A Parameter Estimation Framework for Kinetic Models of Biological Systems

Dissertation

zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften (Dr. rer. nat.)

der Naturwissenschaftlichen Fakultät III
Agrar- und Geowissenschaften, Mathematik und Informatik
(Institut für Informatik)
der Martin-Luther-Universität Halle-Wittenberg

vorgelegt von

Masters in IT Syed Murtuza Baker
geb. am. 11. September, 1981 in Dhaka, Bangladesh

Halle (Saale), im November 2012

Zeitraum
Vorgelegt: 16.05.2012
Verteidigt: 23.11.2012

Gutachter

Dr. Björn H. Junker	Martin Luther University Halle-Wittenberg
Prof. Dr. Falk Schreiber	Martin Luther University Halle-Wittenberg
apl. Prof. Dr. Michael Mangold	Otto-von-Guericke-Universität Magdeburg

Dedication

To my Creator

To my father and mother
without whom I would not be what I am now

To my lovely wife Tabassum
without her this journey would have not been possible

and

To my kids
without whom life would have little meaning

Acknowledgement

During my research stay at the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), I received a lot of help from different individuals. This dissertation would not have been ever possible without their sincere support. First and foremost I would like to give my sincere gratitude to my wonderful supervisor, Dr. Björn H. Junker, for his extreme patience, wonderful guidance and continuous support over the duration of my thesis. The knowledge he imparted to me during all these years are pivotal to make me a researcher. I am extremely fortunate to have him as my supervisor. I am extremely grateful to my mentor Prof. Dr. Falk Schreiber for all the fruitful discussions and to keep me on track during my research. I am extremely thankful to Dr. Hart Poskar who would always pull me out whenever I get stuck in my research. His invaluable suggestions were always helpful to solve critical research problems. My sincere gratitude goes to Dr. Kai Schallau, Dr. Cristian Krach, Dr. Jan Hüge for their intellectual support. I would also like to thank Tiina Liiving, Mathias Franke for congenial discussions. A heartfelt gratitude to Dr. Britt Leps for making our life in Germany so much comfortable, Dr. Astrid Jünker and Andrea Hüge for their assistance when it mattered.

All my family members for their love and support. Specially, my brother Raquib: Thank you for always encouraging me.

All my wonderful friends (of whom there are simply too many to mention by name): Thanks for your friendship and support. It kept me in sound mind.

All the works mentioned in this thesis have been done at IPK Gatersleben. The computational experiments carried out in the thesis profited a lot from the exceptional computing environments at IPK Gatersleben.

Abstract

In order to obtain the correct predictive behaviour of a kinetic model, it is crucial to have an accurate and complete set of parameter values. Lack of information regarding these parameters from wet lab experiments has held back the successful use of such models. Therefore these parameters have to be estimated computationally to feature a complete description of the model. In this thesis I propose a novel parameter estimation framework combining different existing approaches with a newly proposed filtering technique for the successful estimation of these unknown parameter values. The framework includes a constrained extension of the square-root unscented Kalman filter to estimate the parameter values within a biologically meaningful parameter space. This framework is capable of addressing both the issues of structural and practical non-identifiablity before performing the final estimation of the parameters. The constrained square-root unscented Kalman filter (CSUKF), guarantees numerical stability by ensuring positive definiteness of the covariance matrix. The CSUKF takes into consideration one of the common features of biological models, noise. Noise is introduced in two ways, in the system due to the uncertainty in the model and in the measurement data due to the inaccuracy in the method or device used to collect the data. By representing the dynamic system as a state space model, the CSUKF jointly estimate the states and the parameters of the non-linear dynamic systems. This makes it possible for the CSUKF to estimate both the parameter values and the hidden variables. CSUKF uses the general probability theory to estimate the parameter values of biological models where reasoning under uncertainty is essential. An identifiability analysis module is included in the framework to identify the non-identifiable parameters. Wherever possible the problem of non-identifiability is resolved through additional, and/or more accurate, measurement data. To assist in resolving the issue, the framework includes ranking of the parameters, determination of the correlation and functional relationship of non-identifiable parameters with other parameters. Finally, when it is not possible to solve the parameter non-identifiability through standard methods, the informed prior is formulated for the unique estimation of parameters even in the presence of non-identifiability. This framework is successfully applied to estimate parameters for three published biological models, the glycolysis model in yeast, the sucrose accumulation model in sugarcane culm tissue and a gene regulatory network.

Contents

List of Figures

List of Tables

List of Abbreviations

CD	Central Difference
CI	Confidence Interval
CSUKF	Constrained Square-Root Unscented Kalman Filter
DDM	Decoupled direct method
EKF	Extended Kalman Filter
EP	Evolutionary Programming
FBA	Flux Balance Analysis
FDM	Finite Difference Method
FIM	Fisher Information Matrix
GA	Genetic Algorithm
ICUT	Interval Constrained Unscented Transformation
IUKF	Interval Constrained Unscented Kalman Filter
KF	Kalman Filter
MFA	Metabolic Flux Analysis
MLE	Maximum Likelihood Estimator
MMSE	Minimum Mean Squared Error
NI	Non-Identifiable
ODE	Ordinary Differential Equation
PSO	Particle Swarm Optimization
RLS	Recursive Least Squares
SA	Simulated Annealing
SQ-UKF	Square-root Unscented Kalman Filter
UKF	Unscented Kalman Filter
UT	Unscented Transformation

List of Symbols

θ	Parameter vector
N	Stoichiometric matrix
v	Flux vector
$x(t)$	Continuous time state vector
$x(k)$	Discrete time state vector
$y(t)$	Continuous time measurement vector
$y(k)$	Discrete time measurement vector
A	Transition matrix
B	Input matrix
C	Output matrix
w	Process noise
e	Measurement noise
Q	State noise covariance matrix
R	Measurement noise covariance matrix
F	Non-linear continuous state space equation
H	Non-linear continuous measurement equation
f	Non-linear state space equation
h	Non-linear measurement equation
W	Weighting matrix
j	Objective function value
J	Jacobian matrix
Z^a	Sensitivity matrix
Z	Normalized sensitivity matrix
ψ	Parameter of interest in PLE
η	Nuisance parameter in PLE
σ	Confidence bound
$\emptyset^b$	Observability matrix
n	State space dimension, unless specified otherwise
$\mathcal{X}$	Sigma points
L	Lower constrained boundary
U	Upper constrained boundary

S	Sigma point direction matrix
ζ	Step size for sigma point selection
λ	Scaling factor
α	Scaling factor
κ	Scaling factor
W^m	Weight for calculating mean
W^c	Weight for calculating covariance
P	Covariance matrix
O	Orthogonal matrix
G	Upper triangular matrix
V	Square root of covariance matrix
Rs	Residual matrix
X^{ace}	Predictor variable in ACE
Y^{ace}	Response variable in ACE
H^{ace}	Test function in ACE
T^{thr}	Threshold for identifying functional relationship

Chapter 1

Introduction and literature review

1.1 Overview

The study of living organisms is a very important task as it makes it possible for us to understand life on earth. It enables us to recognize the functionality of organisms that have the capacity to grow and reproduce, and to respond to different stimuli. To understand how the cell and its components work, biologists have studied its biochemistry, the DNA and RNA structure, the protein structure and how the translation and transcription shapes this structure along with its functionality in metabolism. Through all of this, a theoretical concept of how the interaction of different elements in a cell form a network has been developed [Klipp 2005]. However, in order to have a better understanding of the mechanisms underlying the different functionalities of a living cell, a systematic approach, investigating the composition and molecular biology of single cells, is required. In 2001, Kitano coined the term for this systematic approach, systems biology [Kitano 2001]. Systems biology helps to better understand the functionality of the cell, how cellular processes are regulated and how cells react to environmental perturbations. In order to have a coordinated study of the cellular network and its interaction, systems biology integrates computational methods with high-throughput experimental techniques [Klipp 2005]. One of the major limitations to this technique is to transform system-level data into systems-level understanding [Bray 2003]. Mathematical models play a central role in addressing this task by integrating experimental data with theoretical approaches [Kitano 2002b]. Systems biology combines experimental design with data collection, data processing and mathematical modelling. Starting with an open question regarding the system under investigation, an initial model is deployed with the knowledge gathered from experimentation. This model is used to make a prediction, which can either be verified or proven wrong with a different set of experimental data. If the model simulation does not agree with this different set of experimental data, the model has to be refined accordingly. This would result in a new prediction which must still agree with the experimental data or further be refined. This iterative process continues until a good agreement is reached between the experimental data and the model predictions.

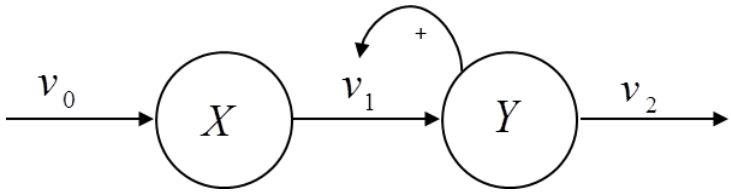
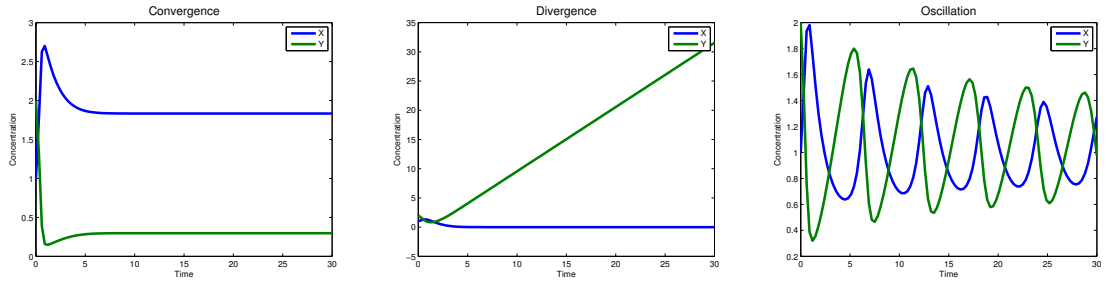


Figure 1.1: Network diagram of Higgins-Sel'kov oscillator. The supply rate of the substrate, X, is v_0, the conversion of substrate to product Y occurs at a rate of v_1 and product Y is consumed at a rate of v_2.

These mathematical models have the potential to provide both an experimentally testable hypothesis as well as a prediction of cellular functionality.

