorylated STAT (x_2) dimerizes (x_3) and is subsequently imported into the nucleus (x_4). Here dissociation and dephosphorylation occurs, which is associated with a delay. The driving input function was approximated by a spline interpolant, while the delay was approximated using a linear chain approximation. The model equations are as follows:

$$
\begin{aligned}
\dot{x_1} &= 2\left(\frac{V_{nucleus}}{V_{cyto}}\right)(p_4 x_{13}) - p_1 x_1 u_1 & \dot{x_8} &= p_4 x_7 - p_4 x_8 \\
\dot{x_2} &= p_1 x_1 u_1 - 2p_2 x_2^2 & \dot{x_9} &= p_4 x_8 - p_4 x_9 \\
\dot{x_3} &= p_2 x_2^2 - p_3 x_3 & \dot{x_{10}} &= p_4 x_9 - p_4 x_{10} \\
\dot{x_4} &= \frac{V_{cyto}}{V_{nucleus}}(p_3 x_3) - p_4 x_4 & \dot{x_{11}} &= p_4 x_{10} - p_4 x_{11} \\
\dot{x_5} &= p_4 x_4 - p_4 x_5 & \dot{x_{12}} &= p_4 x_{11} - p_4 x_{12} \\
\dot{x_6} &= p_4 x_5 - p_4 x_6 & \dot{x_{13}} &= p_4 x_{12} - p_4 x_{13} \\
\dot{x_7} &= p_4 x_6 - p_4 x_7
\end{aligned}
\tag{3.8}
$$

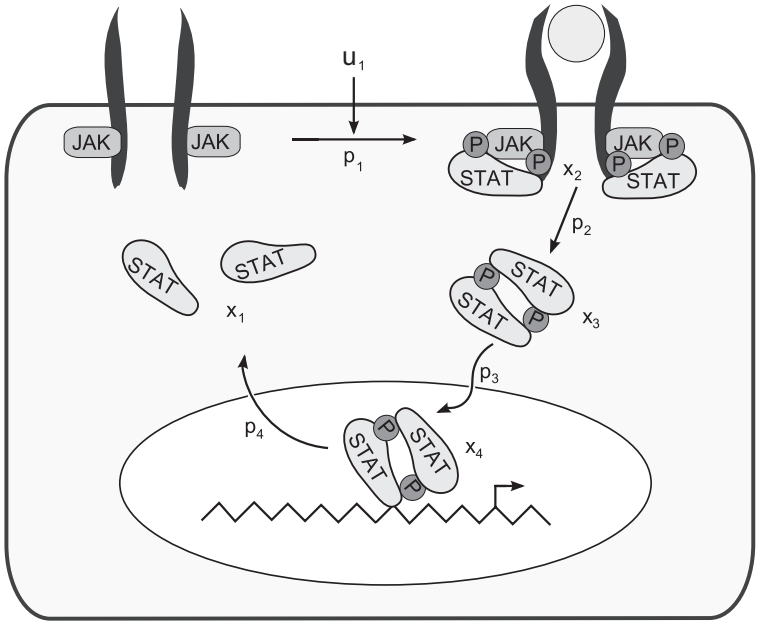

Figure 3.1: Model of the JAK-STAT pathway. In this model u_1 serves as driving input, while the total concentration of STAT ($x_1 + x_2 + 2x_3$) and the total concentration of phosphorylated STAT in the cytoplasm ($x_2 + 2x_3$) were measured. Note that the step from x_4 back to x_1 is associated with a delay.

Data from Swameye et al. were used for inference [14]. Observables were the total concentration of STAT and the total concentration of phosphorylated STAT in the cytoplasm, both reported in arbitrary units (which necessitates two scaling parameters). The initial cytoplasmic concentration of STAT is unknown while all other forms of STAT are assumed zero at the start of the simulation. The vector of unknown parameter values consists of the following elements $\vec{\theta} = \left(p_1, p_2, p_3, p_4, s_1, s_2, x_1^0\right)$.

3.5 Application of the strategy to the JAK-STAT signaling cascade

Multiple modes were detected using a large scale MCMM search with initial conditions based on a log-uniform random sampling between the ranges 10^{-3} and 10^2 ($N = 10000$). After optimization, samples were either accepted or rejected based on the likelihood ratio bound based on the best fit value. The resulting distribution and associated WRSS are shown in Figure 3.2. It is clear that there are at least three local minima in the likelihood. Although all three modes describe the data adequately, they show different prediction results for the unobserved internal state variables of the model.

Subsequently, profile likelihood analysis was performed. To increase confidence in the fact that no acceptable regions of parameter space were missed, pro-

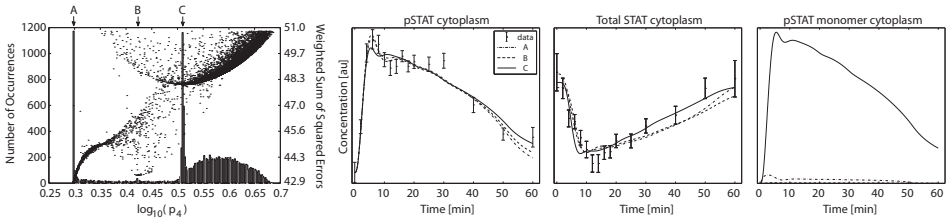

Figure 3.2: Left: Histogram of final parameter values for parameter four post optimization (bars), and the associated weighted residual sum of squares (dots). Note that all of the optimized parameter sets shown are acceptable with respect to the LR ratio. Right: Model predictions from parameter sets taken from location A, B and C for two measured outputs as well as one unmeasured internal state variable.

file likelihoods were started from each mode detected using the MCMM method (Step 1). Subsequently all profile likelihoods were merged, followed by checking whether they covered the full span of acceptable parameter sets obtained from the MCMM. Based on the profile likelihood, shown in the top panel of Figure 3.3, it can be concluded that the model based on first principles is structurally non-identifiable [2]. Plotting the parameter sets associated with the structurally non-identifiable parameters reveals clear relationships between the parameters (see Figure 3.4). The profiles reveal that the initial condition x_1^0 and scaling parameters s_1 and s_2 were structurally related and therefore non-identifiable.

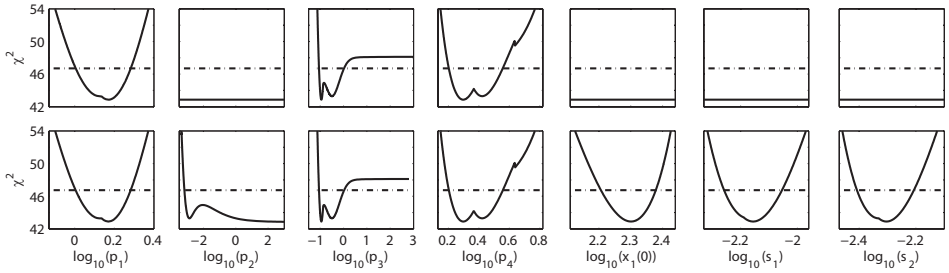

Figure 3.3: Merged profile likelihoods of the JAK-STAT model. Top: Without prior on the initial condition, Bottom: With prior on the initial condition.

Analogously to [2] a Gaussian prior ($\mu = 200nM, \sigma = 20nM$) was specified for the initial condition (which is comparable to assuming that the initial concentration was measured with this accuracy). New profiles were subsequently computed using MAP estimation (by incorporating the prior in the procedure). In this case, the Gaussian prior constrained both the initial condition as well as both scaling factors (see Figure 3.3). Additionally, it can be observed that parameter two is still practically non-identifiable at a significance level of $\alpha = 0.05$.

In the case of JAK-STAT, at least three priors are required to render the model identifiable for all levels of significance. For the initial condition a Gaussian prior ($\mu = 200nM, \sigma = 20nM$) was specified while log-uniform priors with support from 10^{-8} to 10^{2} and 10^{-8} to $10^{1.5}$ were used for parameters p_2 and p_3. The other parameters are assigned unbounded log-uniform prior distributions.

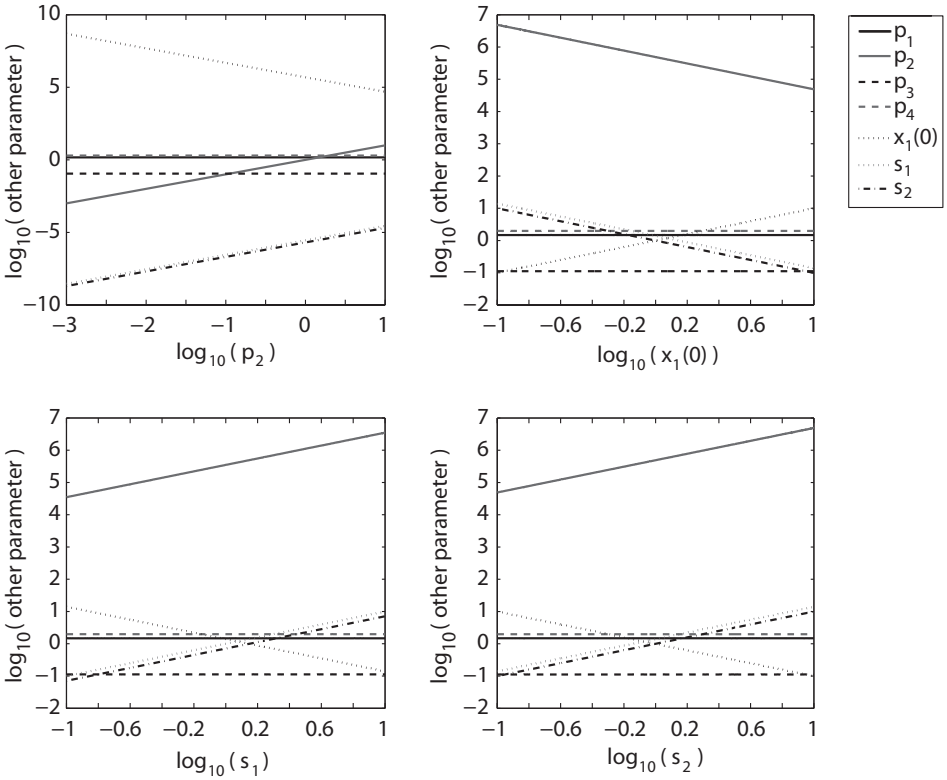

Figure 3.4: Scatter plots of parameters belonging to each profile likelihood

As shown in Figure 3.5, the parameter bounds based on profile likelihood agree well with those based on the MCMC sampling for the identifiable parameters. What can also be observed however is the fact that parameter sets that would be considered likely based solely on the likelihood of the data can still be relatively improbable when the prior probability densities are taken into account. This can be observed for parameter three where the PL path reveals two modes that are almost equally likely, yet show large difference in terms of probability density. The posterior parameter distribution does contain a few samples in this region, but relatively little mass. It is an example of the difference between integration and maximization and indicates that this region corresponds to a sharp ridge in the likelihood. The inferred posterior predictive distribution is shown in Figure 3.6.

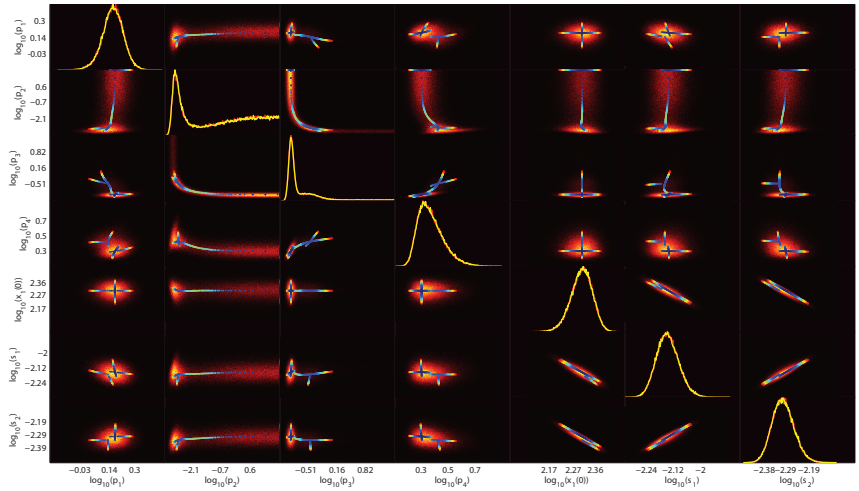

Figure 3.5: Histograms of the posterior distribution. Shown on the diagonal are the marginal (integrated) distributions of the parameters, where different colours indicate different batches of samples. Off the diagonal are the joint probability distributions between two parameters. The correlated nature of several parameters can clearly be observed. Lines indicate profile likelihood trajectories where blue and red corresponds to a good and bad fit respectively. Note that the profile likelihood includes the prior on the scaling.

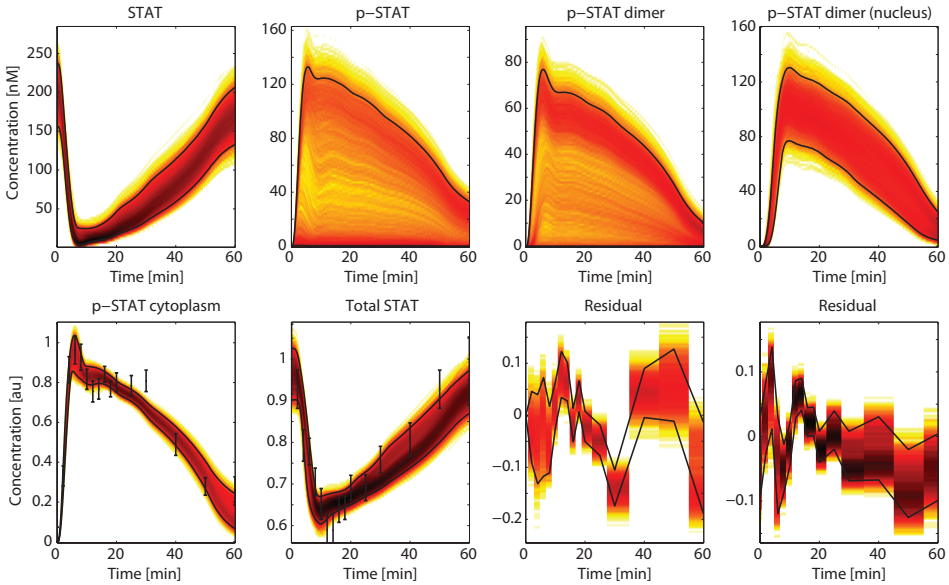

Figure 3.6: Posterior predictive distribution of model predictions (colours) with 95% credible intervals (black lines). Top: Unmeasured internal model predictions. Bottom: Measured model output, data $\pm$ standard deviation and residual distributions.

3.6 Discussion

This chapter describes a strategy for reliably inferring posterior distributions and their predictive counterparts. By classifying different parameters as identifiable or non-identifiable, it is possible to determine which priors are required to be sufficiently informative to ensure that the posterior distribution is proper and can be sampled from. This sampling can subsequently be performed by using an MCMC sampler which incorporates information regarding the local geometry of the posterior density. The strategy enables a comprehensive analysis on the effect of parameter uncertainty on model predictions and enables the modeler to relate these effects to the model parameters.

Given a sufficient amount of data, inference should be relatively insensitive to the assumed priors. As observed in the case of JAK-STAT however, it can be seen that even for a small model, identifiability can be problematic. It is important to realize that in such cases the choice of priors will strongly affect the outcome of the analysis. Regarding the posterior predictive distribution, the assumed priors could be a point of debate. Partial Rank Correlation Coefficient (PRCC) analysis on the predictions revealed that state variable two had a very strong dependence on parameter two (see Appendix 3.7.2). This indicates that predictions regarding state two should be made with care.

Whereas purely likelihood based methods typically revolve around determining properties of all acceptable parameter sets and their associated predictive counterparts, the Bayesian methodology attempts to provide a probabilistic assessment of both the parameter and prediction space. The difference between these two becomes quite clear in the JAK-STAT example, where the probabilistic assessment assigns more probability to the mode with the larger mass (but lower likelihood).

Different approaches for prediction uncertainty analysis based on optimization are proposed in [15–17]. Such methods are useful for probing consistent behavior (termed core predictions) among multiple parameter sets, even in the non-identifiable case. However, they do not result in a probabilistic assessment of the prediction uncertainty. Probing consistent behavior is also the main focus of a workflow proposed by [18] for classifying consistent model behaviors and hypotheses.

Several steps in the proposed approach are computationally challenging and require many model evaluations. Because of this, model simulation time is a primary concern. Many packages have been able to attain significant simulation speed-ups by compiling simulation code, reducing model evaluation time by up to two orders of magnitude (Potters Wheel [19]; COPASI [20]; Sloppy Cell [21]). Additionally, new computational platforms such as General Purpose programming on the Graphical Processing Unit (GPGPU) are being explored [22].

In conclusion, the presented strategy enables the modeler to account for parameter uncertainty when making model predictions. In the case of a fully identifiable model overconfident conclusions that could result from a model described by a single parameter set can be avoided. Regarding non-identifiable models, a practical approach can be adopted where the dependence with respect to the assumed prior distributions can be investigated *a posteriori*. Though this makes

computing the posterior distribution feasible, such an approach underestimates the full uncertainty when the priors are not warranted. Performing the analysis and obtaining a sample from the posterior takes considerably more computational effort than determining a single parameter set. However, once such a sample is obtained, the results can be used for a wide array of model analysis techniques, which more than warrant the additional computational time invested. Relations within this posterior distribution and its predictive counterpart can be extracted. The next chapter will go into more detail regarding how these different relations can actually be used to perform Optimal Experiment Design.

3.7 Appendix

3.7.1 Merging profiles

Merging different profiles corresponding to the same profiled parameter is performed in the following way. First, a list containing all the parameter values that were evaluated is constructed. Subsequently, the WRSS of each individual profile is interpolated for all the points that fall within its bounds. The merged profile is then obtained by determining the minimum value across the different interpolated profiles for each of the points in the list.

Figure 3.7 depicts the separate profiles corresponding to the different modes. Note how the profile shown in solid black fails to leave the local mode for parameter 4, which would suggest its confidence interval is much smaller than it really is. This further underlines the necessity of running profiles for each mode when using local optimization methods.

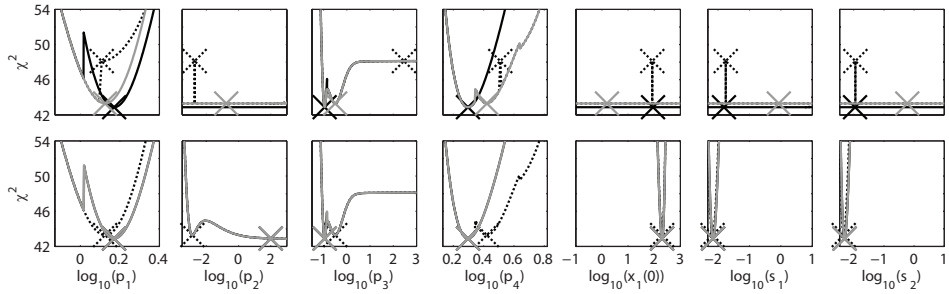

Figure 3.7: Separate profile likelihoods for each mode. Sharp discontinuities indicate a switch between modes. Top: Without prior on the initial condition, Bottom: With prior on the initial condition.

3.7.2 Prior dependence

In order to test the parameter dependence of the different predictions, a Partial Rank Correlation Coefficient is calculated. This method performs a linear regression analysis on rank transformed quantities. The PRCC method computes Pearson correlation coefficients after eliminating the confounding interactions of all the other parameters. This elimination is performed by means of linear regression models. Consider the PRCC between vectors X and Y, while removing the effect of vectors $Z_{1..Q}$. In this case X and Y are regressed against the vectors in Z and the correlation coefficient is computed on their residuals:

$$\xi_{X,i} = \min_{\beta} \left(\sum_{i=1}^{N} \left(X_i - \sum_{j=1}^{Q} \beta_j Z_{j,i} \right)^2 \right)$$

$$\xi_{Y,i} = \min_{\beta} \left(\sum_{i=1}^{N} \left(Y_i - \sum_{j=1}^{Q} \beta_j Z_{j,i} \right)^2 \right)$$

The PRCC between X and Y can be computed as:

$$\rho_{XY/Z} = N \frac{\sum_{i=1}^{N} \xi_{X,i}\xi_{Y,i} - \sum_{i=1}^{N} \xi_{X,i} \sum_{i=1}^{N} \xi_{Y,i}}{\sqrt{N \sum_{i=1}^{N} \xi_{X,i}^2 - \left(\sum_{i=1}^{N} \xi_{X,i}\right)^2} \sqrt{N \sum_{i=1}^{N} \xi_{Y,i}^2 - \left(\sum_{i=1}^{N} \xi_{Y,i}\right)^2}} \tag{3.9}$$

The PRCC provides a measure of whether a monotonic relation exists between the quantities of interest after correcting for all the other variables. The result of this analysis on the JAK-STAT posterior is shown in Figure 3.8. It is clear that s_2 is strongly rank correlated to the second parameter. Since a prior distribution was assumed to artificially constrain parameter two, predictions on s_2 should be made with care.

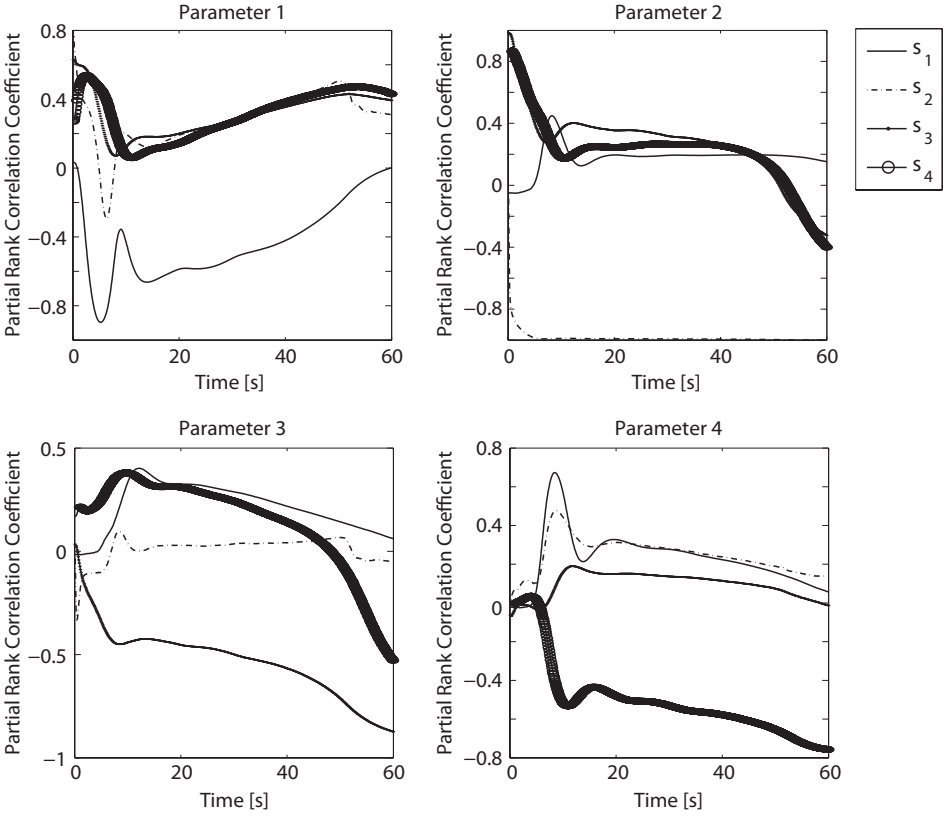

Figure 3.8: Partial Rank Correlation Coefficients between the state variables and parameters. A large absolute value indicates a monotonic relationship between a state variable and parameter. What can be seen is that parameter two and state variable two are highly rank correlated.

References

[1] Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP: **Universally Sloppy Parameter Sensitivities in Systems Biology Models**. *PLoS Comput Biol* 2007, **3**(10):e189.

[2] Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmüller U, Timmer J: **Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood**. *Bioinformatics* 2009, **25**(15):1923.

[3] Jeffreys H: **An Invariant Form for the Prior Probability in Estimation Problems**. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* 1946, **186**(1007):pp. 453–461.

[4] Zwickl D, Holder M: **Model parameterization, prior distributions, and the general time-reversible model in Bayesian phylogenetics**. *Systematic biology* 2004, **53**(6):877.

[5] Geyer C: **Practical markov chain monte carlo**. *Statistical Science* 1992, :473–483.

[6] Calderhead B, Girolami M: **Estimating Bayes factors via thermodynamic integration and population MCMC**. *Computational Statistics & Data Analysis* 2009, **53**(12):4028–4045.

[7] Rannala B: **Identifiability of parameters in MCMC Bayesian inference of phylogeny**. *Systematic Biology* 2002, **51**(5):754.

[8] Gelfand A, Sahu S: **Identifiability, improper priors, and Gibbs sampling for generalized linear models**. *Journal of the American Statistical Association* 1999, :247–253.

[9] Box G, Tiao G: *Bayesian inference in statistical analysis*. Wiley Online Library 1973.

[10] Grandison S, Morris R: **Biological pathway kinetic rate constants are scale-invariant**. *Bioinformatics* 2008, **24**(6):741.

[11] Cowles M, Carlin B: **Markov Chain Monte Carlo convergence diagnostics: a comparative review.** *Journal of the American Statistical Association* 1996, **91**(434).

[12] Gilks W, Gilks W, Richardson S, Spiegelhalter D: *Markov chain Monte Carlo in practice*. Interdisciplinary statistics, Chapman & Hall 1996.

[13] Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf M: **Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems**. *Journal of the Royal Society Interface* 2009, **6**(31):187–202.

[14] Swameye I, Müller T, Timmer J, Sandra O, Klingmüller U: **Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling**. *Proceedings of the National Academy of Sciences* 2003, **100**(3):1028.

[15] Brännmark C, Palmér R, Glad S, Cedersund G, Strålfors P: **Mass and information feedbacks through receptor endocytosis govern insulin signaling as revealed using a parameter-free modeling framework**. *Journal of Biological Chemistry* 2010, **285**(26):20171.

[16] Nyman E, Brannmark C, Palmer R, Brugard J, Nystrom F, Stralfors P, Cedersund G: **A hierarchical whole body modeling approach elucidates the link between in vitro insulin signaling and in vivo glucose homeostasis**. *Journal of Biological Chemistry* 2011.

[17] Cedersund G, Roll J: **Systems biology: model based evaluation and comparison of potential explanations for given biological data**. *FEBS Journal* 2009, **276**(4):903–922.

[18] Gomez-Cabrero D, Compte A, Tegner J: **Workflow for generating competing hypothesis from models with parameter uncertainty**. *Interface Focus* 2011, **1**(3):438.

[19] Maiwald T, Timmer J: **Dynamical modeling and multi-experiment fitting with PottersWheel**. *Bioinformatics* 2008, **24**(18):2037–2043.

[20] Hoops S, Sahle S, Gauges R, Lee C, Pahle J, Simus N, Singhal M, Xu L, Mendes P, Kummer U: **COPASIa complex pathway simulator**. *Bioinformatics* 2006, **22**(24):3067.

[21] Brown KS, Sethna JP: **Statistical mechanical approaches to models with many poorly known parameters**. *Phys. Rev. E* 2003, **68**(2):021904.

[22] Liepe J, Barnes C, Cule E, Erguler K, Kirk P, Toni T, Stumpf M: **ABC-SysBio approximate Bayesian computation in Python with GPU support**. *Bioinformatics* 2010, **26**(14):1797.

A Bayesian Approach to Experimental Design

4

Parts of this chapter are described in:

Vanlier, J. and Tiemann, C.A. and Hilbers, P.A.J. and van Riel, N.A.W. (2012), *A Bayesian approach to targeted experiment design*, *Bioinformatics*, 28(8), 1136-1142.

Abstract

Systems biology employs mathematical modeling to further our understanding of biochemical pathways. Since the amount of experimental data on which the models are parameterized is often limited, these models exhibit large uncertainty in both parameters and predictions. Statistical methods can be used to select experiments that will reduce such uncertainty in an optimal manner. However, existing methods for Optimal Experiment Design (OED) rely on assumptions that are inappropriate when data is scarce considering model complexity. This chapter introduces a novel method to perform OED for models that cope with large parameter uncertainty. The efficacy of new measurements on the uncertainty of selected predictions can be predicted by applying importance sampling on the Posterior Predictive Distribution. These predicted measurement efficacies can subsequently be used for selecting an optimal experiment. The proposed method is demonstrated by applying it to a case where it is shown that specific combinations of experiments result in more precise predictions.

4.1 Introduction

In the previous chapters, it was discussed how computational models can be used to predict (un)measured behavior or system responses and formalize hypotheses in a testable manner. Despite the development of new quantitative experimental techniques, the modeler is often faced with a situation where large regions of parameter space can describe the measured data to an acceptable degree [1–5]. Poorly constrained parameters do not necessarily have to be a problem when the predictions required for testing the hypothesis (which shall be referred to as predictions of interest) are well constrained [6–10]. However, when this is not the case, more data will be required. Optimal Experiment Design methods can be used to determine which experiments would be most useful in order to perform statistical inference. These methods typically involve specifying an optimality or design criterion and finding the experiment which maximizes this criterion. Classical design criteria are often based on linearization around a best fit parameter set [11] and pertain to effectively constraining the parameters [12, 13] or predictions [14]. These design criteria have been assigned different letters of the alphabet as shown in table 4.1. Bayesian analogs to all of these design criteria exist, which reduce to the same forms when the models are linear and Gaussian [15].

Optimality	Description
A	Minimize the average variance of parameter estimates
C	Minimize a linear combination of the model parameters
D	Maximize determinant of the Fisher information matrix
E	Maximize the minimum eigenvalue of the Fisher information matrix
T	Maximize trace of the information matrix
G	Minimize maximum variance over the model predictions
V	Minimize average variance of model predictions

Table 4.1: Different optimality criteria for experiment design

However as shown in Chapter 2, methods based on linearizations and Gaussian parameter distributions are not appropriate when data is scarce considering the model complexity or when the model is strongly non-linear [7]. Additionally, investigators may be interested in what a model predicts for only a small set of predictions of interest without having to constrain the rest. In such cases, it makes little sense to optimize for all the predictions or parameters in the model. This is why a more general framework, free of these approximations, is desired. In this chapter a method for optimal experiment design is proposed which overcomes these issues by adopting an approach which exploits the prediction uncertainties in the model. This method enables the modeler to target experimental efforts to selectively reduce the uncertainty of predictions of interest. Using the approach, multiple experiments can be designed simultaneously, revealing potential benefits that arise from specific combinations of experiments.

To perform inference and experiment design, an error model is required. For ease of notation, the method is demonstrated using a Gaussian error model. Considering M time series of length N_1, N_2 ... N_M, the following equation is obtained for the probability density function of the output data:

$$\begin{aligned} p(\mathbf{y}|\vec{\theta}) &= \prod_{i=1}^{M}\prod_{j=1}^{N_i} p(y_{ij},\vec{\theta}) \\ &= K\exp\left(-\sum_{i=1}^{M}\sum_{j=1}^{N_i}\left(\frac{y_{ij}-y_i(t_j,\vec{\theta})}{\sqrt{2}\sigma_{ij}}\right)^2\right) \end{aligned} \tag{4.1}$$

Again y represents the noiseless system output, while y_{ij} and σ_{ij} indicate the mean and standard deviation of a specific data point. K serves as a normalization constant. Using Bayes' theorem and prior distributions for the parameters, an expression for the posterior probability density of the parameter values can be obtained [16]. Both computational as well as methodological advances have made Markov Chain Monte Carlo (MCMC) an attractive option for obtaining samples from such a distribution [4, 16, 17].

Given a sample of the posterior parameter distribution, predictions can be made by simulating the model for each of the sampled parameter sets. The distribution of such predictions shall be referred to as the Posterior Predictive Distribution (PPD) and reflects their uncertainty. Since all of these predictions are linked via the parameter distributions, the relations between the different predictions can be used for experimental design. By considering the effects of a new measurement on the PPD, it is possible to generate a prediction on the usefulness of a future experiment. The method consists of a number of steps which shall be discussed first. Subsequently the method is demonstrated by applying it to a case study followed by some concluding remarks.

4.2 Approach

To overcome the limitations of existing OED methods, a sampling based approach for experimental design is proposed. In brief, the approach consists of the following steps. First, we sample from the posterior distribution. Using the sampled parameters, PPDs are simulated for the output of interest and each experiment that can be performed in the lab. Subsequently, based on a measurement uncertainty model, expected variance reductions are computed for various combinations of experiments. After sampling a large number of potential experiments, these predicted reductions are used to select the optimal (combination of) experiments.

Step 1. Sample from the posterior parameter distribution

The first step in the analysis consists of obtaining a sample from the posterior parameter distribution of the model conditioned on the available data:

$$p(\vec{\theta}|\mathbf{y^D}) \propto p(\mathbf{y^D}|\vec{\theta})p(\vec{\theta}) \tag{4.2}$$

Where $p(\mathbf{y^D}|\vec{\theta})$ denotes the conditional probability of the data given the model parameters. Here probability reflects a degree of belief and prior knowledge regarding the parameter values is included in the form of prior distributions $p(\vec{\theta})$. A sample from the posterior distribution is obtained with Markov Chain Monte Carlo (MCMC) sampling. As a proposal kernel, an adaptive Gaussian proposal distribution is employed whose covariance matrix is based on a quadratic approximation to the posterior probability density at the current sample point [6].

Step 2. Simulate posterior predictive distributions for all candidate experiments

A Posterior Predictive Distribution (PPD) is a distribution of new observations conditioned on the available data as shown in equation (4.3). Samples from the PPD are obtained by simulating the model (including the addition of measurement noise) for a sample of parameter sets from the posterior parameter distribution. For practical reasons, the addition of the observational noise will be performed in step 3. PPDs are simulated for each candidate experiment and target quantity. This involves enumerating and simulating each potential experiment and storing the associated measurable predictions. These PPDs now form a link between the parameters and different predictions. Since the model and data constrain the dynamics of the system, these impose non-trivial relations between the different predictions. Therefore, the observable quantities of candidate experiments are related to the prediction being targeted. The next step is to exploit the relations within these distributions for experiment selection.

$$p(\mathbf{y_{obs}}|\mathbf{y^D}, \vec{u}_t) = \int p(\mathbf{y_{obs}}|\vec{\theta}, \vec{u}_t)p(\vec{\theta}|\mathbf{y^D})d\vec{\theta} \tag{4.3}$$

Step 3. Predict expected variance reductions

To perform experiment design, a measure of expected measurement efficacy is required. Therefore, the Expected Variance Reduction (EVR) is introduced. Consider an independent new measurement. This new measurement is associated with an error model G, which reflects the uncertainty associated with the new experiment. If this new experiment were to be performed, then the subsequent step would be to incorporate the new data point (and its associated error model) in the likelihood function and perform inference. This new data would subsequently constrain the posterior parameter distribution, hence also affecting the prediction of interest (which cannot be measured directly). This process is illustrated in Figure 4.1.

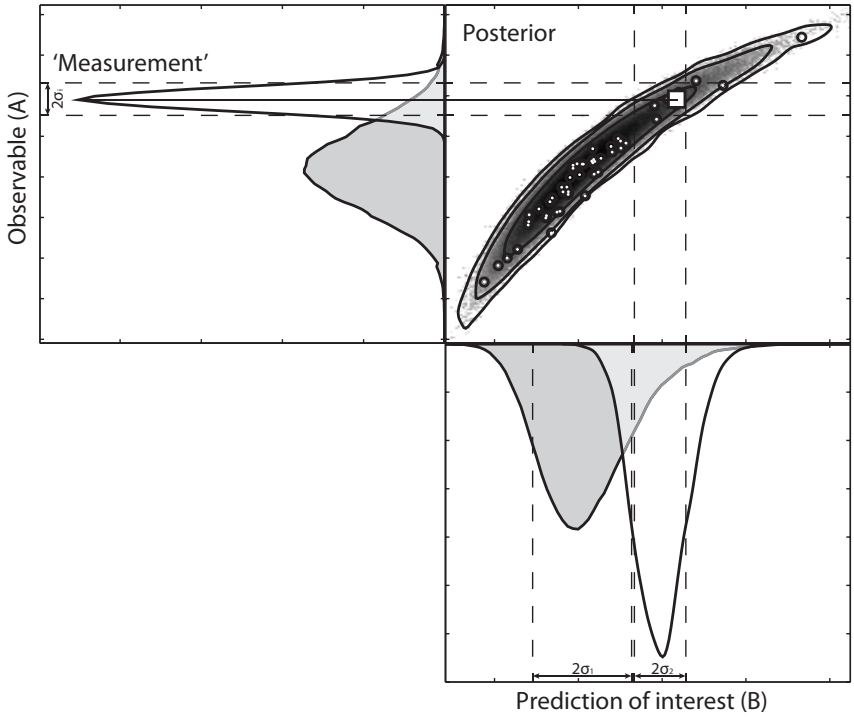

Figure 4.1: Illustration of the effect of adding a new data point on the Posterior Predictive Distribution (PPD). Shown on the top right is the PPD at one specific time point for two predictions with a subset of the samples of the chain indicated with white dots. The square denotes the location of the 'new measurement'. Prediction A refers to a prediction of which a new measurement can be performed (observable), while B denotes the prediction of interest. Here the gray distribution corresponds to the PPD before the new measurement, while the white Gaussian corresponds to the error model of the new measurement. Due to additional constraints imposed by this new measurement in combination with the old data and the model, the distribution on the hypothesis side is also updated in light of the new data point and shown in white.

Before the experiment is performed, the resulting data point is unknown. The PPD provides a predictive distribution of the new data point (shown in gray in Figure 4.1), which also reflects the uncertainty associated with this value. Samples from the PPD can subsequently be substituted as measured values in the error model. A distribution of variances can be obtained by repeating this process for every R^{th} point of the MCMC chain. By averaging the result over the whole chain, the expected variance is implicitly weighted by the distributions of the candidate observables.

Considering that a single MCMC is often computationally demanding, approaching this problem by means of a nested MCMC is not tractable. Therefore an alternative approach is required. Consider the non-normalized densities $\tilde{p}(y|\vec{\theta})$, $\tilde{p}(y_n|\vec{\theta})$ and $\tilde{p}(\vec{\theta})$ respectively corresponding to the density model of the data used to determine the initial posterior distribution, the density model for the new data point, and the parameter prior. Assuming that the noise on the new data point is independent of the existing data points gives $p_N(\vec{\theta}|y, y_n) \propto \tilde{p}(y|\vec{\theta})\tilde{p}(y_n|\vec{\theta})\tilde{p}(\vec{\theta})$. From this it follows that the new normalized posterior is given by:

$$\begin{aligned} p_N(\vec{\theta}|y, y_n) =& \frac{\tilde{p}(y|\vec{\theta})\tilde{p}(y_n|\vec{\theta})\tilde{p}(\vec{\theta})}{\int \tilde{p}(y|\vec{\theta})\tilde{p}(y_n|\vec{\theta})\tilde{p}(\vec{\theta})d\vec{\theta}} \\ =& \frac{\tilde{p}(y|\vec{\theta})\tilde{p}(\vec{\theta})}{\int \tilde{p}(y|\vec{\theta})\tilde{p}(\vec{\theta})d\vec{\theta}} \frac{\tilde{p}(y|\vec{\theta})\tilde{p}(y_n|\vec{\theta})\tilde{p}(\vec{\theta})}{\tilde{p}(y|\vec{\theta})\tilde{p}(\vec{\theta})} \frac{\int \tilde{p}(y|\vec{\theta})\tilde{p}(\vec{\theta})d\vec{\theta}}{\int \tilde{p}(y|\vec{\theta})\tilde{p}(y_n|\vec{\theta})\tilde{p}(\vec{\theta})d\vec{\theta}} \\ =& p(\vec{\theta}|y)\tilde{p}(y_n|\vec{\theta}) \frac{\int \tilde{p}(y|\vec{\theta})\tilde{p}(\vec{\theta})d\vec{\theta}}{\int \tilde{p}(y|\vec{\theta})\tilde{p}(y_n|\vec{\theta})\tilde{p}(\vec{\theta})d\vec{\theta}} = p(\vec{\theta}|y)\tilde{p}(y_n|\vec{\theta}) \frac{Z_1}{Z_2} \end{aligned} \tag{4.4}$$

In this equation Z_1 and Z_2 denote the normalization constants of the old and new posterior respectively. The proportional relation between the two posteriors can be used to compute expected values by re-weighting samples from the old posterior. Rather than running a new MCMC for every sample, Self Normalized Importance Sampling is used on the predictions of the output in order to compute expected values. This is shown in equation (4.5), where samples $\vec{\theta}_i$ and $\vec{\theta}_j$ are taken from the old posterior distribution, T indicates the number of MCMC samples included in the analysis and $z(\vec{\theta})$ indicates the quantity of interest.

$$\begin{aligned} \mathbb{E}_{\vec{\theta}|y,y_n}[z] = \int p_N(\vec{\theta}|y, y_n)z(\vec{\theta})d\vec{\theta} &= \int p(\vec{\theta}|y)\tilde{p}(y_n|\vec{\theta}) \frac{Z_1}{Z_2} z(\vec{\theta})d\vec{\theta} \\ &\approx \sum_{i=1}^{T} \frac{\tilde{p}(y_n|\vec{\theta}_i)}{\sum_{j=1}^{T} \tilde{p}(y_n|\vec{\theta}_j)} z(\vec{\theta}_i) \end{aligned} \tag{4.5}$$

The value of y_n is not known *a priori*. Therefore, expected values are computed for each parameter set in the PPD sample with y_n set to the sampled output prediction. This provides a distribution of expected values for z, where each sample corresponds to one possible outcome of the experiment. The mean of these expected values provides a prediction of the quantity of interest averaged

over the density of the (still unmeasured) data point. The entire approach can succinctly be summarized as:

$$\mathbb{E}_{y_n}\left[\mathbb{E}_{\vec{\theta}|y,y_n}[z]\right] \approx \frac{1}{T}\sum_{r=1}^{T}\sum_{i=1}^{T}\frac{G(t,\vec{u}(t),\vec{\theta}_i,\vec{\theta}_r)}{\sum_{k=1}^{T}G(t,\vec{u}(t),\vec{\theta}_k,\vec{\theta}_r)}z(t,\vec{u}(t),\vec{\theta}_i) \tag{4.6}$$

Here G corresponds to the error model and $\vec{\theta}_i$ refers to the i^{th} parameter vector of the chain. Assuming a Gaussian error model with standard deviation σ for the new measurement of y, the density G is given by:

$$G(t,\vec{u}(t),\vec{\theta}_i,\vec{\theta}_r) = \exp\left(-\frac{1}{2\sigma^2}\left(y(t,\vec{u}(t),\vec{\theta}_i) - y(t,\vec{u}(t),\vec{\theta}_r)\right)^2\right) \tag{4.7}$$

Note that both the input y as well as the output z can be *any* quantity of interest (prediction or parameter) indicating the flexibility of the approach. Since the variance of a variable of interest can be expressed in terms of expected values:

$$Var[z] = \mathbb{E}[z^2] - (\mathbb{E}[z])^2 \tag{4.8}$$

the aforementioned equation can be used to estimate this quantity. Doing so allows calculation of a predicted conditional variance for every sample of the posterior distribution. The variance reductions can then be computed as:

$$VarR = 1 - \frac{\sigma^2_{new}}{\sigma^2_{old}} \tag{4.9}$$

where σ^2_{old} corresponds to the posterior variance without the new measurement and σ^2_{new} corresponds to the expected posterior variance with the new measurement(s) taken into account. In other words, one obtains the expected variance reduction considering the prediction uncertainty. The variance reduction computed by this sampling method is referred to as the Sampled Variance Reduction (SVR).