Mathematical modelling approaches range from basic stoichiometric models to fine grained mechanistic models or kinetic models. Stoichiometric models are time independent static models. These are basic models and their study does not reveal the functional operation of the cell [Nikolaev 2006]. Thus stoichiometric modelling alone is not sufficient to make predictions about the interaction between different cellular components that could be modified to create a network with new functionalities. At the other extreme, kinetic models give the most detailed and complex mathematical representation of a biological system. Kinetic modelling builds on the network stoichiometry incorporating the dynamic interactions between the different components of the network. In a kinetic model, biological systems are usually represented by a system of ordinary differential equations (ODEs). These ODEs represent the change of species concentration over time. These changes in the concentrations are driven through rate laws, which are mathematical expressions that represent the internal



(a) (b) (c)

Figure 1.2: Dynamic behaviour of Higgins-Selḱov oscillator. The value of the parameters are $k_1 = k_2 = 1.1$, and different values of parameter k_3. a) Converging, with $k_3 = 0.6$, b) diverging with $k_3 = 1.6$ and, c) oscillating with $k_3 = 1.1$.

reaction mechanism as a function of species concentrations and parameters. These model parameters play a critical role in describing the dynamics of the model. Therefore to have an accurate model of the system it is necessary to have a complete and accurate set of parameter values.

To illustrate the importance of these parameters to a kinetic model, the Higgins-Selkov oscillator [Sel'kov 1968] is given as an example. This model represents a glycolytic oscillator mainly studied in yeast, describing the dynamic and biochemical properties of the enzyme phosphofructokinase [Hess 1979]. The model describes an enzymatic reaction with substrate inhibition and product activation [Sel'kov 1968]. As depicted in Figure 1.1, the network has one substrate, X, one product, Y and three parameters, k_1, k_2, k_3. The fluxes, v_i, are represented by the following rate laws,

$$\begin{aligned} v_0 &= k_1 \\ v_1 &= k_2[X][Y]^2 \\ v_2 &= k_3[Y] \end{aligned} \tag{1.1}$$

The model consists of two ODEs, that describe the temporal behaviour of the concentrations of substrate X and product Y. Substituting the rate law that corresponds with each flux into the ODEs yields:

$$\begin{aligned} \frac{d[X]}{dt} &= k_1 - k_2[X][Y]^2 \\ \frac{d[Y]}{dt} &= k_2[X][Y]^2 - k_3[Y] \end{aligned} \tag{1.2}$$

To illustrate the significance of even small changes to a single parameter, the network is simulated three times. In each simulation parameters k_1 and k_2 are kept unchanged (at a value of 1.1), while the value for parameters k_3 is set to 0.6, 1.6, and 1.1, respectively. As can be seen in Figure 1.2 this small variation of a single parameter has profound effects on the dynamics of the system. In each case the X and Y are set to start with the same initial concentrations. This clearly illustrates the crucial role of accurate quantitative information for the model parameters in order to have the correct dynamic behaviour of the model [Ingram 2006].

While the importance of having a complete and accurate parameter set is clear, the availability can not always be guaranteed. It is possible to measure only a fraction of the kinetic parameters due to high cost, difficulty and limitations of current technical resources associated with biological experiments. The remaining parameters have to be

estimated through computational methods. These methods determine the values of the parameters in an indirect fashion from the measurement of other quantities. Parameters are then determined so as to minimize an error measurement between the simulated data (as generated by the model) and the corresponding measurement data. In dynamic biological models, the measurement data usually come in the form of time-series data and is often incomplete. Thus it is not uncommon to lack a sufficient amount of measurement data to make an unbiased estimation using traditional parameter estimation methods. Moreover biochemical networks are often multimodal, meaning that they have multiple solutions, due to high non-linearity and their dynamic nature [Sun 2008]. Traditional optimizers have a tendency of getting stuck in local optima instead of finding global optima [Sun 2008]. Furthermore although measurement errors can be incorporated into the observation equation of these optimization methods, they do not consider the process noise within the system equations. Considering these limitations a dynamic recursive estimator with the capability of efficiently handling both the process and measurement noise would be a better choice for the estimation of parameters in biological models.

The Kalman filter (KF) has the capability of using noise-corrupted measurement data and other inaccuracies to estimate the parameter values in a recursive manner. Thus the KF may perform more efficiently in these situations compared to the conventional parameter estimation methods. As the Kalman filter is an extension of the Bayesian solution, it can approximate the probability density function and can cope with multi-modality, asymmetries and discontinuities [Julier 2004]. This filter is very powerful as it can perform estimation even when the precise knowledge of the model is not available or the measurement data is incomplete [Welch 2006], which is often the case with biological models. This means that the Kalman filter can take into consideration both the process noise and the measurement noise. One other common phenomenon in biology that increases the difficulty of the estimation task are the presence of hidden variables, i.e. variables that cannot be measured directly. With the technique of joint state and parameter estimation, KF can function jointly to estimate the parameters as well as these hidden variables. One major drawback of the Kalman filter is that it can only function with linear systems. As a result different non linear extensions of the Kalman filter have been proposed among which the two most widely used are the **extended Kalman filter (EKF)** and the **unscented Kalman filter (UKF)**. Among these two non linear extensions, UKF has the better estimation accuracy and the difference in approach to handling non-linearity [Merwe 2001, Leven 2004] makes UKF a more robust estimation method than EKF [van der Merwe 2004].

Another difficulty that severely limits successful parameter estimation in any kind of model is the model's identifiability property. This property addresses the question of whether it is at all possible to determine a unique solution for an unknown parameter within the constraints of the mathematical model and the available measurement data. For a non-identifiable model, different sets of parameter values would agree equally well with the measurement data. If model parameters are non-identifiable, i.e. not well determined, then consequently the model predictions are unreliable. If this is the case then it may not be possible to address the underlying biological question, thus reducing any value that may be derived from the model. It is reasonable to perform parameter estimation only when the identifiability of the parameters has been ensured. Identifiability depends on the model structure, available measurement data, the corresponding level of error and the optimization method used for estimation. Non-identifiability arising due to the model structure is called structural non-identifiability, while non-identifiability due to measurement data is called practical non-identifiability. Structural non-identifiability is mostly due to over parametrization of the model, manifesting functionally related parameters. For example if θ_1 and θ_2 are functionally related, say $\theta_1.\theta_2 - 20 = 0$, then it is not possible to estimate both parameters simultaneously, as there could be a large number of different values for θ_1 and θ_2 that will satisfy this relationship. However if functional relationships between parameters can be identified it would assist in solving the structural non-identifiability. Before conducting the identifiability analysis it is useful to know the sensitivity of the parameters with respect to the system. A ranking based on sensitivity reveals which parameters have a high impact on the system model and therefore have a high importance in the identifiability analysis. These high-ranking parameters need to be estimated with high confidence to make sure that the model is behaving as expected. This sensitivity based ranking also performs an initial identifiability analysis by determining the linear relationship between the coefficients of the sensitivity matrix. However, parameters that are linearly independent in the sensitivity coefficient can still be non-identifiable, but this cannot be determined during the ranking method. Furthermore it also cannot determine the practical non-identifiability. As a result a detailed identifiability analysis is necessary to find both structural and practical identifiability for all parameters.

To conduct a successful parameter estimation it is crucial to first resolve parameter non-identifiability. If parameters are found to be structurally non-identifiable then there are two approaches to solve this non-identifiability. The first approach is to acquire direct measurement data of the currently non observable variables (the species concentration that was so far not measured directly) or measurement of different combinations of non

observables. This would change the function describing the relationship between the state variables and the observables and can thus define the change of each parameter individually. If no such measurement is possible then the second approach is to measure the specific structurally non-identifiable parameter. When it is not possible to measure all the structurally non-identifiable parameters, then the measurement of high rank parameters needs to be made available whereas the low rank parameters can be set to a nominal value. If none of these measurements are possible then the model has to be changed to combine these functionally related parameters, for example combing two non-identifiable parameters to have an identifiable combination. Practical non-identifiability is caused due to limited measurement data or error in measurement data. To solve this non-identifiability more measurement data or measurement data with less error needs to be incorporated in places where there is high uncertainty in the parameter values. All of these steps need to be performed recursively until all parameters are judged to be both structurally and practically identifiable.

Unfortunately it is not always possible to have more measurement data to solve either structural or practical non-identifiability. Furthermore simplification of the model might significantly limit the model's capability for generating a predictive behaviour and by extension to answer the open question properly. To overcome this limitation a novel solution exploiting the concept of Bayesian inference is incorporated into this framework. In contrast to the frequentist approach (traditional optimization techniques) to estimation, Bayesian inference can be obtained despite the presence of non-identifiability, if an informative prior is available [Rannala 2002, Samaniego 2010]. As the Kalman filter and each of its non-linear variants can be considered to be dynamic Bayesian networks, these property of using informed prior of Bayesian inference can be applied to CSUKF. This will allow CSUKF to have a unique parameter estimation for a model which is non-identifiable from the perspective of likelihood.

To further discuss the issues related to parameter identifiability analysis and their successful determination of accurate values key sub topics are elaborated in the next sections.

1.2 Introduction to biological models

Biologists have long been investigating the components and interactions of single cells. These studies include investigations into the functionality of individual parts of the cell, how different components develop and how they function in different situations. Recently a more systematic view of biological processes has emerged by integrating different fields

of biology in order to have a more detailed picture of cellular functionality [Klipp 2005]. This allows the scientists to make a prediction on the outcome of complex processes such as the effect of light intensity on plant metabolism [Stitt 2010]. These models can be regarded as a ”virtual laboratory”, building up a characteristic equation of the system and giving insight into the functional design principals of cellular activity [Steuer 2006].

Systems biology integrates experimental methods and data processing with mathematical models. Modelling in biology deals with the methods for describing biological phenomena in mathematical terms [Klipp 2005]. These models and computer simulations assist in understanding the intrinsic behaviour of biological systems and make predictions about the changes of the system with external perturbations and environmental interactions. There exist a variety of approaches for mathematical modelling. Different models can be chosen to describe different phenomena of the same system. When addressing a specific biological question it is important to select the appropriate approach. Based on the availability of data and desired analysis, modelling can be broadly categorized into **bottom-up modelling** and **top-down modelling**. In the bottom-up approach the base elements of the system are studied in greater detail and linked together in many levels to form a complete top-level system. In the top-down approach a birds eye view of the system is first formulated and then refined progressively in greater detail until the model specification is reduced to its base elements.