Implementation of the methodology is straightforward. The measurement efficacy is computed by weighting samples of the old posterior. This weighted average then forms a prediction for the situation after the experiment has been performed. Since the method extensively employs importance sampling, it is important to consider the quality of these IS estimates (see Appendix 4.6.1). If the posterior distribution before and after a new experiment is very different, many of the sample weights will be low and a large fraction of the samples will contribute negligibly to the predicted variance. This degeneracy can be monitored by estimating the Effective Sample Size (ESS) defined below [18]:

$$ESS_r = \frac{\left(\sum_{k=1}^{N}G(t,\vec{u}(t),\vec{\theta}_k,\vec{\theta}_r)\right)^2}{\sum_{k=1}^{N}G(t,\vec{u}(t),\vec{\theta}_k,\vec{\theta}_r)^2} \tag{4.10}$$

Measurement noise

New measurements will be affected by noise, which is part of the posterior predictive distribution. This noise can be simulated when generating samples from the predictive distributions. In many cases additive Gaussian errors are assumed as they provide a convenient, and often sufficient description of the measurement errors involved. An additional advantage is that this removes the requirement to explicitly simulate the measurement noise since the following identity can be used:

$$\begin{aligned} p(y_n|\vec{\theta}) &= \int_{-\infty}^{\infty} K p(y_n|y_p) p(y_p|\vec{\theta}) dy_p \\ &= \int_{-\infty}^{\infty} K \exp\left(-\left(\frac{y_n - y_p}{\sqrt{2}\sigma} \right)^2 - \left(\frac{y_p - y(\vec{\theta})}{\sqrt{2}\sigma} \right)^2 \right) dy_p \\ &= K\sqrt{\pi\sigma^2} \exp\left(-\left(\frac{y_n - y(\vec{\theta})}{2\sigma} \right)^2 \right) \end{aligned} \tag{4.11}$$

Here K denotes a normalization constant independent of $\vec{\theta}$. $p(y_p|\theta)$ is the posterior probability density of predictions while $p(y_n|\theta)$ represents the probability density of observing y_n in a new measurement. Furthermore $p(y_t|y_p)$ refers to the error model of the new measurement. Since the sampling uses self-normalization, these constants are irrelevant. From this it follows that in the Gaussian additive case, the measurement noise can be taken into account by multiplying the standard deviation of the measurement error by a factor of $\sqrt{2}$.

Linear variance reduction

When the measurement error models and PPD can reasonably be assumed Gaussian, the variance reduction can be estimated using a Gaussian approximation of the PPD between the output and the measurements of interest. Using this Gaussian approximation, additional measurements hampered by Gaussian additive errors can be taken into account. First the covariance matrix of the PPDs is computed as:

$$\Sigma_{posterior} = cov\left(\begin{bmatrix} z_1 & x_1^1 & \dots & x_1^Q \\ z_2 & x_2^1 & \dots & x_2^Q \\ \vdots & \vdots & \ddots & \vdots \\ z_T & x_T^1 & \dots & x_T^Q \end{bmatrix} \right) \tag{4.12}$$

Where z denotes the output of interest and x_b^a the b^{th} MCMC sample of the a^{th} measurable state (without measurement noise), with Q and T the number of measured points and samples respectively. New measurements would be hampered by additive noise whose contribution can be taken into account by either

explicitly simulating it or, in the case of Gaussian additive noise, adding a diagonal matrix containing the noise variances for each of the measurable components. This noise matrix is given by:

$$\Sigma_{noise} = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & \sigma_1^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_Q^2 \end{bmatrix} \tag{4.13}$$

In the linear case, updates of the covariance matrix do not depend on the mean of the new measurement and therefore the Gaussian distribution can be updated by using the Gaussian identity:

$$\mathbf{C} = \left(\mathbf{A}^{-1} + \mathbf{B}^{-1}\right)^{-1} \tag{4.14}$$

The measurement accuracies of the new measurement (also Gaussian) are given as follows:

$$\Sigma_{meas} = \begin{bmatrix} \infty & 0 & \dots & 0 \\ 0 & \sigma_1^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_Q^2 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 1/\sigma_1^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1/\sigma_Q^2 \end{bmatrix} \tag{4.15}$$

Therefore, the covariance matrix can be updated according to:

$$\Sigma_{new} = \left(\left(\Sigma_{posterior} + \Sigma_{noise}\right)^{-1} + \Sigma_{meas}\right)^{-1} \tag{4.16}$$

The resulting variance of the prediction of interest z can then be obtained as $\Sigma_{new}(1,1)$. It is worth noting that this approximation comes at very little computational cost and does not suffer from sampling degeneracy.

Step 4. Determine experiments with optimal variance reduction

Considering independent measurements, the method can be generalized to multiple experiments. In this case, the probability density model can be obtained by multiplying the error models for each candidate measurement. Subsequently, the space of all candidate measurements is sampled using Monte Carlo sampling. The efficacy of a specific combination of measurements is evaluated by computing the variance reduction, which is defined as equation (4.9). During this sampling stage, additional constraints which arise because of practical considerations can be imposed on the experimental design (by rejecting such samples). An example of this could be the inability to measure certain state variables simultaneously. The optimal experiment is obtained by determining those combinations of measurements which yield the largest predicted variance reduction.

4.3 Computational methods

Algorithms were implemented in MATLAB (Natick, MA). Numerical integration of the differential equations was performed with compiled MEX files using the numerical integrators for stiff systems from the SUNDIALS CVode package (Lawrence Livermore National Laboratory, Livermore, CA). Absolute and relative tolerances were set to 10^{-8} and 10^{-9} respectively. The Gaussian proposal distribution for the MCMC was based on an approximation to the Hessian computed using a Jacobian based on finite differences ($H \approx J^T J$). All available priors were included in the Hessian approximation. After convergence, the chain was thinned to 10000 samples. The Sampled Variance Reduction was computed in parallel using a MEX-based GPU implementation based on OpenCL.

4.4 Results

To demonstrate the method, it is applied to the STAT signaling pathway model [1,19] introduced in the previous chapter. To infer the posterior distribution, data from the paper by Swameye et al [20] were used. Measured quantities were the total concentration of STAT ($x_1 + x_2 + 2x_3$) and the total concentration of phosphorylated STAT in the cytoplasm ($x_2 + 2x_3$), both reported in arbitrary units. The initial cytoplasmic concentration of STAT is unknown while all other forms of STAT are assumed zero at the start of the simulation. Given the data, not all parameters are identifiable [1]. Log-uniform priors were used for the kinetic parameters and a Gaussian distribution ($\mu = 200nM, \sigma = 20nM$) for the initial condition. Parameter two was bounded between ranges, since this parameter was non-identifiable from the data [1]. Two chains were simulated starting at different initial values. These were simulated up to one million parameter sets and convergence was assessed by visually inspecting differences between batches of samples.

4.4.1 Relations present in the Posterior Predictive Distribution

After determining the posterior parameter distribution, PPDs were computed for all model state variables. An example of a time dependent relation between predictive densities is shown in Figure 4.2. The second and third column represent two different snapshots in time. Each dot in the scatter plots corresponds to the simulated values of two state variables for a single sample of the MCMC chain. It can be observed that the state variables are related in a highly non-linear fashion. The associated 2D histograms provide an estimate of the density of these predictions. It can also be seen that the posterior predictive distribution contains multiple modes.

Considering state variable three as observable and state variable four as prediction, while assuming a measurement accuracy of $\sigma = 10/\sqrt{2}$ reveals that a significant decrease in variance can be attained during the rise of state three. Measuring state three at the peak value results in a smaller variance reduction. Considering the relations between the various predictions, the following obser-

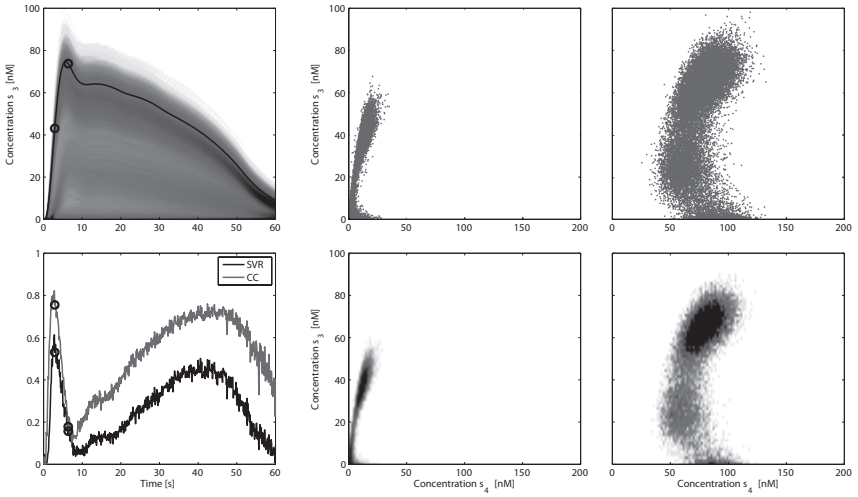

Figure 4.2: Top left: One simulated time course of state 3 superimposed on the PPD. Two time points are indicated with circles. Bottom left: Correlation Coefficient (CC) between state 3 and 4 and Sampled Variance Reduction of state 4 based on a measurement of state 3 (SVR). The relation between the two states at the indicated time points is shown in both a scatter plot as well as a 2D histogram. The former shows the actual samples from the posterior predictive distribution for one point in time. The dots represent simulated values corresponding to parameter sets from the MCMC chain. The histogram serves as a density estimate, where the color indicates the number of samples in a particular region.

vations were made. An experiment is only effective if there is a correlation between the measurement and the prediction of interest. Additionally, the uncertainty in both should be large enough to result in an appreciable variance reduction. Since all predictions of state three start with an initial condition of zero, this implies that the uncertainty at this point is low. Therefore an additional measurement at $t = 0$ would not yield any variance reduction which is also reflected by the fact that the Sampled Variance Reduction starts at a value of zero.

4.4.2 Leave one out experiment

To test the method, OED was performed using only a subset of the data. A posterior distribution was computed where most of the data corresponding to the total amount of cytoplasmic STAT was omitted. Of this observable, only the first data point was included. Subsequently, MCMCs were performed for each omitted time point. For every MCMC run, only one of the omitted data points was included, thereby obtaining post-experiment variance reductions. Subsequently, variance reductions for these same time points were predicted using the proposed SVR method. To ensure that the outcome would be comparable to the experiments, the standard deviations of the omitted experiments were used as measurement accuracy. The results of this analysis are shown in Figure 4.3. The observed variance reductions agree well with the actual reductions obtained.

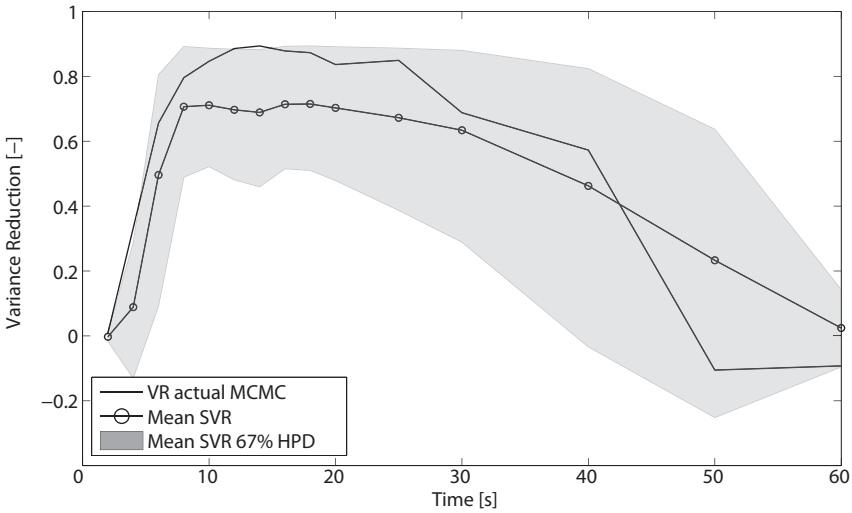

Figure 4.3: Expected variance reduction compared with true variance reduction after the experiment has been performed. The shaded area indicates the 67% Highest Posterior Density of the distribution of variance reductions obtained, while the solid line indicates its mean. Variance reductions obtained after including the actual measurement are shown in circles.

4.4.3 Combinatorial experiment design

In this section, the time to peak of dimerized STAT in the nucleus (state variable s_4) is targeted. A sample from the predictive distribution was computed by determining the time to peak of s_4 for each parameter set sampled from the posterior parameter distribution. All state variables, except s_4 were assumed measurable with an accuracy of $\sigma = 10/\sqrt{2}$. The two sums of states as measured in earlier experiments were also included as potential candidates for new measurements. The experiment space was sampled using a Monte Carlo approach, uniformly sampling the experiment design space.

The result of this sampling is shown in Figure 4.4, where the SVR is shown for several combinations of two measurements. Each axis corresponds to a potential measurement. Different model outputs are separated by grid lines, while the interval between each pair of lines corresponds to an entire time series. The color value indicates the SVR for that specific experiment. Recall that the data set used for inferring the initial posterior distribution contained measurements of two sums of model states. These two observables correspond to output five and six in Figure 4.4. The low variance reductions in the area corresponding to these two observables indicates that additional measurements on output five and six would provide very little additional information. This is not surprising since these observables were already measured before. Performing the experimental design for two measurements revealed that the largest reduction in variance could be obtained by measuring state variable s_1 at an early and late time point. This result underlines the benefit of being able to combine multiple measurements in the OED procedure. Furthermore, the analysis revealed that the

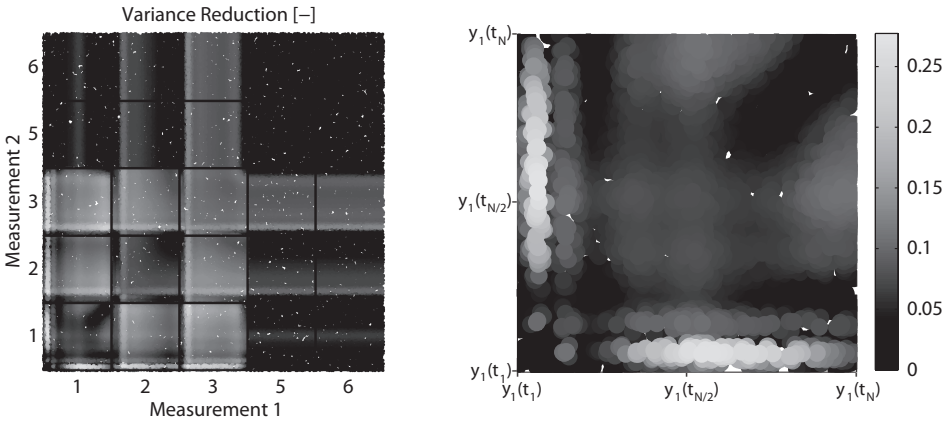

Figure 4.4: Expected variance reductions (SVR) of the peak time of dimerized STAT (x_4) when performing two new measurements. In the left panel each axis represents an experiment, where the different model outputs are numbered. Numbers one to three correspond to the first three states while four and five correspond to the sums of states which were already measured. Each block corresponds to an entire time series. The block corresponding to measurements of state one is enlarged in the right panel.

timing of this first time point is crucial. This also meant that if accurate timing is not possible in the experiment one could consider measuring state variable s_3 and s_1 instead. Here smaller reductions are attained but the timing accuracy required for a reasonable reduction is less stringent. Since both error models are Gaussian, we can also compare the results from the SVR with its linearized form, the LVR. The results of this comparison are shown in Figure 4.5. Qualitatively, these results agree well.

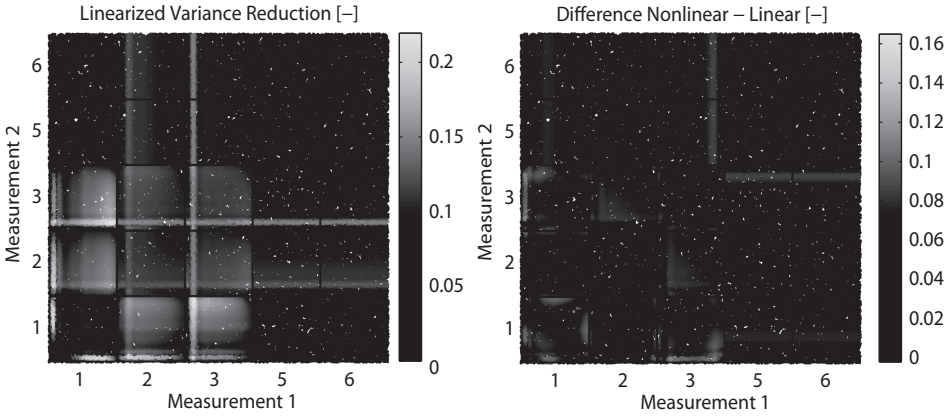

Figure 4.5: Variance reduction of the peak time of dimerized STAT (x_4) with respect to two new measurements. Left: Linear Variance Reduction (LVR). Right: Difference between the variance reduction computed by means of LVR and SVR (see Figure 4.4).

4.5 Discussion and concluding remarks

This chapter described a method to select optimal experiments for reducing the uncertainty of specific predictions. The method is particularly well-suited to cases where data is scarce and the posterior parameter distribution cannot reasonably be assumed Gaussian. By applying it to quantities that depend nonlinearly on model simulations, we have demonstrated that the method is flexible. These particular characteristics make it well suited to Systems Biology research.

For an experiment to be effective, the variables under consideration need to be correlated as well as show sufficient uncertainty in their posterior predictive distribution. In this work, the timing of the new measurement was assumed instantaneous (infinitely accurate). It remains an open but relevant challenge to incorporate temporal inaccuracies in the current framework. It is expected that when timing is more error prone and explicitly accounted for, experiments that are only effective during brief time intervals will be marked less beneficial.

In this framework, the selection of an experimental design is based on the expected value of a distribution of predicted variance reductions. Though basing the design on the expected variance reduction makes sense from a probabilistic point of view, other options could be considered. Since a distribution of potential variance reductions is inferred and has been computed, one could also consider incorporating information regarding the accuracy of this estimate into the selection process. As demonstrated in the leave-one-out experiment it is possible to compute credible intervals on the predicted variances. Such intervals can be incorporated in the decision making process. Finding a sensible trade-off between the expected variance reduction and its inaccuracy remains an open topic for further research.

To obtain the posterior distribution, the parameters are required to be either identifiable or restricted by means of a finite prior distribution. Even for a small model, identifiability can be problematic but easily tested [1]. Given a sufficient amount of data, the posterior distribution should be relatively insensitive to the assumed priors. It is important to verify this *a posteriori*. One option to investigate prior dependence is to vary the assumed priors or determine the effect of a measurement on the quantity that the assumed prior represents. The latter option strongly depends on the initial prior, which should be chosen sufficiently wide to cover all potential parameter regimes.

The method is not limited to any specific family of distributions for the parameters and model predictions. However, strongly tailed distributions can be problematic. Such distributions can lead to unreliable variance estimates due to undersampling the tail of the distribution. Additionally, variances of such distributions are difficult to interpret, as the tail generally dominates the variance estimate. Therefore, it is sensible to visually inspect the actual predictive distributions of the optimal design. Since the number of samples required to get a reliable variance estimate is problem specific, it is recommended to monitor the ESS. Additionally, refer to Chapter 7 for a suitable bias correction for low ESS.

Obtaining samples from the PPD and performing the experimental design is computationally expensive. Regarding inference, model simulation time is a primary concern which can be significantly reduced by using compiled simulation

code (see COPASI [21]; ABC-SysBio [22]; Potters Wheel [23]; Sloppy Cell [24]). Additionally, more efficient sampling methods for obtaining posterior distributions in high dimensional spaces are being developed [4, 19]. Regarding the experiment design, the computational burden can be divided into two contributions. First is sampling the experiment space. Since each experiment constitutes a dimension in experiment space, densely sampling this space for a large number of experiments can become prohibitively time consuming. When designing several experiments simultaneously, it may be required to resort to more sophisticated sampling techniques such as Sequential Monte Carlo methods. In many cases, it is beneficial to perform a fast initial sweep of the experiment space by sampling the LVR. This can then be followed up by computing the actual SVR for those samples that resulted in an appreciable LVR. Finally, profiling the resampling step revealed that the distance calculations for the error model were most time consuming. Since this step exhibits a large degree of parallelism, a GPU OpenCL implementation of the resampling step was implemented. Even on a modest GPU (NVIDIA Quadro FX 580) this resulted in a considerable speedup (see Appendix 4.6.2).

A flexible data-based strategy for optimal experiment design was proposed. Where existing design criteria pertain to effectively constraining specific parameters or target the variance of predictions using model linearization [12–14], this method is not limited to any specific error models or assumptions regarding the parameter distributions. It enables the modeler to select specific predictions of interest that require decreased uncertainty and thereby focuses the experimental efforts towards specific predictive goals. Furthermore, it allows the prediction of interest to be any quantity that can be obtained from simulations and enables the inclusion of multiple different measurements simultaneously in order to elucidate their combinatorial efficacy.

4.6 Appendix

4.6.1 Sampling bias

Consider performing a new measurement as illustrated in Figure 4.1. The estimation of the measurement efficacy involves multiplying samples of the old posterior with weights to estimate quantities that resemble the situation after the experiment has been performed. When computing such a weighted average, it is important to keep track of the quality of the estimation. When the posterior before and after a new experiment is very different, many of these sample weights will be very low and a large fraction of the samples will contribute only negligibly to the estimation of the new variance. This degeneracy is monitored by estimating the Effective Sample Size (ESS).

When the importance sampling distribution is similar to the new posterior, the ESS should scale linearly with the number of included samples. When the values for the ESS are very low then values obtained for the variance reduction can be inaccurate. Since every point of the MCMC is treated as a potential measurement result, this also includes samples farther from the high density region of the posterior. Since the density of samples is lower here, the number of samples that significantly contribute to the new posterior variance estimate is small. In the most extreme case (consider the outermost sample), the expected mean, after incorporating the new measurement, would be biased towards the high density region, while the variance would be underestimated. This worst case scenario is illustrated in Figure 4.6 for a 1D distribution.

Since the aim is not to estimate a single variance, but rather compute an expected value for an entire distribution of variances, this problem is mitigated somewhat. Considering that most of the weighting will take place in the high density region, the estimation error will be reasonable for most samples. It is expected that the method will show slight bias for low numbers of included samples, but that the bias will quickly decrease as the sample size increases. Several factors play a role in this sampling. The number of points included in the sampling step, the dimensionality of the problem, the difference between the variance of the posterior and the new measurement, and, the amount of correlation between measurement and quantity of interest. The bias during the sampling step was investigated by performing tests using multidimensional Gaussians. One example of such a test is shown in Figure 4.7. Here, it can be observed that the bias of the sampling approach is indeed more pronounced for smaller sample sizes. Interestingly, low correlations (associated with low variance reductions) result in slightly more bias. For low ESS, this could potentially produce false positives, therefore this prompted investigation into a bias correction which is derived in Chapter 7. Note however, that the linear method is unbiased, even at low sample sizes. This linearized version depends on the assumption that the PPD is Gaussian however, an assumption which in this experiment holds by design, but is questionable for real PPDs.

The effect of varying the different variances was investigated. There are three variances that play a role. The variance of the old posterior, on the side of the quantity of interest, the variance of the posterior where the measurement will

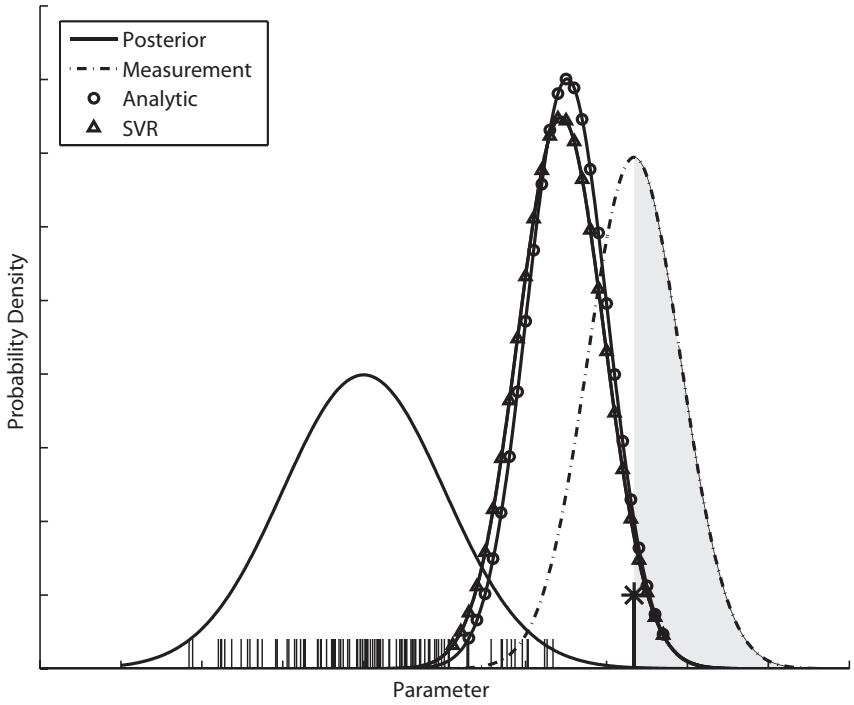

Figure 4.6: Illustration of the effect of a new datapoint on the Posterior Predictive Distribution (PPD). The stems indicate different samples obtained using MCMC. The sample currently under consideration as the new experimental value obtained from the measurement is denoted with a star. The lines correspond to the different distributions based on computed means and variances. Shown are the posterior before incorporating the new datapoint (solid line), the distribution of the measurement (dashed), the distribution based on the true mean and variance of the new posterior (circles) and the distribution based on the mean and variance estimated from resampling as performed in the proposed method (triangles)

take place, and the variance associated with the uncertainty of the new measurement. To investigate these in the non-linear case, an analysis on a 2D banana function was performed. The residual vector used in this analysis was defined as:

$$\vec{r}(\vec{x}) = \left[\sqrt{10}\left(x_2 - x_1^2\right), \sqrt{2} - x_1\right] \tag{4.17}$$

with parameters x_1 and x_2. The associated density function was defined as:

$$C(\vec{x}) = e^{\Sigma_i r_i^2} \tag{4.18}$$

Here, the variances for each sample of the MCMC chain were computed in two ways. First by means of running a new MCMC for each sample of the previous posterior. In this way, we simulate the outcome of future experiments. To test our methodology, we subsequently use the same posterior to predict these

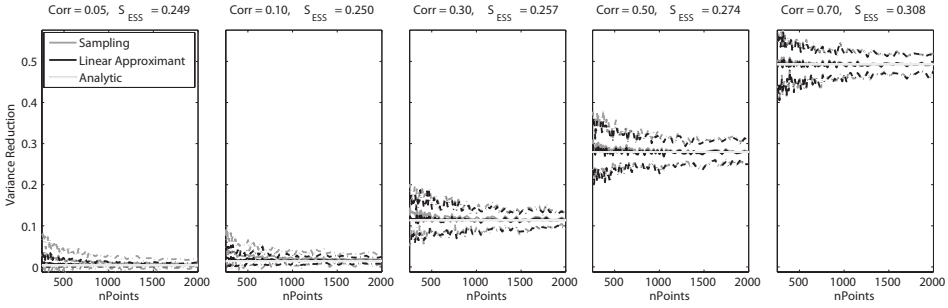

Figure 4.7: Estimated variance reduction as a function of the number of points included in the analysis. The PPD was modeled by a multivariate Gaussian distribution with output variance 5 and observation variances 4 and 3. The measurement accuracy of the new measurement was assumed to be Gaussian with variance 1. All correlation coefficients were set to the same value. Each experiment was repeated 50 times. Shown in gray, black and light gray are the mean variance reductions based on sampling, the linear approximation (which is exact for a Gaussian) and the true analytical solution. Dashed lines indicate 95 percentile bounds. Figure titles indicate used correlation and estimated slope of the Effective Sample Size as a function of the number of sample points.

variances using self normalized importance sampling. The results are shown in Figure 4.8. Although the high and low values differ, the majority of the estimated variances agree well and the mean is still well estimated (see Figure 4.9).

In this chapter, we relied on the ESS being high enough for reliable estimation. When the ESS becomes very small ($N_{ESS} < 50$), bias can occur and expected variances can be underestimated. Interestingly, this is more of an issue in the case where no correlation exists. See Chapter 7 for a bias correction for cases where a low ESS is unavoidable.

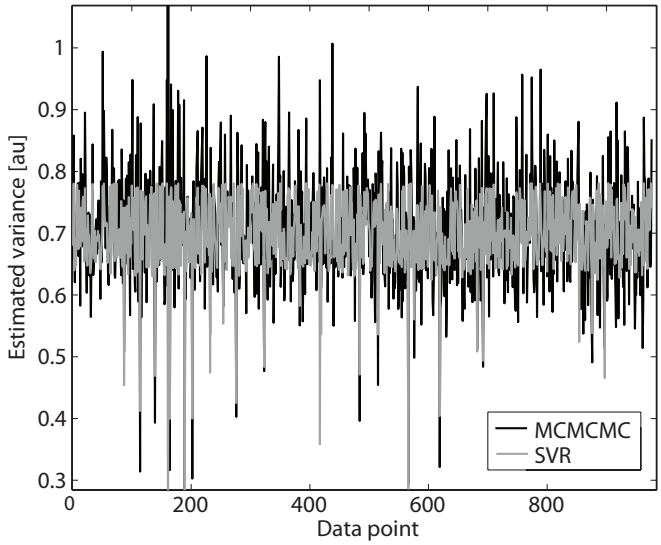

Figure 4.8: Comparison of nested MCMC approach to resampling technique for the banana function. Shown are estimates for the variance based on an MCMC for each sample of the original MCMC versus importance sampling.

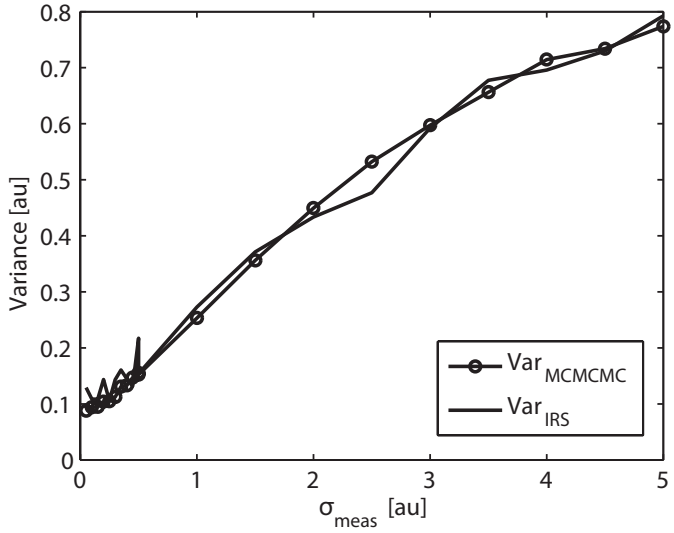

Figure 4.9: Comparison of the variances obtained using the nested MCMC approach to the resampling technique for the experiments performed on the banana function.

4.6.2 OpenCL implementation

Profiling the targeted experiment design code revealed that the biggest computational burden was the computation of distances between the particles (even in the fully vectorized case). Since this computation is identical for all particles, this could straightforwardly be outsourced to hardware designed for parallel processing. OpenCL was used to outsource these calculations to the Graphical Processing Unit (GPU), taking advantage of the parallel processing capabilities of this device. OpenCL was linked against a MEX file to run the method from MATLAB. In brief, working with OpenCL involves writing kernels, which are functions that execute on OpenCL enabled devices. These kernels are compiled at run-time by the graphics driver, which incurs additional overhead. To avoid the overhead of having to build the OpenCL code into GPU binaries at each evaluation, the binary code coming from the graphics driver is returned as an output variable to MATLAB. The resulting binary is passed as an input for subsequent calls to the function. GPU devices have different types of memory, which transfer information at different speeds. Global memory is slow, but useful for large amounts of data. Local registers are fast, but using more local memory restricts the number of threads that can be executed in parallel. In our implementation, the sequence of samples is stored in global memory, while all the computations are performed using local registers. This OpenCL implementation resulted in considerable speedups even on modest graphics hardware (see Figure 4.10).

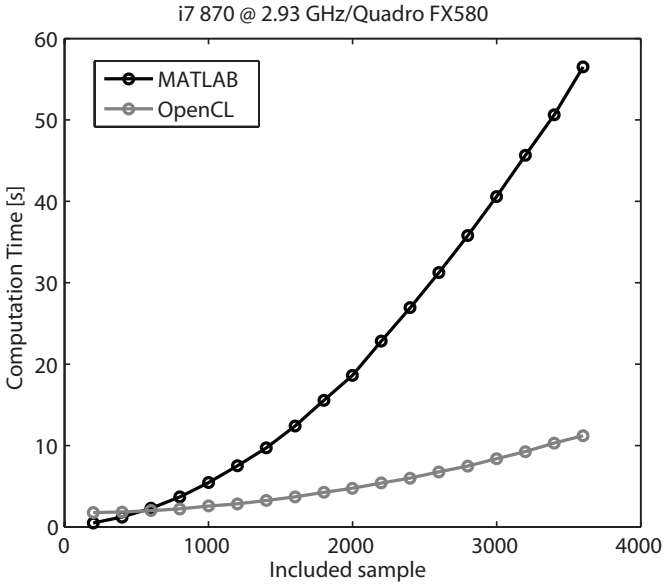

Figure 4.10: Comparison of the vectorized MATLAB implementation (black) and parallelized OpenCL implementation (gray).

References

[1] Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmüller U, Timmer J: **Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood**. *Bioinformatics* 2009, **25**(15):1923.

[2] Hasenauer J, Waldherr S, Wagner K, Allgower F: **Parameter identification, experimental design and model falsification for biological network models using semidefinite programming**. *Systems Biology, IET* 2010, **4**(2):119–130.

[3] Brännmark C, Palmér R, Glad S, Cedersund G, Strålfors P: **Mass and information feedbacks through receptor endocytosis govern insulin signaling as revealed using a parameter-free modeling framework**. *Journal of Biological Chemistry* 2010, **285**(26):20171.

[4] Girolami M, Calderhead B: **Riemann manifold langevin and hamiltonian monte carlo methods**. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2011, **73**(2):123–214.

[5] Calderhead B, Girolami M: **Statistical analysis of nonlinear dynamical systems using differential geometric sampling methods**. *Interface Focus* 2011, **1**(6):821–835.

[6] Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP: **Universally Sloppy Parameter Sensitivities in Systems Biology Models**. *PLoS Comput Biol* 2007, **3**(10):e189.

[7] Kreutz C, Raue A, Timmer J: **Likelihood based observability analysis and confidence intervals for predictions of dynamic models**. *BMC Systems Biology* 2012, **6**:120.

[8] Gomez-Cabrero D, Compte A, Tegner J: **Workflow for generating competing hypothesis from models with parameter uncertainty**. *Interface Focus* 2011, **1**(3):438.

[9] Cedersund G, Roll J: **Systems biology: model based evaluation and comparison of potential explanations for given biological data**. *FEBS Journal* 2009, **276**(4):903–922.

[10] Tiemann C, Vanlier J, Hilbers P, van Riel N: **Parameter adaptations during phenotype transitions in progressive diseases**. *BMC Systems Biology* 2011, **5**:174.

[11] Kreutz C, Timmer J: **Systems biology: experimental design**. *FEBS Journal* 2009, **276**(4):923–942.

[12] Faller D, Klingmüller U, Timmer J: **Simulation methods for optimal experimental design in systems biology**. *Simulation* 2003, **79**(12):717.

[13] Rodriguez-Fernandez M, Mendes P, Banga J: **A hybrid approach for efficient and robust parameter estimation in biochemical pathways**. *Biosystems* 2006, **83**(2-3):248–265.

[14] Casey F, Baird D, Feng Q, Gutenkunst R, Waterfall J, Myers C, Brown K, Cerione R, Sethna J: **Optimal experimental design in an epidermal growth factor receptor signalling and down-regulation model**. *Systems Biology, IET* 2007, **1**(3):190–202.

[15] Chaloner K, Verdinelli I: **Bayesian experimental design: A review**. *Statistical Science* 1995, :273–304.

[16] Klinke D: **An empirical Bayesian approach for model-based inference of cellular signaling networks**. *BMC bioinformatics* 2009, **10**:371.

[17] Geyer C: **Practical markov chain monte carlo**. *Statistical Science* 1992, :473–483.

[18] Del Moral P, Doucet A, Jasra A: **Sequential monte carlo samplers**. *Journal of the Royal Statistical Society: Series B(Statistical Methodology)* 2006, **68**(3):411–436.

[19] Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf M: **Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems**. *Journal of the Royal Society Interface* 2009, **6**(31):187–202.

[20] Swameye I, Müller T, Timmer J, Sandra O, Klingmüller U: **Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling**. *Proceedings of the National Academy of Sciences* 2003, **100**(3):1028.

[21] Hoops S, Sahle S, Gauges R, Lee C, Pahle J, Simus N, Singhal M, Xu L, Mendes P, Kummer U: **COPASIa complex pathway simulator**. *Bioinformatics* 2006, **22**(24):3067.

[22] Liepe J, Barnes C, Cule E, Erguler K, Kirk P, Toni T, Stumpf M: **ABC-SysBio approximate Bayesian computation in Python with GPU support**. *Bioinformatics* 2010, **26**(14):1797.

[23] Maiwald T, Timmer J: **Dynamical modeling and multi-experiment fitting with PottersWheel**. *Bioinformatics* 2008, **24**(18):2037–2043.

[24] Brown KS, Sethna JP: **Statistical mechanical approaches to models with many poorly known parameters**. *Phys. Rev. E* 2003, **68**(2):021904.

Optimal Experimental Design for Model Selection

5

Parts of this chapter are described in:

Vanlier, J. and Tiemann, C.A. and Hilbers, P.A.J. and van Riel, N.A.W. (submitted), *Optimal experiment design for model selection of biochemical networks*

Abstract

Mathematical modeling is often used to formalize hypotheses on how a biochemical network operates. By discriminating between competing models, different hypotheses can be compared. Bayesian model selection offers a way to determine the amount of evidence that data provides to support one model over the other while favoring simple models. In practice, the amount of experimental data is often insufficient to make a clear distinction between competing models. Often, one would like to perform a new experiment to discriminate between competing hypotheses.