Several points need to be considered for the successful construction of a dynamic model [Chou 2008]. First, as most biological models are non-linear, the model must capture this non-linearity. Second, the model should capture the dynamic responses of the system. Third, it requires the ability to capture the interaction at different levels of components. Fourth, stochasticity might come into play when the particle numbers are low and therefore the fundamental laws of kinetics no longer apply. Finally as biological reactions occur across different compartments, the model should be able to capture this compartmentalization. When considering these features together the modelling process becomes extremely complex. Therefore simple methods are needed to strike a balance between the validity of the system, goals of the modelling and mathematical suitability.

Mathematical modelling can be based on a) interactions alone, b) constraints and stoichiometry, and c) detailed reaction mechanisms [Stelling 2004].

1.2.1 Interaction-based modelling

Interaction based modelling is dependent on the most basic feature of any network, that is the pattern of input-output interaction between different components. It does not take into

consideration any type of reaction stoichiometry or velocity. These networks facilitate the system level understanding of sub systems independently by introducing modularity. In this interaction based modelling, scale free network topology paralleling the structure of complex engineered system is described using graph theory [Barabasi 2004]. However, the applications of such networks are limited as they only give topological information.

1.2.2 Stoichiometry-based modelling

A metabolic network is like a set of nodes connected with directed edges and can be represented with matrices. Stoichiometric models translate the network topology into a matrix describing metabolic interactions of the network. Each element's sign represent the inflow (positive) or outflow (negative) of a specific metabolite, and the value denotes the unit of metabolite that is converted by a reaction. If a reaction does not have any impact on a metabolite then the corresponding cell will have a value of zero. This stoichiometric matrix is denoted with N. The reaction rates of the network are, v and the metabolite concentration are, $[X]$. The change of metabolite concentration $[X]$ with respect to time t are denoted with a set of ODEs,

$$\frac{d[X]}{dt} = N \cdot v \tag{1.3}$$

Each ODE represents an individual reaction, taken together they represent the full set of reactions in the model [Heinrich 1996, Stephanopoulos 1998]. This type of modelling is used to calculate the reaction rates of an enzymatic reaction v. If the number of reactions (equations) is equal to the number of metabolites, then they can be easily calculated by assuming a steady state of the reaction. At a steady state the fluxes remain constant and the rate of change of a metabolite concentration is zero. Thus the left hand side of Equation (1.3) becomes zero and the system of ODEs reduces to a system of linear equations which can be solved directly for the unknown fluxes. However in the case where the system is *under-determined*, i.e. there are more unknowns than linearly independent equations, additional information is required to solve in the steady state.

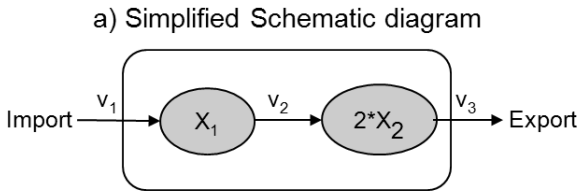
In order to better work with under-determined cases, constraint based tools such as flux balance analysis (FBA) and those including labelling experiments, metabolic flux analysis (MFA), have been developed. In FBA additional constraints are added to the model in order to make the system determined [Stelling 2004, Chou 2008]. These constraints are based on considering additional hypotheses to yield a flux distribution for a specific condition, for eg. maximization of yield/growth rate. Rather than applying predetermined constraints to the system, MFA is a method based on the use of stable

isotopic labelling to provide additional experimental data [Wiechert 2001]. Among different stable isotopes, methods based on ^{13}C isotope labelling are most widely used to explain the cells central metabolism. In ^{13}C MFA the intracellular flux is determined by considering the reversibility of reactions without making any predictions on metabolic energy states [Wiechert 1996]. Although this kind of modelling is useful in making experimentally testable predictions such as the viability of testing the knock-out, by itself it is not sufficient to uniquely calculate intracellular fluxes [Steuer 2008]. Neither does it take into consideration the dynamic nature of the network. Considering the dynamic behaviour displayed in biological systems, the inability to incorporate dynamics into the model limits the applicability of stoichiometry based modelling in predicting network functionality [Steuer 2008].

1.2.3 Kinetic modelling

Kinetic modelling is the most detailed and complex approach in systems biology [Schallau 2010]. It is a mathematical modelling approach that is guided by the detailed mechanism in metabolisms and gene regulation. It describes quantitative dynamics of the network by incorporating kinetic properties with the known stoichiometry of a pathway [Chou 2008]. Such a detailed model helps to formulate a correct, experimentally testable hypothesis, investigate the design principles of cell functions and predict dynamic changes in metabolite and protein concentrations [Steuer 2006].

Construction of a kinetic model. The first step towards developing a kinetic model of a metabolic network is to identify the nodes (*e.g.* metabolites), their connectivity and the individual interaction behind each of these connections. These interactions are represented by enzymatic reactions. Next the network is defined through a system of ODEs defining the rate of change of a species $[X]$, or $\frac{d[X]}{dt}$. Based on the stoichiometry, these ODEs are formulated as the sum of the reaction rates (the flux, v_i of the enzymatic reaction) that are producing the species minus the sum of the reaction rates that are consuming the species. Each reaction rate may be formulated with a rate law, a mathematical description of a type of reaction. For example, the kinetic rate law of an enzymatic reaction in a metabolic network may be formulated as a mass action or Michaelis-Menten rate law. These rate laws are calculated depending on metabolite concentration and parameters such as the binding constant, K_m, or the maximal velocity, V_{max}, of a reaction. For a gene regulatory network they are dependent on the strength of promoter site, transcription factor and ribosomal binding site strengths, together with many other parameters. Once defined, the set of ODEs for the network may be used



b) Reaction Mechanism

r_1: $\rightarrow X_1$
r_2: $X_1 \rightarrow 2X_2$
r_3: $X_2 \rightarrow$

c) Rate Laws

$$v_1 = k_1$$
$$v_2 = \frac{v_{max} \times [X_2]}{K_m + [X_2]}$$
$$v_3 = k_2 \times [X_2]$$

d) Stoichiometric Matrix

$$N = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 2 & -1 \end{bmatrix}$$

e) Differential Equations

$$\frac{d[X_1]}{dt} = v_1 - v_2 = k_1 - \left(\frac{v_{max} \times [X_2]}{K_m + [X_2]} \right)$$
$$\frac{d[X_2]}{dt} = 2 \times v_2 - v_3 = 2 \times \left(\frac{v_{max} \times [X_2]}{K_m + [X_2]} \right) - k_2 \times [X_2]$$

Figure 1.3: Kinetic modelling of a cellular network. **(a)** Schematic diagram of a network where X_1 and X_2 represent the metabolites. **(b)** The reaction mechanism. In reaction r_1, X_1 is imported, in reaction r_2 X_1 is converted to X_2 and in r_3 X_2 is consumed. **c)** The rate laws, As an input v_1 is a zero order kinetic, v_2 is formulated with the Michaelis-Menten rate law and v_3 with the simple mass action formula.The parameters are k_1, k_2, K_m and v_{max} . **(d)** Stoichiometric matrix of the network. Each column represents a flux and each row a metabolite. **(e)** The complete mathematical model of the network with a set of ODEs.

to simulate and predict the behaviour of the network. Figure 1.3 gives an overview of a kinetic model for a simple metabolic network. Figure 1.3 (a) illustrates the simple network composed of a linear pathway with two intermediate substrates. The fluxes are denoted v_i, where v_1 defines the import of X_1, v_2 defines the conversion of X_1 to 2 units of X_2, and v_3 exports X_2. Figure 1.3 (b) describes this network in terms of its reactions. Figure 1.3 (c) gives the specific rate law for each of the fluxes. From the reactions, the stoichiometry of the network is described in Figure 1.3 (d). The full kinetic model is given in Figure 1.3 (e) with the complete set of the ODEs that describe this network. The same principle of kinetic modelling can be applied to other biological networks, such as gene regulatory networks where the metabolites would be replaced with protein and/or

mRNA concentrations.

The complexity of biological networks and the limited knowledge of the underlying mechanisms and their associated parameters has inhibited the success of kinetic modelling. As most biological models are formulated as non-linear ODEs, the parameter optimization techniques are more difficult compared to a linear model. Furthermore they exhibit multi-modality (i.e. multiple optimum solutions, which cause the algorithms to converge to local minima [Singh 2006]). Stiffness, characteristics of differential equations that cause numerical instability unless the step size is taken to be extremely small, adds to the complexity, as the ODEs need to be solved repeatedly during the parameter estimation procedure. Non-identifiability of parameters further increases the difficulty of estimating a non redundant parameter set.

The understanding of complex genotype and phenotype relationships play a vital role in the understanding of the organism. It defines how the actual set of genes of an organism results to its observable expression of characters and traits [Mehmood 2011, Miko 2008]. Due to the number of molecular components in the cell and their non-linear interactions it is only possible to identify this relationship through accurate mechanistic computational modelling [Jamshidi 2009]. Thus the aim of this thesis is to develop a framework to accurately and efficiently estimate the unknown parameter values in order to overcome the bottleneck in kinetic modelling. In this framework I use filtering techniques from control theory in which the biological system is represented as a state-space model. Details of state-space modelling are described in the following section.

1.3 State-space model

State-space modelling is a mathematical representation of a system that originates from control engineering. In this model a physical system is represented as a set of input, output and state variables related by first-order differential equations. The state variables represent the current value of internal elements of the system state at any given time. This system representation of first-order differential equations is collectively known as the **state equation**, and explicitly accounts for state variables. The state equation describes the relationship between the system's input and current state with its future state. The state variables change independent of the output of the system. The state variables are not always measured directly, but rather through the **observation equation** they can be derived from the current system state together with the current measured input/output data. The observation equation shows the relationship between the system state and its input with the output. Together the state equation and the observation equation, forms the

state-space equation which ultimately forms the state-space model. There is no unique formulation of the state-space equations for a particular system, however state-space modelling is a convenient way to model both linear and non-linear dynamic systems with multiple inputs and outputs, as is the case for biological models.