This chapter presents a novel method to perform Optimal Experiment Design to predict which experiments would most effectively allow model selection. The method is based on a k-Nearest Neighbor estimate of the Jensen-Shannon divergence between the multivariate predictive densities of the competing models. The proposed method is evaluated by comparing its outcome to the change in Bayes Factor upon performing the experiments. By applying the method to a few test cases, we show that the method successfully uses predictive differences to enable model selection. Because the design criterion is based on predictive distributions and such distributions can be computed for a wide range of model quantities, the approach is very flexible. The method reveals specific combinations of experiments which improve discriminability, even in cases where data is scarce.

5.1 Introduction

Developing computational models of biochemical networks is complicated by the complexity of their interaction mechanisms [1–8]. Typically, hypotheses on how the system operates are formalized in the form of computational models [9–12]. These models are subsequently calibrated to experimental data using inferential techniques [13–19]. In the previous chapters, we focused on performing uncertainty analysis and experiment design considering only a single model. Often, many alternative models for the same system exist. Gathering sufficient data to draw a distinction between multiple competing models is a challenging task. Often, the uncertainty associated with the predictions hinders the investigator on making a clear distinction between competing models [20–22]. In such cases, additional data is required. Optimal Experiment Design (OED) methods can be used to determine which experiments would be most useful [23]. Existing methods for performing OED aimed at model selection are usually based on best parameter estimates [24–26] or model linearization [27], which are not appropriate when data is scarce with respect to model complexity [28].

In this work, we employ a Bayesian approach using the Posterior Predictive Distribution (PPD), which directly reflects the prediction uncertainty and accounts for model non-linearity and non-Gaussianity of the parameter distribution. Samples from the PPD can be obtained by drawing from the posterior parameter probability distribution and simulating predictions for each parameter set. By simulating a sample from the PPDs for all experimentally accessible moieties and fluxes, differences between models can be explored [29]. Previously, predictive distributions have been used to perform experiment design targeted at reducing the uncertainty of specific predictions [30, 31]. Here, we address the problem of ranking multiple models and present a method which uses samples from these predictive distributions to choose experiments useful for model selection.

In a Bayesian setting, model selection is typically based on the Bayes factor, which measures the amount of evidence data provides for one model over another [32, 33]. For every pair of models, a Bayes factor can be computed, defined as the ratio of their integrated likelihoods. One advantage of the Bayes factor is that it automatically penalizes unnecessary model complexity in light of the experimental data, whereby it reduces the risk of unwarranted model rejections. This penalization occurs because more parameters or unnecessarily wide priors lead to a lower weighting of the high likelihood region. This is illustrated in Figure 5.1. What the Bayesian model selection methodology does not provide, however, is a means to determine which experiments would optimally increase the separation between models.

Determining which measurements to perform in order to optimally increase the Bayes factor in favor of the correct model is a difficult task. We propose a method which allows ranking combinations of new experiments with respect to their efficacy at increasing the Bayes factor in favor of the correct model. Predictions whose distributions do not overlap between competing models are good measurement candidates [34,35]. Often distributions for a single prediction show a large degree of overlap, hampering a decisive outcome. PPDs also contain

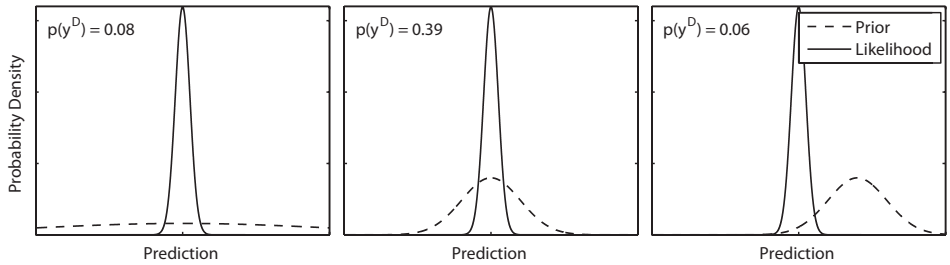

Figure 5.1: Three different examples of integrated likelihoods. Left: Integrated likelihood under wide priors. The mismatch of the prior with respect to the high likelihood region results in low weights for the high likelihood region and therefore low model evidence. This situation is comparable to a case where the model contains too many parameters. A surplus of model parameters leads to a larger parameter space and therefore lower prior probability in the high likelihood region, thus implicitly penalizing complexity not warranted by the data. Middle: A close match between prior and likelihood. Right: A model and prior which do not have sufficient freedom to describe the data very well as evidenced by the low prior probability in the high likelihood region.

information on how model predictions are related to each other. The relations between the different prediction uncertainties depend on both the data and the model. Differences in these inter-prediction relations between competing models can be probed and used (see Figure 5.2). We quantify these differences in predictive distributions by means of the Jensen-Shannon divergence (JSD).

We argue that by measuring those time points at which the models show the largest difference in their predictive distributions, large improvements in the Bayes factors can be obtained. By first applying the methodology on an analytical model it is shown that the JSD is nearly monotonically related to the predicted *change* in Bayes factor. Subsequently, the Jensen-Shannon divergence is computed between predictions of a non-linear biochemical network. Since each model implies different relations between the predictive distributions, certain combinations of predictions lead to more discriminability than others. The JSD serves as a good predictor for effective experiments when we compare it to the Bayes factors obtained after the measurements have been performed. The approach can be used to design multiple experiments simultaneously, revealing benefits that arise from combinations of experiments.

5.2 Methods

Consider biochemical networks that can be modeled using a system of ordinary differential equations. These models comprise of equations $\vec{f}(\vec{x}(t), \vec{u}(t), \vec{p})$, which contain parameters $\vec{p}$ (constant in time), inputs $\vec{u}(t)$, and state variables $\vec{x}(t)$. Given a set of parameters, inputs, and initial conditions $\vec{x}(0)$, these equations can be simulated. Measurements $\vec{y}(t)$ are performed on a subset and/or a combination of the total number of state variables in the model. Measurements are hampered by measurement noise $\vec{\epsilon}$ while many techniques used in biology

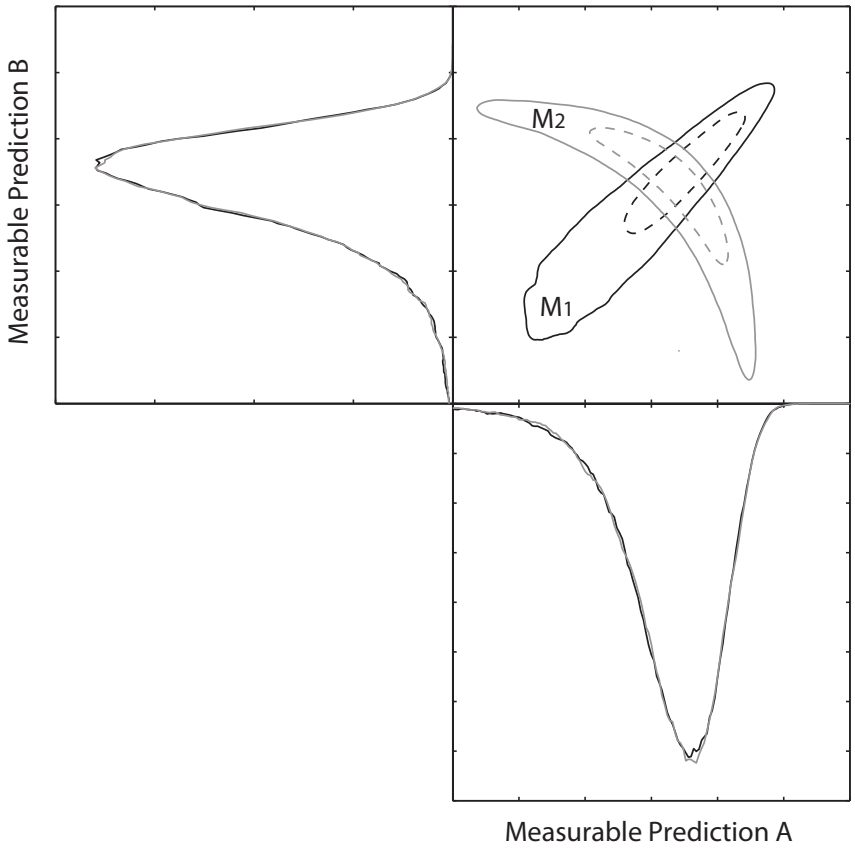

Figure 5.2: An illustrative example of how different models can imply different relations between predictions. On the top right are the 67% (dashed) and 95% (solid) probability contours of the joint probability density function of model M_1 and M_2, while the other two panels show the distribution of that specific prediction. Note how measuring one of the two predictions would yield no additional discriminatory power while measuring both predictions would.

(*e.g.* Western Blotting) necessitate the use of scaling and offset parameters $\vec{q}$ [36]. The vector $\vec{\theta}$ is defined as $\vec{\theta} = \{\vec{p}, \vec{q}, \vec{x}_0\}$, which lists all the required quantities to simulate the model. To perform inference and experiment design an error model is required. Considering R time series of length N_1, N_2 ... N_R hampered by such noise, one obtains the following equation for the probability density function of the output data:

$$p(\mathbf{y^D}|\vec{\theta}, M_i) = \prod_{k=1}^{R} \prod_{j=1}^{N_k} p(y_k^D(t_j), \vec{\theta}, M_i) \tag{5.1}$$

Here M_i indicates a model, the parameters are given by $\vec{\theta}$, while $y_k^D(t_j)$ indicates the value of a data point of state variable k at time j. By specifying prior distributions of the parameters and applying Bayes rule, it is possible to define a posterior distribution over the parameters. After checking maximum *a posteriori* identifiability, a sample from the posterior distribution of parameters can be obtained using Markov Chain Monte Carlo [22, 29, 37]. This sample reflects the uncertainty associated with the parameter values and can be used to simulate different predictions. Posterior Predictive Distributions (PPDs) are defined as distributions of new observations, conditioned on the available data. A sample from these predictive distributions can be obtained by simulating the model for each of the parameter sets drawn from the posterior parameter distribution and adding noise generated by the associated error model. The latter is required since future observations will also be affected by noise.

5.2.1 Model selection

In a Bayesian setting, model selection is often performed using the Bayes Factor [32, 38, 39]. This pivotal quantity in Bayesian model selection expresses the change of relative belief in both models after observing experimental data. By applying Bayes rule to the problem of assigning model probabilities we obtain:

$$P(M|\mathbf{y^D}) = \frac{P(\mathbf{y^D}|M)P(M)}{P(\mathbf{y^D})} \tag{5.2}$$

Here, $P(M|\mathbf{y^D})$ represents the probability of model M given observed data $\mathbf{y^D}$, while $P(M)$ and $P(\mathbf{y^D})$ are the prior probabilities of the model and data, respectively. Rather than explicitly computing the model probability, one usually considers ratios of model probabilities, allowing direct comparison between different models. As the prior model probability can be specified *a priori* (equal if no preference is given), the only quantity that still requires evaluation is $P(\mathbf{y^D}|M)$, which can be obtained by integrating the likelihood function over the parameters:

$$p(\mathbf{y^D}|M) = \int_\Omega p(\mathbf{y^D}|M, \vec{\theta}_M) p(\vec{\theta}_M|M) d\vec{\theta}_M \tag{5.3}$$

The Bayes factor is actually the ratio of these integrated (also named marginal or marginalized) likelihoods and is defined as:

$$B_{12} = \frac{p(\mathbf{y^D}|M_1)}{p(\mathbf{y^D}|M_2)} = \frac{\int_\Omega p(\mathbf{y^D}|M_1, \vec{\theta}_{M_1}) p(\vec{\theta}_{M_1}|M_1) d\vec{\theta}_{M_1}}{\int_\Omega p(\mathbf{y^D}|M_2, \vec{\theta}_{M_2}) p(\vec{\theta}_{M_2}|M_2) d\vec{\theta}_{M_2}} \tag{5.4}$$

where M_1 and M_2 refer to the different models under consideration. One advantage of the Bayes factor is that it automatically penalizes unnecessary model complexity in light of the experimental data (hereby reducing the risk of unwarranted model rejections). This penalization occurs because more parameters, or unnecessarily wide priors, lead to a relatively lower weighting of the high likelihood region, and therefore a lower value for the integrated likelihood. This is illustrated in Figure 5.1.

Bounds can be defined where the Bayes factor value becomes decisive for one model over the other. Typically, a ratio of $100:1$ is considered decisive [32, 40]. Computing the required marginal likelihoods is challenging for non-linear problems where asymptotic approximations to the posterior distribution are not appropriate. Here, one is forced to use more advanced methods such as thermodynamic integration (see Appendix 5.6.1) or annealed importance sampling [32]. Though the Bayes factor is indeed a useful method for model selection, determining what to measure in order to improve the Bayes factor in favor of the correct model is a non-trivial problem. As such, it provides a means to perform model selection, but not experiment design.

5.2.2 Experimental design

Our approach is based on selecting measurements which provide the largest discriminatory power between competing models in terms of their predictive distributions and thereby maximally inform our model comparison. This discriminatory power is quantified by means of the Jensen-Shannon divergence (JSD), as it provides a measure of dissimilarity between the probability density functions. It is defined as the averaged Kullback-Leibler divergence between probability distributions and their mixture:

$$D_{JS} = \sum_{i=1}^{K} p(M_i) D_{KL}\left(p(y|M_i), \sum_{i=1}^{K} p(M_i) p(y|M_i)\right) \tag{5.5}$$

Here, K represents the number of probability densities, $p(M_i)$ the (prior) probability of model M_i and $p(y|M_i)$ the PPD of model M_i. This metric is monotonically related to an upper and lower bound of the classification error rate in clustering problems [41, 42] and is bounded between 0 and 1. In the case where the model that generated the data is in the set of competing models, it is analogous to the mutual information between a new measurement (or sample) coming from a mixture of the candidate models and a model classifier (see Appendix 5.6.2). Mutual information has been considered before in the context of experimental design for constraining predictions or parameters of interest [30], but not in the setting of model selection. Though appealing for its properties, estimating

the Jensen-Shannon divergence for one or more experiments requires integration over the predictive densities since:

$$D_{KL}(P,Q) = \int_X p(x) \log_2 \left(\frac{p(x)}{q(x)} \right) dx \tag{5.6}$$

Here P and Q are defined as random variables with p and q their associated densities. Considering that only a sample of the PPDs is available, it is required to obtain a density estimate suitable for integration. Density estimation can be approached in two ways: by Kernel Density Estimation (KDE), or by k-Nearest Neighbor (kNN) density estimation. In Kernel Density Estimation (KDE), an estimate of the density is made by centering a normalized kernel with bandwidth h on each sample and computing the weighted average:

$$p(\vec{\theta}) = \frac{1}{Nh} \sum_{i=1}^{N} K \left(\frac{\vec{\theta} - \vec{\theta}_i}{h} \right) \tag{5.7}$$

This process subsequently provides a density for all the samples with which the relevant computations can be performed. These kernels typically have a bandwidth parameter (here denoted by h) which is estimated by means of cross validation [43, 44]. A large h will result in a loss of resolution, while a small h results in larger variances. For well behaved, low dimensional distributions with fairly uniform density, KDE is often employed and performs well. Considering the strongly non-linear nature of both the parameter and predictive distributions, a Gaussian kernel with constant covariance is not appropriate. As the number of dimensions of the problem increase, more and more weights in the KDE become small and estimation accuracy is negatively affected [45]. Additionally, cross-validation is a computationally expensive procedure to perform for each experimental candidate.

With k-Nearest Neighbor (kNN) density estimation, density is estimated by computing the volume required to include the k nearest neighbors of the current sample [44–46]:

$$p(\vec{\theta}) = \frac{1}{N} \frac{k}{\rho_k(\vec{\theta})^d v_d} \tag{5.8}$$

In this equation $\rho_k(\vec{\theta})$ represents the distance to the k^{th} nearest neighbor, d the number of dimensions and v_d the volume of the unit ball in $\mathbb{R}^d$. Furthermore, N denotes the number of included samples and v_d is given by:

$$v_d = \frac{\pi^{d/2}}{\Gamma(d/2+1)} \tag{5.9}$$

Here, Γ corresponds to the Gamma function. The advantage of using the kNN estimate is that this estimator adapts to the (local) sampling density, adjusting its volume where sampling is sparse (see Appendix 5.6.3). Consider $\vec{y}_j^{M_i}$, a vector of predictions simulated with model M_i and parameter set $\vec{\theta}_j$, where each element of the vector corresponds to a different model prediction. A model prediction is

defined as a quantity which can be computed by supplying model M_i with parameter set $\vec{\theta}_j$ (*e.g.* a predicted value at a certain time point, a difference between predictions or an area under some predicted curve). As these predictions will be considered as potential measurements, these should typically be quantities that could potentially be measured directly. The set of these vectors of predicted values coming from model M_i shall be referred to as Ω_{M_i}. Inserting these quantities, the kNN estimate of the JSD becomes:

$$\begin{aligned} D_{js} = & \frac{1}{2N_{M_1}} \sum_{i=1}^{N_{M_1}} \log_2 \left(\frac{2N_{M_2} r_k \left(\vec{y}_i^{M_1}, \Omega^{M_2} \right)^d}{N_{M_2} r_k \left(\vec{y}_i^{M_1}, \Omega^{M_2} \right)^d + (N_{M_1} - 1) r_k \left(\vec{y}_i^{M_1}, \Omega^{M_1} \backslash \vec{y}_i^{M_1} \right)^d} \right) \\ + & \frac{1}{2N_{M_2}} \sum_{i=1}^{N_{M_2}} \log_2 \left(\frac{2N_{M_1} r_k \left(\vec{y}_i^{M_2}, \Omega^{M_1} \right)^d}{N_{M_1} r_k \left(\vec{y}_i^{M_2}, \Omega^{M_1} \right)^d + (N_{M_2} - 1) r_k \left(\vec{y}_i^{M_2}, \Omega^{M_2} \backslash \vec{y}_i^{M_2} \right)^d} \right) \end{aligned} \tag{5.10}$$

Here, d corresponds to the number of elements in $\vec{y}_j^{M_i}$ (the number of predictions included), and $r_k \left(x_i, \Omega^{M_j} \right)$ corresponds to the Euclidean distance to the k^{th} nearest neighbor of x_i in Ω_{M_j}. Note that the backslash indicates excluding an element from the set. This equation provides the JSD for *one* combination of experiments. Using this equation, the JSD can then be computed for all possible combinations of experiments by assembling different vectors of predictions $\vec{y}_j^{M_i}$. These JSD estimates can then be used to rank how well these experiments would discriminate between the models. A larger value for the JSD indicates a more informative experiment. The final step involves sampling several combinations of measurements and determining the set of experiments which have the largest JSD. The complete methodology is depicted in Figure 5.3.

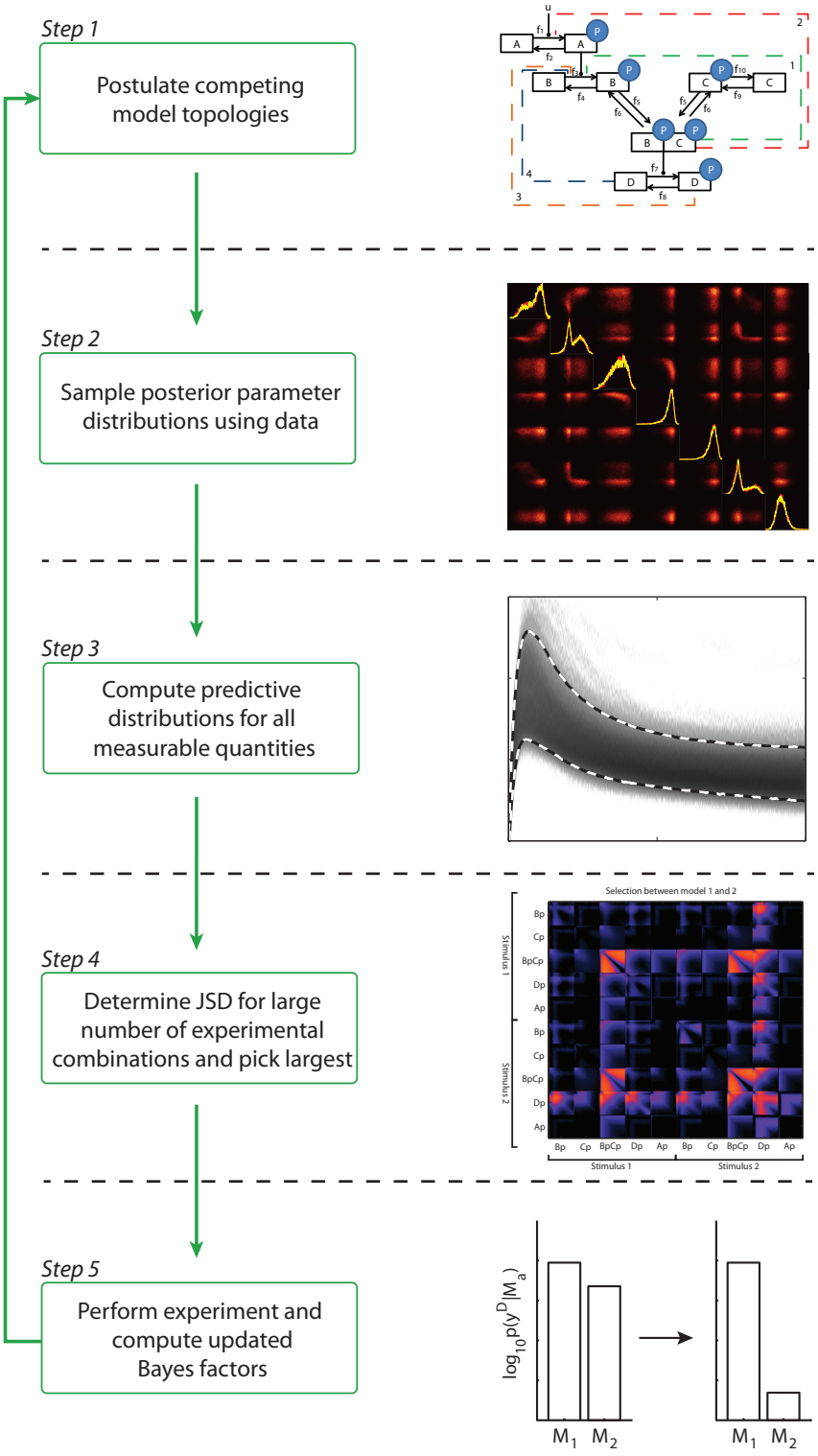

Figure 5.3: Different steps of the proposed methodology.

5.2.3 Numerical experiments

To show the applicability of the method, a series of simulation studies are performed. Since in this case, it is known by design which model generated the data, it is possible to compare to the Bayes factor pointing to the correct model. After generating an initial data set using the true model, the PPDs for each of the competing models are sampled. Subsequently, these predictions are used to compute JSD estimates between the different models. To test whether the JSD estimate can be used to compare different potential experiments, the new experimental data is subsequently included and the JSD compared to the change in Bayes factor in favor of the correct model. A large change in Bayes factor indicates a useful experiment.

Analytical models

First, we apply the method to a number of linear regression models. Linear regression models are models of the form:

$$y(t) = \sum_{i=1}^{L} \theta_i B_i(t) + \epsilon \tag{5.11}$$

Here, $\vec{\theta}$ represents a parameter vector and $\mathbf{B}$ constitutes a design matrix with basis functions $B_i(t)$. Since these models are linear in the parameters, this allows evaluation of an analytical solution. Given that σ is known, the prior distribution over the parameters is a Gaussian with standard deviation ξ, the mean and covariance matrix of the posterior distribution are given by:

$$\begin{aligned} \mu &= \left(\mathbf{B}^T\mathbf{B} + \frac{\sigma^2}{\xi^2}\mathbf{I} \right)^{-1} \mathbf{B}^T\mathbf{y} \\ \mathbf{\Sigma} &= \sigma^2 \left(\mathbf{B}^T\mathbf{B} + \frac{\sigma^2}{\xi^2}\mathbf{I} \right)^{-1} \end{aligned} \tag{5.12}$$

Furthermore, the marginal likelihood $p(y|M)$ used to compute the Bayes factor can be computed analytically as:

$$(2\pi)^{-\frac{m}{2}} |\mathbf{\Omega}|^{-\frac{1}{2}} e^{-\frac{1}{2}\mathbf{y}^T\mathbf{\Omega}^{-1}\mathbf{y}} \tag{5.13}$$

with

$$\mathbf{\Omega} = \sigma^2\mathbf{I} + \xi^2\mathbf{B}\mathbf{B}^T \tag{5.14}$$

Using linear models avoids the difficult numerical integration commonly required to compute the Bayes factor and makes it possible to perform overarching Monte Carlo studies on how these Bayes factors adjust upon including new experimental data. The analytical expressions make it possible to compare the JSD to distributions of the actual Bayes factors for model selection.

5.2.4 Non-linear biochemical networks

A series of artificial models based on motifs often observed in signaling systems [47, 48] were implemented (see Appendix 5.6.4 for the model equations). Artificial data was simulated for M_1 and subsequently inference was performed for all four competing topologies. The difference between each of the models was the origin and point of action of the feedback mechanism (see Figure 5.4).

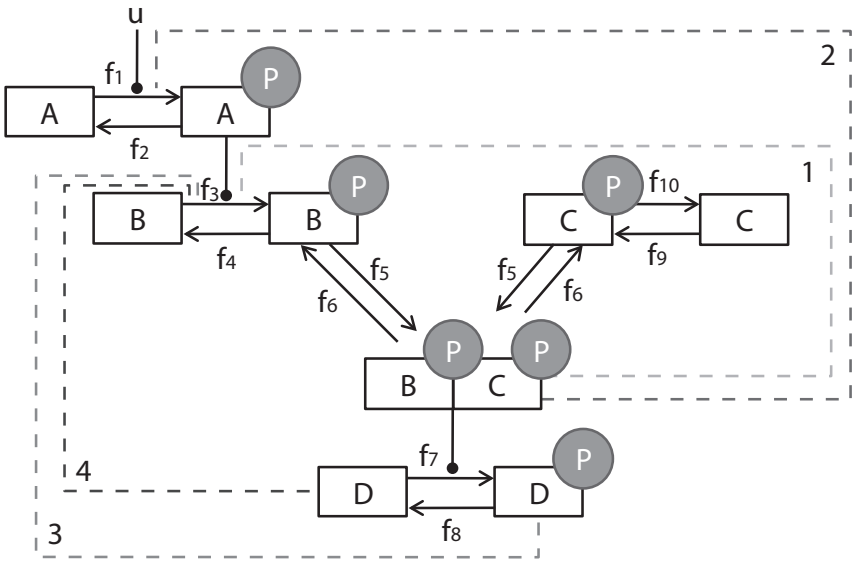

Figure 5.4: Artificial models used to test the method. The dashed lines indicate the different hypotheses regarding the negative feedback mechanisms in each of the models. Here feedback 1 corresponds to the true data generating model. Data of Bp and Dp was used for inference. Here p refers to a phosphorylated species.

Each of the artificial models was able to describe the measured data to an acceptable degree. For the prior distributions on the parameters, a Gamma distribution with $\alpha = 1$ and $\beta = 3$ was used. This prior is relatively non-informative (allowing a large range of parameter values) while not being so vague that the simplest model is always preferred (Lindley's paradox). The true data was simulated using M_1. Observables were Bp, of which three replicates were measured, and Dp, of which two replicates were measured. These were measured at $t = [0, 2, 5, 10, 20, 40, 60, 100]$. All replicates were simulated with additive Gaussian white noise with a standard deviation of 0.03. The parameter values corresponding to the true system were obtained by running Monte Carlo simulations until a visible overshoot above the noise level was observed. Parameter inference was performed using population MCMC with the noise variance σ^2 as a free parameter. Bayes factors were computed using thermodynamic integration (see Appendix 5.6.1).

5.3 Implementational details

All algorithms were implemented in MATLAB (Natick, MA). Numerical integration of the differential equations was performed with compiled MEX files using numerical integrators from the SUNDIALS CVode package (Lawrence Livermore National Laboratory, Livermore, CA). Absolute and relative tolerances were set to 10^{-8} and 10^{-9}. MCMC was performed using a population MCMC approach using $N_T = 40$ chains with a temperature schedule given by $T_n = \left(\frac{N_T}{n}\right)^4$. This permitted usage of thermodynamic integration to compute the Bayes factors between the non-linear models. The Gaussian proposal distribution for the MCMC was based on a Hessian approximation using a Jacobian obtained by simulating the sensitivity equations. After convergence, the chain was thinned to 10000 samples. Since the number of experiments designed simultaneously (and therefore the number of elements of each prediction vector) was reasonably small ($N_{samples} >> 2^k$), the kNN search was performed using k-d trees. The figures in this paper were determined using $k = 10$.

5.4 Results

5.4.1 Analytically tractable models A

A series of experiments were performed using linear regression models. To demonstrate the method, consider the following four competing models, where model three is used to generate the data:

$$\begin{aligned} y_{M_1} &= \theta_1 t \\ y_{M_2} &= \theta_1 t + \theta_2 t^2 \\ y_{M_3} &= \theta_1 t + \theta_2 t^2 + \theta_3 sin\left(\frac{1}{5}t^3\right) \\ y_{M_4} &= \theta_1 t + \theta_2 t^2 + \theta_3 sin\left(\frac{1}{5}t^3\right) + \theta_4 sin\left(2t\right) t \end{aligned} \tag{5.15}$$

The presence of sine waves in M_3 and M_4 elicits particularly noticeable patterns in the optimal experiment design matrices. Using M_3, D equidistantly sampled time points were generated as data (including Gaussian additive noise σ). To make sure that the model selection was unsuccessful *a priori*, these were sampled in a region where the models roughly predict the same behavior. Initially, the Bayes factors were $\log_{10}(B_{13}) = 2.0439$ (decisive), $\log_{10}(B_{23}) = -0.0554$ (pointing to the wrong model) and $\log_{10}(B_{43}) = 0.4658$ (not worth more than a bare mention). PPDs were generated for each of the models and used to compute credible predictive intervals that enclose 95% of the predictive density. The aim of the design is to successfully select between the models after performing new experiments. Since the outcome of the experiment is not known *a priori*, samples from predictive distributions are used to compute an expected change in Bayes factor. For each sample of the predictive distribution, the change in Bayes factor

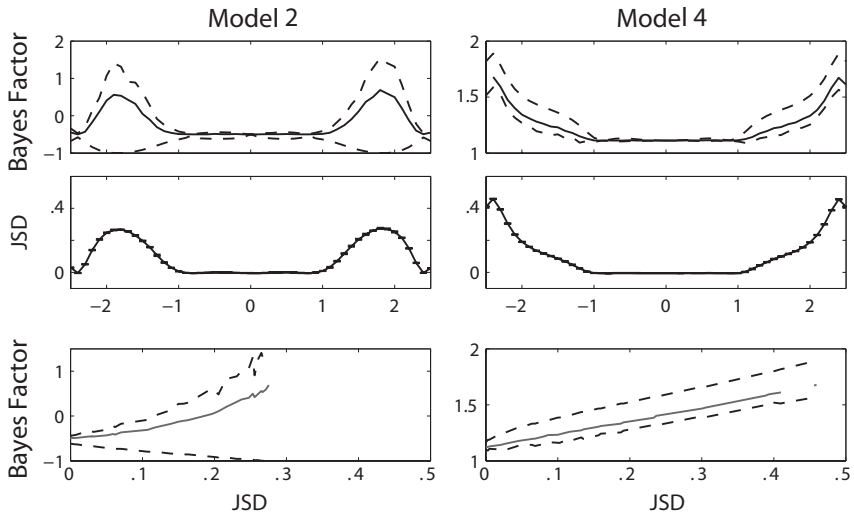

Figure 5.5: Comparing the *Jensen-Shannon Divergence* to the Bayes factor updates. Top row: Bayes factor change in favor of the correct model with associated credible intervals for two predictions. Second row: *Jensen-Shannon divergence* between the relevant Posterior Predictive Distributions. Bottom row: Relation between the Bayes factor updates and the JSD.

in favor of the true underlying model is computed. Hence this procedure results in a distribution of Bayes factors.

$$\Delta(B_{ab}) := \mathbb{E}_{y_n^D}\left[\log_{10}\left(\frac{p(\mathbf{y^D}, \mathbf{y_n^D}|M_a)}{p(\mathbf{y^D}, \mathbf{y_n^D}|M_b)}\right)\right] - \log_{10}\left(\frac{p(\mathbf{y^D}|M_a)}{p(\mathbf{y^D}|M_b)}\right) \tag{5.16}$$

Here, the expectation is taken with respect to new realizations of the data y_n^D. Predicted experimental data are simulated in two ways. Either by using the correct model with the true parameter values and adding measurement noise, which shall be referred to as ΔB_{ab}^T. Or by generating samples from the posterior predictive distribution of the correct model (ΔB_{ab}^B), where the B stands for current state of belief. The change in Bayes factor (in favor of the correct model) was compared to the Jensen-Shannon divergence between the posterior predictive distributions of competing models. Large predicted changes indicate that the experiment would result in a successful selection. As for the JSD, a large value indicates a large divergence between the joint predictive distributions, marking the measurement as useful. See Figure 5.6 for an example of the analysis results.

The JSD agrees well with the actual Bayes factor updates when considering the current state of belief B_{ab}^B as shown in Figure 5.6. Interestingly, all the designs with a high JSD are effective for discriminating between the true model and its competitors without having to specify a true model *a priori*. The patterns that arise in the matrix of combinatorial efficacies are different depending on how the predictive samples were generated. In reality, the parameter values of the true model are not known. If the uncertainty currently associated with the

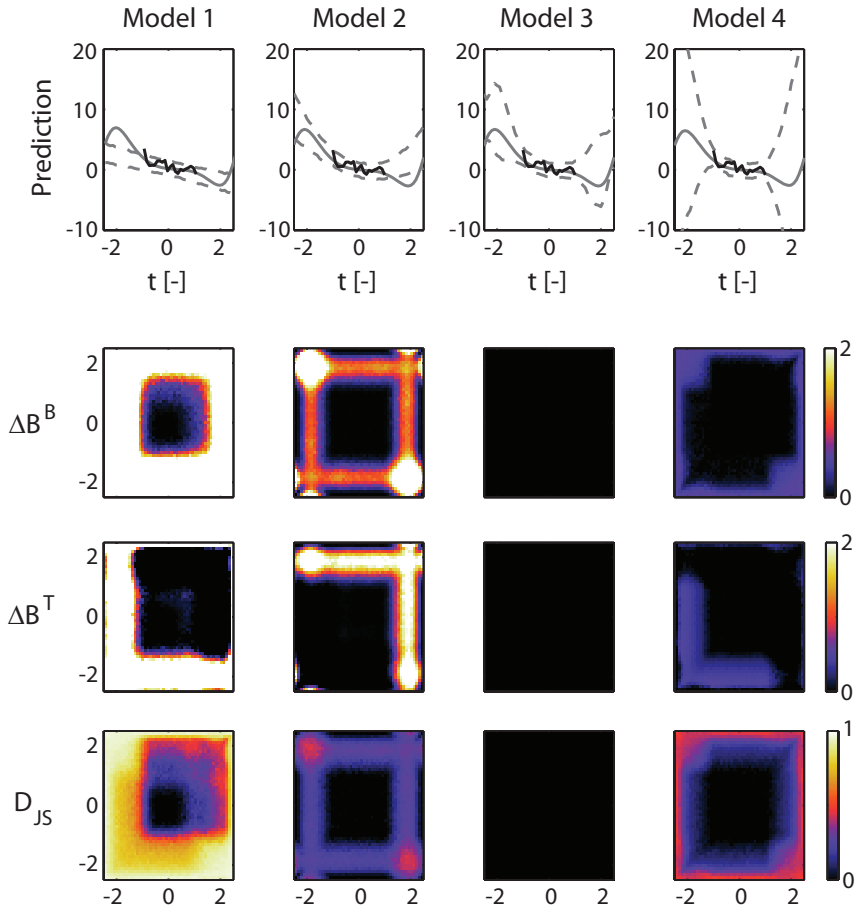

Figure 5.6: Comparing the *Jensen-Shannon Divergence* to the Bayes factor updates for regression models. Top row: Blue lines indicate the 'true' system response. Dashed lines indicate the credible intervals of the Posterior Predictive Distribution for that specific model. The red line indicates the data incorporated prior to the analysis. Second row: Bayes factor change for the support over model 3 over the model corresponding to that column after incorporating two datapoints simulated from the posterior distribution of model true (average of 100 repetitions of this simulated experiment). Third row: Bayes factor change for the support over model 3 over the model corresponding to that column after incorporating two datapoints simulated with the model true with the true parameters (average of 100 repetitions of this simulated experiment). Bottom row: Jensen-Shannon Divergence between the Posterior Predictive Distributions of Model 3 and the model that column number corresponds to. Note that the graphs for model 3 are black by definition.

parameters is taken into account (by simulating from the PPD rather than true model plus noise), then the efficacy matrix seen in the second row of Figure 5.6 is obtained. Knowing the true parameter value breaks the symmetry (see the third row of Figure 5.6). This makes sense, since the simulation corresponding to the true parameters falls within all the PPDs for negative t, and outside of the PPDs for positive t. In practice, the true parameter values are not known and the estimate based on the posterior samples provides the best possible estimate that can be attained considering the current parameter uncertainty. Plotting the relationship between the updated Bayes factors B^B_{ab} upon a new experiment and the corresponding JSD typically reveals a monotonic relationship that underlines its usefulness as a design criterion (see Figure 5.5 for two typical examples).

5.4.2 Analytically tractable models B

Subsequently, a Monte Carlo study was performed where a large number of models were randomly generated from basis functions and compared. Consider linear models constructed from the following basis functions:

$$\begin{aligned} c_1(t) &= t \\ c_2(t) &= (.25t)^2 \\ c_3(t) &= (.1t)^3 \\ c_4(t) &= \sin(\alpha t^2) \\ c_5(t) &= e^{-|\beta t^2|} \\ c_6(t) &= \sin(\gamma t^3) \\ \alpha &\propto U[.1, .6] \\ \beta &\propto U[.05, .1] \\ \gamma &\propto U[.1, .6] \end{aligned} \tag{5.17}$$

Here $U[a, b]$ indicates a sample from a uniform distribution between a and b which is drawn upon basis function construction. Regression models are assembled from these components by shuffling these basis functions in a random order. The reordered basis functions are referred to as C_i. The true model contains the first three reordered basis functions, while the different competing models are defined as follows:

$$\begin{aligned} M_1(t) &= \theta_1 C_1(t) \\ M_2(t) &= \theta_1 C_1(t) + \theta_2 C_2(t) \\ M_3(t) &= \theta_1 C_1(t) + \theta_2 C_2(t) + \theta_3 C_3(t) \\ M_4(t) &= \theta_1 C_1(t) + \theta_2 C_2(t) + \theta_3 C_3(t) + \theta_4 C_4(t) \\ M_5(t) &= \theta_1 C_1(t) + \theta_2 C_2(t) + \theta_3 C_5(t) \\ M_6(t) &= \theta_1 C_1(t) + \theta_2 C_3(t) + \theta_3 C_5(t) \end{aligned} \tag{5.18}$$

Note that model four is an over-parameterized version of the model that generated the data, while the last two models are wrong. The new measurement is computed by sampling from the true model posterior predictive and adding

Comparison	Spearman CC
M_3 over M_1	$.90 \pm .22$
M_3 over M_2	$.93 \pm .12$
M_3 over M_4	$.88 \pm .11$
M_3 over M_5	$.91 \pm .10$
M_3 over M_6	$.94 \pm .04$

Table 5.1: Spearman correlation coefficients between expected Bayes factor and JSD based on 75 random models.

measurement noise. The mean of the resulting distribution of Bayes factors is computed for every potential measurement. To compare with, JSDs are computed for the same predictions.

The results reported here were based on 100 random models. Data at time points $t_{old} = [-1, -.8, \ldots, .8, 1]$ was used for initial inference. The new potential measurements were located at $t_{new} = [-2.5, -2.3, \ldots, 2.3, 2.5]$. The true parameters were drawn from a normal distribution with a standard deviation of 2. The standard deviation of the noise was uniformly sampled between .1 and 3.1, while the standard deviation of the noise on the new measurement was set to .3. The prior distribution on the parameters had width 10. The expected Bayes factors were based on 100 samples from the posterior distribution. Computing the Spearman correlation coefficient (a measure of monotonicity) between the expected Bayes factor and JSD resulted in high average correlation coefficients (see Table 5.1). This provides an additional indication that the JSD can serve as a good predictor for the increase in Bayes factor.

5.4.3 Nonlinear artificial models

The predictive distributions for the different models are shown in Figure 5.7. PPDs were simulated for two experimental conditions. One set where the stimulus u was set to the value 1, and another where u was set to 2. These sets mimic two different concentrations of signaling molecule.

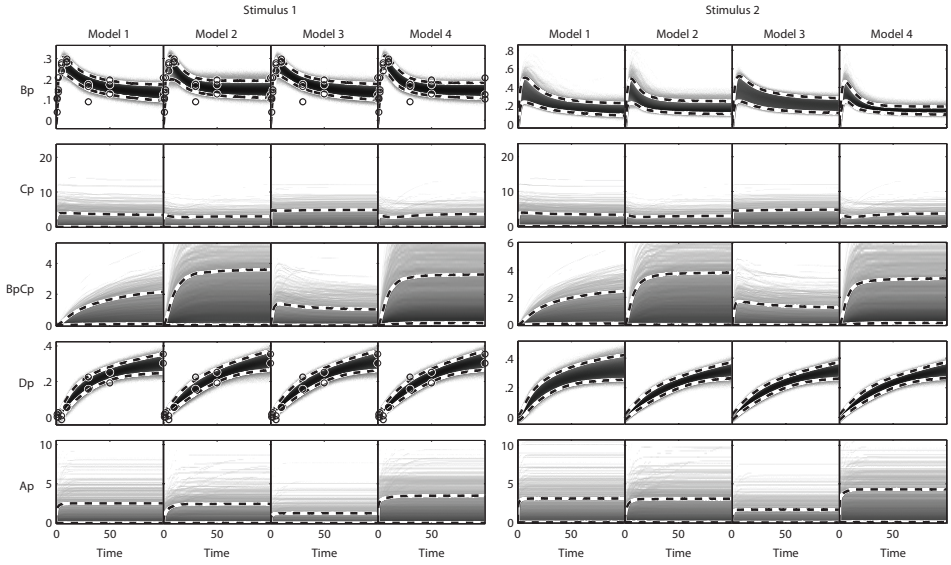

Figure 5.7: Predictions of the various nonlinear ODE models. The first five predictions correspond to the same experimental condition as during the original inference (stimulus 1) while the second five predictions correspond to a different stimulus (stimulus 2). Note that the differences between the different distributions are barely visible.

To test the effect of measuring multiple predictions, divergence estimates were computed for a large number of combinations of two measurements. The results are shown in Figure 5.8. Note the bright squares corresponding to the concentration of $BpCp$ in each of the models. These high efficacies are not surprising, considering that the PPDs for these concentrations are very different for the different models (See Figure 5.7). Also noticeable is that many of the experiments on the same predictions reveal dark diagonals within each tile. Measuring the same thing twice usually adds fewer predictive constraints than measuring at two different time points, unless this second measurement is performed in a different condition (see how the diagonal lights up on the combination measuring $BpCp$ in condition 1 and 2 when selecting between models 3 and 4). The information contained in such a matrix is very valuable when it comes to selecting from a small list of experiments. For example, considering the current predictive distributions, model 2 and 4 can barely be distinguished. This implies that an entirely different experiment is required for distinguishing models 2 and 4.

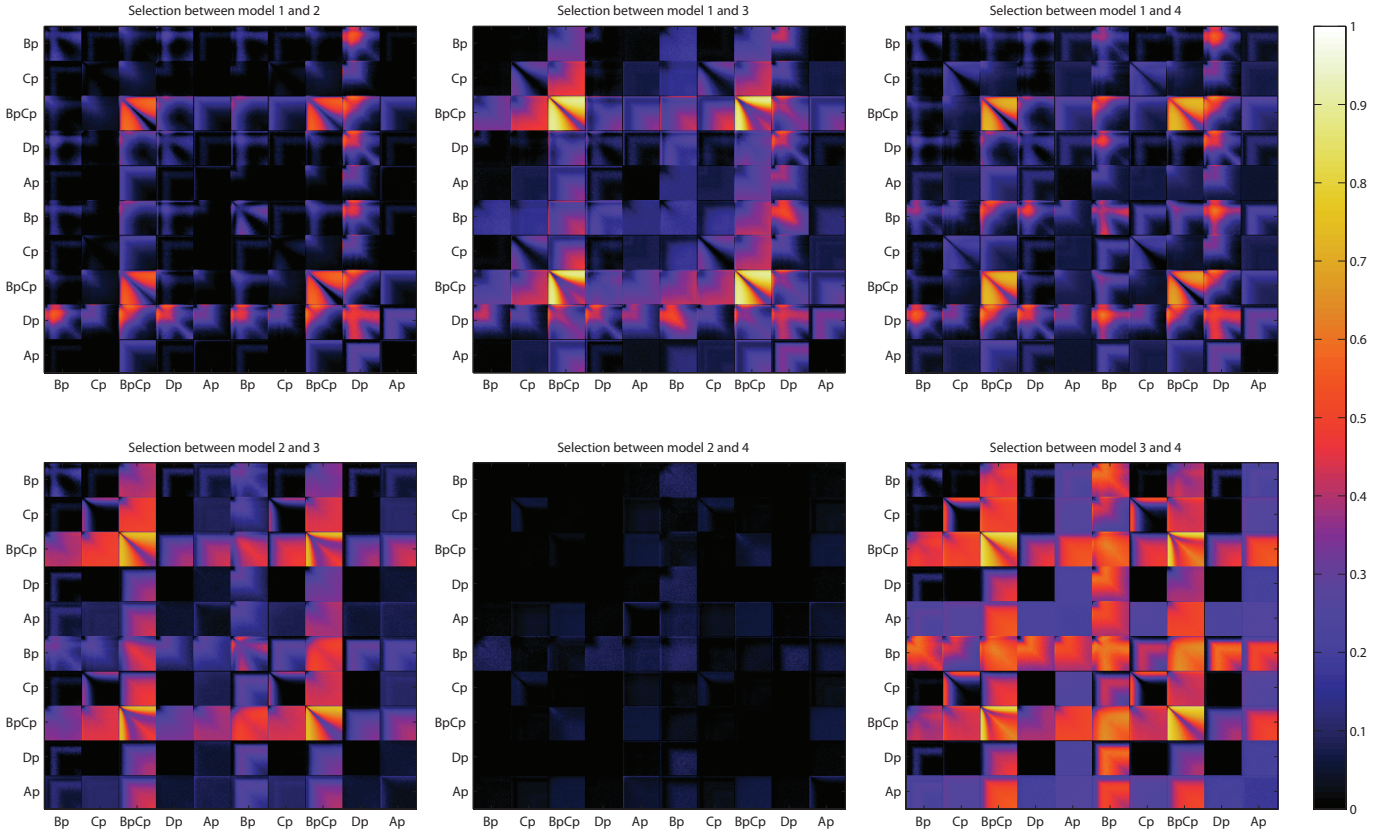

Figure 5.8: Jensen-Shannon divergences for each of the models. Each axis represents a single measurement. Each tile corresponds to a combination of two state variables where the space within each tile corresponds to the actual time point at which the state variable is measured. The first five predictions correspond to the same experimental condition as during the original inference ($u = 1$) while the second five predictions correspond to a different stimulus ($u = 2$).

Subsequently, *in silico* experiments were performed by simulating new data from the artificial true model and determining the Bayes factor change upon including this data. Bayes factors were estimated using thermodynamic integration (see Appendix 5.6.1 for details). For the models considered here, each set of four marginal likelihoods took about 6 days of wall-clock time on an Intel i7 CPU (2.93 GHz) with MATLAB R2010a. The following 'experiments' were performed

- 1. Steady state Cp and $BpCp$ concentration
- 2. Bp and Dp during the peak in the second condition ($u = 2$)
- 3. Steady state Cp

According to the JSD estimates, experiment 1 should differentiate between M_1 and M_3 (JSD ≈ 0.49), but not between M_1 and M_2 or M_1 and M_4. Experiment 2 should give discriminatory power for all models ($J_{12} \approx 0.48, J_{13} \approx 0.54, J_{14} \approx 0.61$). And finally, experiment 3 should not provide any additional discriminatory power at all. The results of these analyses are shown in Table 5.2. As predicted, experiment 1 leads only to an increase in discriminatory power between model M_1 and M_3. Experiment 2 improves the discriminatory power between all the models while experiment 3 even reveals a decrease in discriminatory power for model 1 and 2. Noteworthy is the large variance observed for experiment 3, which is likely related to the large variance in the steady state predictions of Cp. Again, the predictions based on the JSD are well in line with the Bayes factors obtained.

J_{12}	ΔB_{12}	J_{13}	ΔB_{13}	J_{14}	ΔB_{14}
0.03	0.06 ± 0.19	0.49	0.32 ± 0.39	0.05	-0.07 ± 0.36
0.48	0.26 ± 0.14	0.54	0.72 ± 0.36	0.61	0.43 ± 0.38
-0.06	-0.49 ± 0.73	-0.01	-0.35 ± 0.68	-0.04	0.32 ± 0.54

Table 5.2: JSD and change in Bayes Factors denoted as mean $\pm$ standard deviation for each of the reported experiments (n=3)

5.5 Discussion

This chapter described a method applicable to performing experiment design with the aim of differentiating between various pathway hypotheses. We show by means of a simulation study on analytically tractable models that the JSD is approximately monotonically related to the expected change in Bayes factor in favor of the model that generated the data (considering the current uncertainty in its parameters). The applicability to non-linear models of biochemical reaction networks was demonstrated by applying it to models based on motifs previously observed in signaling networks [47, 48]. Experiments were designed for distinguishing between different feedback mechanisms.

Though forecasting a predictive distribution of Bayes factors has been suggested [49], the implicit penalization of model complexity could have adverse consequences. The experiment design could suggest a measurement where the probability densities of two models overlap, leading to implicitly penalizing the more complex model followed by subsequent selection (of the simpler model). Though a successful selection occurs, such an experiment would not provide additional insight however. Additionally, computing the predictive distributions of Bayes factors required for this approach is computationally intractable. By focusing on differences in predictive distributions, it is possible to pinpoint where the different models predict different states of nature. Aside from their usefulness in model selection, such predictive differences could also be attributed to the different mechanisms present in the model. This allows for follow-up studies to investigate whether these are either artefactual or true system behaviour.

A complicating factor in this method is the computational burden. The largest challenge to overcome is to obtain a sample from the posterior parameter distribution. Running MCMC on high dimensional problems can be difficult. Fortunately, recent advances in both MCMC [19, 50] as well as approximate sampling techniques [38, 51–53] allow sampling parameter distributions of increasingly complex models. The bottleneck in computing the JSD resides in searching for the k^{th} nearest neighbor. A subproblem which occurs in many different situations and for which computationally faster solutions exist [54, 55]. Additionally, the number of potential combinations of experiments increases exponentially with the number of experiments designed. It is clear that this rapidly becomes infeasible for large numbers of experiments. However, it is not necessary to fill the entire experimental matrix and techniques such as Sequential Monte Carlo sampling could be considered as an alternative to more effectively probe this space. We refer the reader to Appendix 5.6.5 for a proof of principle implementation of such a sampler.

A more theoretical point of debate is the weighting of each of the models in the mixture distribution used to compute the JSD. It could be argued that it would be more sensible to weight models according to their model probabilities by determining the integrated likelihoods of the data that is already available before performing experiment selection. The reason for not doing this is two-fold. Firstly, the computational burden this adds to the experimental design procedure is significant. More importantly however, the implicit weighting in favor of parsimony could strongly affect the design by removing models which are considered unnecessarily complex at this stage of the analysis. When designing new experiments, the aim is to obtain measurements that allow for optimal discrimination between the predictive distributions under the different hypotheses. Optimal discrimination makes it sensible to consider the models equally probable initially.

The method has several advantages that are particularly useful for modeling biochemical networks. Because the method is based on sampling from the posterior parameter probability distribution, it is particularly suitable when insufficient data is available to consider Gaussian parameter probability distributions or model linearizations. Additionally, it allows incorporation of prior knowledge in the form of prior parameter probability distributions. This is useful when the

available data contains insufficient constraints to result in a well defined posterior parameter distribution. Because the design criterion is based on predictive distributions and such distributions can be computed for a wide range of model quantities, the approach is very flexible. In biochemical research, *in vivo* measurements are often difficult to perform and practical limitations of the various measurement technologies play an important role. In many cases, measurements on separate components cannot be performed, and measurements result in derived quantities. Fortunately, in the current framework such measurements can be used directly, since distributions of such experiments can be simulated.

Moreover, the impact of specific combinations of experiments can be assessed by including them in the design simultaneously which reveals specific combinations of measurements that are particularly useful. This way, informative experiments can be distinguished from non-informative ones and the experimental efforts can be targeted to discriminate between competing hypotheses.

5.6 Appendix

5.6.1 Thermodynamic integration

To estimate the Bayes factor, the marginal likelihood needs to be computed. For a single model this quantity looks as follows:

$$p(\mathbf{y^D}|M) = \int p(\mathbf{y^D}|M,\vec{\theta})p(\vec{\theta}|M)d\vec{\theta} \tag{5.19}$$

One approach to determine this marginal likelihood is to draw random samples $\vec{\theta}$ from the prior and average the likelihood values for those samples. However, due to the fact that most samples will be taken in regions of low likelihood, this method is highly inefficient. Another approach would be to employ importance sampling. One tempting approach here is to use the posterior density function as an importance sampling function leading to the following Monte Carlo Estimate:

$$\left(\frac{1}{N}\sum_{i=1}^{N} p(\mathbf{y^D}|M,\vec{\theta}_i)^{-1}\right)^{-1} \tag{5.20}$$

but this estimate suffers from both numerical instability as well as severe bias [56]. The alternative and more stable approach is to use *thermodynamic integration*, where the problem is recast into a problem involving integration over various intermediate distributions which are defined as follows:

$$p_T(\vec{\theta}|\mathbf{y^D}) = \frac{p(\mathbf{y^D}|M,\vec{\theta})^T p(\vec{\theta})}{z(T)} \tag{5.21}$$

with

$$z(T) = \int p(\mathbf{y^D}|M,\vec{\theta})^T p(\vec{\theta}|M)d\vec{\theta} \tag{5.22}$$

Consider the following:

$$\int_0^1 \frac{d}{dT}\ln(z(T))dT = \int_0^1 \frac{1}{z(T)}\frac{d}{dT}z(T)dT = \ln(Z(1)) - \ln(Z(0)) \tag{5.23}$$

It can be seen that $Z(1)$ corresponds to the marginal likelihood, while $Z(0)$ corresponds to the integrated prior (which is equal to 1). Consider

$$z(T) = \int e^{\ln(p(\mathbf{y^D}|M,\vec{\theta}))T} p(\vec{\theta}|M)d\vec{\theta} \tag{5.24}$$

from which it follows that

$$\ln(Z(1)) = \int_0^1 \int \frac{p(\mathbf{y^D}|M,\vec{\theta})^T p(\vec{\theta}|M)}{z(T)} \ln(p(\mathbf{y^D}|M,\vec{\theta}))d\vec{\theta}dT \tag{5.25}$$

Therefore, after obtaining samples at different intermediate distributions spaced between $T = 0$ and $T = 1$, this integration can be performed numerically by summing the expected value of the log-likelihoods at different temperatures:

$$\hat{p}(\mathbf{y}^{\mathbf{D}}|M) = \sum_{q=1}^{T_{max}} \frac{1}{N_{samples}} \sum_{i=1}^{N_{samples}} \ln\left(p(\mathbf{y}^{\mathbf{D}}|M, \vec{\theta}_i^{T_q})\right) \tag{5.26}$$

Here, T_q represents the temperature and $\vec{\theta}_i^{T_q}$ are samples associated with chain q. $N_{samples}$ indicates the number of samples per chain. The temperature schedule used in this study is given by:

$$T_q = \left(\frac{N_T}{q}\right)^{\gamma} \tag{5.27}$$

with $N_T = 40$ and $\gamma = 4$. For further information see [56].

5.6.2 Mutual information

Mutual information is a quantity that measures the mutual dependence between two random variables and is defined as:

$$I(X,Y) = \int_Y \int_X p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right) dxdy \tag{5.28}$$

It measures the amount of information shared between two random variables. A large mutual information implies that knowing the value of one of the variables reduces the uncertainty in the other. If two random variables are independent, their mutual information is zero. If entropy

$$H(X) = \int_X p(x) \log(p(x)) dx \tag{5.29}$$

is considered as a measure of the uncertainty associated with a random variable, the mutual information can be expressed as follows:

$$I(X,Y) = H(X) - H(X|Y) \tag{5.30}$$

Mutual information reflects how much the uncertainty in one random variable is reduced by knowing about another. Assuming that the models are equally probable *a priori*, averaging the predictive distribution of a new measurement over the different models corresponds to the additive mixture of their predictive densities. The Jensen-Shannon Divergence between two predictive densities can be rewritten as the Mutual Information between this mixture of densities and a model classifier:

$$\begin{aligned}
&D_{js}\left(Y|M_1, Y|M_2\right) \\
&= \sum_{m\in\{M_1,M_2\}} p(m) \int_Y p(y|M=m) \log \left(\frac{p(y|M=m)}{\sum_{m\in\{M_1,M_2\}} p(m)p(y|M=m)} \right) dy \\
&= \int_Y \sum_{m\in\{M_1,M_2\}} p(m)p(y|M=m) \log \left(\frac{p(y|M=m)}{\sum_{m\in\{M_1,M_2\}} p(m)p(y|M=m)} \right) dy \\
&= \int_Y \sum_{m\in\{M_1,M_2\}} p(y,m) \log \left(\frac{p(y,m)}{p(y)p(m)} \right) dy \\
&= I(Y,M)
\end{aligned} \tag{5.31}$$

Informally, measuring a quantity that maximizes this mutual information maximizes the reduction of uncertainty with respect to determining which of the two distributions the new measurement came from. This supports the JSD as a quantity useful for model selection.

5.6.3 Density estimation methods

Density estimation can be approached in three different ways. Either by specifying a discretization of the joint predictive distribution (binning), using kNN nearest neighbor density estimation or Kernel Density Estimation (Parzen windowing). Each of these methods contains a size or bandwidth parameter, which can be difficult to specify. For Gaussian distributions standard rules of thumb exist, but when joint density functions become more complex, these rules of thumbs are no longer appropriate. Figure 5.9 shows some estimates based on the different methods. Note how all JSD estimates strongly depend on the bandwidth parameter. This parameter is often difficult to specify, especially when predictions show different ranges. Note how the JSDs based on kNN quickly converge to a stable estimate when $k > 2$. We chose kNN as a density estimator because JSD estimates based on it are reasonably stable, but do not require choosing a problem specific bandwidth.

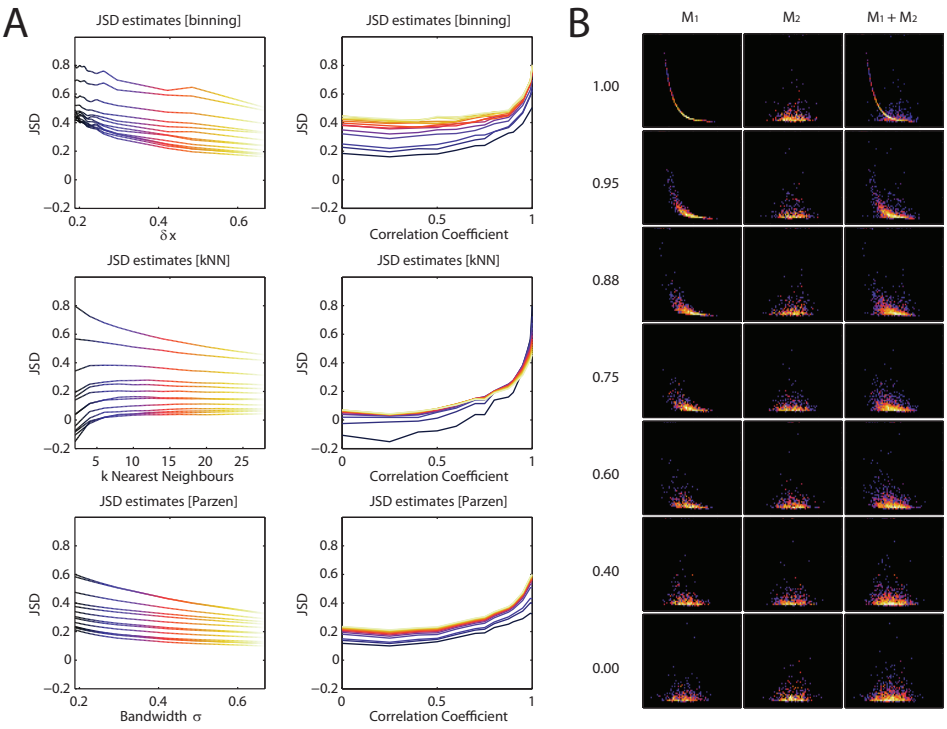

Figure 5.9: A: JSD computed using different density estimation techniques using 1000 samples. The colors in the left and right graph of panel A correspond to the same bandwidth settings. B: Different correlations used to produce the JSD estimates. These results were generated using a normal distribution with varying correlation coefficient and then transforming one of the predictions using an exponential function. Such joint probability structures resemble the ones previously observed in posterior predictive distributions of biochemical networks.

5.6.4 Model equations

Fluxes		Stoichiometry	
$f_1 =$	$k_1[A]$	$[\dot{A}] =$	$-f_1 + f_2$
$f_2 =$	$k_2[Ap]$	$[\dot{Ap}] =$	$f_1 - f_2$
$f_3 =$	$k_3 \frac{1}{k_9 + [BpCp]}[B][Ap]$	$[\dot{B}] =$	$-f_3 + f_4$
$f_4 =$	$k_4[Bp]$	$[\dot{Bp}] =$	$f_3 - f_4 - f_5 + f_6$
$f_5 =$	$k_5[Bp][Cp]$	$[\dot{C}] =$	$-f_9 + f_{10}$
$f_6 =$	$k_6[BpCp]$	$[\dot{Cp}] =$	$-f_5 + f_6 - f_{10} + f_9$
$f_7 =$	$k_7[D]$	$[\dot{BpCp}] =$	$f_5 - f_6$
$f_8 =$	$k_8[Dp]$	$[\dot{D}] =$	$-f_7 + f_8$
$f_9 =$	$k_{10}[C]$	$[\dot{Dp}] =$	$f_7 - f_8$
$f_{10} =$	$k_4[Cp]$		

Table 5.3: Expressions for Model 1

Fluxes		Stoichiometry	
$f_1 =$	$k_1 \frac{1}{k_9 + [BpCp]}[A]$	$[\dot{A}] =$	$-f_1 + f_2$
$f_2 =$	$k_2[Ap]$	$[\dot{Ap}] =$	$f_1 - f_2$
$f_3 =$	$k_3[B][Ap]$	$[\dot{B}] =$	$-f_3 + f_4$
$f_4 =$	$k_4[Bp]$	$[\dot{Bp}] =$	$f_3 - f_4 - f_5 + f_6$
$f_5 =$	$k_5[Bp][Cp]$	$[\dot{C}] =$	$-f_9 + f_{10}$
$f_6 =$	$k_6[BpCp]$	$[\dot{Cp}] =$	$-f_5 + f_6 - f_{10} + f_9$
$f_7 =$	$k_7[D]$	$[\dot{BpCp}] =$	$f_5 - f_6$
$f_8 =$	$k_8[Dp]$	$[\dot{D}] =$	$-f_7 + f_8$
$f_9 =$	$k_{10}[C]$	$[\dot{Dp}] =$	$f_7 - f_8$
$f_{10} =$	$k_4[Cp]$		

Table 5.4: Expressions for Model 2

Fluxes		Stoichiometry	
$f_1 =$	$k_1[A]$	$[\dot{A}] =$	$-f_1 + f_2$
$f_2 =$	$k_2[Ap]$	$[\dot{Ap}] =$	$f_1 - f_2$
$f_3 =$	$k_3 \frac{1}{k_9 + [Dp]}[B][Ap]$	$[\dot{B}] =$	$-f_3 + f_4$
$f_4 =$	$k_4[Bp]$	$[\dot{Bp}] =$	$f_3 - f_4 - f_5 + f_6$
$f_5 =$	$k_5[Bp][Cp]$	$[\dot{C}] =$	$-f_9 + f_{10}$
$f_6 =$	$k_6[BpCp]$	$[\dot{Cp}] =$	$-f_5 + f_6 - f_{10} + f_9$
$f_7 =$	$k_7[D]$	$[\dot{BpCp}] =$	$f_5 - f_6$
$f_8 =$	$k_8[Dp]$	$[\dot{D}] =$	$-f_7 + f_8$
$f_9 =$	$k_{10}[C]$	$[\dot{Dp}] =$	$f_7 - f_8$
$f_{10} =$	$k_4[Cp]$		

Table 5.5: Expressions for Model 3

Fluxes		Stoichiometry	
$f_1 =$	$k_1[A]$	$[\dot{A}] =$	$-f_1 + f_2$
$f_2 =$	$k_2[Ap]$	$[\dot{Ap}] =$	$f_1 - f_2$
$f_3 =$	$k_3 \frac{1}{k_9 + [D]}[B][Ap]$	$[\dot{B}] =$	$-f_3 + f_4$
$f_4 =$	$k_4[Bp]$	$[\dot{Bp}] =$	$f_3 - f_4 - f_5 + f_6$
$f_5 =$	$k_5[Bp][Cp]$	$[\dot{C}] =$	$-f_9 + f_{10}$
$f_6 =$	$k_6[BpCp]$	$[\dot{Cp}] =$	$-f_5 + f_6 - f_{10} + f_9$
$f_7 =$	$k_7[D]$	$[\dot{BpCp}] =$	$f_5 - f_6$
$f_8 =$	$k_8[Dp]$	$[\dot{D}] =$	$-f_7 + f_8$
$f_9 =$	$k_{10}[C]$	$[\dot{Dp}] =$	$f_7 - f_8$
$f_{10} =$	$k_4[Cp]$		

Table 5.6: Expressions for Model 4

5.6.5 Sampling bigger design matrices

When designing multiple experiments, each additional experiment constitutes an additional direction in experimental design space. For the 2D examples in the main text, filling this matrix entirely was not a problem. When designing a larger number of experiments, sampling this entire space rapidly becomes infeasible due to the curse of dimensionality. This is when one must resort to specialized sampling methods. As a proof of principle, we present a sampler based on ideas from Sequential Importance Resampling to sample these spaces. The sampler works by executing a number of steps

- 1. Sample Q particles from a uniform distribution over the design space. When the number of samples is reasonably small, one can use Latin hypercube sampling to improve the coverage of the design space.
- 2. Evaluate JSDs for each of the samples and clip any negative values for the JSD to zero.
- 3. Compute cumulative distribution function CDF_{js} for the JSDs and normalize to a maximum of 1.
- 4. Generate νQ samples from a uniform distribution $X_i \sim U[0,1]$. Generate particles for next iteration by sampling those particles which correspond to values where $X_i > CDF_{js}$ for the first time. Generate the other $(1-\nu)\,Q$ particles from a uniform distribution over the design space.
- 5. Perturb particles using a perturbation kernel until none of the particles correspond to samples that were already evaluated. Note that during this step, we take into account the various symmetries that exist in these matrices such that we do not compute the same JSD twice. These symmetries exist, since all permutations of the same combination of experiments correspond to the same value for the JSD (*e.g.* $J_{12} = J_{21}$).
- 6. Goto 2 until desired number of iterations has been reached.

This algorithm is straightforward to implement and results in regions with high values for the JSD to be explored preferentially. Optionally, one can factor in an experimental cost in the sampling criterion by multiplying the associated JSD before computing the CDF. Note how in this implementation every iteration involves a small fraction of the particles (ν) being sampled from the prior. This is to avoid missing specific regions in experiment design space, that correspond to high JSDs, in longer runs. It allows the algorithm to find new regions of high JSD once the initial ones are densely sampled. An additional advantage is that this approach allows real time monitoring of intermediate results from the experimental matrix.

The sampling was compared to a full determination of the design matrix in the 2D case. The comparison was performed in two ways. Firstly, by determining the maximal JSD found during the sampling. Secondly, by determining the fraction of JSD accounted for when comparing to the full sampling procedure. This fraction is computed by dividing the sum of the JSDs for the sampled points

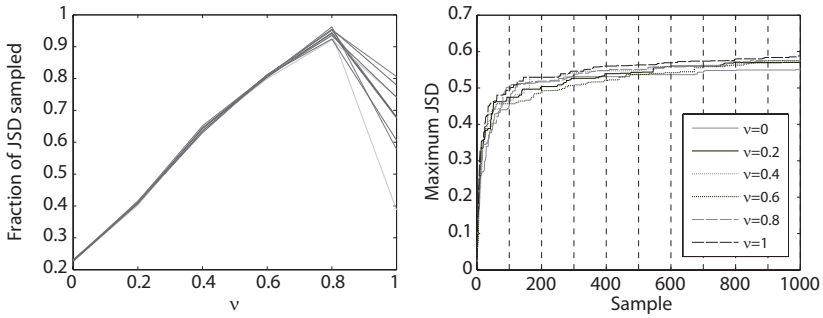

Figure 5.10: Sampling of the experiment design matrix in the 2D case. Left: Fraction of total JSD found versus various values for ν. Shown are ten independent samplings. Right: Maximum JSD found up to a certain iteration for different values of ν.

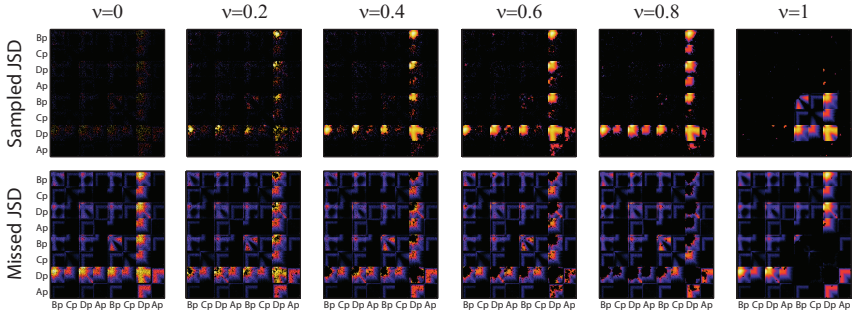

Figure 5.11: Sampling of Jensen-Shannon divergences for various combinations of observables. Shown are the results for 100 iterations using 100 particles. This is 12% of the evaluations required to sample the entire matrix. Top: JSD sampled using the importance resampling approach. Bottom: Regions not sampled.

by the sum over the JSD values for the entire matrix. These two quantities are displayed in Figure 5.10 for $Q = 100$ particles over 100 iterations.

Note how the maximum value found for the JSD does not seem to be very sensitive to the value chosen for ν, while the fraction is strongly affected. For most sampling runs, high values for ν turned out to be better, though the variability between different runs increases drastically for $\nu = 1$. The actual design matrices corresponding to the different values for ν are shown in Figure 5.10. Note how the high JSD regions are sampled more densely for higher ν, yet for $\nu = 1$, some regions are completely missed.

Despite that the efficiency of the sampling likely depends on the posterior predictive distributions (the relative size of the dense regions) and the dimensionality of the problem, we expect that the improvement over simple random sampling ($\nu = 0$) remains.

References

[1] Tiemann C, Vanlier J, Hilbers P, van Riel N: **Parameter adaptations during phenotype transitions in progressive diseases**. *BMC Systems Biology* 2011, **5**:174.

[2] van Riel NA, Tiemann CA, Vanlier J, Hilbers PA: **Applications of analysis of dynamic adaptations in parameter trajectories**. *Interface Focus* 2013, **3**(2).

[3] Schmitz J, Van Riel N, Nicolay K, Hilbers P, Jeneson J: **Silencing of glycolysis in muscle: experimental observation and numerical analysis**. *Experimental physiology* 2010, **95**(2):380–397.

[4] Schilling M, Maiwald T, Hengl S, Winter D, Kreutz C, Kolch W, Lehmann W, Timmer J, Klingmüller U: **Theoretical and experimental analysis links isoform-specific ERK signalling to cell fate decisions**. *Mol Syst Biol* 2009, **5**:334.

[5] Borisov N, Aksamitiene E, Kiyatkin A, Legewie S, Berkhout J, Maiwald T, Kaimachnikov N, Timmer J, Hoek J, Kholodenko B: **Systems-level interactions between insulin–EGF networks amplify mitogenic signaling**. *Mol Syst Biol* 2009, **5**:256.

[6] Cedersund G, Roll J, Ulfhielm E, Danielsson A, Tidefelt H, Strålfors P: **Model-based hypothesis testing of key mechanisms in initial phase of insulin signaling**. *PLoS Comput Biol* 2008, **4**(6):799–806.

[7] Koschorreck M, Gilles E: **Mathematical modeling and analysis of insulin clearance in vivo**. *BMC Syst Biol* 2008, **2**:43.

[8] Schoeberl B, Eichler-Jonsson C, Gilles E, Müller G: **Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors**. *Nat Biotechnol* 2002, **20**(4):370–375.

[9] Jeneson J, Westerhoff H, Kushmerick M: **A metabolic control analysis of kinetic controls in ATP free energy metabolism in contracting skeletal muscle**. *American Journal of Physiology-Cell Physiology* 2000, **279**(3):C813–C832.

[10] Wu F, Jeneson J, Beard D: **Oxidative ATP synthesis in skeletal muscle is controlled by substrate feedback**. *American Journal of Physiology-Cell Physiology* 2007, **292**:C115–C124.

[11] Groenendaal W, Schmidt K, von Basum G, van Riel N, Hilbers P: **Modeling glucose and water dynamics in human skin**. *Diabetes technology & therapeutics* 2008, **10**(4):283–293.

[12] Vanlier J, Wu F, Qi F, Vinnakota K, Han Y, Dash R, Yang F, Beard D: **BISEN: Biochemical simulation environment**. *Bioinformatics* 2009, **25**(6):836–837.

[13] Vanlier J, Tiemann C, Hilbers P, van Riel N: **Parameter uncertainty in biochemical models described by ordinary differential equations**. *Mathematical Biosciences* 2013.

[14] Klinke D: **An empirical Bayesian approach for model-based inference of cellular signaling networks**. *BMC bioinformatics* 2009, **10**:371.

[15] Taylor H, Barnes C, Huvet M, Bugeon L, Thorne T, Lamb J, Dallman M, Stumpf M, et al.: **Calibrating spatio-temporal models of leukocyte dynamics against in vivo live-imaging data using approximate Bayesian computation**. *Integrative Biology* 2012, **4**(3):335–345.

[16] Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmüller U, Timmer J: **Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood**. *Bioinformatics* 2009, **25**(15):1923.

[17] Hasenauer J, Waldherr S, Wagner K, Allgower F: **Parameter identification, experimental design and model falsification for biological network models using semidefinite programming**. *Systems Biology, IET* 2010, **4**(2):119–130.

[18] Brännmark C, Palmér R, Glad S, Cedersund G, Strålfors P: **Mass and information feedbacks through receptor endocytosis govern insulin signaling as revealed using a parameter-free modeling framework**. *Journal of Biological Chemistry* 2010, **285**(26):20171.

[19] Girolami M, Calderhead B: **Riemann manifold langevin and hamiltonian monte carlo methods**. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2011, **73**(2):123–214.

[20] Cedersund G, Roll J: **Systems biology: model based evaluation and comparison of potential explanations for given biological data**. *FEBS Journal* 2009, **276**(4):903–922.

[21] Müller T, Faller D, Timmer J, Swameye I, Sandra O, Klingmüller U: **Tests for cycling in a signalling pathway**. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 2004, **53**(4):557–568.

[22] Calderhead B, Girolami M: **Statistical analysis of nonlinear dynamical systems using differential geometric sampling methods**. *Interface Focus* 2011, **1**(6):821–835.

[23] Tegnér J, Compte A, Auffray C, An G, Cedersund G, Clermont G, Gutkin B, Oltvai Z, Stephan K, Thomas R, et al.: **Computational disease modeling–fact or fiction?** *BMC systems biology* 2009, **3**:56.

[24] Skanda D, Lebiedz D: **An optimal experimental design approach to model discrimination in dynamic biochemical systems**. *Bioinformatics* 2010, **26**(7):939.

[25] Steiert B, Raue A, Timmer J, Kreutz C: **Experimental Design for Parameter Estimation of Gene Regulatory Networks**. *PloS one* 2012, **7**(7):e40052.

[26] Lages NF, Cordeiro C, Silva MS, Freire AP, Ferreira AE: **Optimization of time-course experiments for kinetic model discrimination**. *PloS one* 2012, **7**(3):e32749.

[27] Casey F, Baird D, Feng Q, Gutenkunst R, Waterfall J, Myers C, Brown K, Cerione R, Sethna J: **Optimal experimental design in an epidermal growth factor receptor signalling and down-regulation model**. *Systems Biology, IET* 2007, **1**(3):190–202.

[28] Kreutz C, Raue A, Timmer J: **Likelihood based observability analysis and confidence intervals for predictions of dynamic models**. *BMC Systems Biology* 2012, **6**:120.

[29] Vanlier J, Tiemann C, Hilbers P, van Riel N: **An integrated strategy for prediction uncertainty analysis**. *Bioinformatics* 2012, **28**(8):1130–1135.

[30] Liepe J, Filippi S, Komorowski M, Stumpf MP: **Maximizing the Information Content of Experiments in Systems Biology**. *PLOS Computational Biology* 2013, **9**:e1002888.

[31] Vanlier J, Tiemann C, Hilbers P, van Riel N: **A Bayesian approach to targeted experiment design**. *Bioinformatics* 2012, **28**(8):1136–1142.

[32] Vyshemirsky V, Girolami M: **Bayesian ranking of biochemical system models**. *Bioinformatics* 2008, **24**(6):833–839.

[33] Schmidl D, Hug S, Li WB, Greiter MB, Theis FJ: **Bayesian model selection validates a biokinetic model for zirconium processing in humans**. *BMC systems biology* 2012, **6**:95.

[34] Mélykúti B, August E, Papachristodoulou A, El-Samad H: **Discriminating between rival biochemical network models: three approaches to optimal experiment design**. *BMC systems biology* 2010, **4**:38.

[35] Flassig R, Sundmacher K: **Optimal design of stimulus experiments for robust discrimination of biochemical reaction networks**. *Bioinformatics* 2012, **28**(23):3089–3096.

[36] Kreutz C, Rodriguez M, Maiwald T, Seidl M, Blum H, Mohr L, Timmer J: **An error model for protein quantification**. *Bioinformatics* 2007, **23**(20):2747.

[37] Raue A, Kreutz C, Theis F, Timmer J: **Joining forces of Bayesian and frequentist methodology: A study for inference in the presence of non-identifiability**. *Phil. Trans. Roy. Soc. A* 2012.

[38] Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf M: **Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems**. *Journal of the Royal Society Interface* 2009, **6**(31):187–202.

[39] Burnham KP, Anderson DR: *Model selection and multi-model inference: a practical information-theoretic approach*. Springer 2002.

[40] Good IJ: **Weight of evidence: A brief survey**. *Bayesian statistics* 1985, **2**:249–269.

[41] Daunizeau J, Preuschoff K, Friston K, Stephan K: **Optimizing experimental design for comparing models of brain function**. *PLoS computational biology* 2011, **7**(11):e1002280.

[42] Lin J: **Divergence measures based on the Shannon entropy**. *Information Theory, IEEE Transactions on* 1991, **37**:145–151.

[43] Härdle W, Werwatz A, Müller M, Sperlich S: **Introduction**. *Nonparametric and semiparametric models* 2004, :1–18.

[44] Kraskov A, Stögbauer H, Grassberger P: **Estimating mutual information**. *Physical Review E* 2004, **69**(6):066138.

[45] Budka M, Gabrys B, Musial K: **On accuracy of PDF divergence estimators and their applicability to representative data sampling**. *Entropy* 2011, **13**(7):1229–1266.

[46] Boltz S, Debreuve E, Barlaud M: **High-dimensional statistical distance for region-of-interest tracking: Application to combining a soft geometric constraint with radiometry**. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, IEEE 2007:1–8.

[47] Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M: **KEGG: Kyoto encyclopedia of genes and genomes**. *Nucleic acids research* 1999, **27**:29–34.

[48] Bevan P: **Insulin signalling**. *Journal of Cell Science* 2001, **114**(8):1429–1430.

[49] Trotta R: **Forecasting the Bayes factor of a future observation**. *Monthly Notices of the Royal Astronomical Society* 2007, **378**(3):819–824.

[50] Calderhead B, Girolami M: **Statistical analysis of nonlinear dynamical systems using differential geometric sampling methods**. *Interface Focus* 2011, **1**(6):821–835.

[51] Calderhead B, Girolami M, Lawrence N: **Accelerating Bayesian inference over nonlinear differential equations with Gaussian processes**. *Advances in neural information processing systems* 2009, **21**:217–224.

[52] Toni T, Stumpf M: **Simulation-based model selection for dynamical systems in systems and population biology**. *Bioinformatics* 2010, **26**:104–110.

[53] Liepe J, Barnes C, Cule E, Erguler K, Kirk P, Toni T, Stumpf M: **ABC-SysBio approximate Bayesian computation in Python with GPU support**. *Bioinformatics* 2010, **26**(14):1797.