1.3.1 Linear State-space modelling

Linear dynamic systems are mathematically represented with linear differential equations. For a system with inputs, $u \in \mathbb{R}^m$, outputs, $y \in \mathbb{R}^q$, and state variables, $x \in \mathbb{R}^n$, the state-space definition for a linear time invariant system (a system whose output does not depend explicitly on time) can be written as,

$$\dot{x}(t) = Ax(t) + Bu(t) \tag{1.4}$$

$$y(t) = Cx(t) + Du(t) \tag{1.5}$$

where $A \in \mathbb{R}^{n\times n}$ is the state transition matrix determining the dynamics of the system. $B \in \mathbb{R}^{n\times q}$ is the input matrix, determining how the system input u effects the system change. If the state change is independent of either the current state or the system output, then the corresponding matrix, A or B, would be zero. $C \in \mathbb{R}^{m\times n}$ is the output matrix that defines the relationship between the system input and the system output. Finally $D \in \mathbb{R}^{m\times q}$ is the feed-forward matrix. In the absence of any direct feedthrough to the system model, D is zero. Given the state vector and the input at time t Equation (1.4) determines the rate of change of state and Equation (1.5) determines the output of the system. In other words, given an initial state $x(t_0)$ and input $u(t)$ for $t_0 \leq t < t_f$, the system output $y(t)$ and the state $x(t)$ for $t_0 \leq t < t_f$ can be computed.

In order to perform the state-space calculation with a digital computer, the system needs to be discretized. Discretization transfers the process of a continuous model into its discrete counterpart. This is generally performed as a first step to ensure that the model is suitable for numerical implementation. Using discrete data sets, the discretized versions of Equation (1.4) and (1.5) are

$$x(k) = A_d x(k-1) + B_d u(k) \tag{1.6}$$

$$y(k) = Cx(k) + Du(k) \tag{1.7}$$

where k denotes the iteration number. The ”d” subscript for matrix A and B is used to denote the discrete form of these matrices, which are calculated through piecewise

operations as

$$A_d = e^{AT} \tag{1.8}$$

$$B_d = \int_0^T e^{A\tau} d\tau B \tag{1.9}$$

where T is the sampling interval. The C and D matrix remain the same in the discrete system as in continuous system. If the continuous-time state transition matrix A is non-singular, then its inverse can be used to define B_d

$$B_d = A^{-1}(A_d - I)B \tag{1.10}$$

where I represents the identity matrix. The matrices in a continuous time system and a discrete system are different due to the difference in the underlying equation representing the system. Linear differential equations are used to represent the continuous-time system where as discrete systems are described by difference equations.

The model represented by these equations does not consider the noisy nature of the system or the measurement. To introduce noise into the state-space equations a stochastic description of the method is needed. The following equation is formulated in order to represent the state-space model in a discrete-time process that is governed by the linear stochastic difference equation

$$\begin{aligned} x(k) &= A_d x(k-1) + B_d u(k) + w(k) \\ y(k) &= Cx(k) + Du(k) + e(k) \end{aligned} \tag{1.11}$$

$w(k)$ is the process noise and $e(k)$ is the measurement noise. Both of these noises are assumed to be Gaussian white noise (a discrete-time stochastic process whose terms are uncorrelated, Gaussian and all with zero mean), having the probability density function [Welch 2006]

$$p(w) \sim N(0, Q) \tag{1.12}$$

$$p(e) \sim N(0, R) \tag{1.13}$$

where Q is the process noise covariance matrix and R is the measurement noise covariance matrix.

1.3.2 Non-linear State-Space modelling

The dynamics of linear systems are restricted. Typically they do not represent real life applications accurately, as most real life applications tend to be highly non-linear. Therefore non-linear state-space models are needed to describe more realistic processes with complex dynamic behaviour. Non-linear continuous time state-space models can be described using a set of non-linear system equations, and represented as

$$\dot{x}(t) = F(x(t), u(t)) + w(t) \tag{1.14}$$

$$y(t) = H(x(t), u(t)) + e(t) \tag{1.15}$$

This general form is similar to the linear state-space representation except that the state transition process and observation processes are defined with functions instead of linear combinations (matrices). The state transition is defined by a non-linear function F, where the current state and exogenous known input vector are the parameters of the function. The observation equation that maps the time series measurement from the states variables and exogenous input is defined by function H, the non-linear observation function. Similar to the linear stochastic system, a zero mean uncorrelated Gaussian process noise with covariance Q is added to the state transition function to describe the uncertainties in the model. For uncertainty in the data uncorrelated Gaussian measurement noise with covariance R is added to the observation of the state through function H [van der Merwe 2004]. Larger entries in the Q matrix denote uncertainty in the accuracy of the model and larger entries in the diagonal of matrix R denotes uncertainty in the measurement [Lillacci 2010]. For a system with no exogenous input the state-space equation would be

$$\begin{aligned} \dot{x}(t) &= F(x(t)) + w(t) \\ y(t) &= H(x(t)) + e(t) \end{aligned} \tag{1.16}$$

This non-linear state-space representation is suitable for modelling real life scenarios, in particular biological phenomena. Throughout this thesis all biological systems are represented in this form.

1.3.3 Non-linear state-space modelling of a biological network

The non-linear dynamics of biological systems are efficiently described using a non-linear state-space representation. This representation formulates the processes of a

biological network with a set of ordinary differential equations(ODEs), a convenient and powerful method to capture the dynamics of a biological system [dÁlchè Buc 2010]. To develop this state-space representation it is required to identify the variables which are categorized as input, output and state variables as well as their components and interactions. State equations define the evolution of these state variables over time. In a biological network these state variables generally represent the concentration of species, such as metabolites or proteins. The state equations represent the rate laws through which these concentration changes over time. Though these state variables are essential for the dynamics of the system, they are not always directly accessible directly via measurements, and thus are called hidden variables. The observation function relates the state variables with the output variables through which they can be observed. These observation equations describe how the concentration of these species are observed through different measurements. One other important factor in these systems is the parameters that defines the rate coefficient such as maximal enzyme activity or affinity of the enzyme of a metabolic network. Many of these parameters are hard to measure in wet-lab experiments and therefore need to be estimated *in silico*. One of the main objectives of this thesis is to estimate these parameter values. Taking these parameters into consideration, the biological network can then be represented through a continuous-time non-linear state-space model as

$$\begin{aligned} \dot{x} &= F(x, \theta, u, t) + w(t), \qquad x(t_0) = x_0 \\ y &= H(x, u) + e(t) \end{aligned} \tag{1.17}$$

where $x = [x_1, x_2, x_3, \ldots, x_n]$ represents the species concentration (state vector), t is the time, $\theta = [\theta_1, \theta_2, \ldots \theta_m]$ represents the kinetic parameters [Quach 2007], $w(t)$ is the process noise and $e(t)$ is the measurement noise. The initial state vector (i.e. species concentration), is denoted x_0. The observation equation, H, when added to the noise $e(t)$, produces the observed or output variable, y, from the current state, x. It is often the case in real life applications that the process and measurement noises are additive [van der Merwe 2004]. For this reason, both the process and measurement noise are considered additive in the model rather than incorporating them into the state equations.

With deterministic models of a biological system it is often the case that the continuous time process is estimated using discrete time measurement data. The continuous time state can be cast into a discrete time non-linear state-space equation

through the function f and $x(t_k) = x(k)$ for $k \geq 0$ [Quach 2007],

$$f_k(x(k); \theta, u) = x(k) + \int_{t_k}^{t_{k+1}} F(x(\tau), u, \theta) d\tau \tag{1.18}$$

$$x(k+1) = f_k(x(k), u, \theta) + w(k) \tag{1.19}$$

Throughout this work it is this discrete form of states-space models that is used in the estimation of the state variables.

1.4 Parameter Estimation

The motivation behind developing data-driven models in biological systems is to shed light on the different functionality of living cells and how they can be influenced towards certain behaviour [Balsa-Canto 2008]. These models provide a complete view of dynamic interactions between different intracellular pathways. In order to have a correct predictive behaviour it is important to have a full and accurate set of parameters. However models in systems biology are disproportionate in the relatively small amount of available data compared to the relatively large number of parameters in the rate laws [Liebermeister 2006]. As a result very few parameters can be obtained either from literature or experimental data. Even when parameters are available they tend to be determined with large uncertainty or under environmental conditions different to the current experiment. In some cases determination of these parameters *in vivo* is just impracticable [Diego 2010]. Therefore the values of these parameters need to be predicted computationally. Successful and accurate estimation of these parameter values is a critical part of modelling in systems biology [Ashyraliyev 2009, Sanders 2009].

Parameter estimation in biological models is challenging as the available experimental data are sparse and corrupted with noise. Furthermore the value of model parameters may range over many orders of magnitude, the estimation method might get stuck in local minima or it might roam around a very flat parameter space. What is required is an efficient parameter estimation method having the capability of dealing with these problems.

Depending on the characteristics of experimental data, parameter estimation methods can be categorized in to two types, **steady state** and **time series**. A system is in steady state when the system properties are no longer changing, that is, the rate of change of the system state over time is zero or $\frac{dx}{dt} = 0$. Data gathered at this stage is called steady state data. Parameter estimation methods using steady state data are based on experiments where they try to measure the response of a system after a small perturbation around

steady state. In biological systems it is easier to experimentally measure the steady state data *in vivo*. However, due care needs to be taken to ensure that the system has reached steady state before data is collected. With steady state data it is often not possible to accurately estimate parameters as the data does not describe the dynamics of the system. Its wide use is mainly due to the comparative experimental simplicity.

Time series data gives a sequence of data points measured at specific time intervals during the transitional period of a dynamic system. This time series data is collected before the system reaches steady state, i.e. $\frac{dx}{dt} \neq 0$. Time series data gives a more detailed picture of the system, as it describes the dynamic relationship between different intracellular biochemical processes. Parameter estimation algorithms, especially the recursive methods, perform with much higher accuracy when they estimate parameters based on time series data, as parameter values can be adjusted at each time step to more properly represent the dynamics of the measurement data. The flip side is that it is much more difficult to measure this data from biological experiments compared to steady state data. However with the ongoing development of high throughput experimental devices and methods, such comprehensive data is becoming available. These tools can even generate time-series data under different conditions such as with different gene knock-out experiments [Chou 2008], making time series data the *de facto* standard for *in silico* parameter estimation.