[54] Arefin A, Riveros C, Berretta R, Moscato P: **GPU-FS-kNN: A Software Tool for Fast and Scalable kNN Computation Using GPUs**. *PloS one* 2012, **7**(8):e44000.

[55] Garcia V, Debreuve E, Barlaud M: **Fast k nearest neighbor search using GPU**. In *CVPR Workshop on Computer Vision on GPU*, Anchorage, Alaska, USA 2008.

[56] Calderhead B, Girolami M: **Estimating Bayes factors via thermodynamic integration and population MCMC**. *Computational Statistics & Data Analysis* 2009, **53**(12):4028–4045.

Computational Modeling of Diacylglycerol Transferase (DGAT)

6

Abstract

Triglycerides (TG) are the main form of storing metabolic energy and fatty acids. Excessive accumulation of TG or hypertriglyceridemia has been implicated as an important risk factor for various diseases. Production of TG occurs via two major pathways which converge into a final reaction where fatty acids (FA) and diacylglycerol (DAG) are bound into TG. Diacylglycerol acyltransferases (DGAT) are membrane bound enzymes which are primarily responsible for catalyzing this acylation of DAG. Mathematical modeling is applied to integrate different sources of experimental data and investigate the contributions of the different DGAT enzymes present in the underlying biological system. Taking into account the qualitative understanding from literature, a network topology is proposed and subsequently used to formulate a mathematical model. The fact that each of the data sets required different but unknown normalizations and that the fluxes at the boundary of the model were underdetermined, turned out to be detrimental for the predictive power of the model. Nevertheless, integrating different sources of data into a mathematical model managed to consolidate a number of predictions that seemed inconsistent at first. The difference resided in the amount of upregulation performed in the different studies. Additionally, we may conclude that future experiments should focus on measuring fluxes at the boundary of the model.

6.1 Background

Triglycerides (TG) are the main form of storing metabolic energy and fatty acids. Though essential for normal physiology, excessive accumulation of TG or hypertriglyceridemia has been implicated as an important risk factor for various diseases such as atherosclerosis, hepatic steatosis and obesity [1, 2]. The production of TG occurs via two major pathways: The glycerol phosphate and the monoacylglycerol pathway which converge into a final reaction where fatty acids (FA) and diacylglycerol (DAG) are bound into TG [3]. Diacylglycerol acyltransferases (DGAT) are membrane bound enzymes which are primarily responsible for catalyzing this acylation of DAG [4]. Newly synthesized TG is either stored in cytoplasmic droplets or secreted as very low density lipoprotein (VLDL) particles. Two different DGAT enzymes have been identified that are both highly expressed in liver [5]. However, the role and relative contribution of each enzyme is not well understood quantitatively. Nevertheless, their relative activities may have a significant impact on the development of hypertriglyceridemia and hepatic steatosis. For this reason, they have also been marked as potential drug targets [6].

Several experiments have been performed to elucidate the effects of the two DGAT proteins. Decreasing DGAT2 activity using antisense oligonucleotides (ASO) resulted in decreased liver TG and VLDL TG secretion in both wild type and DGAT1 deficient mice [7]. In this study, the largest fraction of DGAT activity was accounted for by DGAT1, yet liver TG levels were markedly lower in the DGAT2 mice treated with ASO. In two independent studies, DGAT2 activity accounted for only a small fraction of the total activity, yet inhibition of DGAT2 strongly reduced the incorporation of *de novo* synthesized FA both *in vitro* [8] and *in vivo* [9]. This suggests that DGAT2 primarily acts upstream of DGAT1, esterifying newly formed fatty acids, while DGAT1 primarily re-esterifies recycled fatty acids.

In this work, the aim is to use mathematical modeling to integrate different sources of experimental data and to investigate the contributions of the different DGAT enzymes. The reason for this is two-fold. Firstly, to consolidate some of the contradictory observations reported in literature and unravel the functional roles of each of the enzymes. Secondly, once sufficiently constrained, the model can be integrated in the liver model presented in Chapter 7. This would allow use of experimental data involving the use of inhibitors and/or knockouts of one or both DGAT enzymes when performing inference on the overarching model.

6.2 Methods

The system is modeled using a system of ordinary differential equations. The model comprises of equations $\vec{f}(\vec{x}(t), \vec{u}(t), \vec{p})$ which contain parameters $\vec{p}$ (constant in time), inputs $\vec{u}(t)$, and state variables $\vec{x}(t)$. Given a set of parameters, inputs and initial conditions $\vec{x}(0)$, these equations can subsequently be simulated. To estimate these parameters, measurements $y^D_{i,j}$ of observables $\vec{y}(t)$ were gathered from literature. These measurements were acquired for different knock-

outs and up- or downregulations of each of the DGAT enzymes. The advantage of incorporating data sets obtained during different perturbations is that the contribution of each of the DGAT enzymes may change. The aim of including these perturbations is to disentangle the individual DGAT contributions. Certain data sets required scaling and offset parameters $\vec{q}$ to be incorporated in the mathematical model. Additionally, the amount of up- and downregulation of the DGAT enzymes in the various conditions had to be estimated. We define $\vec{\theta}$ as $\vec{\theta} = \{\vec{p}, \vec{q}, \vec{x}_0\}$, which lists all the parameters required to simulate the model. For more information on the scaling and offset parameters, see the section on experimental data. Maximum Likelihood Estimation was performed by minimizing the following least squares criterion

$$\chi^2(\vec{\theta}) = \sum_{i=1}^{M} \sum_{j=1}^{N_i} \left(\frac{y_{i,j}^D - y_i(t_j, \vec{\theta})}{\sigma_{i,j}} \right)^2 \tag{6.1}$$

where $y_i(t_j, \vec{\theta})$ and $y_{i,j}^D$ are the model output and data point corresponding to the j^{th} time point of output i.

6.2.1 Misfit analysis

The quality of a model fit can be assessed by determining whether the model residuals are in agreement with the assumed error model and do not show a systematic trend [10]. When the model insufficiently describes the data, it can be beneficial to elucidate which data sets provide contradictory constraints before deciding to proceed. The approach we apply here is to re-estimate parameters iteratively while artificially reducing the measurement uncertainty associated with the data that the model fails to describe. To perform the analysis, the objective function is divided in two parts, the internal χ^2_{int} and external χ^2_{ext} sum of squares, where the former refers to the sum of squared errors of the measurements that are currently not described by the model, while the latter refers to the measurements that are adequately described. The following augmented objective function is then used to perform the misfit analysis

$$\chi^2_M(\vec{\theta}) = \sum_{i=1}^{M} \sum_{j=1}^{N_i} \left(1 + \lambda \mathbf{1}_{int} \frac{\chi^2_{ext}}{\chi^2_{int}} \right) \left(\frac{y_{i,j}^D - y_i(t_j, \vec{\theta})}{\sigma_{i,j}} \right)^2 \tag{6.2}$$

Where $\lambda > 0$ and $\mathbf{1}_{int}$ is an indicator function which is 1 if measurement point $y_{i,j}^D$ is part of the internal sum of squares and 0 if it is part of the external sum of squares. Both χ^2_{int} and χ^2_{ext} are updated at the end of each optimization. This process is repeated until χ^2_{int} becomes statistically acceptable or converges. By performing this analysis and analyzing the different predictions along the misfit profile, one quickly obtains an idea of which data sets are providing contradictory constraints under the current model.

6.2.2 Profiles

Once model predictions sufficiently describe the experimental data, confidence intervals can be obtained using the Profile Likelihood method [11,12]. Parameter confidence intervals can be computed for each parameter by forcing one parameter to change, while finding the region for which the inequality (6.3) continues to hold. While performing this traversal, the other parameters are continually re-optimized, hereby tracing a path through parameter space.

$$\chi^2(\vec{\theta}_{PL}) - \chi^2(\vec{\theta}_{opt}) \leq \chi^2_{1-\alpha,1} \tag{6.3}$$

Similarly, confidence intervals can be obtained for the predictions by augmenting the experimental data with an additional point which has to be satisfied by the simulation and continually reoptimizing the parameters [13]. Initially, this point is based on the simulation belonging to the optimal parameter values. Similar to the Profile Likelihood, this additional constraint is then shifted followed by subsequent parameter optimization. This process is continued until the sum of squared errors exceeds the likelihood ratio threshold. The resulting profile is known as a prediction profile likelihood (PPL).

$$\chi^2_{PPL}(z) = \min_{\vec{\theta},\vec{\theta}\in\{\vec{\theta}|Q(\vec{\theta})=z\}} \left[\chi^2(\vec{\theta})\right] \tag{6.4}$$

Here, Q refers to a function which takes a parameter vector and produces a simulation point of the prediction that is being profiled.

6.2.3 Experimental data

The availability of quantitative experimental data determines the level of detail at which certain biological processes can be integrated into the mathematical model. In this work, we aim to consolidate different sources of experimental data into a single consistent dataset for parameter estimation.

The first dataset (Liu et al) involved knockdown of DGAT2 using DGAT2 gene-specific antisense oligonucleotide (ASO) in both C57BL/6 wild-type mice and DGAT1 knockouts (DGAT1 KO) [7]. Concentration measurements of VLDL TG after intraperitoneal injection of $400\mu L$ P407 (1 mg/g) solution in sterile PBS were used (Table 6.1). The non-ionic detergent P407 inhibits TG hydrolysis by lipoprotein lipase. The VLDL TG production rate is then typically calculated

Time	WT $[\mu M]$	WT/ASO $[\mu M]$	DG1KO $[\mu M]$	DG1KO/ASO $[\mu M]$
0 h	858 ± 980	4656 ± 2695	10904 ± 1102	21623 ± 1715
1 h	1286 ± 980	2757 ± 1348	5635.6 ± 1348	9005 ± 2083
2 h	1225 ± 1225	7167 ± 2573	13722 ± 2450	18683 ± 3186
4 h	1348 ± 980	3553 ± 1960	7902 ± 1838	10842 ± 5146

Table 6.1: VLDL TG after injection with P407 reported as $mean \pm sem$ (n=5). ASO refers to 60 mg/kg antisense oligonucleotide.

Condition	mRNA level	VLDL TG [%]
WT 25 mg/kg ASO	$.4476 \pm .3211$	74.9
WT 40 mg/kg ASO	$.3252 \pm .0742$	71.1
WT 60 mg/kg ASO	$.1924 \pm .0358$	48.3
DG1KO 60 mg/kg ASO	$.2343 \pm .0655$	NA

Table 6.2: Fractional mRNA decrease reported as $mean \pm sem$ (n=5) and fractional dose response of VLDL TG production.

Condition	VLDL TG $[\mu M/h]$	$TG_C[\mu mol/g]$	$TG_{ER} + DAG\ [\mu mol/g]$
WT	7024 ± 829	17.71 ± 6.35	4.06 ± 0.61
DG1+	6736 ± 630	36.47 ± 14.12	4.23 ± 0.74
DG2+	6598 ± 480	45.09 ± 11.66	4.31 ± 0.60

Table 6.3: VLDL TG production in wild type and during DGAT overexpression. Lipid concentrations in wild type and during DGAT overexpression expressed in μmol per gram liver.

from the increase of plasma VLDL TG over time [14]. Additionally a dose response curve of VLDL TG secretion versus ASO concentration was included (Table 6.2). Liver cDNA was made by reverse-transcription using total RNA and used for real-time PCR to quantify DGAT2 mRNA levels. These were resampled using a Monte Carlo approach to obtain mean and standard deviations of fractional MRNA decrease (Table 6.2).

The second data set (Millar et al) involved short term overexpression of DGAT1 and DGAT2 [15]. This data set included measurements of VLDL secretion, and, cytoplasmic and membrane lipids (Table 6.3). Note however, that the lipid measurements effectively measure the backbone of the TG molecules, therefore the data is only used to constrain the sum of DAG and TG in the endoplasmic reticulum (ER).

Additionally, a TG measurement for the WT was included from [16] which was 13.06 ± 1.65 μmol per gram liver. The fourth data set (Yamazaki et al) was measured during upregulation of DGAT1 and DGAT2 [17]. Here measurements of VLDL TG production (Table 6.4), liver lipids and DGAT activities towards the different compartments (Table 6.5) were measured under all three experimental conditions.

Time	WT $[\mu M]$	DG1+ $[\mu M]$	DG2+ $[\mu M]$
0 h	1367.9 ± 956.4	1367.9 ± 597.9	1128.6 ± 836.8
1 h	7511.1 ± 956.6	9663.5 ± 837.0	7211.9 ± 478.3
2 h	12458.4 ± 2989.3	15447.9 ± 2391.8	13773.9 ± 2032.8
3 h	19917.1 ± 2391.5	24760.0 ± 3228.4	19020.1 ± 2869.4
4 h	23609.1 ± 3348.0	30963.0 ± 2630.6	25940.6 ± 4065.9

Table 6.4: VLDL TG after injection with P407 $mean \pm sem(n = 8)$ in μM.

Condition	Hepatic Lipid $[\mu mol/g]$	Ratio
WT	65.69 ± 13.33	1.401 ± 1.27
DG1+	128.05 ± 13.33	4.231 ± 1.30
DG2+	202.79 ± 7.61	-2.160 ± 1.40

Table 6.5: Hepatic lipid reported as $mean \pm sem$ (n=5) in μmol per gram liver and fractional production to each compartment $\ln\left(\frac{V_{overt}}{V_{latent}}\right)$ in the wild type and during DGAT upregulation.

The DGAT activity was measured by determining the incorporation of labeled DAG into TG. Studies revealed that *de novo* TG are incorporated via DGAT2 and not DGAT1 [8]. Since the *de novo* flux is based on unlabeled monoacylglycerol, it was effectively not measured. Additionally, the activity data was expressed relative to the amount of protein, which required additional scaling factors for each condition. To avoid including another (non-identifiable) scaling factor, the data was incorporated as a ratio-metric constraint between overt and latent activities. Assuming log-normal distributions for the activities, we determined the mean and standard deviation of these ratios in log-space by employing a Monte Carlo approach resampling parametric distributions fitted to the data.

Since most of the experimental data is expressed in units per gram liver, the computational model employs these units. VLDL TG levels and VLDL secretion parameters are represented in μM and $\mu M/hr$ however. Consequently, a scaling factor is required for the ratio g_{liver}/V_{plasma}. The former is an estimated parameter, bounded between 0.9 and 2 grams [18], while the latter is fixed at $1mL$ [16].

6.2.4 Computational model

A mathematical multi-compartment model of the diacylglycerol transferase system was proposed. The model contains three compartments representing the cytoplasm, endoplasmic reticulum (ER) and ER membrane where the DGAT enzymes reside. Diacylglycerol and fatty acids in the membrane compartment are explicitly modeled where influx and degradation are estimated quantities. Triglycerides are produced from these metabolites and transported to the cytoplasmic and endoplasmic compartments. Triglycerides in the endoplasmic compartment are converted into nascent produced VLDL particles associated with a rate k_{VLDL}. The complete model, schematically shown in Figure 6.1 comprises of a system of coupled ordinary differential equations.

Each of the DGAT proteins is associated with a rate, V_{DG1} and V_{DG2} respectively. Since it is unclear how much each of the enzymes contributes to the cytosolic and endoplasmic production flux, the direction in which the DGAT enzymes operate was not fixed in the design. Rather than making each DGAT produce triglycerides into a specific compartment, we introduced fraction parameters γ_{DG1} and γ_{DG2} for each of the DGAT enzymes. This fraction pertains to the fraction of TG that is pumped into the cytosol. These fraction parameters were subsequently estimated along with the other model parameters. Recent tracer studies identified that *de novo* fatty acid synthesis occurs via DGAT2 [8, 9].

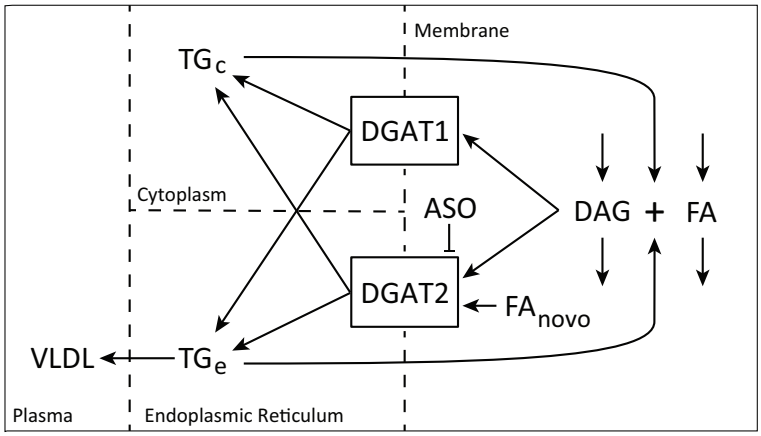

Figure 6.1: The DGAT model structure.

Therefore an additional parameter representing *de novo* TG synthesis is also included in the DGAT2 flux. Finally, the reverse fluxes of TG from the cytosolic and endoplasmic compartment to DAG and FA in the membrane are associated with estimated rate constants k_{cbwd} and k_{ebwd}.

To use the dose-response data [7], a minimal mRNA transcription model was used. Since the ASO treatment operated at a time scale of weeks and the VLDL experiments were on the time scale of hours, the transcription model was assumed to be in equilibrium (quasi steady state). The mRNA model is comprised of the following equations

$$\frac{d[MRNA]}{dt} = k_1 - (k_2 + k_4 + k_5[ASO])[MRNA] \tag{6.5}$$

$$\frac{d[DG2]}{dt} = k_2[MRNA] - k_3[DG2] \tag{6.6}$$

Setting the right hand side to zero results in the following expression for the steady state concentration of DGAT2

$$[DG2]_{SS} = \frac{k_a}{k_b + [ASO]} \tag{6.7}$$

Enzyme activity was modeled as being proportional to the amount of enzyme available. The different parameters can be combined into phenomenological parameters V_{DG2} and k_{ASO}, which are subsequently estimated along with the other model parameters. This leads to the following expression for the DGAT rate.

$$V_{DG2}^{app} = \frac{V_{DG2}}{k_{ASO} + [ASO]} \tag{6.8}$$

Additional parameters were added to simulate the upregulation of the different DGAT enzymes. These parameters represent the fractional increase of enzyme activity when simulating these experimental perturbations.

Flux	Expression	Description
f_{DG1}	$V_{DG1}[DAG][FA]$	DGAT1 flux
$f_{DG2,a}$	$V_{DG2}\left(k_{DG2}+[ASO]\right)^{-1}[DAG][FA]$	DGAT2 flux
$f_{DG2,b}$	$V_{DG2}\left(k_{DG2}+[ASO]\right)^{-1}[DAG]FA_{denovo}$	DGAT2 *de novo* flux
f_{cbwd}	$k_{cbwd}[TG_c]$	TG flux from cytosol
f_{ebwd}	$k_{ebwd}[TG_e]$	TG flux from ER
f_{VLDL}	$k_{VLDL}[TG_e]$	VLDL production flux
f_{DAG}^{in}	DAG_{in}	DAG influx
f_{DAG}^{out}	$k_{DAG}[DAG]$	DAG outflux
f_{FA}^{in}	FA_{in}	FA influx
f_{FA}^{out}	$k_{FA}[FA]$	FA outflux

Table 6.6: Flux expressions for the DGAT model.

$$\begin{aligned} V_{DG1}^{*} &= V_{DG1}\left(1+DG1_{up}\right) \\ V_{DG2}^{*} &= V_{DG2}\left(1+DG2_{up}\right) \end{aligned} \tag{6.9}$$

The full list of flux expressions is shown in Table 6.6. Which combine into the following differential equations for the model state variables

$$\begin{aligned} \frac{d[TG_c]}{dt} &= \gamma_{DG1}f_{DG1}+\gamma_{DG2}\left(f_{DG2,a}+f_{DG2,b}\right)-f_{cbwd} \\ \frac{d[TG_e]}{dt} &= \left(1-\gamma_{DG1}\right)f_{DG1}+\left(1-\gamma_{DG2}\right)\left(f_{DG2,a}+f_{DG2,b}\right)-f_{ebwd}-f_{VLDL} \\ \frac{d[DAG]}{dt} &= -f_{DG1}-\left(f_{DG2,a}+f_{DG2,b}\right)+f_{ebwd}+f_{cbwd}+f_{DAG}^{in}-f_{DAG}^{out} \\ \frac{d[FA]}{dt} &= -f_{DG1}-\left(f_{DG2,a}\right)+f_{ebwd}+f_{cbwd}+f_{FA}^{in}-f_{FA}^{out} \end{aligned} \tag{6.10}$$

6.3 Results

The first step in model analysis is fitting the model outputs to the experimental data. Optimization was initiated from widely dispersed initial values to reduce the chance of missing local minima. This multiple minimization approach failed to provide an adequate description of all the data simultaneously. Misfit analysis revealed that there was a systematic discrepancy between the constraints provided by the activity data measured by Yamazaki et al and the rest of the data (see Figure 6.2). Since these measurements were obtained by integrating production over a fixed period of time *in vitro*, it is possible that the activity measured in this manner was not sufficiently representative for the steady state condition *in vivo*. Alternatively, one could consider incorporating the ratio by simulating the experiment dynamically, but the initial conditions in the experimental medium are unknown. This means a dynamic simulation of this experiment would require the

addition of several free parameters. These additional parameters rendered the actual activity non-identifiable (by being related to the initial conditions). Therefore, this data did not provide additional model constraints and was omitted from further analysis. After omission of the activity data, the model simulations were in statistically acceptable agreement with the data (χ^2 test with significance level $\alpha = 0.05$).

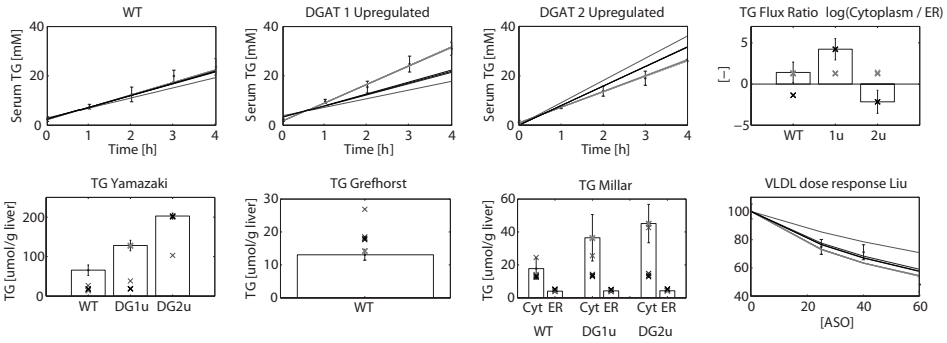

Figure 6.2: Comparison of the trade-off made between the different data sets. The different simulations are denoted by lines and crosses, where darker lines and crosses indicate a higher weighting of the activity data. Note how the model fails to describe the measured TG flux ratios.

Profile Likelihood analysis was performed for each of the parameters to investigate how well the model parameters were constrained. As shown in Figure 6.3, most parameters are either non-identifiable or characterized by very wide bounds. This is not unexpected, since most fluxes can be compensated by unmeasured fluxes at the boundary of the model. The relevant prediction profiles (the individual fluxes of the two DGAT enzymes) were also non-identifiable. Such non-identifiabilities imply that a Bayesian analysis would be very sensitive to the assumed prior distributions and in this case, realistic priors or bounds are not available for the majority of the non-identifiable model parameters. The fractions resembling the destination (cytosol or endoplasmic reticulum) of TG produced by either DGAT enzyme turned out to be correlated, implying that there is too little information available to adequately disentangle the two enzymes. This is not surprising considering that the only difference between the two is the *de novo* flux of fatty acids which turned out to be non-identifiable.

Ranked correlation coefficients between the different parameters were computed for each of the profiles. Subsequently thresholding these at a correlation coefficient of 0.95 revealed distinct patterns of correlations between the parameters corresponding to each profile. This analysis revealed that the fact that the *de novo* fatty acid production was not measured compromises the ability to estimate the rate constant of DGAT2. Since the measurement of the membrane TG effectively only provides an upper bound for the TG in the ER, the VLDL rate constant is merely constrained at its lower bound. Similarly, the FA sink is entangled with both the ER TG concentration (and therefore the backward flux) and FA influx. Interestingly, this structural relation does not appear in the profiles belonging to the DAG rate constants, which are merely related to each other.

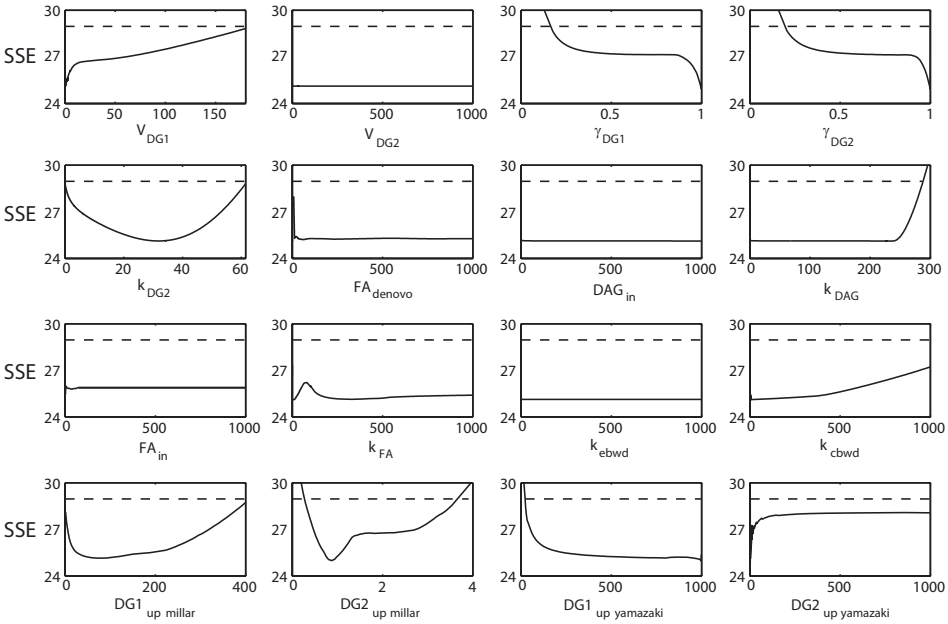

Figure 6.3: Profile Likelihoods with respect to various kinetic parameters of interest in the DGAT model. Note how the likelihood profiles corresponding to most of the parameters flatten out and do not reach the bound (dashed line) required for identifiability.

PPLs were determined for each of the state and flux predictions, from which it could be concluded that these were non-identifiable. Since these fluxes were ill determined, integration in the overarching model is not yet reasonable at this point.

6.4 Discussion

Earlier works revealed no effects of short-term over-expression of DGAT1 or DGAT2 on VLDL secretion [15]. Whereas another study showed that over-expression of DGAT1 resulted in increased secretion of TG via VLDL particles and an increase of TG within the lumen of the ER. Overexpression of DGAT2 gave neither of these effects, but greatly increased the TG present in the cytosol [17]. Though these observations may seem contradictory at first, the model is able to describe both these observations and show that they are not inherently inconsistent. The difference resides in the amount of upregulation that is actually being performed. Although the actual value of the constants that control the amount of upregulation are poorly identifiable from the data in the Yamazaki data set, the parameter sets consistent with the data consistently predict a larger upregulation than in the Millar data. Misfit analysis revealed that, considering the current model, the activity measurement provided constraints contradictory to the rest of the data.

The fact that the fluxes at the boundary of the model were underdetermined

and that most of the data was acquired in steady state turned out to be detrimental for the predictive power of the model. Due to these issues, data mined from literature can (as it has in this case) result in a large number of non- or poorly identifiable parameters and predictions. As such, this scenario represents a fairly typical example of a first generation model based on literature data and furthermore underlines the importance of reporting the predictive power of a model. Additionally, the fact that two seemingly contradictory observations could be consolidated by modelling them with different amounts of upregulation revealed that considerable care must be taken when comparing the results from different biological studies. Conclusions drawn between studies may depend on unknown and unmeasured factors, confounding the quantities of interest. By formalizing the conceptual model in terms of mathematical equations before doing the measurements, it is possible to increase the probability that the experiment will be informative. Though the biological insights this chapter delivers are somewhat limited, the model could still be used to guide future experiments and determine what quantities should be measured to infer the relative contributions of each of the enzymes and/or develop treatments based on targeting either enzyme. Important to consider is the contribution of the *de novo* fatty acid production flux, which should differentiate the relative contributions of each of the enzymes. Once sufficiently constrained at the model boundaries, the model can be integrated in the liver model presented in Chapter 7 and used to incorporate measurement data obtained during manipulation of the DGAT enzymes.

References

[1] Cullen P: **Evidence that triglycerides are an independent coronary heart disease risk factor**. *The American journal of cardiology* 2000, **86**(9):943–949.

[2] Ginsberg H: **New perspectives on atherogenesis role of abnormal triglyceride-rich lipoprotein metabolism**. *Circulation* 2002, **106**(16):2137–2142.

[3] Bell R, Coleman R: **Enzymes of glycerolipid synthesis in eukaryotes**. *Annual review of biochemistry* 1980, **49**:459–487.

[4] Yen C, Stone S, Koliwad S, Harris C, Farese Jr R: **Thematic review series: glycerolipids. DGAT enzymes and triacylglycerol biosynthesis**. *Journal of lipid research* 2008, **49**(11):2283–2301.

[5] Cases S, Smith S, Zheng Y, Myers H, Lear S, Sande E, Novak S, Collins C, Welch C, Lusis A, et al.: **Identification of a gene encoding an acyl CoA: diacylglycerol acyltransferase, a key enzyme in triacylglycerol synthesis**. *Proceedings of the National Academy of Sciences* 1998, **95**(22):13018.

[6] Zhao G, Souers A, Voorbach M, Falls H, Droz B, Brodjian S, Lau Y, Iyengar R, Gao J, Judd A, et al.: **Validation of diacyl glycerolacyltransferase I as a novel target for the treatment of obesity and dyslipidemia using a potent and selective small molecule inhibitor**. *Journal of medicinal chemistry* 2008, **51**(3):380–383.

[7] Liu Y, Millar J, Cromley D, Graham M, Crooke R, Billheimer J, Rader D: **Knockdown of acyl-CoA: diacylglycerol acyltransferase 2 with antisense oligonucleotide reduces VLDL TG and ApoB secretion in mice**. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids* 2008, **1781**(3):97–104.

[8] Wurie H, Buckett L, Zammit V: **Diacylglycerol acyltransferase 2 acts upstream of diacylglycerol acyltransferase 1 and utilizes nascent diglycerides and de novo synthesized fatty acids in HepG2 cells**. *FEBS Journal* 2012.

[9] Qi J, Lang W, Geisler J, Wang P, Petrounia I, Mai S, Smith C, Askari H, Struble G, Williams R, et al.: **The use of stable isotope-labeled glycerol and oleic acid to differentiate the hepatic functions of DGAT1 and-2**. *Journal of Lipid Research* 2012, **53**(6):1106–1116.

[10] Cedersund G: **Conclusions via unique predictions obtained despite unidentifiability–new definitions and a general method**. *FEBS Journal* 2012, **279**(18):3513–3527.

[11] Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmüller U, Timmer J: **Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood**. *Bioinformatics* 2009, **25**(15):1923.

[12] Vanlier J, Tiemann C, Hilbers P, van Riel N: **An integrated strategy for prediction uncertainty analysis**. *Bioinformatics* 2012, **28**(8):1130–1135.

[13] Kreutz C, Raue A, Timmer J: **Likelihood based observability analysis and confidence intervals for predictions of dynamic models**. *BMC Systems Biology* 2012, **6**:120.

[14] Millar JS, Cromley DA, McCoy MG, Rader DJ, Billheimer JT: **Determining hepatic triglyceride production in mice: comparison of poloxamer 407 with Triton WR-1339**. *Journal of lipid research* 2005, **46**(9):2023–2028.

[15] Millar J, Stone S, Tietge U, Tow B, Billheimer J, Wong J, Hamilton R, Farese Jr R, Rader D: **Short-term overexpression of DGAT1 or DGAT2 increases hepatic triglyceride but not VLDL triglyceride or apoB production**. *Journal of lipid research* 2006, **47**(10):2297–2305.

[16] Grefhorst A, Elzinga B, Voshol P, Plösch T, Kok T, Bloks V, van der Sluijs F, Havekes L, Romijn J, Verkade H, et al.: **Stimulation of lipogenesis by pharmacological activation of the liver X receptor leads to production of large, triglyceride-rich very low density lipoprotein particles**. *J Biol Chem* 2002, **277**(37):34182–34190.

[17] Yamazaki T, Sasaki E, Kakinuma C, Yano T, Miura S, Ezaki O: **Increased very low density lipoprotein secretion and gonadal fat mass in mice overexpressing liver DGAT1**. *Journal of Biological Chemistry* 2005, **280**(22):21506–21514.

[18] Jones L, Nielsen M, Britton R: **Genetic variation in liver mass, body mass, and liver: body mass in mice**. *Journal of animal science* 1992, **70**(10):2999.

Optimal Experimental Design for Identifying Progressive Adaptations in Biological Systems

7

Parts of this chapter are described in:

Tiemann, C.A.* and Vanlier, J.* and Hilbers, P.A.J. and van Riel, N.A.W. (submitted), *Optimal experiment design to constrain predictions of progressive adaptations in biological systems*

Abstract

Unraveling long-term adaptations in biological systems is complicated by the multilevel aspects of such systems and the time-scale on which they occur. These complications are exacerbated by the fact that insufficient information on the network structure and interaction mechanisms is available to explicitly formulate models of the involved processes. For this reason, classical methods for modeling and optimal experiment design are not applicable. We propose and demonstrate a new method for experimental design, which we apply on a model of hepatic lipid and plasma lipoprotein metabolism describing pharmacological activation of the liver X receptor (LXR). Our method captures the modulating effects of the genome and proteome on the metabolic level using time-dependent descriptions (or trajectories) of the model parameters. By generating bootstrap replicates of the data, a distribution of parameter trajectories is generated. The non-linear relations in this distribution are subsequently probed to determine which experiments would lead to more constrained predictions. The proposed method enabled us to rank different experiments according to their efficacy at constraining biliary cholesterol excretion. When designing for reduced uncertainty in the biliary excretion of cholesterol, a non-invasive experiment appeared to be highly effective. Data corresponding to the proposed experiment was subsequently included. This lead to a 52% reduction in the uncertainty of the determined adaptations in cholesterol excretion. Additionally, obtained reductions in flux uncertainty were well in line with predictions obtained in the design.

7.1 Introduction

Improved understanding of biochemical networks and their molecular adaptations during diseases or interventions, is one of the driving ambitions of computational biology. Computational modeling allows us to integrate various sources of experimental data with models that can be simulated. By calibrating models to data and subsequently making predictions, conceptual understanding can be tested in a quantitative manner [1–3]. One application of particular interest are the adaptations that occur during progressive diseases, *e.g.*, diabetes type 2, metabolic syndrome and cardiovascular diseases [1]. Unraveling such long term adaptations is complicated by the multilevel aspects of the underlying system and the time-scale on which they occur. Whereas classical models in computational biology are typically constructed to simulate processes at a single level, *e.g.*, transcriptome, proteome, or metabolome level [4–9], progressive diseases are often the result of a combination of processes and changes which occur on multiple levels and time-scales. Identifying these changes is difficult since explicit formulation of the processes at these levels is not realistic due to insufficient information on the network structure and interaction mechanisms.

Recently, a computational approach named ADAPT (Analysis of Dynamic Adaptations in Parameter Trajectories) was proposed which addresses the aforementioned issues [1, 10] by capturing the modulating effects on the metabolic level using time-dependent descriptions (or trajectories) of the model parameters. These trajectories are obtained by determining which dynamic changes in the model parameters are required to describe data acquired at different time points. The trajectories form an hypothesis on how the various model parameters, fluxes and concentrations change during this adaptation. Typically, large ranges of parameter values correspond to model simulations that have an acceptable agreement with the data [2, 11–14]. These parameter (and therefore also parameter trajectory) uncertainties are probed by means of a parametric bootstrap of the data. Parameter uncertainties do not necessarily have to be a problem as long as the model predictions of interest are still well constrained [1, 15–18]. When this is not the case, more data is required. Improving such uncertain predictions is a relevant step to further our understanding of the molecular adaptations that drive disease progression. As these uncertainties depend on both model equations and the available data, it is often not evident which experiments would constrain such predictions most effectively. Statistical methods can be applied in order to select experiments that will reduce uncertainty in an optimal manner. In this work, we focus on V-optimality, which strives to minimize prediction variances. Classical methods for experimental design are often based around models with constant parameters [19–21], where all the time dependence is captured by the dynamical model. However, when parameters depend on time, such methods are not appropriate.

We propose a method for selecting optimal experiments, which is based on predicting variance reductions for various combinations of experiments. The method is based on simulating new experiments by weighting the determined parameter trajectories and hereby exploits relations between different parameters and predictions [3]. The proposed method enables the modeler to focus

experimental efforts towards reducing the uncertainty of specific predictions.

In the present study, the approach is applied to a case study that concerns pharmacological activation of the liver X receptor (LXR) in mice for a period of three weeks. LXR plays a central role in the control of cellular lipid and cholesterol metabolism, and its activation promotes the cellular efflux, transport, and excretion of cholesterol from the body. Consequently, LXR reduces the risk of plaque formation in arteries and is therefore considered a potential drug target for the prevention or treatment of atherosclerosis [22–24]. However, LXR also induces a wide range of other adaptations in lipid and sterol metabolism, *e.g.*, the excessive accumulation of triglycerides in the liver and the production of enlarged lipoproteins. The molecular mechanisms inducing these adaptations are not fully understood, which complicates the clinical application of LXR agonists [23, 24]. Results from earlier computational analyses indicated that, despite a large amount of data on the metabolic level, many predictions are still highly uncertain [25]. In this work, we apply the proposed method to further constrain these predictions.

7.2 Methods

The methods section is divided into two parts. First, we give a brief overview of ADAPT, which is used to describe the modulations at the metabolic level [1, 10]. Subsequently, we discuss how these parameter trajectories can be used for optimal experiment design. The entire approach is briefly depicted in Figure 7.1 and shall be discussed below.

7.2.1 Obtaining a continuous description of the data

The first step in our methodology is to obtain a time dependent description of the data. Continuous dynamic descriptions of the experimental data were obtained by calculating cubic smoothing splines that describe the dynamics of the experimental data. The uncertainty associated with the experimental data is propagated to these interpolants by means of sampling a parametric uncertainty model [18, 26, 27]. Since data variability scaled with magnitude, multiplicative measurement errors were assumed [28]. To convert these to additive error, both data and model simulations are log-transformed and inferences are performed in logarithmic space [29]. New replicates of the experimental data are generated using Gaussian distributions based on the transformed data. Subsequently, splines are generated for each of the data realizations, thereby obtaining a sample of data interpolants $\vec{d}(t)$. These splines are subsequently used for parameter trajectory estimation.