In this thesis I model the biological processes using non-linear state-space modelling, where the system is described by the (ODE) presented in Equation (1.17). This type of model describes the evolution of a cellular component over time. At the molecular level the variables, or states, are the concentration of a chemical component such as metabolite or protein concentration. The parameters represent the components which remain constant over the dynamics of the state such as the binding affinity of the Michaelis-Menten rate law (K_m) or inhibition constant (K_i). Parameter estimation tries to estimate the unknown value of these constants. In recent years the development of an efficient parameter estimation technique has been the thrust of increasing research [Moles 2003, Schaber 2011, Lillacci 2010, dÁlchè Buc 2010, Quach 2007, Ashyraliyev 2009, Balsa-Canto 2008, Chou 2008, Rodriguez-Fernandez 2006a, Diego 2010, Sun 2008, Moles 2003, Rodriguez-Fernandez 2006b, Schenkendorf 2011, Jia 2011]. In this section I will describe two of the main approaches to *non-linear optimization* in addition to the *filtering approach*.

1.4.1 Non-linear Optimization

In non-linear optimization the parameter estimation method is formulated with differential algebraic constraints where the objective function is derived as the weighted distance between the model prediction and experimental data [Moles 2003, Baker 2010].

$$j = \int_0^t (y_{mes}(t) - y_{pre}(\theta, t))^T W(t)(y_{mes}(t) - y_{pre}(\theta, t))\mathrm{d}t \tag{1.20}$$

The task is to minimize the objective function, j, over the time series data where θ is the vector of parameters that the optimization problem is supposed to estimate. The objective function is a weighted sum of squared errors, where the error is given as the difference between the experimental or measured data, y_{mes}, and the simulated value y_{pre}, at the start of the current iteration. A weighting matrix, W, is used to vary the effect of individual differences. The inverse of the standard deviation of the measurement data is commonly used for the weights, reducing the effect of less precise measurement data from the objective function calculation. The minimization of the objective function is an inverse problem, that is, it tries to estimate the parameters from the measured data compared to a forward approach where all model information is used to generate measurement data. Figure 1.4 gives an overview of this optimization. The optimizing function receives measurement data from the measurement unit and simulated data from the biological model, generated using the current estimate of the parameter values. The optimization algorithm tries to minimize Equation (1.20), by iteratively updating the estimated parameter values, until the specified stop criteria has been reached.

Optimization methods are classified as either *local* or *global*. Local methods find a solution that is optimal within a neighbouring set of solutions but might not be the global optima. Due to their non-linear and constrained nature, biological models are quite often multimodal, meaning that multiple local minima exist. If sufficiently close to some minima, local methods will quickly converge. However, they often get stuck in local minima and are unable to find better optima, let alone global minima [Rodriguez-Fernandez 2006a]. Global optimization methods search throughout the parameter space for the best minimum value of the objective function. The downside of global optimization algorithms is that there is no proof of convergence [Ashyraliyev 2009].

Next I will briefly describe these two general classes together with some of the more widely used algorithms.

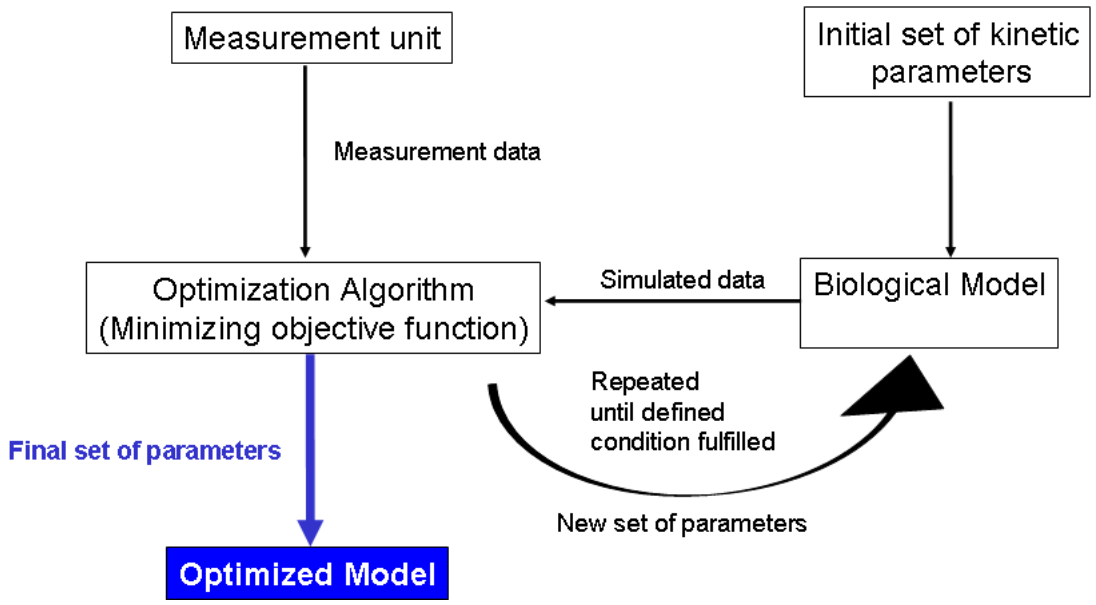


Figure 1.4: General scheme of the optimization method. Simulated data comes from the model, and measurement data comes from the measurement unit. The optimization is repeated until a pre-defined condition is met.

1.4.1.1 Local Optimization

The two most common methods of local optimization are the gradient search method and the direct search method. The gradient search method finds an optima when the calculated gradient of the objective function approaches zero, i.e. a plateau. The direct search method tries to find the minimum by bracketing the solution and rejecting any point that is not adjacent to the lowest functional values.

Gradient search method. Gradient based methods are all descent methods. These methods search for the local minimum of the objective function, j, by taking steps proportional to the negative of the gradient of that function,$-\nabla j$ at the current position [Ashyraliyev 2009]. In doing so it first finds a descent direction $dy = -\nabla j$ and takes a step αdy towards that direction, where α is the step size [Kaj 2004, Wang 2008, Ashyraliyev 2009]. This step size can vary with each iteration. The selection of α, i.e. how far the step should be in the direction of dy, is determined through a line search. This method starts from an arbitrary initial point $y(0)$ and then follows the gradient until it reaches the solution. This iterative procedure can be described mathematically as

$$y(k+1) = y(k) - \alpha_k \nabla j(y_k) \tag{1.21}$$

Newton's Method. Newton's method also known as Newton-Raphson method, is an iterative 2^{nd} order numerical method that attempts to find the optima of a real valued function. It calculates the Hessian (2^{nd} order derivatives) along with the gradient of the function to find the critical point at which the gradient of the function is zero. Thus this method is analogous to Newton root finding method.

Gradient based non-linear least square problem. In the case of minimizing a sum of square errors of a function, within a gradient based search method, least square can be applied. This method attempts to minimize the norm of the objective function. For solving such problems *Newton's* method is replaced with the *Gauss Newton* method so that the Hessian is not needed.

Levenberg-Marquardt method (LM). This method combines the advantages of steepest descent method with the Newton method [Marquardt 1963]. It tries to find the local minima of a function expressed in the form of sum of squares. It uses the minimization along the direction of the gradient to obtain its operating stability and uses a quadratic model to speed up the process of convergence to the vicinity of the minimum.

1.4.1.2 Global optimization

Global optimization methods can be broadly classified as deterministic, stochastic and metaheuristic. Deterministic methods can guarantee global minima for relatively small problem sizes. However as computational complexity increases rapidly for large problems, no deterministic method can find the global minimum in finite time for such problems. As biological models have large number of unknown parameters, these methods do not perform well in such large scale problems [Rodriguez-Fernandez 2006a].

Stochastic methods are probabilistic in nature and can locate the vicinity of global solutions in modest computational time. These methods can be successfully paired with local methods [Banga 2003].

Lastly, metaheuristic methods are used to iteratively improve the candidate solution with regard to a given measure of quality. These methods diversify the search over the solution space by swapping between different fitness landscapes. This swapping is done through intensification where the search is performed carefully around good solutions and diversification, which deals with the search of unvisited regions. Many of the algorithms in this class involve some form of stochasticity to accelerate the search [Hoos 2004]. Though metaheuristic-based methods do not guarantee that a global optimization is ever found, they are the most successful when balancing computational complexity and the overall convergence of the algorithms [Rodriguez-Fernandez 2006a].

In this thesis a comparison is made between the most widely used global optimization algorithms and the most widely used non-linear extensions of Kalman Filter including the proposed filtering method. These global optimization algorithms, *simulated annealing (SA)*, *genetic algorithm (GA)*, *evolutionary programming (EP)* and *particle swarm optimization (PSO)* will be briefly described in the reminder of this section. All of these algorithms with the exception of SA are metaheuristic-based, while SA is stochastic based. The description of the Filtering techniques can be found in section 1.5.1.1 and onwards.

Simulated Annealing (SA). Simulated Annealing is a stochastic optimization algorithm proposed by Kirkpatrick *et al.* (1983) [Kirkpatrick 1983]. The name derives from annealing in metallurgy, a physical process involving the heating and controlled cooling of a material to increase the energy of its crystals to yield the desired properties. In this method each point of the search space corresponds to a state of some physical system, and the objective function captures the internal energy of the system in that state. The algorithm starts with a high temperature indicating equal probability of all states or parameter vectors distributed over the parameter space. Then the temperature is slowly decreased based on a cooling schedule to come up with a new distribution of parameters. By allowing heating, i.e., additional energy in the system that allows uphill move, this method eliminates the problem of getting stuck in local minima.

Evolutionary Computation (EC). Evolutionary computation is a class of algorithms based on the principles of biological evolution. In EC each potential solution of the parameter vector is considered as the population and the new population is generated through reproduction, selection, mutation and survival of the fittest [Eiben 2008].

The algorithm starts by generating random parameter vectors (a population of random individuals). The fitness of each of these individuals is measured through their corresponding objective function (higher is better). The selection process assigns each individual a probability to indicate their chance of being selected in the next generation relative to their fitness values, a higher fitness value results in a higher probability of selection. Crossover and mutation are then used to create new individuals. The crossover or recombinant operator selects the parents and makes the crossover between them to generate children (the new population). Mutation operates on one candidate and mutates it to make sure that the new population is not too similar to the parent population. In this manner a new set of candidate solutions are created. These new candidates then compete for their place with the old ones to be selected in the next generation, i.e. survival of the fittest. Two of the most popular algorithms in this class are, evolutionary programming

(EP) and genetic algorithm (GA).

The main difference between evolutionary programming and genetic algorithms is in how each goes about generating a new population. EP uses only mutation whereas GA uses both crossover and mutation for generating new candidate solutions.