7.2.2 Mathematically modeling the metabolic network

Model development is mostly driven by two factors, the scientific question at hand, and, the amount and quality of the available experimental data. Whereas the former dictates which processes must be present, the latter determines the level of detail with which various processes can be incorporated. In this case, the

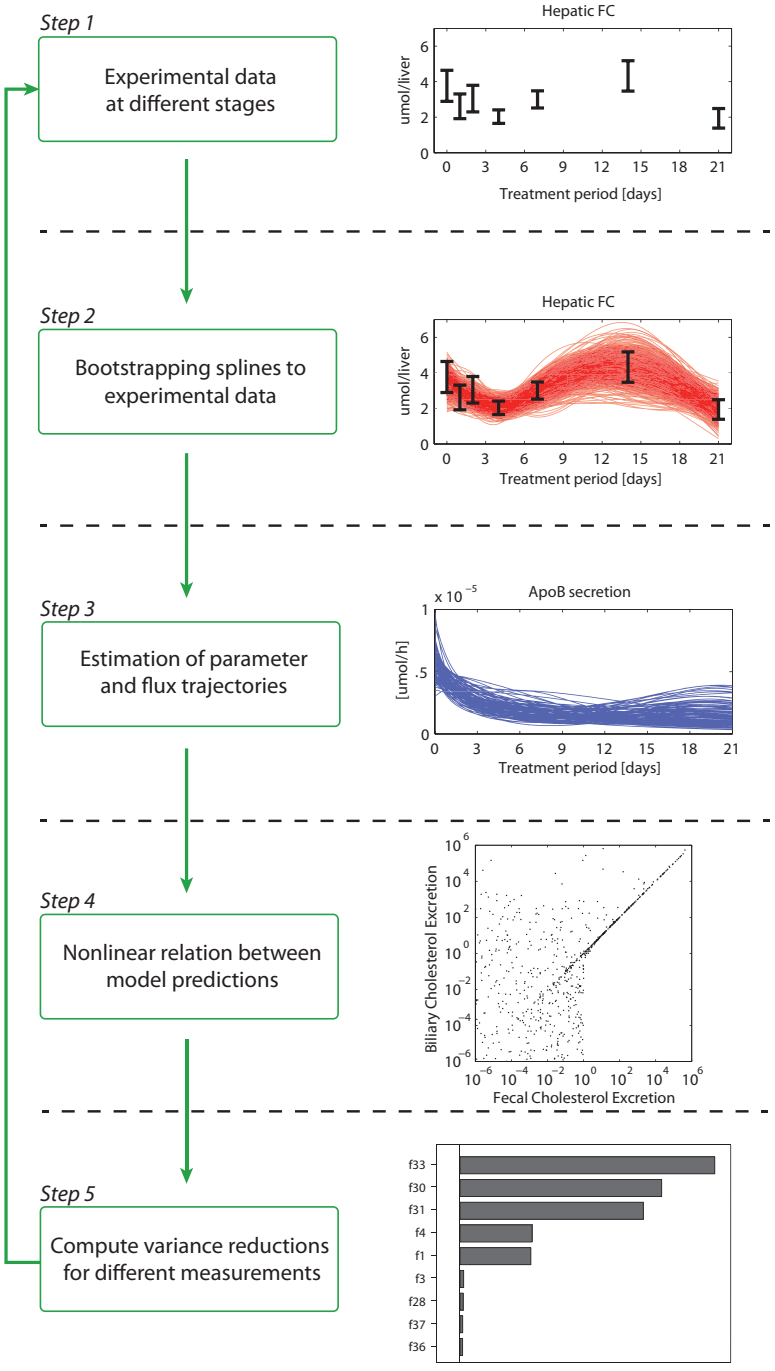

Figure 7.1: Overview of the optimal experiment design method.

experimental data mostly consisted of measurements at the metabolome level and therefore modeling efforts were mostly focused on the pathways at this level. The metabolic level is modeled using ordinary differential equations which provide dynamical relationships between the metabolites in time:

$$\dot{\vec{x}}(t) = \vec{f}(\vec{x}(t), \vec{\theta}, \vec{u}(t)) \tag{7.1}$$
$$\vec{y}(t) = \vec{g}(\vec{x}(t), \vec{\theta}, \vec{u}(t)) \tag{7.2}$$
$$\vec{x}(0) = \vec{x}_0 \tag{7.3}$$

Here, $\dot{\vec{x}}$ is a vector of time derivatives of molecular species $\vec{x}$. The initial concentrations are given by $\vec{x}_0$. The vector $\vec{y}$ represent the model outputs. Both mappings $\vec{f}$ and $\vec{g}$ depend on kinetic parameters $\vec{\theta}$ and optional model inputs $\vec{u}(t)$.

7.2.3 Modulating effects on the metabolic network

The system of ordinary differential equations relate the various concentrations at the metabolic level. Rather than explicitly modeling the processes at the other levels (*e.g.* genome, proteome), their effects on the metabolic fluxes are captured by introducing a time-dependency of the metabolic parameters $\vec{\theta}(t)$ [1]. This time-dependency is implemented by dividing the simulation into steps and re-estimating the vector of parameters $\vec{\theta}(t + \delta t)$ at each time step.

$$\vec{X}[n] = \vec{x}(\Delta t, \vec{\theta}[n]) \quad \text{with} \quad \vec{x}(0) = \vec{X}[n-1] \tag{7.4}$$
$$\vec{Y}[n] = \vec{g}(\vec{X}[n], \vec{\theta}[n], \vec{u}) \tag{7.5}$$
$$\vec{X}[0] = \vec{x}_{ss}(\vec{\theta}[0]), \quad \vec{\theta}[0] \in \hat{\Theta}_0 \tag{7.6}$$

$\vec{X}$ and $\vec{Y}$ represent the internal model states and measurable model output predictions at the discretized time steps $0, \Delta t, ..., N\Delta t$ respectively. Each time step, the value at the previous time step $\vec{X}[n-1]$ is used as an initial condition for ODE simulation $\vec{x}(t)$, while the previous parameter set is used as initial parameter set for optimization. The term $\vec{x}_{ss}(\vec{\theta}[0])$ refers to the steady state solution corresponding to the initial parameters $\vec{\theta}[0]$. These initial parameters are drawn from the collection $\hat{\Theta}_0$ which contains parameter sets optimized for the untreated phenotype. An optimized parameter set $\hat{\vec{\theta}}[n]$ is defined as follows:

$$\hat{\vec{\theta}}[n] = \arg\min_{\vec{\theta}[n]} \left(\chi^2_d(\vec{\theta}[n]) + \lambda_r \chi^2_r(\vec{\theta}[n]) \right) \tag{7.7}$$

Here, χ^2_d represents the fidelity to the data and corresponds to the sum of squared differences between the data interpolants corresponding to a single spline and the associated model outputs, while χ^2_r represents a regularization term that minimizes parameter fluctuations. Here λ_r is a constant which determines the strength of the regularization term. The constant λ_r is chosen in

such a way that χ_r^2 reduces while biasing the sum of squared differences χ_d^2 as little as possible [1]. Here χ_d^2 and χ_r^2 are given by:

$$\chi_d^2(\vec{\theta}[n]) = \sum_{i=1}^{N_y} \left(\frac{Y_i[n] - d_i(n\Delta t)}{\sigma_i(n\Delta t)} \right)^2 \tag{7.8}$$

$$\chi_r^2(\vec{\theta}[n]) = \sum_{i=1}^{N_p} \left(\frac{\theta_i[n] - \theta_i[n-1]}{\Delta t} \frac{1}{\theta_i[0]} \right)^2 \tag{7.9}$$

where N_y is the number of observable quantities and, N_p is the number of parameters. This process is subsequently repeated for each set of splines, resulting in a sample of parameter trajectories. We shall refer to the distribution of these as Parameter Trajectory Distribution (PTD). The corresponding model outputs shall be denoted as $\mathbf{y}$ with elements $y_{i,j}(t)$. Here, the first index i refers to the output, while the latter j refers to the spline replicate the estimate was based on. For the sake of clarity, we omit denoting the dependence on time t.

7.2.4 Using parameter trajectories for experiment design

The PTD is a distribution of predicted trajectories conditioned on the available data. It forms a link between data, the parameter trajectories and the various predictions. The dynamics of the system and the effects of modulations from the proteome and transcriptome levels are constrained by the model, the experimental data and the regularizing constraint that ensures minimal adaptation. Therefore, the model, regularization and data implicitly impose non-trivial relations between the different predictions and parameter modulations. Since all of these quantities are linked, this also means that predictions corresponding to candidate experiments are related to our prediction of interest. The prediction we are interested in (biliary cholesterol excretion) is only accessible by means of an invasive procedure. The objective here is to find measurable quantities which relate to our prediction of interest and use these relations for experiment design.

Consider a new measurement. The additional data point that would be obtained is associated with an error model G which reflects the uncertainty associated with the new data point. Incorporating this additional data would subsequently lead to an additional constraint on the PTD, hence also affecting the prediction of interest. Measurement efficacy will be evaluated by determining its effect on the variance V_{post} of the prediction of interest. To this end, we predict the variance of the prediction of interest z after the new measurement data is added. Consider a new measurement of prediction n leading to a data point with a mean d and standard deviation σ. By weighting the different replicates in the PTD according to the new data point, we can obtain an estimate for the variance after the measurement were to be included. The predicted bias-corrected weighted variance after including the new data point is given by:

$$V_{post}\left(\mathbf{y}, z, n, d, \sigma\right) = \frac{ESS_n}{ESS_n - 1} \left(\sum_{i=1}^{N} w_{n,i} y_{z,i}^2 - \left(\sum_{i=1}^{N} w_{n,i} y_{z,i} \right)^2 \right) \tag{7.10}$$

where the individual weights are given by:

$$w_{n,i} = w_{n,i}(\mathbf{y}, d, \sigma) = \frac{G\left(y_{n,i}, d, \sigma\right)}{\sum_{k=1}^{N} G\left(y_{n,k}, d, \sigma\right)} \tag{7.11}$$

Here N corresponds to the number of spline replicates. The Effective Sample Size (ESS) used in the bias correction term (see Appendix 7.5 for further information) is given by:

$$ESS_n = \left(\sum_{i=1}^{N} w_{n,i}^2\right)^{-1} \tag{7.12}$$

Assuming Gaussian measurement errors, G is given by:

$$G\left(x, d, \sigma\right) = \exp\left(-\frac{(x-d)^2}{2\sigma^2}\right) \tag{7.13}$$

Subsequently, variance reduction V_{red} is calculated as:

$$V_{red}(\mathbf{y}, z, n, d, \sigma) = 1 - \frac{V_{post}\left(\mathbf{y}, z, n, d, \sigma\right)}{V_{pre}(y_z)} \tag{7.14}$$

where $V_{pre}(y_z)$ is the variance of the prediction of interest y_z before experiment design. *A priori*, it is unknown what the measured value d will be. However, by consecutively repeating this estimation procedure for all the replicates (substituting the predicted value corresponding to the current replicate as the measured value), a sample of predicted variances is obtained. This sample can subsequently guide the experimental design as it gives information on both the accuracy of the variance estimate and its expected value. Using this approach, it is possible to compute an expected variance reduction after the new measurement(s).

$$SVR(\mathbf{y}, z, n, \sigma_n) = 1 - \frac{1}{N}\sum_{i=1}^{N}\left(\frac{V_{post}\left(\mathbf{y}, z, n, y_{n,i}, \sigma_n\right)}{V_{pre}(y_z)}\right) \tag{7.15}$$

7.2.5 Implementation details

The mathematical model and parameter estimation routines were implemented in MATLAB (2010b, The MathWorks, Natick, Massachusetts). The ordinary differential equations were integrated using compiled MEX files based on numerical integrators from the SUNDIALS CVode package (2.6.0, Lawrence Livermore National Laboratory, Livermore, California) [30, 31]. Absolute and relative tolerances were set to 10^{-12}. The MATLAB non-linear least-squares optimization method LSQNONLIN, which uses an interior reflective Newton method, was used to estimate model parameters [32]. Both termination tolerances for the change in objective function and parameter estimates were set to 10^{-10}. The MATLAB function CSAPS was used to calculate the cubic smoothing splines.

7.3 Results

We presented a generally applicable computational approach to improve predictions of molecular adaptations by predicting the efficacy of additional experiments. In the present section, we shall demonstrate the method by applying it to a model of hepatic lipid and plasma lipoprotein metabolism upon pharmacological treatment of mice by LXR agonist T0901317. To demonstrate the capabilities of the proposed method and to test its applicability, the design is focused on model outputs that can experimentally be verified. A beneficial effect of LXR activation is the increased excretion of cholesterol from the body. The biliary system is an important route that facilitates the transport of cholesterol from the liver to the intestinal lumen, which precedes its subsequent excretion from the body. However, at present, the adaptation of the biliary cholesterol excretion flux upon LXR activation, and its contribution to the whole-body excretion of cholesterol, cannot be predicted accurately. In the following sections, optimal experiment design shall be performed to reduce the variance of the corresponding flux in the computational model (f_{29}).

7.3.1 Experimental data

A data set of C57BL/6J mice treated with T0901317 for $0, 1, 2, 4, 7, 14$, and 21 days was acquired and included in the computational analyses. A detailed description of the experimental materials is included in Appendix 7.5. In brief, the set contains quantitative measures of hepatic triglyceride, free cholesterol, and cholesterylester levels, as well as the fractional contribution of *de novo* lipogenesis. Furthermore, data of plasma concentrations of triglyceride, total cholesterol, HDL-cholesterol, and free fatty acids (FFA) were included. Data was obtained providing information on the production and clearance rates of VLDL particles, as well as the diameter and triglyceride/cholesterol composition ratio of these particles. Additional data for the wild-type phenotype was included, containing flux measurements of the hepatic uptake of cholesterol [33]. Information on the dietary intake of cholesterol and bile acids was obtained from [34].

7.3.2 Computational model

A mathematical multi-compartment model was constructed [1, 25], that integrates metabolic processes involved in hepatic and intestinal lipid metabolism, as well as plasma lipoprotein metabolism (see Figure 7.2). The mathematical model contains five compartments representing the liver, intestine, intestinal lumen, periphery, and blood plasma. The liver compartment includes the production, utilization and storage of triglycerides and cholesterols, as well as the mobilization of these metabolites to the endoplasmic reticulum where they are incorporated into nascent produced VLDL particles. The VLDL particles are secreted in the plasma where they provide nutrients for peripheral tissues. Moreover, the model includes the hepatic uptake of free fatty acids and reverse transport of cholesterol via HDL. In addition to the original model by Tiemann *et al* [25], we added processes involved in the efflux of cholesterol and bile acids to the intestinal lumen

where they are reabsorbed or excreted from the body. A detailed description of the model, including equations, is presented in Appendix 7.5.

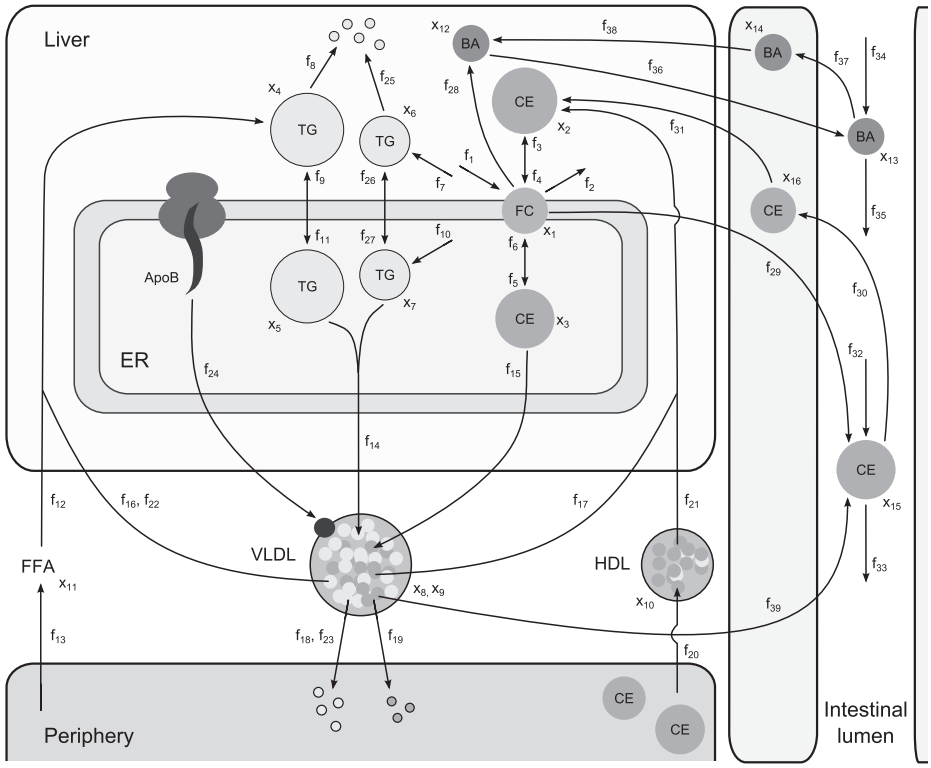

Figure 7.2: Multi-compartment model of lipid and cholesterol metabolism. A mathematical multi-compartment model was constructed which integrates metabolic processes involved in hepatic and intestinal lipid metabolism, as well as plasma lipoprotein metabolism. The mathematical model contains five compartments representing the liver, intestine, intestinal lumen, periphery, and blood plasma. The liver compartment includes the production, utilization and storage of triglycerides and cholesterols, as well as the mobilization of these metabolites to the endoplasmic reticulum where they are incorporated into nascent produced VLDL particles. The VLDL particles are subsequently secreted in the plasma where they provide nutrients for peripheral tissues. The model furthermore includes the hepatic uptake of free fatty acids from the plasma, and the reverse cholesterol transport pathway, *i.e.*, the net transport of cholesterol from peripheral tissues back to the liver via HDL. The model includes processes involved in the efflux of cholesterol and bile acids to the intestinal lumen where they are reabsorbed or excreted from the body. ApoB, apolipoprotein B; BA: bile acid; CE, cholesterylester; ER, endoplasmic reticulum; FFA, free fatty acid; FC, free cholesterol; HDL, high-density-lipoprotein; TG, triglyceride; VLDL, very low density lipoprotein

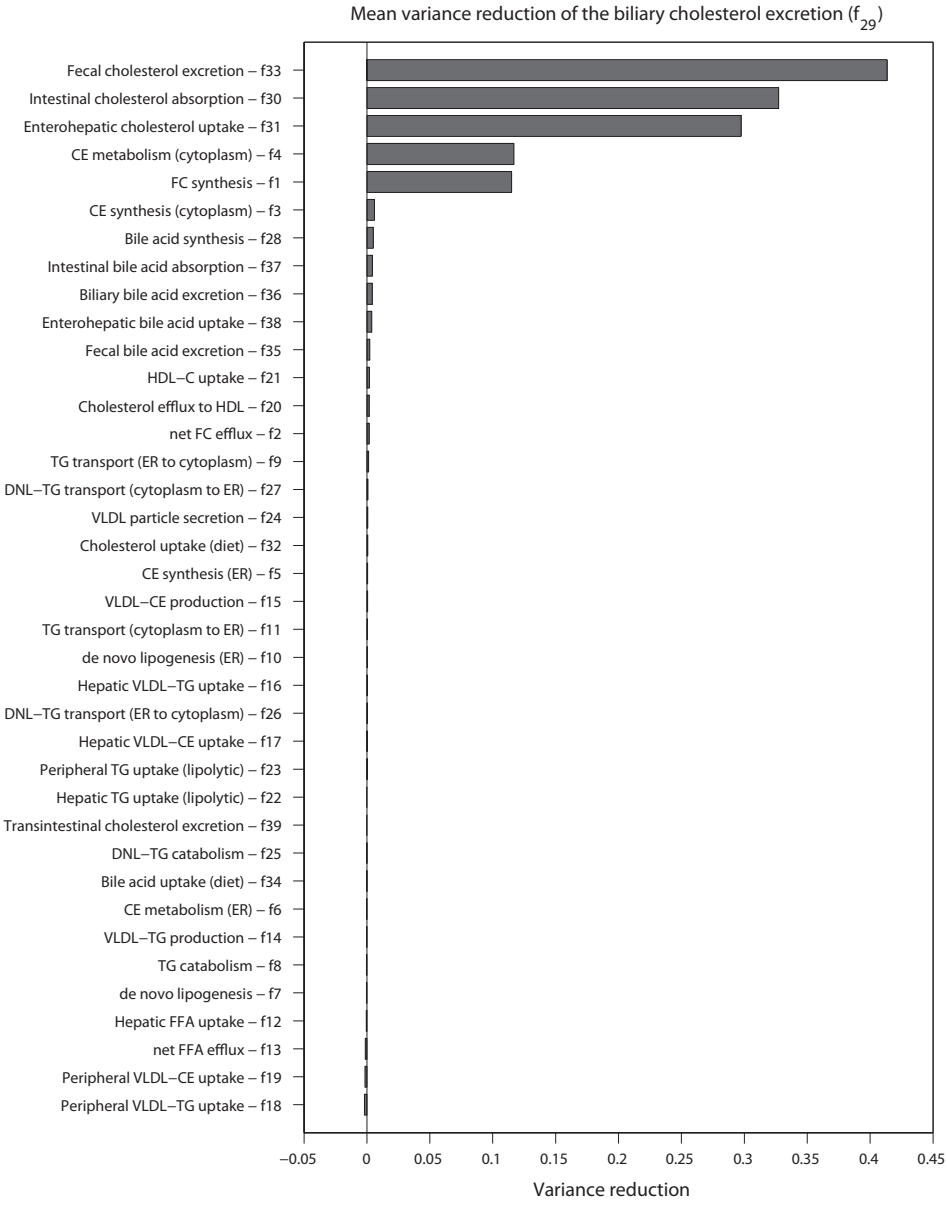

Figure 7.3: Experimental design was performed targeting a variance reduction of the biliary cholesterol excretion (f_{29}). Presented are the mean variance reductions for the different fluxes included in the mathematical model. The optimal quantity to measure would be flux f_{33} which represents the fecal cholesterol excretion.

7.3.3 Identification of the optimal quantity to measure

Experimental design was performed targeting a variance reduction of the biliary cholesterol excretion (f_{29}). We restricted the analysis to model fluxes. We design for the mean variance reduction over the entire duration of the treatment. The standard deviation of a new measurement was set to 0.5 [$\ln(\mu mol/h)$], which was based on the observed variability in similar experiments [35]. Figure 7.3 shows the mean variance reduction of f_{29} when considering measuring different fluxes in the model. From this figure, we can conclude that overall, the optimal quantity to measure would be flux f_{33}, which represents the fecal cholesterol excretion. An additional attractive aspect of measuring this flux is that the experiment would be non-invasive. After selecting measurements that could potentially lead to large variance reductions, time courses of expected reductions can be computed to determine which time points are optimal measurement candidates. An example is shown in Figure 7.4, which reveals that for this output flux (f_{29}), the expected variance reduction is fairly insensitive to the measurement time point. Here, the intervals indicate the central 67% of the distribution of variance reductions obtained.

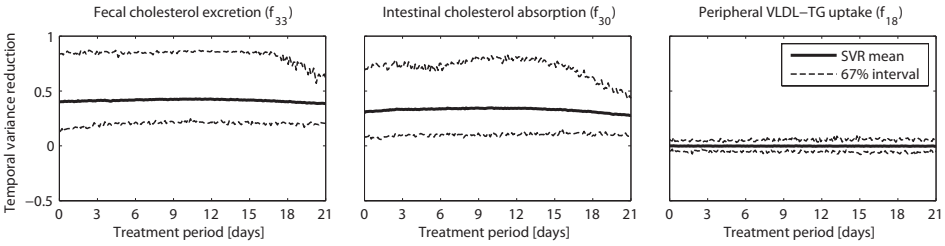

Figure 7.4: Time courses of expected variance reductions were computed to determine whether measuring particular time points result in larger variance reductions. Note that each figure title indicates the measured flux, while the variance reductions reported correspond to f_{29}. The intervals indicate the central 67% of the distribution of variance reductions obtained via the SVR method.

Subsequently, we investigated the efficacy of adding more than one additional measurement. When considering measurements on fecal matter, increasing the sampling rate is easier to achieve than to add a completely different experimental procedure. The predicted effects of including multiple measurements of f_{33} are shown in Figure 7.5. The number of additional measurements was varied from 2 to 6. The first two measurements correspond to the wild type and the end time of the treatment. Since dynamics are usually more pronounced early on during treatments, measurements were added in such a way that the early stage of the intervention was sampled more densely. This was accomplished by incrementally adding each additional measurement in the middle of the leftmost time interval. The left panel displays the temporal variance reductions for different numbers of additional measurements. The right panel displays the mean variance reduction over the entire time series.

As shown in Figure 7.5, 95% of the attainable variance reduction (52%) is

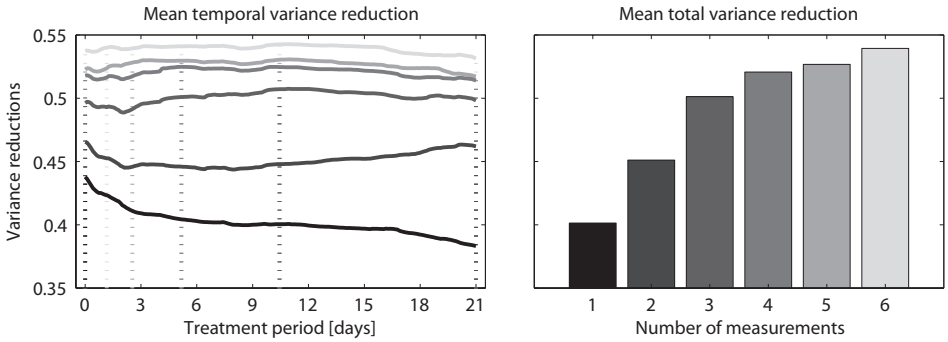

Figure 7.5: The effect of including multiple measurements of f_{33} on the variance reduction was investigated. The number of additional measurements was varied from 2 to 6. Measurements were added in such a way that the early stage of the intervention was sampled more densely. This was accomplished by incrementally adding each additional measurement in the middle of the leftmost time interval (indicated by the dotted lines). The left panel displays the temporal variance reductions for different numbers of additional measurements. The right panel displays the mean variance reduction over the entire time series.

already attained at the fourth measurement (compared to 54% with six measurements). Note how the effect of a single measurement progressively reduces in time. This is most apparent in the transition from only including the wild type, to including the time point at 21 days. Gradually increasing the number of time points improves the variance reduction over the rest of the time course.

7.3.4 Integration of additional measurements

The computational analysis revealed that including time course measurements of the fecal cholesterol excretion (f_{33}) would optimally reduce the prediction variance of the biliary cholesterol excretion (f_{29}). The optimal experiments were performed in C57BL/6J mice treated with T0901317 for $0, 1, 7$ and 14 days [35]. Subsequently, the additional data was included and the model analysis was repeated. Figure 7.6 shows the resulting distribution of the model predictions of f_{29}, before (left) and after (right) including the experiments. Note that before the additional experiments were included, these fluxes could not be predicted accurately. After incorporating the new data, a large reduction in uncertainty was obtained. Furthermore, these constrained predictions are in good agreement with data on f_{29} of untreated mice and mice treated with T0901317 for 14 days [35]. Note that latter data was not included in the optimization procedure, but serves as model validation. Also note that the experiments required to obtain a direct measurement of this flux are highly invasive and require more effort than the measurements on fecal matter.

To validate the methodology, we computed the expected change in variance for the other fluxes upon including the chosen experiments. To get an idea of the uncertainty associated with the true decrease in variance, we resampled the post-

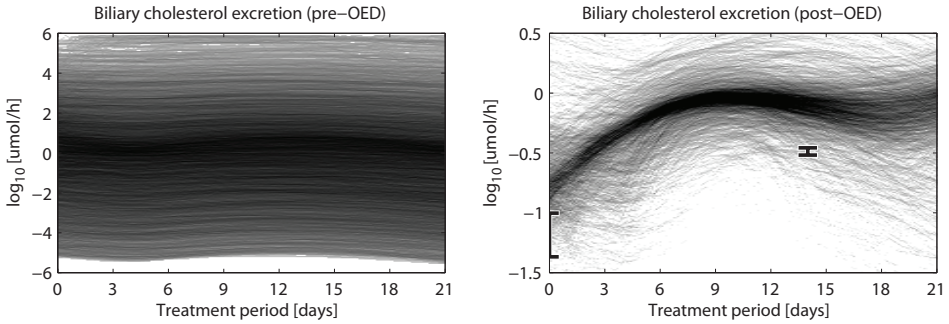

Figure 7.6: Data was acquired of the fecal cholesterol excretion (f_{33}) in C57BL/6J mice treated with T0901317 for 0, 1, 7, and 14 days [35]. Subsequently, the additional data was included and the model analysis was repeated. Plots indicate trajectory densities (2-dimensional histogram) of the biliary cholesterol excretion (f_{29}). Note how, before performing the experiments, the adaptation upon LXR activation could not be predicted accurately. After including the selected experiments, a large reduction in uncertainty is obtained. Furthermore, these constrained predictions are in good agreement with data of mice treated with T0901317 for 14 days [35]. Note that latter data was not included in the optimization procedure but serves as model validation.

experiment distribution. Different subsets are drawn from the post-experiment distribution and used to compute the variance. As shown in Figure 7.7, the predicted changes in variance were in good agreement with the true changes.

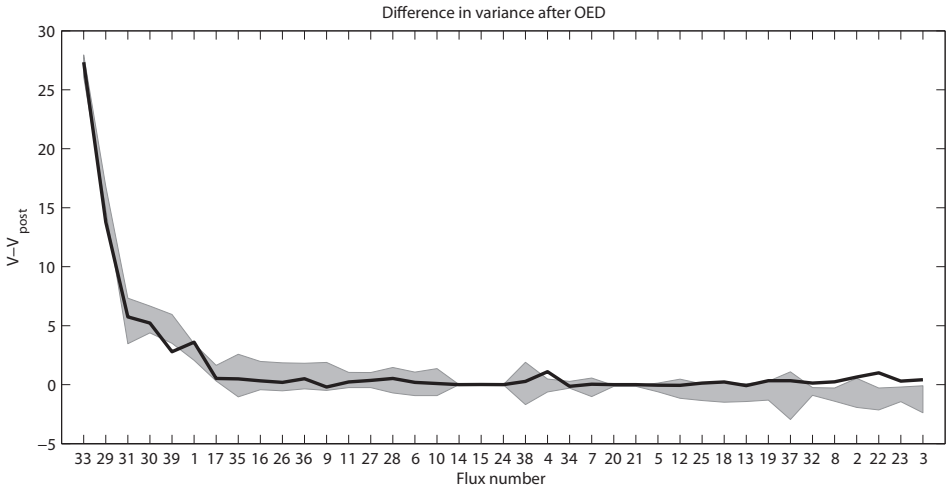

Figure 7.7: Change in variance of the fluxes which were not targeted. Grey: True change in variance. Black: Change in variance as predicted by the proposed methodology. Note that the x-axis is sorted according to the mean change in true variance.

7.3.5 Refining predictions of hepatic cholesterol clearance

LXR agonists promote the excretion of cholesterol from the body. Figure 7.6 (right part) shows that this adaptation is partly caused by a progressive increase of the biliary cholesterol excretion during the first week of the treatment. A topic for further research could be to study the processes and mechanisms that drive the increased biliary cholesterol excretion upon LXR activation. One hypothesis is that the increased hepatobiliary cholesterol excretion is accomodated by an increased uptake of cholesterol by the liver from the plasma. In this section, we aim to design an experiment to more accurately determine how much cholesterol is cleared by the liver from the plasma (7.16) over the treatment period.

$$F_C = V_{plasma}(f_{17} + f_{21}) + f_{31} \tag{7.16}$$

To investigate which measurements should be performed to decrease our uncertainty, we set up a 2D matrix representing combinations of potential flux measurements. We sample this matrix using an iterative approach. The sampling is initialized by randomly sampling the experiment matrix. Subsequently, the positions of a fraction $\nu N_{samples}$ of these samples are resampled proportionally to the determined variances. These positions are then perturbed and the variances at the new positions are computed. The remaining $(1-\nu)N_{samples}$ positions are discarded and replaced with randomly sampled positions. This procedure iteratively samples regions associated with higher variance reductions more densely. As shown in Figure 7.8, we see that specific fluxes are more optimal to measure than others. Note how measurements of the enterohepatic cholesterol uptake (f_{31}) would be most beneficial. Other cholesterol fluxes in- and out of the intestinal lumen (f_{29} and f_{30}) would also result in a large variance reduction. Finally, similar reductions can be attained by measuring both cholesteryl ester metabolism (f_4) and cholesteryl ester synthesis (f_3) in the cytoplasm simultaneously.

7.4 Conclusions and discussion

Progressive diseases and pharmacological interventions typically affect processes at various different levels. In many cases, the amount of topological information available is insufficient to specify a dynamic model. In this work, we presented a method to design experiments for identifying adaptations in metabolic networks. It is well suited to situations where insufficient information is available to mechanistically model the processes that drive these adaptations. The method works in two distinct steps. The first step employs a methodology known as ADAPT [25] to iteratively estimate time-dependent model parameters for a large number of simulated data replicates. ADAPT links the time scale on which the adaptations occur (typically long) with short-term dynamics. Whereas larger models are typically plagued by large topological and parametric uncertainties, lumped approximations are not always considered as useful for unraveling mechanistic detail. ADAPT tries to balance the two by incorporating a fairly detailed model of the measurable mass fluxes while estimating phenomenological changes of its kinetic parameters. As such, it

enables the incorporation of data at both these levels, allowing information to be carried over between different phenotypical snapshots [10]. The second step in our approach uses these time dependent parameter trajectory distributions to predict measurement efficacies to ascertain which experiment would lead to an optimal reduction of uncertainty. By iteratively refining the different trajectories, parameter adaptations can be identified. These can then serve as a tool to derive hypotheses on the interactions between various components of the system. Because the design criterion is based on simulated distributions of predictions and these can be computed for a wide range of model quantities, the approach is very flexible. Practical limitations of various measurement technologies often play an important role. Measurements often result in derived quantities rather than the model states themselves. Fortunately, the current approach allows incorporation of such measurements directly, since distributions of such experiments can be simulated. The method provides a means to design experiments which help elucidate the underlying mechanisms that drive the response to a treatment intervention.

Aside from its direct applicability, the approach also provides avenues for future research. Rather than computing spline interpolants, it may be possible to use parametric probability models such as Gaussian Processes [36]. A Gaussian Process is a collection of random variables defined by a mean and a covariance function which relate the different points in the collection. Using the available data in combination with Bayes Rule, the parameters of such a GP model are updated to a posterior distribution, from which samples can be drawn. Once such a Gaussian process is calibrated to data, it can be used to specify a distribution over the state derivatives. This means that simulating the system of ODEs would no longer be necessary since the derivatives could be used directly to infer a posterior parameter distribution. Similar use of GPs has previously been shown to result in drastic speedups in performing parameter inference on parameter-constant ODEs [37]. In this work, we used prediction variance as a measure for efficacy. One could argue that for strongly tailed distributions, variance has limited descriptive power. In such cases, transformation of the PTD before performing the experimental design may be desirable. Estimating variance reductions involves computing weights of the different samples. Here, it is sensible to compute some metric that reflects how many samples had an appreciable contribution to the estimate by determining the Effective Sample Size (ESS) [38]. Though we introduced a bias correction to account for bias at low effective sample sizes, extremely low values for the ESS still indicate imprecise variance estimates.

Relevant applications of the proposed method are the investigation of metabolic pathways in relation to progressive diseases such as Type 2 Diabetes and cardiovascular disease. In the present study, the method was applied to a multi-compartment model of lipid and cholesterol metabolism to study metabolic adaptations induced upon pharmacological activation of LXR in mice. The main focus was to gain additional insight in the beneficial effects of LXR activation on cholesterol metabolism, *i.e.*, the stimulation of the cellular efflux, transport, and excretion of cholesterol from the body. The biliary system plays an important role in this by facilitating the transport of cholesterol and bile acids

from the liver to the intestinal lumen. However, at present, the adaptation of the biliary cholesterol excretion flux upon LXR activation could not be predicted accurately (Figure 7.6, left part). Therefore, additional experiments were performed to reduce the variance of this flux (f_{29}). Here, the computational analysis revealed that the fecal cholesterol excretion (f_{33}) would be the optimal quantity to measure (Figure 7.3). An attractive aspect of measuring this flux is that the experiment would be non-invasive. We showed that the predicted reductions agreed well with the variances obtained after incorporating additional experimental data (Figure 7.7). One could argue that the predicted efficacy of including data on f_{33} is an expected result, as f_{29} and f_{33} both directly influence the cholesterol content in the intestinal lumen (x_{15}). However, note that including data on fluxes f_{32} and f_{39} (also acting on x_{15}), or fluxes f_2, f_3, f_5, and f_6 (acting on shared metabolite x_1) would not result in a significant variance reduction. The efficacy of including data on f_{33} is illustrated in Figure 7.6 (right part). The newly constrained prediction is in good agreement with independently measured data on f_{29} of untreated mice and mice treated with T0901317 for 14 days [35]. Furthermore, the computational analysis provided the additional insight that the biliary cholesterol excretion increased during the first week of the treatment and subsequently stabilized upon prolonged treatment. The method we described provides a means to design experiments for reducing specific prediction uncertainties in dynamical systems where many of the regulatory mechanisms are poorly known. By reducing the uncertainties in these adaptations, we showed that it is possible to iteratively refine hypotheses on how the metabolic state changes during treatment intervention. Once sufficiently refined, such trajectory distributions may help disentangle the effects regulatory mechanisms have on various parts of the system, using only data on their collective consequences.

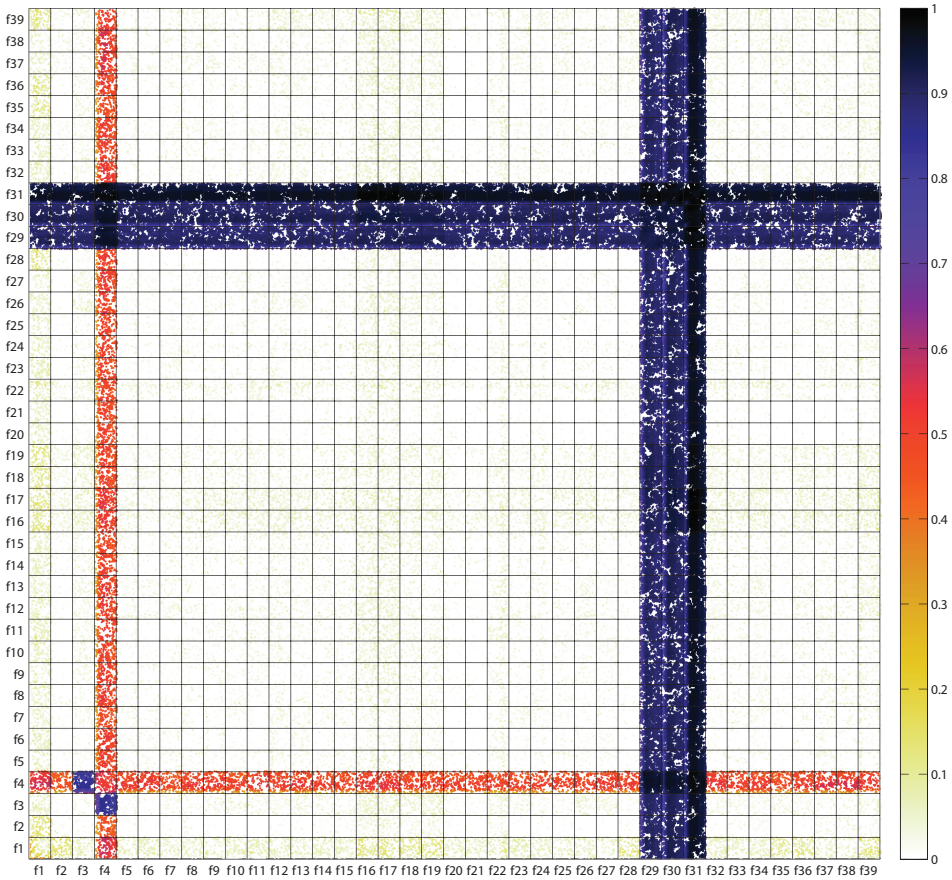

Figure 7.8: Expected variance reduction for the cholesterol clearance from the plasma when considering two measurements. Each axis corresponds to a different experiment and the color value indicates the expected variance reduction. Each subdivision indicates a different measured flux, while the space within each subdivision corresponds to the different time points.

7.5 Appendix

Mathematical model

A mathematical model was constructed, which integrates metabolic processes involved in hepatic and intestinal lipid metabolism, as well as plasma lipoprotein metabolism (Figure 7.2). The model contains five compartments representing the liver, intestine, intestinal lumen, periphery, and blood plasma. The liver compartment includes the production, utilization and storage of triglycerides and cholesterols, as well as the mobilization of these metabolites to the endoplasmic reticulum where they are incorporated into nascent produced VLDL particles. The VLDL particles are subsequently secreted in the plasma where they provide nutrients for peripheral tissues. The model includes the hepatic uptake of free fatty acids from the plasma, and the reverse cholesterol transport pathway, *i.e.*, the net transport of cholesterol from peripheral tissues back to the liver via HDL. Moreover, the model includes processes involved in the efflux of cholesterol and bile acids to the intestinal lumen where they are reabsorbed or excreted from the body.

State	Name	Description
x_1	x_{FC}	Hepatic free cholesterol
x_2	$x_{CE_{cyt}}$	Hepatic cholesteryl ester (cytosol)
x_3	$x_{CE_{ER}}$	Hepatic cholesteryl ester (ER)
x_4	$x_{TG_{cyt}}$	Hepatic triglyceride (cytosol)
x_5	$x_{TG_{ER}}$	Hepatic triglyceride (ER)
x_6	$x_{TGdnl_{cyt}}$	Hepatic *de novo* triglyceride (cytosol)
x_7	$x_{TGdnl_{ER}}$	Hepatic *de novo* triglyceride (ER)
x_8	$x_{TG_{VLDL}}$	Plasma VLDL-triglyceride
x_9	$x_{C_{VLDL}}$	Plasma VLDL-cholesterol
x_{10}	$x_{C_{HDL}}$	Plasma HDL-cholesterol
x_{11}	x_{FFA}	Plasma free fatty acid
x_{12}	$x_{BA_{hep}}$	Hepatic bile acids
x_{13}	$x_{BA_{lum}}$	Intestinal lumen bile acids
x_{14}	$x_{BA_{int}}$	Intestinal bile acids
x_{15}	$x_{C_{lum}}$	Intestinal lumen cholesterol
x_{16}	$x_{C_{int}}$	Intestinal cholesterol

Table 7.1: Description of the state variables included in the model.