Particle Swarm Optimization (PSO) A second biologically inspired optimization method is particle swarm optimization (PSO) proposed by Kennedy and Elberhart (1995) [Eberhart 1995]. This algorithm imitates the social behaviour and movement dynamics of insects, birds and fish. Like EC, PSO randomly generates parameter vectors as candidate solutions (a random population) and dubs them as particles. These particles have a position and velocity in the search space. The algorithm makes changes to the position and velocity, so that the particles move around the search space in order to optimize the solution of a problem. The movement of the particles is influenced by knowledge of their own best position combined with information of their neighbour's best position. This makes the swarm tend to move towards the global optimum.

Other global optimization methods A large number of global optimization methods have been developed. Banga *et al.* [Banga 2003] provides a good overview of many of these algorithms.
A number of hybrid approaches have been developed to combine the advantages of both the global and local optimization algorithms. These approaches combine global optimization methods with local ones. They utilizes the global optimization method to determine the vicinity of the global minima and then use local search methods to quickly converge to the minima [Balsa-Canto 2008, Hedar 2002, Pedamallu 2008]. For example, SA takes a long time to reach the global minima, but it reaches to the vicinity of the global optima fairly quick. When SA is combined with other local optimization algorithms it yields a considerable saving in computational effort.

Bayesian methods have also drawn a lot of interest in parameter estimation as they can extract information from noisy or uncertain data [Lillacci 2010]. The advantages of such methods are that they can estimate a whole probability distribution of a parameter rather than making a point estimation.

1.5 Parameter estimation as non-linear sate estimation

Recently parameter estimation has been addressed in the framework of control theory using state observers [Lillacci 2010]. In this way the parameter estimation problem is formulated as a state estimation problem and filtering techniques are applied to perform this state estimation [Denis 2003]. Details on filtering techniques and their estimation strategies can be found in the next section.

Non-linear state estimation is the method of estimating the hidden variables of a system. Hidden variables are those that are not directly observed, but rather inferred from observed or measured variables. An optimal set of these hidden states is thus derived from a set of noisy and incomplete measurement data where the process model and/or the measurement model are non-linear. Whenever a state estimation is required from noisy measurement data, a state estimator is applied to fuse the measurement data from different sources to derive an optimal estimate of this state [Julier 1997].

The state-space definition can be extended to facilitate simultaneous state and parameter estimation by treating the parameters to be estimated as augmented states [Jazwinski 1970]. These parameters are constant values in the model, i.e. their rate of change with respect to time is 0. In treating these parameters as an additional state vector (with a 0 rate of change), it is possible to extend the state-space definition to include them [Lillacci 2010]. Thus by treating the parameters as functions of time instead of constants, the parameter estimation problem is converted into a state estimation problem which can be addressed within the framework of control theory. The extended system is described by

$$\left.\begin{aligned} \dot{x} &= F(x,\theta,u,t) + w(t), \qquad x(t_0) = x_0 \\ \dot{\theta} &= 0 \\ \theta(t) &= \theta_0, \qquad t \geq t_0 \\ y(t) &= H(x,t) + e(t) \end{aligned}\right\} \tag{1.22}$$

In the equations, F and H represent the non-linear state and observation function, respectively, both corrupted with noise. The state vector x is then augmented to make $x^{aug} = [\theta_1 \ldots \theta_l \quad x_1 \ldots x_m]$ having the dimension $n = l + m$ where l is the state dimension and m is the parameter dimension. For simplicity the augmented state vector x^{aug} will be refereed to as x from this point unless otherwise specified. This state extension tries to identify the initial value for θ that when used would generate the observed output y. As the rate of change for θ is 0, $\theta(t) = \theta_0$ for all $t \geq t_0$. To make the representation of the final non-linear state-space suitable for numerical evaluation and implementation with

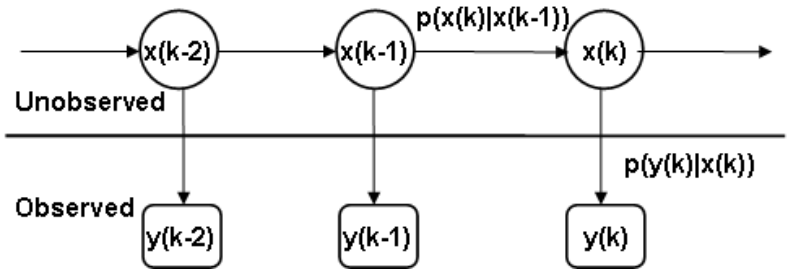


Figure 1.5: Graphical representation of a dynamic state-space model with their probabilistic inference

discrete data, the non-linear state-space model is discretized. The function f represents the discretized form of the state function in Equation (1.18). As the rate of change for the parameters are zero, they are not integrated rather taken directly as $f(\theta) = \theta_0$. The measurement function h takes the discrete state from f to calculate the value of the observables. The final non-linear discrete state-space equation is

$$\begin{aligned} x(k) &= f(x(k-1), u) + w(k) \\ y(k) &= h(x(k), u) + e(k) \end{aligned} \tag{1.23}$$

In Equation (1.23) the augmented state-space variable, x, represents both the parameters and the state variables of the model and k is the iteration number. As mentioned earlier, the augmented state variable will be referred to as the state variable and simply denoted $x(k)$. A non-linear state estimation problem is mainly addressed in probabilistic view [Schön 2006]. As discussed in the literature [van der Merwe 2004], a probabilistic inference (estimation based on probability) of a state in a non-linear discrete dynamic system can be described using a dynamic state-space model such as the one shown in Figure 1.5.

The state variable $x(k)$, with an initial probability density function (pdf) $p(x(0))$, evolves over time as a partially observed first order Markov process according to the conditional probability distribution $p(x(k)|x(k-1))$ [van der Merwe 2004]. The observations $y(k)$ are conditionally independent of other observations given the state variable $x(k)$ at the current step. These observations are generated according to the conditional pdf $p(y(k)|x(k))$. When a system is represented with a state-space model as in Equation (1.23), then the state transition density, $p(x(k)|x(k-1))$, would be completely specified by f and the process noise pdf. The observation function, h, and observation noise pdf specifies the observation probability density, $p(y(k)|x(k))$. In a Bayesian

framework, the posterior density of the state given all the observations $p(x(k)|y(1:k))$, yields the complete solution for the probabilistic inference problem. It calculates the optimal estimate of the state through the conditional mean as $\hat{x}(k) = E[x(k)|y(1:k)]$. For an optimal state estimation from a non-linear model it is necessary to have a complete description of conditional probability density [Julier 1997, Kandepu 2007]. Generally different filtering techniques are being applied for the propagation of this probability density function. These filtering techniques are central to the statistical problem of state-space modelling and try to estimate the states using the probability density function [Koyama 2010].

1.5.1 Filtering techniques

As the name implies a filter is anything that can be used to extract usable information from noisy data. As applied to state estimation, filtering techniques address the problem of estimating the hidden states from noisy measurement data. In kinetic models of a biological network, these filters can be used to estimate metabolite concentrations that are not directly measurable due to a lack of appropriate methods or instruments to measure the concentrations. Filters can also be used to estimate the parameters of these kinetic models when represented with an augmented state-space equation.

Having a stochastic state process and a related observation process at discrete time steps $k = 0, 1, 2, \ldots n$, the filtering technique aims at estimating the state, $x(k)$, by applying a Bayesian approach given the sequence of observations, $y(1), \ldots, y(k)$. In other words filtering tries to find the posterior distribution of states given all observation data, $p(x(k)|y(1:k))$. When the state follows a Markov process having a transition density $p(x(k)|x(k-1))$ with initial density $p(x(0))$ [Koyama 2010], the Bayes formula for recursive filtering becomes:

$$p(x(k)|y(1:k)) = \frac{p(y(k)|x(k))p(x(k)|y(1:k-1))}{\int p(y(k)|x(k))p(x(k)|y(1:k-1))dx(k)} \tag{1.24}$$

The solution of Equation (1.24) is extremely general and incorporates multi-modality, discontinuities and asymmetries. However it requires the propagation of the full pdf. But as the form of the pdf is not restricted, it cannot be described using a finite number of parameters [Julier 2004] which is applicable for any form of the pdf. Furthermore, it is extremely complicated to evaluate the multi-dimensional integral for the normalizing constant [Press 1992, van der Merwe 2004]. For this reason the application of this approach has been limited, typically requiring some kind of approximation. To avoid this

inherent limitation of this method, alternative estimators have been developed. Among these, the most widely used is the Kalman filter (KF) [Julier 2004].

1.5.1.1 Kalman filter (KF)

The Kalman filter was proposed by R. E. Kalman in 1960 and describes the recursive solution to the discrete data linear filtering problem [Kalman 1960]. The Kalman filter provides an optimal minimum mean squared error (MMSE) estimate for linear system dynamics and observation models [Julier 1997]. It uses only the first two moments (mean and covariance) to update its state estimate. Although this gives a simple representation of the states, the compromise between computational complexity and representational flexibility provides sufficient information for most types of operational activities [Julier 2004]. As both the mean and covariance are linearly transformable quantities, they can be maintained effectively when transformed through linear state-space and measurement equations [Julier 2004].

The Kalman filter estimates the state of a process in a manner described as feedback control. It first estimates the process state at one point in time and then obtains feedback in the form of noisy measurements. This process of estimation is divided into two sets of equations, *time update* equations and *measurement update* equations, i.e. a predictor corrector process [Welch 2006]. The time update equations, the predictor, calculate the *a priori* estimate of the next time step by projecting forward the mean and covariance estimate of the state from the current time step. The measurement update equations, the corrector, calculate the *a posteriori* estimate by incorporating the measurement data into the a priori estimate for a specific time step.

For a discrete time controlled process the state and measurement equations are defined as

$$x(k) = Ax(k-1) + Bu(k-1) + w(k-1) \tag{1.25}$$

$$y(k) = Cx(k) + e(k) \tag{1.26}$$

where w and e represents the process and measurement noise, A is the state transition matrix, B the input and C the output matrix. This system considers that there is no feedthrough in the system and therefore matrix D is zero. Detail explanation of matrix A, B, C are given in section 1.3.1. For the Kalman filter the initial estimate of the state

and error covariance are:

$$\hat{x}(0) = E[x(0)] \tag{1.27}$$
$$P(0) = E[(x(0) - \hat{x}(0))(x(0) - \hat{x}(0))^T] \tag{1.28}$$

here $x(0)$ is the value of the state-space vector and $P(0)$ the covariance at iteration 0. The expectation operator E is taken to calculate the average of the random variable corresponding to its probabilities. Then the predictor-corrector equations are:

1) time update (predictor) Before the measurement data is available the a priori mean and covariance of the state is calculated as

$$\hat{x}^-(k) = A\hat{x}(k-1) + Bu(k-1) \tag{1.29}$$
$$P^-(k) = AP(k-1)A^T + Q \tag{1.30}$$

where Q represents the process noise covariance.

2) measurement update (corrector) When the measurement data $y(k)$ is available the Kalman gain K is first calculated followed by the posterior state and covariance estimate as follows

$$K(k) = P^-(k)H^T(HP^-(k)H^T + R)^{-1} \tag{1.31}$$
$$\hat{x}(k) = \hat{x}^-(k) + K(k)(y(k) - H\hat{x}^-(k)) \tag{1.32}$$
$$P(k) = (I - K(k)H)P^-(k) \tag{1.33}$$

Steps 1 and 2 are then repeated to iteratively improve the estimate of the state $x(k)$.

The Kalman Filter only works when both the process and measurement models are linear. Unfortunately this is not the case with most real life applications. Instead system dynamics and observation equations are non-linear which necessitated the development of non-linear extensions to the Kalman filter [Julier 1997]. The two most widely used non-linear extension are the extended Kalman filter and the unscented Kalman filter.

1.5.1.2 Extended Kalman Filter (EKF)

The extended Kalman filter (EKF) can estimate the states when the state dynamics and/or the observation dynamics are non-linear. The EKF does this by utilizing the Kalman filter after linearizing the non-linear system around the previous mean and covariance estimate [Ribeiro 2004]. Although EKF has been the *de facto* standard for non-linear

dynamic systems, it has several drawbacks [Julier 2004]. The linearized approximation can be very poor if the error propagation cannot be well estimated with the linearized transformation. Determining this approximation can be very difficult as it depends on the transformation, the current mean and the covariance value. Furthermore the Jacobian is required to exist, which is not guaranteed. For a discontinuous system or a system with a singularity it would not be possible to calculate the Jacobian and the filtering technique fails. Additionally EKF does not account the probabilistic spread of the prior state distribution when linearizing the process and observation model. This results in an inaccurate approximation of the posterior statistics when the non-linearity in the prior distribution is high [van der Merwe 2004]. Finally, even when available the Jacobian calculation is computationally expensive and error prone.

1.5.1.3 Unscented Kalman Filter (UKF)

The unscented Kalman filter (UKF) takes a different approach in state estimation for the non-linear systems [Merwe 2001]. It does so without having to calculate any linearization steps and having an equivalent performance to the KF applied to a linear system [Julier 1997]. Instead of calculating the Jacobian (as required by EKF) the UKF uses a statistical linearization technique that utilizes a linear regression by drawing 'n' points from the prior distribution of the random variable [van der Merwe 2004]. It is more accurate than EKF as it considers the spread of the random variable [van der Merwe 2004]. At the core of the UKF is the *unscented transformation (UT)* which uses a set of deterministically chosen weighted points, called *sigma points*, to parametrize the mean and covariance of a probability distribution [Julier 1995, Julier 1996, Julier 1997, Julier 2000, Julier 2004, van der Merwe 2004].

The main idea behind UT is that it is easier to approximate a Gaussian probability distribution with a fixed number of parameters than to approximate an arbitrary non-linear function [Julier 1996]. This leads to a parameterization that captures the same mean and covariance while permitting the direct propagation of information through an arbitrary non-linear transformation. This is achieved by generating a discrete distribution having the same first and second moments using a minimum number of points where each of the points can be directly transformed [Julier 1996, Julier 1997]. The statistics of the transformed points can be used to calculate the mean and covariance of the non-linear transformation. For an n-dimensional Gaussian distribution having the covariance matrix P, the sigma point selection scheme calculates the minimum number of sigma points based on the column or row of the symmetric matrices $\pm\sqrt{nP}$. This set of sigma points has zero mean but the current state mean $\bar{x}$ value can be added with each of

these sigma points to generate $2n$ symmetric points matching the prior distribution of the current state $\bar{x}$. This method differs from the particle filter algorithm in that the sigma points are chosen deterministically, not randomly. One of the problems with this selection scheme is that as the dimension of the state-space (n) increases, the radius of the sphere bounding all sigma points also increases. While this does not affect the accuracy of the mean and covariance estimates, it does affect the cost of sampling non-local effects where locality is defined through the probabilistic spread of x summarized by its covariance [van der Merwe 2004, Julier 2002]. If the non-linearity is severe, this may lead to difficulties. One solution is a scaling scheme where by the sigma points are chosen as $\pm\sqrt{(n+\lambda)P}$, $\lambda = \alpha^2(n+\kappa) - n$ [van der Merwe 2004], where α and κ are the two scaling parameter.

Thus the UKF uses the UT to select the sigma points to calculate the mean and covariance of the non-linear transformed state random variables and then the Kalman filter is used to make the recursive estimation of the state.

In UKF the noise sources can be augmented into the equation or can be considered additive. For the biological problem the process noise and measurement noise are considered to be purely additive as it is widely believed that white noise need not be augmented into the system state and non-augmented UKF yields similar results [Wan 2001]. However, when additive noise is considered, a new set of sigma points has to be redrawn to efficiently incorporate the effect of additive process noise [Wu 2005]. The covariance of these noise sources are added to the state covariance with a simple additive procedure [van der Merwe 2004]. To estimate an n dimensional state vector, the complexity of UKF is $O(n^3)$. With the case of additive noise the equations for UKF are as follows

Initialization

$$\text{State vector:} \quad \hat{x}(0) = \mathbb{E}[x(0)] \tag{1.34}$$

$$\text{Covariance matrix:} \quad P(0) = \mathbb{E}[(x(0) - \hat{x}(0))(x(0) - \hat{x}(0))^T] \tag{1.35}$$

Weight Initialization:

$$W_0^m = \frac{\lambda}{n+\lambda}, \qquad W_0^c = \frac{\lambda}{n+\lambda} + (1 - \alpha^2 + \beta) \tag{1.36}$$

$$W_i^m = W_i^c = \frac{\lambda}{2(n+\lambda)}, \quad i = 1\ldots, 2n \tag{1.37}$$

For $k \in \{1, \ldots, T\}$

Sigma point calculation

$$\mathcal{X}(k-1) = \Big[\hat{x}(k-1) \quad \hat{x}(k-1) + \sqrt{(n+\lambda)P(k-1)} \quad \hat{x}(k-1) - \sqrt{(n+\lambda)P(k-1)}\Big] \tag{1.38}$$

Time update

$$\mathcal{X}^a(k|k-1) = f[\mathcal{X}(k-1), u(k-1)] \tag{1.39}$$

$$\hat{x}^-(k) = \sum_{i=0}^{2n} W_i^m \mathcal{X}_i^a(k|k-1) \tag{1.40}$$

$$P^-(k) = \sum_{i=0}^{2n} W_i^c [\mathcal{X}_i^a(k|k-1) - \hat{x}^-(k)][\mathcal{X}_i^a(k|k-1) - \hat{x}^-(k)]^T + Q \tag{1.41}$$

Sigma points are redrawn accounting for the effect of additive process noise.

$$\mathcal{X}^y(k|k-1) = [\hat{x}^-(k) \quad \hat{x}^-(k) + \sqrt{(n+\lambda)P^-(k)} \quad \hat{x}^-(k) - \sqrt{(n+\lambda)P^-(k)}] \tag{1.42}$$

$$\mathcal{Y}(k|k-1) = h[\mathcal{X}^y(k|k-1)] \tag{1.43}$$

$$\hat{y}^-(k) = \sum_{i=0}^{2n} W_i^m \mathcal{Y}_i(k|k-1) \tag{1.44}$$

Measurement update equations

$$P^{yy}(k) = \sum_{i=0}^{2n} W_i^c [\mathcal{Y}_i(k|k-1) - \hat{y}^-(k)][\mathcal{Y}_i(k|k-1) - \hat{y}^-(k)]^T + R \tag{1.45}$$

$$P^{xy}(k) = \sum_{i=0}^{2n} W_i^c [\mathcal{X}_i(k|k-1) - \hat{x}^-(k)][\mathcal{Y}_i(k|k-1) - \hat{y}^-(k)]^T \tag{1.46}$$

$$\mathcal{K}(k) = P^{xy}(k)(P^{yy}(k))^{-1} \tag{1.47}$$

$$\hat{x}(k) = \hat{x}^-(k) + \mathcal{K}(k)(y(k) - \hat{y}^-(k)) \tag{1.48}$$

$$P(k) = P^-(k) - \mathcal{K}(k)P^{yy}(k)\mathcal{K}^T(k) \tag{1.49}$$

In the equations Q and R are respectively the process and measurement noise covariance matrices. The three parameters that need to be determined are κ, α, β. To guarantee that the covariance matrix remains positive definite, κ is choosen to be non-negative. As the specific value of κ is otherwise not crucial it is possible to select $\kappa = 0$. The

spread of the sigma points has a lower bound to guarantee positive semi-definiteness of the covariance matrix. At the same time, increase of this spread with the dimension of the state-space will also cause problems due to the sampling of non-local features of a highly non-linear model [Orderud 2006]. Therefore a tuning parameter α is used to arbitrarily control the spread of the sigma points, while guaranteeing positive semi-definiteness of the covariance matrix. β is used to incorporate the knowledge of the higher order moments of the distribution and needs to be a non-negative number. For a Gaussian distribution the optimal value is $\beta = 2$ [van der Merwe 2004].

1.6 Identifiability Analysis

Given a mathematical model together with available input and output data, identifiability analysis aims at finding out whether it is really possible to estimate the value of the unknown parameters uniquely [Quaiser 2009]. Identifiability analysis is particularly significant as it determines the extent to which a particular model is suitable for the deduction of data as performed by a parameter estimation algorithm [Cobelli 1980]. It is therefore reasonable to limit parameter estimation to only those parameters that are identifiable. Identifiability analysis can be divided into two categories, structural identifiability and practical identifiability. This section starts with the discussion of structural identifiability analysis, for which several notions have been proposed in the literature. A brief review of the most widely used notions are given below.