Model equations

The mathematical model contains sixteen metabolic species $\vec{x}$ (Table 7.1) interlinked by forty-one flux interactions (Table 7.2). The flux equations are based on mass-action kinetics. The ordinary differential equations are given by:

$$\begin{aligned}
\frac{d[x_{FC}]}{dt} &= F_{FC_{prod}} + F_{CEdef_{cyt}} + F_{CEdef_{ER}} - F_{FC_{met}} - F_{CEfor_{cyt}} - F_{CEfor_{ER}} \\
&\quad - F_{BAprod_{hep}} - F_{Cprod_{lum}} \\
\frac{d[x_{CE_{cyt}}]}{dt} &= F_{CEfor_{cyt}} + F_{Cupt_{hep}} - F_{CEdef_{cyt}} + V_{plasma}\left(F_{CEupt_{hep}} + F_{CEupt_{HDL}}\right) \\
\frac{d[x_{CE_{ER}}]}{dt} &= F_{CEfor_{ER}} - F_{CEdef_{ER}} - F_{VLDL-CE} \\
\frac{d[x_{TG_{cyt}}]}{dt} &= F_{TGfor_{cyt}} - F_{TGfor_{ER}} - F_{TGmet_{cyt}} \\
&\quad + V_{plasma}\left(\frac{F_{FFA_{upt}}}{3} + F_{TGupt_{hep}} + F_{TGhyd_{hep}}\right) \\
\frac{d[x_{TG_{ER}}]}{dt} &= F_{TGfor_{ER}} - F_{TGfor_{cyt}} - F_{VLDL-TGndnl} \\
\frac{d[x_{TGdnl_{cyt}}]}{dt} &= F_{TGdnl_{cyt}} - F_{TGdnlmet_{cyt}} + F_{TGdnlfor_{cyt}} - F_{TGdnlfor_{ER}} \\
\frac{d[x_{TGdnl_{ER}}]}{dt} &= F_{TGdnl_{ER}} + F_{TGdnlfor_{ER}} - F_{TGdnlfor_{cyt}} - F_{VLDL-TGdnl} \\
\frac{d[x_{TG_{VLDL}}]}{dt} &= \frac{F_{VLDL-TG}}{V_{plasma}} - F_{TGupt_{hep}} - F_{TGupt_{per}} - F_{TGhyd_{hep}} - F_{TGhyd_{per}} \\
\frac{d[x_{C_{VLDL}}]}{dt} &= \frac{F_{VLDL-CE}}{V_{plasma}} - F_{C_{TICE}} - F_{CEupt_{hep}} - F_{CEupt_{per}} \\
\frac{d[x_{C_{HDL}}]}{dt} &= F_{CEfor_{HDL}} - F_{CEupt_{HDL}} \\
\frac{d[x_{FFA}]}{dt} &= F_{FFA_{prod}} - F_{FFA_{upt}} \\
\frac{d[x_{BA_{hep}}]}{dt} &= F_{BAprod_{hep}} - F_{BAprod_{lum}} + F_{BAupt_{hep}} \\
\frac{d[x_{BA_{lum}}]}{dt} &= F_{BAprod_{lum}} + F_{BA_{diet}} - F_{BA_{fecal}} - F_{BAupt_{int}} \\
\frac{d[x_{BA_{int}}]}{dt} &= F_{BAupt_{int}} - F_{BAupt_{hep}} \\
\frac{d[x_{C_{lum}}]}{dt} &= F_{Cprod_{lum}} + V_{plasma}F_{C_{TICE}} + F_{C_{diet}} - F_{C_{fecal}} - F_{Cupt_{int}} \\
\frac{d[x_{C_{int}}]}{dt} &= F_{Cupt_{int}} - F_{Cupt_{hep}}
\end{aligned}$$

The blood plasma volume, given by V_{plasma}, was assumed to be 1 mL [39].

Table 7.2: Overview and description of the fluxes included in the mathematical model.

Flux	Name	Equation	Description
f_1	F_{FCprod}	p_1	Hepatic *de novo* synthesis of free cholesterol
f_2	F_{FCmet}	$p_2[x_{FC}]$	Net hepatic catabolism of free cholesterol
f_3	$F_{CEfor_{cyt}}$	$p_3[x_{FC}]$	Hepatic synthesis of cholesteryl ester (cytosol)
f_4	$F_{CEdef_{cyt}}$	$p_4[x_{CE_{cyt}}]$	Hepatic conversion of cholesteryl ester (cytosol) to free cholesterol
f_5	$F_{CEfor_{ER}}$	$p_5[x_{FC}]$	Hepatic synthesis of cholesteryl ester (ER)
f_6	$F_{CEdef_{ER}}$	$p_6[x_{CE_{ER}}]$	Hepatic conversion of cholesteryl ester (ER) to free cholesterol
f_7	$F_{TGdnl_{cyt}}$	p_7	Hepatic *de novo* synthesis of triglyceride (cytosol)
f_8	$F_{TGmet_{cyt}}$	$p_8[x_{TG_{cyt}}]$	Hepatic catabolism of triglyceride (cytosol)
f_9	$F_{TGfor_{cyt}}$	$p_9[x_{TG_{ER}}]$	Hepatic transport of triglyceride from the ER to the cytosol
f_{10}	$F_{TGdnl_{ER}}$	p_{10}	Hepatic *de novo* synthesis of triglyceride (ER)
f_{11}	$F_{TGfor_{ER}}$	$p_{11}[x_{TG_{cyt}}]$	Hepatic transport of triglyceride from the cytosol to the ER
f_{12}	$F_{FFA_{upt}}$	$p_{12}[x_{FFA}]$	Hepatic uptake of free fatty acid
f_{13}	$F_{FFA_{prod}}$	p_{13}	Net efflux of free fatty acid from peripheral tissues to plasma
f_{14}	$F_{VLDL-TG}$	$p_{14}([x_{TG_{ER}}] + [x_{TGdnl_{ER}}])$	Hepatic secretion rate of VLDL-triglyceride
f_{15}	$F_{VLDL-CE}$	$p_{15}[x_{CE_{ER}}]$	Hepatic secretion rate of VLDL-cholesterol
f_{16}	$F_{TGupt_{hep}}$	$p_{16}[x_{TG_{VLDL}}]$	Hepatic uptake of triglyceride via whole-particle uptake
f_{17}	$F_{CEupt_{hep}}$	$p_{16}[x_{C_{VLDL}}]$	Hepatic uptake of cholesterol via whole-particle uptake
f_{18}	$F_{TGupt_{per}}$	$p_{17}[x_{TG_{VLDL}}]$	Peripheral uptake of triglyceride via whole-particle uptake
f_{19}	$F_{CEupt_{per}}$	$p_{17}[x_{C_{VLDL}}]$	Peripheral uptake of cholesterol via whole-particle uptake
f_{20}	$F_{CEfor_{HDL}}$	p_{20}	Peripheral efflux of cholesterol to HDL particles
f_{21}	$F_{CEupt_{HDL}}$	$p_{21}[x_{C_{HDL}}]$	Hepatic uptake of HDL-cholesterol
f_{22}	$F_{TGhyd_{hep}}$	$p_{18}[x_{TG_{VLDL}}]$	Hepatic uptake of triglyceride via lipolytic enzymes
f_{23}	$F_{TGhyd_{per}}$	$p_{19}[x_{TG_{VLDL}}]$	Peripheral uptake of triglyceride via lipolytic enzymes

Continued on next page

Table 7.2 – continued from previous page

Flux	Name	Equation	Description
f_{24}	$F_{apoB_{prod}}$	p_{22}	Hepatic secretion rate of apolipoprotein B
f_{25}	$F_{TGdnlmet_{cyt}}$	$p_8[x_{TGdnl_{cyt}}]$	Hepatic catabolism of *de novo* triglyceride (cytosol)
f_{26}	$F_{TGdnlfor_{cyt}}$	$p_9[x_{TGdnl_{ER}}]$	Hepatic transport of *de novo* triglyceride from the ER to the cytosol
f_{27}	$F_{TGdnlfor_{ER}}$	$p_{11}[x_{TGdnl_{cyt}}]$	Hepatic transport of *de novo* triglyceride from the cytosol to the ER
f_{28}	$F_{BAprod_{hep}}$	$p_{23}[x_{FC}]$	Hepatic bile acid synthesis
f_{29}	$F_{Cprod_{lum}}$	$p_{27}[x_{FC}]$	Biliary cholesterol excretion
f_{30}	$F_{Cupt_{int}}$	$p_{28}[x_{C_{lum}}]$	Intestinal cholesterol absorption
f_{31}	$F_{Cupt_{hep}}$	$p_{29}[x_{C_{int}}]$	Hepatic cholesterol uptake (enterohepatic)
f_{32}	$F_{C_{diet}}$	p_{32}	Cholesterol intake via diet
f_{33}	$F_{C_{fecal}}$	$p_{33}[x_{C_{lum}}]$	Fecal cholesterol excretion
f_{34}	$F_{BA_{diet}}$	p_{30}	Bile acid intake via diet
f_{35}	$F_{BA_{fecal}}$	$p_{31}[x_{BA_{lum}}]$	Fecal bile acid excretion
f_{36}	$F_{BAprod_{lum}}$	$p_{24}[x_{BA_{hep}}]$	Biliary bile acid excretion
f_{37}	$F_{BAupt_{int}}$	$p_{25}[x_{BA_{lum}}]$	Intestinal bile acid absorption
f_{38}	$F_{BAupt_{hep}}$	$p_{26}[x_{BA_{int}}]$	Hepatic bile acid uptake (enterohepatic)
f_{39}	$F_{C_{TICE}}$	$p_{34}[x_{C_{VLDL}}]$	Transintestinal cholesterol excretion
f_{40}	$F_{VLDL-TGndnl}$	$p_{14}[x_{TG_{ER}}]$	Hepatic secretion rate of non *de novo* VLDL-triglyceride
f_{41}	$F_{VLDL-TGdnl}$	$p_{14}[x_{TGdnl_{ER}}]$	Hepatic secretion rate of *de novo* VLDL-triglyceride

Calculation of the VLDL particle diameter

The following approach was used to calculate nascent VLDL particle diameters (D_{VLDL}). Since each VLDL particle contains one apolipoprotein B particle, the number of triglyceride and cholesterylester molecules per VLDL particle can be determined by correcting the specific lipid fluxes for the number of apolipoprotein B proteins. The core volume of a VLDL particle was subsequently determined assuming a molecular volume of 946.84 mL/mol for triglyceride (TG_{mv}) and a molecular volume of 685.48 ml/mol for cholesterylester (CE_{mv}) [40]. A core radius (R_c) was calculated from the core volume assuming a spherical shape of the VLDL particles. Furthermore, the particle membrane accounts for an additional two nanometers (R_s) [41].

$$D_{VLDL} = 2\,(R_c + R_s) \tag{7.17a}$$

$$R_c = \sqrt[3]{\frac{3V_c}{4\pi}} \tag{7.17b}$$

$$V_c = 10^{21}\,\frac{TG_{cnt}\cdot TG_{mv} + CE_{cnt}\cdot CE_{mv}}{N_A} \tag{7.17c}$$

$$TG_{cnt} = \frac{F_{VLDL-TG}}{F_{apoB_{prod}}} \tag{7.17d}$$

$$CE_{cnt} = \frac{F_{VLDL-CE}}{F_{apoB_{prod}}} \tag{7.17e}$$

Where N_A is the constant of Avogadro.

Calculation of de novo lipogenesis

The fractional contribution of *de novo* lipogenesis was calculated as follows in the computational model:

$$FC_{DNL}(t) = \frac{[x_{TGdnl_{cyt}}](t) + [x_{TGdnl_{ER}}](t)}{[x_{TG_{cyt}}](t) + [x_{TG_{ER}}](t) + [x_{TGdnl_{cyt}}](t) + [x_{TGdnl_{ER}}](t)} \tag{7.18}$$

Calculation of the VLDL catabolic rate

The VLDL catabolic rate was calculated as follows in the computational model:

$$CR_{VLDL}(t) = \frac{p_{16}(t) + p_{17}(t)}{p_{16}(t_0) + p_{17}(t_0)} \tag{7.19}$$

Experimental procedures

Animals and experimental design

Male C57Bl/6J mice (Charles River, L'Arbresle Cedex, France) were housed in a light- and temperature-controlled facility (lights on 6:30 AM-6:30 PM, 21 °C) and fed a standard laboratory chow diet (RMH-B, Abdiets, Woerden, The Netherlands) containing T0901317 (0.015% wt/wt; $\sim$ 50 mg/kg) for 1, 2, 4, 7, 14, or 21 days. Untreated controls received non-supplemented laboratory chow. After 1 day, 7 days, and prior to sacrifice on day 14, a small blood sample was taken from 4-h fasted (8-12 AM) mice by tail bleeding to evaluate plasma lipoprotein profiles. All animals had free access to drinking water. During the final 24-hours of the treatment period, the different groups of mice received sodium [1-^{13}C]-acetate (99 atom %, Isotec/Sigma-Aldrich, St. Louis, MO) via the drinking water (2%). On the last treatment day, 4-h fasted (8-12 AM) animals were sacrificed by cardiac puncture under isoflurane anaesthesia. Livers were quickly removed, freeze-clamped and stored at -80 °C. Blood was centrifuged (4000xg for 10 minutes at 4 °C) and plasma was stored at -20 °C. Experimental procedures were approved by the Ethics Committee for Animal Experiments of the University of Groningen.

Liver and plasma metabolites and plasma lipoprotein analysis

Plasma lipoproteins were separated by fast protein liquid chromatography (FPLC) gel filtration using a superose 6 column (GE Healthcare, Uppsala, Sweden) [42]. Triglyceride contents of the collected FPLC fractions were determined using a commercially available kit (Roche Diagnostics, Mannheim, Germany). Plasma non-esterified fatty acid profiles were analyzed as previously described [43]. Frozen liver was homogenized in ice-cold PBS. Hepatic triglyceride and total cholesterol contents were assessed using commercial available kits (Roche Diagnostics) after lipid extraction [44]. Hepatic triglyceride fractions were obtained from lipid extracts using Isolute SPE NH_2 columns (Biotage AB, Uppsala, Sweden) [45,46].

Fractional contribution of *de novo* lipogenesis

Hepatic triglyceride fractions obtained by lipid extraction were hydrolyzed, and the free fatty acids were extracted and derivatized [47]. The fatty acid-mass isotopomer distributions were determined by GC-MS and used in mass isotopomer distribution analysis (MIDA) to calculate fractional palmitate and oleate synthesis rates from *de novo* lipogenesis [47].

Quantification of VLDL-TG production rates

Separate groups of mice were injected intravenously with Triton WR1339 (0.5 g/kg body weight) as a 125 mg/mL solution in PBS after a 4-hour fast (8-12 AM). Blood samples were drawn by retro-orbital bleeding into heparinized tubes at 0, 30, 60, 120, and 240 min after injection. After the last blood draw, animals

were sacrificed by cardiac puncture under isoflurane anaesthesia. Blood was centrifuged (10 minutes, 4000xg) to obtain plasma. Plasma triglyceride levels and triglyceride production rates were determined as described [48]. Nascent VLDL ($d< 1.006$) was isolated from the final plasma sample of each animal using a Optima TM LX tabletop ultracentrifuge (Beckman Instruments Inc., Palo Alto, CA) at 108,000 rpm for 150 minutes.

Determination of nascent VLDL composition and particle size

Triglyceride, cholesterol and phospholipid concentrations of the nascent VLDL particles were determined using commercially available kits (Roche Diagnostics and Wako Chemicals). Protein concentrations were determined using the BCA Kit (Pierce, Rockford, IL). VLDL particle diameter D was estimated according to [49] using the following formula: $D = 60 \cdot ((0.211 \cdot TG/PL) + 0.27))$, where D is given in [nm].

Quantification of VLDL catabolic rates

VLDL was isolated from pooled plasma of fasting healthy human subjects by ultracentrifugation ($d< 1.006$). VLDL was iodinated using the iodine monochloride method [50]. Free iodine was removed by passing over a PD-10 column (GE Healthcare, Diegem, Belgium) followed by extensive dialysis against PBS. More than 95% of the VLDL radioactivity was precipitable by trichloroacetic acid, and less than 6% of the radioactivity was associated with the lipid fraction of VLDL. Separate groups of mice were injected with 0.5 Ci of 125I-VLDL via the tail vein after a 4-hour fast (8-12 AM). Blood samples were taken by retro-orbital bleeding after 0, 15, 30, 60, 120 and 240 minutes and plasma radioactivity was determined using a Cobra II γ counter (Packard Instruments, Downers Grove, IL). Plasma decay curves for the tracer were generated by dividing plasma radioactivity at each time point by the radioactivity present at the initial 1min time point. Fractional catabolic rates were calculated from the area under the plasma disappearance curves fitted to a bicompartmental model using SAAM-II (version 1.2.1; SAAM Institute, University of Washington, Seattle, WA) [42].

Immunoblotting procedures

Protein concentrations in liver homogenates containing protease inhibitors (Complete; Roche Diagnostics) were determined using the BCA Kit (Pierce). Volumes of VLDL containing equal amounts of triglyceride were pooled and lipids were extracted with methanol and cold ether. The remaining VLDL proteins were subjected to SDS-PAGE. Apolipoprotein B100 and apolipoprotein B48 were determined using antibodies against antimouse apoB raised in rabbit (Biodesign, Saco, ME). Horseradish peroxidase-conjugated antirabbit antibodies from donkey (Amersham Pharmacia Bioscience, GE Healthcare) was used as a secondary antibody for all immunoblots. Protein bands were detected using SuperSignal West Pico Chemiluminescent Substrate System (Pierce). Band-densities were determined by using a Gel Doc XR system (Biorad, Hercules CA, USA).

Monte Carlo sampling of data interpolants

To enable the estimation of dynamic trajectories of metabolic parameters and fluxes, continuous dynamic descriptions of the experimental data were used as input for the computational approach. For this purpose, cubic smoothing splines were calculated that describe the dynamic trend of the experimental data. To account for experimental and biological uncertainties a collection of splines was calculated using a Monte Carlo approach. Random samples based on the experimental data were generated assuming Gaussian distributions with means and standard deviations of the data. Subsequently, for each generated sample a cubic smoothing spline was calculated. An overview of the experimental data is presented in Figure 7.9. A darker color represents a higher density of trajectories in that specific region and time point. The data is represented by means $\pm$ standard deviations, with an exception for the experimental data obtained via FPLC measurements. These measurements were performed on pooled mice plasma and are represented by the white dots. Measures of spread used for the Monte Carlo sampling of these quantities were estimated based on similar experiments that were performed [35]. An overview of the quantities that were experimentally observed and its relation to corresponding model components is presented in Table 7.3. Note that model output y_{13} was only experimentally observed for the untreated phenotype [33].

Table 7.3: Overview of the quantities that were measured and its relation to corresponding model components.

Measurement	Output	Equation
Hepatic triglyceride	y_1	$[x_{TG_{cyt}}] + [x_{TG_{ER}}] + [x_{TGdnl_{cyt}}] + [x_{TGdnl_{ER}}]$
Hepatic cholesteryl ester	y_2	$[x_{CE_{cyt}}] + [x_{CE_{ER}}]$
Hepatic free cholesterol	y_3	$[x_{FC}]$
Plasma total cholesterol	y_4	$[x_{C_{VLDL}}] + [x_{C_{HDL}}]$
HDL-cholesterol	y_5	$[x_{C_{HDL}}]$
Plasma triglyceride	y_6	$[x_{TG_{VLDL}}]$
Plasma free fatty acid	y_7	$[x_{FFA}]$
VLDL TG/C ratio	y_8	$\frac{TG_{cnt}}{CE_{cnt}}$
VLDL diameter	y_9	D_{VLDL}
VLDL-TG production	y_{10}	$F_{VLDL-TG}$
VLDL catabolic rate	y_{11}	CR_{VLDL}
De novo lipogenesis	y_{12}	FC_{DNL}
Hepatic HDL-C uptake	y_{13}	$F_{CEupt_{HDL}}$

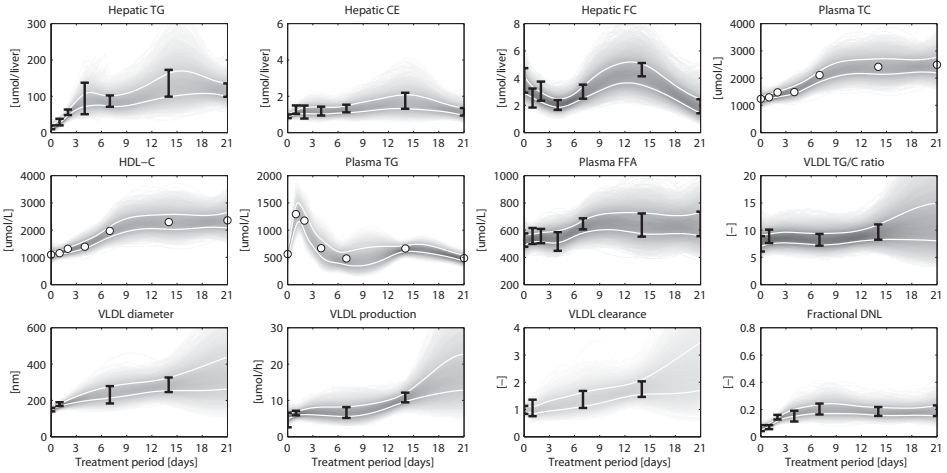

Figure 7.9: Experimental data and interpolants. An overview of the experimental data, as well as corresponding $2D$ histograms of the splines that were used as input for ADAPT, is presented. Data is represented by means $\pm$ standard deviations (N=5-6), with an exception for the experimental data obtained via FPLC measurements. These measurements were performed on pooled mice plasma and are represented by the white dots. The white lines enclose the central 67% of the interpolant density at each time point.

Bias correction weighted variance

The bias correction for the weighted variance can be derived as follows

$$
\begin{aligned}
E\left[\sigma_b^2\right] &= E\left[\sum_i w_i (x_i - \hat{\mu})^2\right] \\
&= E\left[\sum_i w_i \left((x_i - \mu) - (\hat{\mu} - \mu)\right)^2\right] \\
&= E\left[\sum_i w_i \left(x_i - \mu\right)^2\right] + E\left[-2\sum_i w_i (x_i - \mu)(\hat{\mu} - \mu) + \sum_i w_i (\hat{\mu} - \mu)^2\right] \\
&= \sigma^2 - E\left[(\hat{\mu} - \mu)^2\right] \\
&= \sigma^2 - E\left[\left(\sum_i w_i (x_i - \mu)\right)^2\right] \\
&= \sigma^2 - E\left[\sum_i \sum_j w_i w_j (x_i - \mu)(x_j - \mu)\right] \\
&= \sigma^2 - E\left[\sum_i \sum_j w_i w_j \left(x_i x_j - (x_i + x_j)\mu + \mu^2\right)\right] \\
&= \sigma^2 - \sum_i \sum_j w_i w_j \left(E[x_i x_j] - E[(x_i + x_j)]\mu + \mu^2\right)
\end{aligned}
\tag{7.20}
$$

For $i \neq j$ the double summation cancels out since

$$E[x_i x_j] = E[x_i]E[x_i] = E[x_i]^2 = \mu^2 \tag{7.21}$$

which leads to

$$\mu^2 - 2\mu^2 + \mu^2 = 0 \tag{7.22}$$

For $i = j$

$$E[x_i x_i] = E[x_i^2]. \tag{7.23}$$

therefore

$$\sigma_b^2 = \sigma^2 - \sum_i w_i^2 \left(E[x_i^2] - E[(x_i)]^2\right) = \sigma^2 - \sum_i w_i^2 \sigma^2 \tag{7.24}$$

The unbiased weighted variance is given by

$$\sigma^2 = \frac{1}{1 - \sum_i w_i^2} \sigma_b^2 \tag{7.25}$$

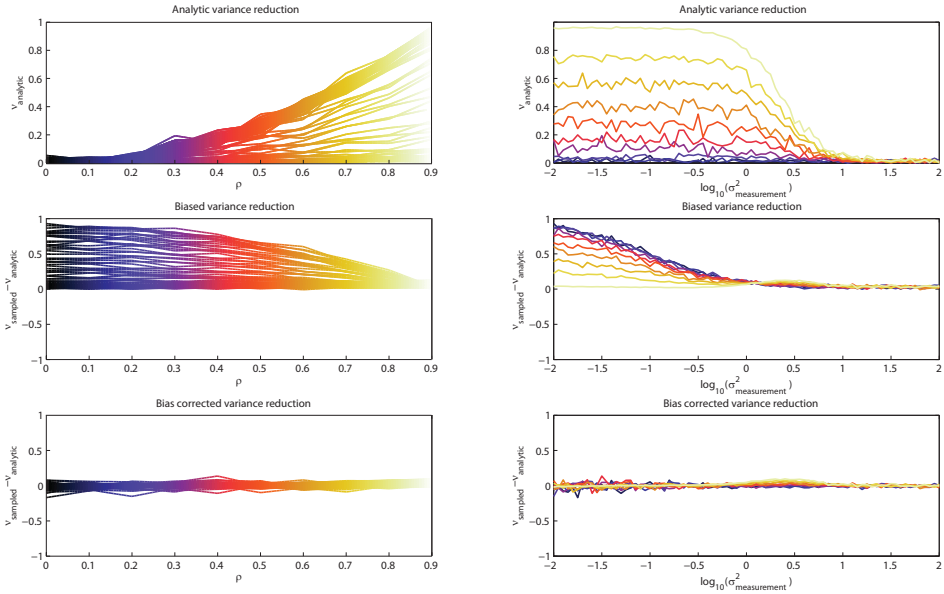

Figure 7.10: On the top row, variance reductions based on the analytic solution are shown. On the second row, the difference between the analytic and sampled solution are shown for the biased estimates. The bottom row shows the difference to the analytic solution after bias correction.

With the effective sample size (ESS) defined as:

$$ESS = \left(\sum_{i=1}^{N} w_i^2 \right)^{-1} \tag{7.26}$$

This results in:

$$\sigma^2 = \frac{ESS}{ESS - 1} \sigma_b^2 \tag{7.27}$$

To test the bias correction, expected variance reductions were computed for several multivariate Gaussians. Each experiment, a distribution of 500 samples was simulated using means $[5, 10, 11]$ and standard deviations $[40, 10, 20]$. The correlation coefficients between different components were $[0.1, \rho_{1,3}, 0.5]$ where the correlation coefficient between observable 1 and 3 was varied. Accuracies for the measurement of state 3 were also varied. Differences between analytical and sampled estimates are depicted in Figure 7.10.

As shown in Figure 7.10, without the correction, bias is more pronounced when there is little correlation and little measurement uncertainty. Without correction for the low ESS, these result in artificially large estimates for the variance reduction. On the bottom row the unbiased estimate is shown.

References

[1] Tiemann C, Vanlier J, Hilbers P, van Riel N: **Parameter adaptations during phenotype transitions in progressive diseases**. *BMC Systems Biology* 2011, **5**:174.

[2] Vanlier J, Tiemann C, Hilbers P, van Riel N: **An integrated strategy for prediction uncertainty analysis**. *Bioinformatics* 2012, **28**(8):1130–1135.

[3] Vanlier J, Tiemann C, Hilbers P, van Riel N: **A Bayesian approach to targeted experiment design**. *Bioinformatics* 2012, **28**(8):1136–1142.

[4] Schmitz J, Van Riel N, Nicolay K, Hilbers P, Jeneson J: **Silencing of glycolysis in muscle: experimental observation and numerical analysis**. *Experimental physiology* 2010, **95**(2):380–397.

[5] Schilling M, Maiwald T, Hengl S, Winter D, Kreutz C, Kolch W, Lehmann W, Timmer J, Klingmüller U: **Theoretical and experimental analysis links isoform-specific ERK signalling to cell fate decisions**. *Mol Syst Biol* 2009, **5**:334.

[6] Borisov N, Aksamitiene E, Kiyatkin A, Legewie S, Berkhout J, Maiwald T, Kaimachnikov N, Timmer J, Hoek J, Kholodenko B: **Systems-level interactions between insulin–EGF networks amplify mitogenic signaling**. *Mol Syst Biol* 2009, **5**:256.

[7] Cedersund G, Roll J, Ulfhielm E, Danielsson A, Tidefelt H, Strålfors P: **Model-based hypothesis testing of key mechanisms in initial phase of insulin signaling**. *PLoS Comput Biol* 2008, **4**(6):799–806.

[8] Koschorreck M, Gilles E: **Mathematical modeling and analysis of insulin clearance in vivo**. *BMC Syst Biol* 2008, **2**:43.

[9] Schoeberl B, Eichler-Jonsson C, Gilles E, Müller G: **Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors**. *Nat Biotechnol* 2002, **20**(4):370–375.

[10] van Riel NA, Tiemann CA, Vanlier J, Hilbers PA: **Applications of analysis of dynamic adaptations in parameter trajectories**. *Interface Focus* 2013, **3**(2).

[11] Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmüller U, Timmer J: **Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood**. *Bioinformatics* 2009, **25**(15):1923.

[12] Brännmark C, Palmér R, Glad S, Cedersund G, Strålfors P: **Mass and information feedbacks through receptor endocytosis govern insulin signaling as revealed using a parameter-free modeling framework**. *Journal of Biological Chemistry* 2010, **285**(26):20171.

[13] Girolami M, Calderhead B: **Riemann manifold langevin and hamiltonian monte carlo methods**. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2011, **73**(2):123–214.

[14] Calderhead B, Girolami M: **Statistical analysis of nonlinear dynamical systems using differential geometric sampling methods**. *Interface Focus* 2011, **1**(6):821–835.

[15] Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP: **Universally Sloppy Parameter Sensitivities in Systems Biology Models**. *PLoS Comput Biol* 2007, **3**(10):e189.

[16] Kreutz C, Raue A, Timmer J: **Likelihood based observability analysis and confidence intervals for predictions of dynamic models**. *BMC Systems Biology* 2012, **6**:120.

[17] Gomez-Cabrero D, Compte A, Tegner J: **Workflow for generating competing hypothesis from models with parameter uncertainty**. *Interface Focus* 2011, **1**(3):438.

[18] Cedersund G, Roll J: **Systems biology: model based evaluation and comparison of potential explanations for given biological data**. *FEBS Journal* 2009, **276**(4):903–922.

[19] Faller D, Klingmüller U, Timmer J: **Simulation methods for optimal experimental design in systems biology**. *Simulation* 2003, **79**(12):717.

[20] Rodriguez-Fernandez M, Mendes P, Banga J: **A hybrid approach for efficient and robust parameter estimation in biochemical pathways**. *Biosystems* 2006, **83**(2-3):248–265.

[21] Casey F, Baird D, Feng Q, Gutenkunst R, Waterfall J, Myers C, Brown K, Cerione R, Sethna J: **Optimal experimental design in an epidermal growth factor receptor signalling and down-regulation model**. *Systems Biology, IET* 2007, **1**(3):190–202.

[22] Oosterveer M, Grefhorst A, Groen A, Kuipers F: **The liver X receptor: Control of cellular lipid homeostasis and beyond: Implications for drug design**. *Prog Lipid Res* 2010, **49**(4):343–352.

[23] Grefhorst A, Elzinga B, Voshol P, Plösch T, Kok T, Bloks V, van der Sluijs F, Havekes L, Romijn J, Verkade H, et al.: **Stimulation of lipogenesis by pharmacological activation of the liver X receptor leads to production of large, triglyceride-rich very low density lipoprotein particles**. *J Biol Chem* 2002, **277**(37):34182–34190.

[24] Grefhorst A, Parks E: **Reduced insulin-mediated inhibition of VLDL secretion upon pharmacological activation of the liver X receptor in mice**. *J Lipid Res* 2009, **50**(7):1374–1383.

[25] Tiemann C, Vanlier J, Oosterveer M, Groen A, Hilbers P, van Riel N: **Parameter Trajectory Analysis to Identify Treatment Effects of Pharmacological Interventions**. *PLoS Computational Biology* in press.

[26] Joshi M, Seidel-Morgenstern A, Kremling A: **Exploiting the bootstrap method for quantifying parameter confidence intervals in dynamical systems**. *Metabolic Engineering* 2006, **8**(5):447–455.

[27] DiCiccio T, Efron B: **Bootstrap confidence intervals**. *Statistical Science* 1996, :189–212.

[28] Limpert E, Stahel W, Abbt M: **Log-normal distributions across the sciences: keys and clues**. *BioScience* 2001, **51**(5):341–352.

[29] Kreutz C, Rodriguez M, Maiwald T, Seidl M, Blum H, Mohr L, Timmer J: **An error model for protein quantification**. *Bioinformatics* 2007, **23**(20):2747.

[30] van Riel N: **Speeding up simulations of ODE models in Matlab using CVode and MEX files** 2012.

[31] Hindmarsh A, Brown P, Grant K, Lee S, Serban R, Shumaker D, Woodward C: **SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers**. *ACM T Math Software* 2005, **31**(3):363–396.

[32] Coleman T, Li Y: **An Interior Trust Region Approach for Nonlinear Minimization Subject to Bounds**. *SIAM J Optimiz* 1996, **6**:418–445.

[33] Xie C, Turley S, Dietschy J: **ABCA1 plays no role in the centripetal movement of cholesterol from peripheral tissues to the liver and intestine in the mouse**. *Journal of lipid research* 2009, **50**(7):1316–1329.

[34] van de Pas N, Woutersen R, van Ommen B, Rietjens I, de Graaf A: **A physiologically-based kinetic model for the prediction of plasma cholesterol concentrations in the mouse**. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids* 2011, **1811**(5):333–342.

[35] Grefhorst A, Oosterveer M, Brufau G, Boesjes M, Kuipers F, Groen A: **Pharmacological LXR activation reduces presence of SR-B1 in liver membranes contributing to LXR-mediated induction of HDL-cholesterol**. *Atherosclerosis* 2012.

[36] Kirk PDW, Stumpf MPH: **Gaussian process regression bootstrapping: exploring the effects of uncertainty in time course data**. *Bioinformatics* 2009, **25**(10):1300–1306.

[37] Calderhead B, Girolami M, Lawrence N: **Accelerating Bayesian inference over nonlinear differential equations with Gaussian processes**. *Advances in neural information processing systems* 2009, **21**:217–224.

[38] Del Moral P, Doucet A, Jasra A: **Sequential monte carlo samplers**. *Journal of the Royal Statistical Society: Series B(Statistical Methodology)* 2006, **68**(3):411–436.

[39] Rand M: **Handling, restraint, and techniques of laboratory rodents**. *Department of Animal Care, University of Arizona* 2001.

[40] Teerlink T, Scheffer P, Bakker S, Heine R: **Combined data from LDL composition and size measurement are compatible with a discoid particle shape**. *J Lipid Res* 2004, **45**(5):954–966.

[41] Miller A, Smith L: **Activation of lipoprotein lipase by apolipoprotein glutamic acid**. *J Biol Chem* 1973, **248**(9):3359–3362.

[42] Tietge UJ, Maugeais C, Cain W, Grass D, Glick JM, de Beer FC, Rader DJ: **Overexpression of secretory phospholipase A2 causes rapid catabolism and altered tissue uptake of high density lipoprotein cholesteryl ester and apolipoprotein AI**. *Journal of Biological Chemistry* 2000, **275**(14):10077–10084.

[43] Duran M, Bruinvis L, Ketting D, de Klerk J, Wadman S: **Cis-4-decenoic acid in plasma: a characteristic metabolite in medium-chain acyl-CoA dehydrogenase deficiency.** *Clinical chemistry* 1988, **34**(3):548–551.

[44] Bligh E, Dyer WJ: **A rapid method of total lipid extraction and purification**. *Canadian journal of biochemistry and physiology* 1959, **37**(8):911–917.

[45] Kaluzny M, Duncan L, Merritt M, Epps D: **Rapid separation of lipid classes in high yield and purity using bonded phase columns.** *Journal of Lipid Research* 1985, **26**:135–140.

[46] Hoving EB, Jansen G, Volmer M, Van Doormaal JJ, Muskiet FA: **Profiling of plasma cholesterol ester and triglyceride fatty acids as their methyl esters by capillary gas chromatography, preceded by a rapid aminopropyl-silica column chromatographic separation of lipid classes**. *Journal of Chromatography B: Biomedical Sciences and Applications* 1988, **434**(2):395–409.

[47] Oosterveer M, Van Dijk T, Tietge U, Boer T, Havinga R, Stellaard F, Groen A, Kuipers F, Reijngoud D: **High fat feeding induces hepatic fatty acid elongation in mice**. *PLoS One* 2009, **4**(6):e6066.

[48] Gautier T, Tietge UJ, Boverhof R, Perton FG, Le Guern N, Masson D, Rensen PC, Havekes LM, Lagrost L, Kuipers F: **Hepatic lipid accumulation in apolipoprotein CI-deficient mice is potentiated by cholesteryl ester transfer protein**. *Journal of lipid research* 2007, **48**:30–40.

[49] Fraser R: **Size and lipid composition of chylomicrons of different Svedberg units of flotation.** *Journal of lipid research* 1970, **11**:60–65.

[50] Fidge NH, Poulis P: **Studies on the radioiodination of very low density lipoprotein obtained from different mammalian species**. *Clinica Chimica Acta* 1974, **52**:15–26.

8 Future Perspective

8.1 General perspective

Over the recent years, Systems Biology has seen a striking increase in the complexity of the models under investigation. Since close collaborations between modelers and experimentalists are still rare, the available data suitable for dynamic modeling has not followed this trend. The time and resource consuming nature of data generation is still a bottleneck. Especially when considering the additional effort that is required to make it suitable for modeling. The incorporation of a new data set typically involves introducing additional parameters and simulation schemes to mimic the experimental procedure performed to obtain the data [1]. Often data is expressed as ratios, is measured in steady state or requires scaling factors in order to match observations with model outputs [2,3]. Furthermore, as shown in Chapter 6, lack of knowledge regarding the experimental conditions can render data ill suited for inference due to the fact that incorporating the data requires the addition of so many additional unknown parameters that the data hardly provides any constraints on the other parameter values. The situation is further exacerbated by the fact that the biological questions we wish to answer necessitate models of a certain complexity. Consequently, there is often a mismatch between the information content in the data and the complexity of the model whose parameters we wish to infer. Such a mismatch can result in large uncertainties in the predictions.

Interestingly, this gives rise to various approaches for mathematical modeling. Some physically based modelers believe that all model structures and associated parameters can be defined *a priori* and that additional parameter estimation or uncertainty analysis is not necessary or even undesirable since the predictions are based on a physically correct model of reality. The model parameters are often taken from literature data or estimated from different experiments. In this paradigm, predictions which deviate from their observed counterparts constitute a failure of the model and result in extending or changing the model. These approaches seem a remnant from an extremely reductionist paradigm whose aim was to characterize the individual components and subsequently reconstitute the system from its parts. Though appealing because of its simplicity, such a standpoint does not seem maintainable considering several recent publications on the effects of parameter uncertainty in these models [2,4–9]. These (and other) publications revealed that even medium-sized models are often plagued by highly uncertain predictions. Rejecting a model based on such highly uncertain predictions is unwarranted and results in unnecessarily complex models. Moreover, parameter values in literature are rarely reported along with an assessment of their identifiability. An additional complication with this approach is that enzymes and proteins tend to behave differently *in vivo* than *in vitro* [10]. Whether this is merely attributable to missing interaction mechanisms [11], post-translational modifications, the composition of the experimental medium [12,13] or variations between cells [14] is unclear.

Another popular approach is to make subjective decisions on which parameters are considered adjustable and perform parameter estimation and model sensitivity analyses over predefined physiological ranges [1,3,15,16]. This form of uncertainty analysis is highly pragmatic and can work well when one ensures

that all uncertainties relevant to the problem under investigation are sufficiently probed, and, the assumptions regarding the physiological ranges are justified. The effects of assumptions such as the adjustability of a parameter or the consequences of changing the predetermined ranges or assumed prior distributions are rarely reported in literature. More often than not, the information reported in scientific literature contains insufficient information to reproduce analyses, which makes subsequent follow-up work unnecessarily cumbersome and error prone.

Finally, there is a third group of modelers who quantify the uncertainty associated with their models and data, propagating these uncertainties to the various predictions [14, 17–19]. In some cases, such uncertainties can even be averaged over multiple models. These approaches are generally hampered by a significantly larger computational burden when compared with more straightforward point estimates or sensitivity analyses. Though recent advances in computational resources and methodology have made these approaches somewhat more tractable, such uncertainty analyses are often performed at the expense of mechanistic detail and have mostly been performed on models that either describe systems on a small scale or which are fairly rough approximations to reality. Rough approximations, which in terms of revealing mechanistic detail, are not always considered as useful as the original (albeit uncertain) model.

The scientific literature which deals with the methodology required to perform uncertainty analysis is often rather technical and specific research problems require specialized approaches. The threshold for deciding to invest in gaining the required expertise to apply such methodologies is high for a group which is mostly interested in studying a biological process or phenomenon. Especially considering that the short term payoff (more certain conclusions and more informative experiments) is not immediately evident. Though recent years have seen some improvements in terms of deciding on standards for model-based conclusions in computational biology, there is still no consensus on what is considered acceptable. One could argue that the field needs more accessible and user-friendly methods (or packages) since diagnosing and tackling both numerical and theoretical issues still requires a great deal of practical experience and patience. Some software packages exist (*e.g.* BioBayes, ABC-Sysbio, PottersWheel) but user documentation is typically sparse and solutions to potential pitfalls are rarely documented. Consequently, it seems more realistic to tackle biochemical problems in multidisciplinary teams containing at least one expert on inference and prediction.

Any form of uncertainty analysis contains subjective decisions and it is unrealistic to demand complete objectivity. Depending on the formalism of choice, necessary choices range from prior distributions to assumptions regarding the observational errors or which topologies and reaction mechanisms to include in the analysis. An important issue for future research is that such assumptions and their associated motivations have to be made explicit, so that their validity can be discussed and follow-up studies can be performed. Hopefully projects such as *MIASE*, a guideline for reporting simulation experiments [20], will lead to full disclosure of such information. This way, approaches that turn out to be effective in practice can be adopted by others and a consensus on required standards for

different types of analyses can be reached.

Despite the difficulties it poses, the future of the field of Systems Biology depends on accurately exploring and reporting the magnitude and consequences of the uncertainty present in our mathematical models, inferences and predictions. It is also crucial to find a means to communicate how these uncertainties affect the information we can extract from additional experimental data to experimental partners, so that they can measure or control a sufficient number of boundary conditions (*i.e.* experimental conditions) to improve the informativeness of the acquired data. We have shown that the statistical uncertainties in our mathematical models can typically be probed and used to design experiments which effectively reduce those uncertainties relevant to the question being answered. It has also become clear that specific combinations of experiments are vastly more informative than others due to the relations that exist between the various uncertainties. As demonstrated, knowledge of the uncertainties in the model can drastically improve how informative a new set of measurements will be, and, can save both time and money.

This thesis provided some practical approaches particularly suitable for modeling biochemical systems. Various aspects of the proposed methodologies would benefit from additional research, however. The following sections list a few avenues that can be explored in the future.

8.2 DGAT model

Chapter 6 presented a first generation model of the diacylglycerol transferase system. The available data was not informative enough to make precise predictions, however. Neither the relative contribution of the DGAT enzymes or the steady state levels of DAG and FA in the cytosol and endoplasmic reticulum could be predicted with finite uncertainty bounds. During the analysis, it became apparent that better characterization of the different experimental conditions is required to be able to incorporate the activity data. Additionally, rough estimates of the model boundary conditions would be beneficial for inference. The DGAT enzymes form the last and most important step in the production of triglycerides. Having a well parameterized mechanistic model of triglyceride production could be used to improve the predictions of the model presented in Chapter 7. Including a mechanistic model of the DGAT enzymes would allow incorporation of different DGAT perturbation experiments when performing inference on the overarching model.

8.3 Sequential importance sampling

The approach presented in Chapter 4 can suffer from under-sampling issues during the importance sampling step. This is especially an issue when multiple highly informative experiments are planned simultaneously. Recent advances in Sequential Monte Carlo sampling which involve resampling such a distribution and then subsequently perturbing and reweighting the different samples could potentially be used to increase the effective sample sizes at a reasonable cost.

In Chapters 4, 5 and 7, the design of experiments was separated from sampling from the posterior predictive or parameter trajectory distribution. One advantage of this approach is that searching for the optimal experiment amongst a list of experimental candidates does not require any additional model simulations. Moreover, algorithms based on sequential importance sampling could be considered for the overarching problem of finding the best experiment. As a proof of principle, a sampler based on ideas from Sequential Importance Sampling was implemented and presented in the Appendix of Chapter 5. Even with relatively simplistic perturbation kernels, improved sampling efficiences were attained. Nevertheless, finding kernels better suited to these types of problems and determining optimal settings that work well in higher dimensions provide an interesting avenue for future research.

8.4 Better approximations to the posterior predictive

The Posterior Predictive Distribution contains a wealth of information regarding the relations between model predictions, parameters and data. The methodology discussed in Chapters 2 and 3 can aid in obtaining a representative sample from this distribution. Subsequently, this sample was used in Chapter 4 to predict distributions of variance reductions. Although this is indeed a measure of prediction spread, this quantity is less descriptive when considering very poorly constrained, non-symmetric distributions. Another issue with these distributions is that the tail is typically sampled rather sparsely. One avenue for future research is finding suitable approximating functions for these probability densities in high dimensions. Aside from the beneficial effect it would have on the model selection procedure presented in Chapter 5, it would also allow extension of the method in Chapter 4 to other quantities for determining information gain such as the Kullback-Leibler divergence upon a new measurement or the Mutual Information between different predictions.

Finding such density approximations and studying their applicability to study the relations between uncertainties in the model remains an open problem however. In low dimensions, and for well behaved distributions, Kernel Density Estimation (KDE) is often employed (see Chapter 5). Here the probability density at a point is estimated by considering a sum of probability kernels K placed on the individual samples. Consider the extreme case where two predictions are functionally related versus an uncorrelated pair. It is clear that the former will be more susceptible to bias as the structure is sharper and therefore more sensitive to the use of diffuse kernels for density estimation. One avenue to explore could be kernel density estimation on Riemannian manifolds [21] for the parameters and extending such methodology to the predictions via local linearization. Alternatively one could consider estimators based on k-Nearest Neighbours (kNN) such as the ones used in Chapter 5. Though these typically tend to scale with the local density of the distribution quite naturally, kNN estimators suffer from slower convergence than their KDE counterparts [22].

8.5 Parameter trajectory distributions

In Chapter 7, a methodology was introduced for performing experiment design in cases where specific regulatory mechanisms are unknown. By iteratively re-estimating parameters in time, distributions of parameter values with respect to time are obtained. Currently, this is performed by combining a Monte Carlo re-sampling of the data in combination with successive optimization. Although this was sufficient for the presented case, where all the parameters were continually reoptimized, it is problematic when attempting to keep certain parameters fixed. This is problematic since the optimization is performed locally, while a fixed parameter should be estimated over the entire time course. Future work could make use of a fully probabilistic approach by using Gaussian Processes to obtain time dependent distributions of the data. Smoothing could be handled more implicitly by specifying specific covariance functions, while non-identifiability could be mediated by assuming appropriate prior distributions on the parameters. Since the Gaussian Process also specifies a distribution over the state derivatives, simulating the system of ODEs would no longer be necessary since the relation between the states and their derivatives are known. Similar use of GPs has previously been shown to result in drastic speedups in performing parameter inference on parameter-constant ODEs [23]. Another alternative would be to parameterize the parameter trajectories using certain low-order approximations and use classical inferential techniques to obtain posterior parameter trajectory distributions.

8.6 Concluding remarks

Unraveling the mechanisms that drive biological systems is a challenging task. Trying to infer mathematical models from experimental data is fraught with challenges that often require subjective decisions. Though new experimental and computational technologies are helping us to more reliably and objectively infer system properties and make predictions, both biological variability and system complexity complicate analyses of larger systems considerably. It seems that there is still a gap between groups working on methodology and groups involved in the development of new models and measurement technologies. Hopefully, the contributions presented in this thesis will help consolidate these somewhat.

References

[1] Wu F, Jeneson J, Beard D: **Oxidative ATP synthesis in skeletal muscle is controlled by substrate feedback**. *American Journal of Physiology-Cell Physiology* 2007, **292**:C115–C124.

[2] Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmüller U, Timmer J: **Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood**. *Bioinformatics* 2009, **25**(15):1923.

[3] Groenendaal W, Schmidt K, von Basum G, van Riel N, Hilbers P: **Modeling glucose and water dynamics in human skin**. *Diabetes technology & therapeutics* 2008, **10**(4):283–293.

[4] Cedersund G, Roll J: **Systems biology: model based evaluation and comparison of potential explanations for given biological data**. *FEBS Journal* 2009, **276**(4):903–922.

[5] Vanlier J, Tiemann C, Hilbers P, van Riel N: **An integrated strategy for prediction uncertainty analysis**. *Bioinformatics* 2012, **28**(8):1130–1135.

[6] Brännmark C, Palmér R, Glad S, Cedersund G, Strålfors P: **Mass and information feedbacks through receptor endocytosis govern insulin signaling as revealed using a parameter-free modeling framework**. *Journal of Biological Chemistry* 2010, **285**(26):20171.

[7] Girolami M, Calderhead B: **Riemann manifold langevin and hamiltonian monte carlo methods**. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2011, **73**(2):123–214.

[8] Calderhead B, Girolami M: **Statistical analysis of nonlinear dynamical systems using differential geometric sampling methods**. *Interface Focus* 2011, **1**(6):821–835.

[9] Bachmann J, Raue A, Schilling M, Böhm M, Kreutz C, Kaschek D, Busch H, Gretz N, Lehmann W, Timmer J, et al.: **Division of labor by dual feedback regulators controls JAK2/STAT5 signaling over broad ligand range**. *Molecular systems biology* 2011, **7**.

[10] Teusink B, Passarge J, Reijenga C, Esgalhado E, van der Weijden C, Schepper M, Walsh M, Bakker B, van Dam K, Westerhoff H, et al.: **Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry**. *European Journal of Biochemistry* 2000, **267**(17):5313–5329.

[11] van Eunen K, Kiewiet JA, Westerhoff HV, Bakker BM: **Testing biochemistry revisited: how in vivo metabolism can be understood from in vitro enzyme kinetics**. *PLoS computational biology* 2012, **8**(4):e1002483.

[12] Beard DA, Qian H: *Chemical biophysics: quantitative analysis of cellular systems*. Cambridge University Press 2008.

[13] Vanlier J, Wu F, Qi F, Vinnakota K, Han Y, Dash R, Yang F, Beard D: **BISEN: Biochemical simulation environment**. *Bioinformatics* 2009, **25**(6):836–837.

[14] Kalita MK, Sargsyan K, Tian B, Paulucci-Holthauzen A, Najm HN, Debusschere BJ, Brasier AR: **Sources of cell-to-cell variability in canonical nuclear factor-κB (NF-κB) signaling pathway inferred from single cell dynamic images**. *Journal of Biological Chemistry* 2011, **286**(43):37741–37757.

[15] Schmitz J, Van Riel N, Nicolay K, Hilbers P, Jeneson J: **Silencing of glycolysis in muscle: experimental observation and numerical analysis**. *Experimental physiology* 2010, **95**(2):380–397.

[16] van Eunen K, Bouwman J, Daran-Lapujade P, Postmus J, Canelas A, Mensonides F, Orij R, Tuzun I, van den Brink J, Smits G, et al.: **Measuring enzyme activities under standardized in vivo-like conditions for systems biology**. *FEBS Journal* 2010, **277**(3):749–760.

[17] Xu T, Vyshemirsky V, Gormand A, von Kriegsheim A, Girolami M, Baillie G, Ketley D, Dunlop A, Milligan G, Houslay M, et al.: **Inferring signaling pathway topologies from multiple perturbation measurements of specific biochemical species**. *Science Signalling* 2010, **3**(113).

[18] Becker V, Schilling M, Bachmann J, Baumann U, Raue A, Maiwald T, Timmer J, Klingmuller U: **Covering a broad dynamic range: information processing at the erythropoietin receptor**. *Science Signalling* 2010, **328**(5984):1404.

[19] Bachmann J, Raue A, Schilling M, Becker V, Timmer J, Klingmüller U: **Predictive mathematical models of cancer signalling pathways**. *Journal of internal medicine* 2012, **271**(2):155–165.

[20] Waltemath D, Adams R, Beard DA, Bergmann FT, Bhalla US, Britten R, Chelliah V, Cooling MT, Cooper J, Crampin EJ, et al.: **Minimum information about a simulation experiment (MIASE)**. *PLoS computational biology* 2011, **7**(4):e1001122.

[21] Pelletier B: **Kernel density estimation on Riemannian manifolds**. *Statistics & probability letters* 2005, **73**(3):297–304.

[22] Budka M, Gabrys B, Musial K: **On accuracy of PDF divergence estimators and their applicability to representative data sampling**. *Entropy* 2011, **13**(7):1229–1266.

[23] Calderhead B, Girolami M, Lawrence N: **Accelerating Bayesian inference over nonlinear differential equations with Gaussian processes**. *Advances in neural information processing systems* 2009, **21**:217–224.

Markov Chain Monte Carlo Sampler

A

This chapter contains some of the implementational details of the Markov Chain Monte Carlo (MCMC) sampler used in this thesis. MCMC is a technique used to obtain samples from probability distributions known only up to a normalizing factor. Given that $p(\vec{\theta})$ is a non-negative integrable function, the Metropolis Hastings algorithm will provide a sequence of samples (also known as chain) whose equilibrium distribution is proportional to $p(\vec{\theta})$ using only evaluations of $p(\vec{\theta})$. A sufficient condition to ensure that $p(\vec{\theta})$ is the equilibrium distribution is that the sampler satisfies detailed balance (A.1) where $\pi(.)$ corresponds to the invariant distribution and $k(x, y)$ corresponds to the distribution used for proposing the next step (the proposal distribution). This property ensures that the chain is reversible (two sides are equal).

$$\pi(x)k(x, y) = \pi(y)k(y, x) \tag{A.1}$$

The Metropolis algorithm ensured detailed balance by using only symmetric proposals (where the probability density of going from x to y is equal to the probability density of going from y to x, *i.e.* the proposal distribution does not change). A generalization by Hastings lead to an additional term in the acceptance probability which ensures detailed balance for non-symmetric proposal distributions. The resulting algorithm, named the Metropolis-Hastings algorithm proceeds by iteratively performing a number of steps:

- 1. Generate a sample $\vec{\theta}_{n+1}$ from a proposal distribution based on the current state.
- 2. Compute the likelihood $p(\mathbf{y^D}|\vec{\theta}_{n+1})$ at $\vec{\theta}_{n+1}$ and calculate $\tilde{p}(\vec{\theta}_{n+1}|\mathbf{y^D}) = p(\mathbf{y^D}|\vec{\theta}_{n+1})p(\vec{\theta}_{n+1})$, where $p(\vec{\theta}_{n+1})$ refers to the prior probability density function.
- 3. Draw a random number γ from a uniform distribution between 0 and 1 and accept the new step if $\gamma < \min\left(\frac{\tilde{p}(\vec{\theta}_{n+1}|\mathbf{y^D})Q(\vec{\theta}_{n+1}\rightarrow\vec{\theta}_n)}{\tilde{p}(\vec{\theta}_n|\mathbf{y^D})Q(\vec{\theta}_n\rightarrow\vec{\theta}_{n+1})}, 1\right)$.

The ratio of Q is known as the Hastings correction and ensures detailed balance. It corrects for the fact that the probability density going from parameter set $\vec{\theta}_n$ to $\vec{\theta}_{n+1}$ and $\vec{\theta}_{n+1}$ to $\vec{\theta}_n$ is unequal when the proposal distribution depends on the current parameter set. It is defined as the ratio between the proposal densities associated with going from iteration n to $n+1$ and $n+1$ to n. The apparent simplicity of the algorithm makes it conceptually attractive.

Proposals

To accelerate convergence, our implementation adapts to the local geometry of the problem by taking larger steps in directions where the posterior probability density does not change much. Implementationally, we employ an adaptive Gaussian proposal distribution whose covariance matrix is based on a quadratic approximation to the logarithm of the probability density function [1]. This matrix is computed by taking the inverse of an approximation to the Hessian matrix.

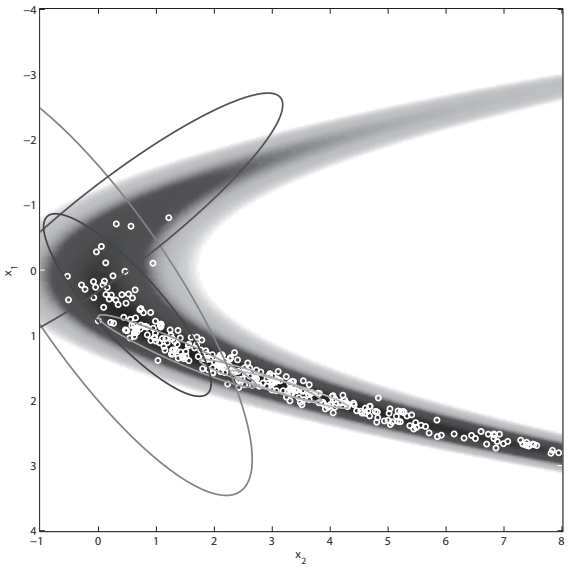

Figure A.1: Snapshot of the different proposal kernels as the Markov Chain Monte Carlo progresses. Note how it adapts to the local curvature.

.

As depicted in Figure A.1, this approach allows the sampler to follow the curvature of the probability density function naturally.

The d-dimensional Gaussian distribution is characterized by a positive definite covariance matrix $\mathbf{\Sigma}$ and a vector of mean values $\vec{\mu}$:

$$G(\vec{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}\sqrt{|\mathbf{\Sigma}|}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^T \mathbf{\Sigma}^{-1}(\vec{x}-\vec{\mu})} \tag{A.2}$$

To sample from this Gaussian proposal distribution, we compute a decomposition such that $\mathbf{R}\mathbf{R}^T = \mathbf{\Sigma}$. Subsequently we draw a vector $\vec{z}$ of d independent normal variates. Therefore the expression for the next iteration becomes $\vec{\theta}_{n+1} = \vec{\theta}_n + \mathbf{R}\vec{z}$. Since the proposal distribution depends on the current state (asymmetric proposals), it needs to be corrected for the imbalance in proposal densities using the Hastings correction, which can be calculated for a multivariate Gaussian proposal distribution as:

$$Q(\vec{\theta}_n \to \vec{\theta}_{n+1}) = \frac{1}{\sqrt{|\mathbf{\Sigma_n}|}} e^{-\frac{1}{2}(\vec{\theta}_{n+1}-\vec{\theta}_n)^T \mathbf{\Sigma_n}^{-1}(\vec{\theta}_{n+1}-\vec{\theta}_n)} \tag{A.3}$$

Calculating the Hessian of the log-likelihood is costly and numerically challenging, which is why an approximation based on the sensitivities of the model residuals is used. Depending on the model, these can either be computed by solving the sensitivity equations and applying the chain rule on the functions describing the observables or by means of finite differences (for which *strict tolerances* are required to ensure reliable derivatives). Analogously to the approximate Hessian of the log-likelihood, the Hessian of the log-prior is also approxi-

mated by its linear sensitivities. Though exact expressions for the Hessian of the logarithm of the prior can be derived for a wide range of prior distributions, they can result in the Hessian no longer being positive definite which is problematic when computing the decomposition required for generating samples.

When the noise variance of the error model is unknown, it has to be treated as a parameter and estimated from the data. This additional parameter requires explicit inclusion of the constant K which normalizes the likelihood. This is incorporated as an additional element in the vector of squared residuals. The approximated Hessian is subsequently decomposed using the singular value decomposition:

$$\mathbf{H} = \mathbf{U}\mathbf{S}\mathbf{V}^T \tag{A.4}$$

where $\mathbf{S}$ is a diagonal matrix containing the singular values and $\mathbf{U}$ the matrix of singular vectors. Large singular values correspond to singular vectors which point in directions where the probability density function changes rapidly (*i.e.* well constrained), while low values correspond to poorly constrained directions in parameter space. In practical cases, some directions in parameter space can be so poorly constrained that this leads to a (near) singular Hessian (some singular values near zero). As a result, the proposal distribution (which is based on an pseudo inverse of this matrix) will become extremely elongated in these directions. Consequently, proposals are generated where parameters take on extreme values and acceptance ratios decline due to either integration failures or rejections (due to low probability). One approach to avoid such numerical difficulties is to set singular values below a certain threshold to a specific minimal threshold (prior to inversion). Informally, this threshold can be interpreted as a circular region in which the Gaussian approximation to the log-posterior density is considered a reasonable approximation. The covariance matrix of the proposal distribution is computed directly from the SVD using:

$$\mathbf{R} = \frac{s\sqrt{T}}{\sqrt{d}}\mathbf{V}\sqrt{\mathbf{S}^{-1}} \tag{A.5}$$

Here s corresponds to a problem specific (tuned) scaling factor, T to the temperature and d to the number of parameters. The inverse required for the Hastings correction can subsequently be computed as:

$$\mathbf{\Sigma}^{-1} = \frac{N_{dim}}{s^2 T}\mathbf{V}\mathbf{S}\mathbf{V}^{\mathbf{T}} \tag{A.6}$$

Since the determinant only appears in ratios and we consider MCMC algorithms where the number of parameters do not change, the factors cancel out. Therefore the ratio is computed as a product of the ratios of the singular values.

$$det(\mathbf{\Sigma}) = \left(\frac{N_{dim}}{s^2 T}\right)^{-N_{dim}} \prod \frac{1}{S_{ii}} \tag{A.7}$$

Parameter representation

To deal with the large difference in scale between the various parameters, we typically consider log-transformed model parameters. Since the prior distribution is not invariant of model parameterization, a correction for this transformation is required. The transformation between parameters can be described by the matrix of partial derivatives with respect to the equations which transform the parameters from one parameterization to another (the Jacobian of the transformation). Similar to a change of variables in integration, the correction factor required for the prior distribution is given by the absolute value of the determinant of the Jacobian of the transformation. This corrects for the stretching and compression of the distribution due to the reparameterization.

$$p(f(\vec{\theta})) = p(\vec{\theta}) \left| \begin{bmatrix} \frac{df(\theta_1)}{d\theta_1} & \cdots & \frac{df(\theta_1)}{d\theta_n} \\ \vdots & \ddots & \vdots \\ \frac{df(\theta_n)}{d\theta_1} & \cdots & \frac{df(\theta_n)}{d\theta_n} \end{bmatrix} \right| \tag{A.8}$$

In the case where we perform the MCMC in logarithmic space, we obtain the following expression:

$$\frac{|J(\theta^a)|}{|J(\theta^b)|} = \prod_i^{N_{pars}} \left(\frac{\theta_i^b}{\theta_i^a} \right) \tag{A.9}$$

which should be included in the acceptance probability. For the Hessian based approach, the proposals can subsequently be generated using the following equation:

$$\vec{\theta}_{n+1} = \vec{\theta}_n e^{N(0, \mathbf{\Sigma}_{\ln})} \tag{A.10}$$

Where the Hessian approximation in log-space is computed by applying the chain rule:

$$\frac{\delta^2 L(\vec{\theta})}{\delta \ln \theta_i \delta \ln \theta_j} = \frac{\delta^2 L(\vec{\theta})}{\delta \theta_i \delta \theta_j} \theta_i \theta_j \tag{A.11}$$

Metropolis Coupled MCMC

To improve mixing in multi-modal problems, we implemented Metropolis Coupled MCMC (MC^3) [2]. In MC^3, multiple chains are simulated simultaneously, while allowing them to interact between parameter updates. In our implementation, each chain can optionally run at a different temperature T:

$$p_T(\mathbf{y}^{\mathbf{D}}|\vec{\theta}) = p(\mathbf{y}^{\mathbf{D}}|\vec{\theta})^{\frac{1}{T}} \tag{A.12}$$

One advantage of using multiple temperatures is that the likelihood function will flatten out for higher temperatures. Consequently, chains at higher temperatures are able to traverse the parameter space more freely. Interactions between the chains are implemented in the form of switch moves. These are performed

by either randomly selecting two adjacent temperatures and computing a switch probability based on a Metropolis-Hastings acceptance probability [2]:

$$\alpha < \min\left(\frac{p(\mathbf{y^D}|\vec{\theta}_{n+1})^{\frac{1}{T_n}}\, p(\mathbf{y^D}|\vec{\theta}_n)^{\frac{1}{T_{n+1}}}}{p(\mathbf{y^D}|\vec{\theta}_n)^{\frac{1}{T_n}}\, p(\mathbf{y^D}|\vec{\theta}_{n+1})^{\frac{1}{T_{n+1}}}}\right) \tag{A.13}$$

or by assigning one 'mother-chain' which is involved in every interaction [3]. In the current implementation, regular MCMC updates are performed per parameter group while exchange moves enable the sampler to switch the parameters between two groups.

Autocorrelation

An MCMC sampler will generate samples that are often highly correlated. To get an idea of how well the chain is mixing, one can plot the autocorrelation function of an observable or parameter. Such correlations decrease approximately exponentially suggesting that the chain decorrelates as it progresses. Informally, a sample drawn a certain number of iterations later is no longer closely related to the initial one. Slowly decaying autocorrelations are a warning sign that the chain is mixing slowly for that observable or parameter. Estimates of the effective sample size (ESS) are obtained using the initial monotone sequence estimator [4]. The ESS is computed from the empirical autocorrelation $\rho(t)$. First, the sum of adjacent pairs of autocorrelations $\phi(t) = \rho(2t) + \rho(2t+1)$ is defined. Subsequently monotonicity is enforced by setting each subsequent $\phi(t)$ to the minimum of the preceding ones. Then b is chosen such that $\phi(t)$ is positive for $1...b$. The ESS is then given by:

$$ESS = \frac{M}{-1 + 2\sum_{t=0}^{b}\phi(t)} \tag{A.14}$$

where M denotes the number of samples in the MCMC chain. Because of the highly correlated nature of MCMC chains, results are often thinned (*i.e.* using only every m^{th} iteration). Thinning can be considered when computing predictions using samples obtained via MCMC is costly compared to running the chain [4] and/or the autocorrelation function decreases slowly.

To validate the results from our MCMC, we typically compared inferences made with our implementation to results obtained with another MCMC named MMALA. MMALA is based on Langevin diffusion on a Riemannian Manifold [5]. The timestep was chosen in such a manner that we obtained an acceptance rate of about 50%. We could see no systematic difference between the posterior distributions obtained with the different methods.

References

[1] Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP: **Universally Sloppy Parameter Sensitivities in Systems Biology Models**. *PLoS Comput Biol* 2007, **3**(10):e189.

[2] Calderhead B, Girolami M: **Estimating Bayes factors via thermodynamic integration and population MCMC**. *Computational Statistics & Data Analysis* 2009, **53**(12):4028–4045.

[3] Rigat F, Mira A: **Parallel hierarchical sampling: A general-purpose interacting Markov chains Monte Carlo algorithm**. *Computational Statistics & Data Analysis* 2011.

[4] Geyer C: **Practical markov chain monte carlo**. *Statistical Science* 1992, :473–483.

[5] Girolami M, Calderhead B: **Riemann manifold langevin and hamiltonian monte carlo methods**. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2011, **73**(2):123–214.

Summary

Uncertainty Analysis in Systems Biology

Biochemical systems are complex systems comprised of several highly interlinked components. Understanding such systems is essential in order to predict their behavior under specific conditions and unravel the mechanisms that drive multifactorial diseases. Computational models are often employed to formalize our understanding in a testable manner. These are models based on mathematical equations that can be solved using computer software. This thesis pertains to models based on ordinary differential equations (ODEs). These models contain parameters and state variables embedded in a system of equations that describes their evolution in time. To simulate mathematical models, parameter values are required. These parameter values are typically obtained by calibrating models to reality using experimental data. Typically, only a few of the system components are experimentally accessible and measurements are performed with finite accuracy. Considering the complexity of the models under investigation, this can result in large uncertainties in the parameters and predictions.

Quantifying and reducing this uncertainty effectively is the main topic of this thesis. Chapter 2 provides an introduction to the different methods for uncertainty quantification available in literature. Subsequently, Chapter 3 describes a strategy which helps to avoid problems that can occur when performing uncertainty analysis on models comprised of ODEs. In Chapter 4, a method is presented that uses the result from Chapter 3 to select new experiments in such a way that specific uncertainties are reduced in an optimal manner. This is achieved by exploiting relations between the different uncertainties which have been implicitly imposed by the model topology and the available data. This methodology allows the investigator to target specific model predictions and select those experiments which optimally reduce the uncertainty. Additionally, we can predict the effect of performing multiple experiments simultaneously. In some cases, the topology of the network (and therefore also the model equations) are also uncertain. In Chapter 5, a method is proposed which can be used to select experiments in such a manner that the discriminatory power between different models is maximized. The aim of these experiments is to enable the researcher to select which model most closely resembles reality.

Chapter 6 details modeling of a system related to the production of triglycerides in the liver. Here, some issues that arise when using literature data in a Systems Biology approach are discussed. Chapter 7 shows an application of the methodology presented in Chapter 4. Here parameters are no longer considered fixed in time. The fact that certain interaction mechanisms between model components are missing is accounted for by allowing parameters to vary in time. It is shown that the methodology can also be used to successfully improve specific predictions of interest. Additionally, an example of an inexpensive measurement giving nearly as much information as a costly and invasive measurement is provided.

Nederlandse Samenvatting

Onzekerheidsanalyse in Systeem Biologie

Biochemische netwerken zijn complexe systemen bestaande uit meerdere stoffen die onderling interacties aangaan. Het begrijpen van dergelijke systemen en het kunnen voorspellen van hun gedrag onder bepaalde omstandigheden is van essentieel belang om multifactoriële ziektes te begrijpen. Kennis over een systeem kan concreet en toetsbaar gemaakt worden met behulp van computationele modellen. Dit zijn modellen gebaseerd op mathematische vergelijkingen die opgelost kunnen worden met behulp van een computer. Dit proefschrift behandelt modellen die gebaseerd zijn op differentiaalvergelijkingen. Deze modellen bestaan uit vergelijkingen en parameters die de verschillende stoffen in het model aan elkaar relateren. Om deze modellen te simuleren zijn parameterwaarden nodig. Deze parameterwaarden worden typisch verkregen door het model, met behulp van metingen, te ijken op de realiteit. Niet alle componenten kunnen echter gemeten worden. Bovendien worden metingen altijd gedaan met een eindige nauwkeurigheid. Deze feiten gecombineerd met de benodigde modelcomplexiteit zorgen ervoor dat er grote onzekerheden ontstaan in de voorspellingen die door het model gemaakt kunnen worden. Deze onzekerheden kwantificeren en zo efficiënt mogelijk reduceren zijn de hoofdonderwerpen van dit proefschrift.

Hoofdstuk 2 betreft een korte beschrijving van de verschillende methodes voor onzekerheidsanalyse in de literatuur. In Hoofdstuk 3 volgt een strategie waarmee bepaalde problemen die zich bij de individuele methodes kunnen voordoen voorkomen kunnen worden. Het resultaat hiervan is een distributie van parameters die hun onzekerheid weerspiegelt en gebruikt kan worden om modelvoorspellingen te genereren. In Hoofdstuk 4 wordt gebruik gemaakt van deze distributies om nieuwe experimenten te selecteren. Deze methode bepaalt welke experimenten optimaal zijn om selectief onzekerheden te reduceren. Hierbij wordt gebruik gemaakt van de relaties tussen de onzekerheden van de verschillende modelvoorspellingen die door de data en het model zijn opgelegd. Een voordeel is dat deze niet beperkt zijn tot het voorspellen van enkel concentraties of parameters. Ook is het mogelijk om het combinatoriële voordeel van meerdere experimenten in kaart te brengen. Vaak bestaan er meerdere hypotheses (modellen) over hoe het systeem werkt. In Hoofdstuk 5 wordt een methode voorgesteld die voorspellingen aanwijst die het onderscheidingsvermogen tussen verschillende modellen zo groot mogelijk maakt. Het vervolgens uitvoeren van deze experimenten dient de onderzoeker dan in staat te stellen om het meest geschikte model te kiezen.

Hoofdstuk 6 betreft een model van de productie van triglyceriden in de lever. Hierbij wordt gedemonstreerd wat voor problemen zich kunnen voordoen bij het gebruik van literatuur data. Hoofdstuk 7 betreft een applicatie van de methode gepresenteerd in Hoofdstuk 4. Hierbij wordt rekening gehouden met het feit dat bepaalde interacties tussen componenten in het model nog niet expliciet gemodelleerd kunnen worden. Ondanks grote onzekerheden elders in het model, kan het specifiek richten op bepaalde voorspellingen toch leiden tot grote verbeteringen. In deze studie blijkt een relatief eenvoudige meting net zo nuttig te zijn als een kostbare invasieve meting.

Dankwoord

Na al dat schrijfwerk ben ik nu toch eindelijk bezig met de laatste paar pagina's. Tijd om d'r een einde aan te breien en toch eens de mensen te bedanken zonder wie dit proefschrift er niet geweest zou zijn.

Peter, bedankt dat je mij de middelen en vrijheid hebt gegeven om dit proefschrift tot iets eigens te maken. Het was niet altijd even makkelijk, en ik heb me menig moment afgevraagd of wat meer sturing niet fijn of nodig zou zijn. Maar uiteindelijk is het iedere keer toch gelukt om weer iets nieuws te verzinnen. Achteraf gezien was de gegeven vrijheid, met slechts hier en daar een kritische vraag, duwtje of opmerking, misschien toch wat ik nodig had. Bedankt voor je vertrouwen en interesse in m'n werk.

Natal, onze gesprekken zijn vaak nuttig geweest voor mij om dingen op een rij te zetten. Je bent goed in het zoeken naar positieve aspecten van bepaalde aanpakken en werk zo te presenteren dat deze aspecten bijzonder in het daglicht gezet worden. Ook kijk ik met een glimlach terug naar de uitstapjes naar de verschillende congressen en workshops.

Jens Timmer, Mark Girolami and Klaas Nicolay, I am honored that you agreed to be members of the core-committee for my PhD defense. Thank you for your interest in my work and for thoroughly reading my dissertation.

Christian, ik heb altijd graag met je samengewerkt. Je bent vaak voor mij een klankbord geweest, zowel voor nieuwe ideëen als andere niet-wetenschap gerelateerde zaken. Ook was je goed in het ietwat nuanceren van mijn af en toe wat cynische blik op 'het veld'. Zowel je werk als je persoonlijkheid zijn behoorlijk van invloed geweest op mijn doen en laten tijdens deze promotie en ik heb vaak met je moeten lachen. Thanks man!

Andere Joep, je vaak praktische insteek heeft me vaak geholpen in te zien dat de toepassing van bepaalde technieken niet vergeten moet worden. Een techniek kan nog zo mooi zijn, als het gebruik te ingewikkeld is gaat niemand deze gebruiken. Rik, ook al waren we het niet altijd met elkaar eens, ik heb altijd graag met je gepraat. Zowel on- als off-topic. Fianne, jij ook bedankt en succes met het afmaken van het boekje (al heb ik er het volste vertrouwen in dat dit lukt!). Alon, thanks for keeping me enthused (and well informed) about running! Ceylan, Huili, I've always enjoyed discussing cultural differences between our countries and mispronouncing both your last names. Doing a PhD is sometimes stressful (funfun!) and all of you have kept me relatively sane during these past 4+ years. The drinks with pizza and random banter on our little island of PhD students will not be forgotten quickly.

Jeroen, als ik aan enthousiaste wetenschappers denk, denk ik meteen aan jou. Hierdoor maakte je af en toe wat grote sprongen, en twijfelde ik hier en daar aan de realiseerbaarheid van een aantal van je ideëen. Desondanks, hebben zowel je wetenschappelijke enthousiasme als je kijk op het leven een ontzettend

motiverende invloed op mij gehad. Kalyan, thanks for an awesome time during my internship in the USA. Your hospitality made me feel very welcome. Next time I am in the US, we'll have to go for some injera with honey wine again!

Huub en Koen, jullie waren dan wellicht niet mijn promotie- of afstudeerbegeleiders, stiekem heb ik altijd veel van jullie geleerd. Vooral de helderheid waarmee jullie zaken uitlegden en de duidelijkheid waarmee jullie aangaven tot waar jullie kennis reikte waren fijn. Of course, for the entire BIOMIM crew, thanks for the good times! The lab outings, the conversations at the lunchtable and other things that made my time here more enjoyable.

Steve, Sophie, Hakim, thanks for listening to me grumble every now and then! It's much appreciated.

Tommy, en ook de rest van de massiv kru, jullie ook bedankt natuurlijk. De menige gare avondjes en verschillende parties heb ik altijd erg gewaardeerd. Zal ook zeker nog wel vaker op bezoek komen in Helden city! En Tommy, ik verwacht wel de eerste beta tester van je spel te worden he!

Pap, mam, ik wil jullie bedanken voor jullie goede zorg, raad en steun. Zowel voor, tijdens, als (hoop ik nog lang) na mijn promotie. Zonder jullie had ik hier niet gestaan. Het is niet altijd even makkelijk geweest en gedurende het afronden van dit werk heb ik waarschijnlijk ook jullie geduld vaak op de proef gesteld. Desondanks hebben jullie me altijd gesteund. Het is bijzonder fijn om te weten dat ik altijd op jullie kan rekenen en dat jullie altijd voor me klaar staan. En mam, Duitsland is echt zo ver niet hoor!

Thijs, Melanie, Cathelijne en Stefan, bedankt dat jullie er ook altijd voor me waren.

En tenslotte, Rachel, ik ga dit toch in het Nederlands doen, aangezien ik er het volste vertrouwen in heb dat je Nederlands tegen deze tijd zo goed is dat je dit zelf kunt lezen. Ik zou je graag vertellen hoeveel je voor mij betekent, maar dit kan ik nauwelijks in woorden uitdrukken. Je hebt mij een kant van het leven laten zien die mij bijzonder dierbaar is. Ik kijk altijd weer uit naar die momenten waarop ik weer van je lieve glimlach mag genieten. Je onuitputtelijke positieve instelling en vrolijk karakter zijn bijzonder en ik hoop samen met jou nog heel wat jaartjes door te mogen brengen.

Oh, en Melanie? Bedankt voor de kaft he!

Groetjes,
Joep.

Curriculum Vitae

Joep Vanlier was born on February 17th 1985 in Helden, The Netherlands. Here he grew up and graduated from secondary school (atheneum) at the Bouwens van der Boijecollege in 2003. Subsequently, he enrolled in the Biomedical Engineering program at the Eindhoven University of Technology where he obtained his bachelor's degree in 2006 (*cum laude*). This degree was followed up by a master's program at the same university. During his internship, he developed tools for analyzing force production and oxygen consumption during electrical stimulation in isolated murine skeletal muscle. Joep carried out an internship at the Department of Physiology in group of prof.dr. Daniel Beard at the Medical College of Wisconsin, USA. There he worked on a parser that generates models using thermodynamic databases and small kinetic submodels. These models explicitly account for reaction thermodynamics, effects of ionic strength, temperature, compartmentalization, electrophysiology and buffering of multiple ions and metabolites. In 2009, he graduated (*cum laude*) within the groups Computational Biology (prof.dr. Peter Hilbers) and Biomedical NMR (prof.dr. Klaas Nicolay), under the supervision of dr. Jeroen Jeneson. His masters project revolved around the quantification of ^{31}P Nuclear Magnetic Resonance spectra during exercise. He developed a MATLAB toolbox for quantifying data hampered by baseline and poor field homogeneity. Uncertainty and quantification continued to play a large role during his PhD research, which was carried out under the supervision of dr.ir. Natal van Riel and prof.dr. Peter Hilbers at the Eindhoven University of Technology.

List of journal publications

International refereed journal publications

Vanlier J., Tiemann C.A., Hilbers, P.A.J., van Riel, N.A.W., *Optimal experiment design for model selection of biochemical networks* (manuscript submitted)

Vanlier J., Tiemann C.A., Hilbers, P.A.J., van Riel, N.A.W., *Parameter uncertainty in biochemical models described by ordinary differential equations* (in press), Mathematical Biosciences

Tiemann C.A., **Vanlier J.**, Hilbers, P.A.J., van Riel, N.A.W., *Optimal experiment design to constrain predictions of progressive adaptations in biological systems* (manuscript submitted)

Tiemann C.A., **Vanlier J.**, Hilbers, P.A.J., van Riel, N.A.W., *Parameter trajectory analysis to identify treatment effects of pharmacological interventions* (in press), PLoS Computational Biology

van Riel N.A.W., Tiemann C.A., **Vanlier J.**, Hilbers P.A.J., *Applications of analysis of dynamic adaptations in parameter trajectories* (2013), Interface Focus; 3(2)

Vanlier J., Tiemann C.A., Hilbers, P.A.J., van Riel, N.A.W., *A Bayesian approach to targeted experiment design* (2012), Bioinformatics; 28(8):1136-1142

Vanlier J., Tiemann C.A., Hilbers, P.A.J., van Riel, N.A.W., *An integrated strategy for prediction uncertainty analysis* (2012), Bioinformatics; 28(8):1130-1135

Schmitz J.P.J., **Vanlier J.**, van Riel N.A.W., Jeneson J.A.L., *Computational modeling of mitochondrial energy transduction* (2011), Critical Reviews in Biomedical Engineering, 39(5):363-377

Tiemann C.A., **Vanlier J.**, Hilbers, P.A.J., van Riel, N.A.W., *Parameter adaptations during phenotype transitions in progressive diseases* (2011), BMC Systems Biology; 5(1):174

Vanlier J., Wu F., Qi F., Vinnakota K.C., Han Y., Dash R.K., Yang, F., Beard D.A., *BISEN: Biochemical Simulation Environment* (2009), Bioinformatics; 25(6):836-837