A model is said to be globally or structurally identifiable if a unique value can be found for each parameter so that the model reproduces the simulated or measured value(s), under ideal, i.e. noise free conditions. Thus structural identifiability is directly related to the model structure. Few methods have been developed for the structural identifiability of non-linear models [Balsa-Canto 2010], and are further categorized into two sub-groups [Geffen 2008]. In the first group, the problem of identifiability analysis is treated as a problem of observability analysis, considering parameter estimation problem as a state estimation problem [Denis-Vidal 2001, Xia 2003, Pohjanpalo 1978, Vajda 1989, Ljung 1994]. The second group of methods address structural identifiability by checking for linear or non-linear functional relationships between parameters through simulation, optimization and parameter estimation [Hengl 2007, Quaiser 2009, Müller 2002].

In 1989, Vajda *et al.* extended the approach of similarity transformation to non-linear systems theory and used local state isomorphism to verify necessary conditions for structural identifiability [Vajda 1989]. In 1994, Ljung *et al.* demonstrated how structural

identifiability can be analyzed using methods from differential algebra. In this work the problem of structural identifiability was reduced to a question of whether it is possible to rearrange the model structure into a linear regression [Ljung 1994]. Pohjanpalo developed a widely used power series method, assuming that the derivatives of the observations with respect to time are unique and can be represented by a Maclaurin series expansion [Pohjanpalo 1978]. Quaiser *et al.* developed an eigenvalue based method that checks whether there is a functional relationship between the parameters by analysing the covariance matrix [Quaiser 2009]. For identifiability analysis Hengl *et al.* used the mean optimal transformation approach to determine functional relationships between parameters and concluded that functionally related parameters are not identifiable [Hengl 2007].

In control theory, the structural identifiability of a model is mainly addressed in terms of models' observability [Lillacci 2010]. Observability tries to answer the question of whether it is possible to know the dynamic behaviour of the state-space variable given only the system output. A system is observable if for any initial state $x(0)$ and any final time point $t > 0$ the initial state can be uniquely determined from the system output data [Simon 2006]. If the initial state $x(0)$ can be determined from system output, then starting from $x(0)$, any state of the variable may be determined recursively. In control theory, the parameters are represented as state variables (albeit constant in time) $\frac{d\theta}{dt} = 0$. Observability analysis is performed with the introduction of the *observability matrix*. To generate the observability matrix a linear time-invariant, discrete-time system in the state-space is considered as $x(k+1) = Ax(k)$ where $x(0) = unknown$ and measurement equation $y(k) = Cx(k)$ where, $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^p$. n measurement values would be sufficient, to determine the initial (n dimensional) state vector $x(0) = x_0$ from the measurement data. The following sequence of equations formulate the observability matrix

$$\left.\begin{aligned} y(0) &= Cx(0) \\ y(1) &= Cx(1) = CAx(0) \\ y(2) &= Cx(2) = CAx(1) = CA^2x(0) \\ &\vdots \\ y(n-1) &= Cx(n-1) = CA^{n-1}x(0) \end{aligned}\right\} \tag{1.50}$$

The matrix form will then be

$$\begin{bmatrix} y(0) \\ y(1) \\ y(2) \\ \vdots \\ y(n-1) \end{bmatrix}^{(np)\times 1} = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{n-1} \end{bmatrix}^{(np)\times n} \qquad x(0) = O^b x(0) \tag{1.51}$$

where O^b is the observability matrix. As per the linear algebra rule, a system of linear algebraic equations with n unknowns, has a unique solution if the rank of the system matrix is n. According to that definition the initial condition x_0 is uniquely determined if the rank of the observability matrix O^b is n, i.e. $rank(O^b) = n$. In the non-linear case system, the observation space O^b_s is defined as the space containing all repeated lie derivatives (a lie derivative defines the change of a vector field along another vector field).

For a linear system the initial state, x_0, can be uniquely determined and thus the system is observable if the rank of O^b is full. Similarly, a non-linear system is locally observable at x_0 if O^b has rank n [Karl 2010].

The main drawback with the methods in structural identifiability is poor scalability with the problem size. The resulting equations quickly grow to be too complicated even for relatively small problems [Chis 2011b]. Furthermore, observability analysis is limited to structural non-identifiability, providing no analysis of practical non-identifiability.

Practical non-identifiability depends on the amount and accuracy of the measurement data. This identifiability analysis determines whether a unique solution of the parameters is possible given the erroneous measurement data. [Guedj 2007] used the Fisher information matrix for the evaluation of the impact of the experimental condition and then a statistical approach to deal with the issue of practical non-identifiability. Brockmann *et al.* used local sensitivity functions for practical identifiability analysis of a biofilm model [Brockmann 2008].

However none of these methods have been shown to be sufficient for both structural and practical identifiability analysis on all types of biological models. Raue *et al.* proposed profile likelihood based identifiability analysis using the likelihood based confidence interval. Among the available methods for identifiability analysis in large models this has been one of the most efficient [Raue 2009].

In this thesis the parameter estimation problem has been considered as a state estimation problem where the general approach to address parameter identifiability is

in the framework of observability. However, biological models tend to be quite large where the observability analysis is not feasible due to its highly complex nature. In this thesis this parameter identifiability is addressed with a profile likelihood based approach instead of an observability based approach, which substantially simplifies the identifiability analysis.

1.7 Dissertation overview

One of the major challenges in the computational modelling of biological systems is the determination of model parameters. As only a fraction of these parameters can be measured experimentally, the rest have to be estimated through computational methods. In spite of the advancements in the development of different parameter estimation methods to date, and despite individual advantages, none have emerged as a fully general purpose estimator in terms of accuracy, efficiency and reliability for estimating parameters of biological models. Although filtering techniques, specifically UKF, have proven to be efficient for parameter estimation in a large number of applications [Li 2004, Zheng 2009, Quach 2007, Qi 2011, Malcolm 2009, Zhu 2008, Rajeswari 2009, Zheng 2009], they still suffer from numerical instability unless the covariance matrix is positive definite. Furthermore there is no general technique for the introduction of constraints into the estimation of UKF. A Square-Root variation of UKF solves the numerical stability but still does not have the capability of adding constraints. At the same time, identifiability analysis has been addressed in control theory from the perspective of observability, within the framework of Kalman filter and its non-linear variants. This analysis is very complicated due to the size and non-linear structure of biological models. Furthermore, this method can only identify structural non-identifiability and does not shed light on practical non-identifiability. This creates a need for an efficient identifiability technique when using such filtering techniques.

Therefore to develop an improved method for parameter identifiability analysis and estimation using Kalman filter theory, this work proposes a complete framework of parameter estimation. This framework consists of a modified version of the square root unscented Kalman filter, one that is more effective and can consider the biological reality of the parameter space. This algorithm is named the constrained square-root unscented Kalman filter (CSUKF). In addition to the development of a new estimation algorithm the framework also incorporates a data based identifiability analysis approach. This approach is a profile likelihood based identifiability analysis proposed by Raue (2009). This approach is complementary to CSUKF, making it more mathematically tractable with

increasing model complexity than observability based analysis. Necessary modifications to CSUKF were made to ensure a successful incorporation of this identifiability analysis. In order to facilitate the solution of parameter non-identifiability, methods to rank the parameters through sensitivity analysis, identify the linearly correlated parameters and finding functional relationship between the parameters have been included in the framework. Different properties of CSUKF are used to aid in calculating all these functions. These analysis helps to identify points where more measurement data would help in solving the parameter non-identifiability. However, situations may arise where no more measurement of parameter or species concentration is possible. Simplification of the model through lumping the parameters might also not be possible as they are vital for the predictive behaviour of the model. In that case, the informative prior is used to ensure that an estimation can be obtained despite the fact that a model is not identifiable in the frequentists approach. In contrast to the frequentist approach to estimation, Bayesian inference can be obtained despite the presence of non-identifiability if an informative prior is available [Rannala 2002, Samaniego 2010]. As the Kalman filter and its non-linear variants can be considered as a dynamic Bayesian network, this identifiability property of Bayesian inference, given the informative prior can also be applied to CSUKF to have a unique parameter estimation values for a model which is non-identifiable from the perspective of likelihood. The complete framework is then implemented in MATLAB and applied to three published biological model to estimate their parameter values.

The thesis is organized as follows: chapter 2 is one of the core chapters of this dissertation and covers the theoretical development of the complete framework in detail. Section 2.2 derives the numerically efficient and stable variant of UKF that incorporates constraints into state estimation, called CSUKF. All the mathematical details and the complete algorithm as well as it applicability for state estimation of a biological model are described in this section. Section 2.3 discusses the theoretical aspects of the orthogonal based ranking method that is employed by the framework to rank the model parameters according to their importance. The profile likelihood based identifiability analysis is investigated in section 2.4. This section also explains the necessary modifications made to CSUKF for use with this identifiability analysis. Section 2.5 focuses on the methods to determine the correlation and functional relationship between parameters which will help in solving the non-identifiability of the system. In cases when it is not possible to solve the non-identifiability of the parameters, they are treated with informed prior in Bayesian inference in order to have a unique estimation, which is detailed in section 2.6. In chapter 3 the major results and their general implication on the biological model

are discussed by applying the framework for parameter estimation on three previously published biological models. Finally in chapter 4 the conclusion is drawn and a direction for future research are given.

Chapter 2

The parameter estimation framework

Parameter estimation is one of the daunting tasks in computational modelling. This is even more so when it comes to modelling biological systems, due to: high interconnectivity, non-linearity and a lack of measurement data both in terms of quality and quantity [Kitano 2002a]. The main objective of this thesis is to develop a complete parameter estimation framework based around a novel non-linear filtering technique. The filtering technique is a constrained extension of the square-root unscented Kalman filter (SR-UKF). A very important and often overlooked aspect to parameter estimation is the identifiability of the parameters. This novel filtering technique in combination with a variety of conventional algorithms are combined into a single framework to analyse the parameter identifiability. This identifiability analysis suggests possible solutions to the parameter non-identifiability problem. Thus the complete framework ensures a successful estimation of the unknown parameter values after resolving the non-identifiability.

At the core of this framework lies the novel parameter estimation method, the constrained square-root unscented Kalman filter or CSUKF. This is a modified version of the (SR-UKF) and addresses parameter estimation in the form of state estimation, while handling the constraints within the state estimation. Profile likelihood based identifiability analysis is integrated with CSUKF as part of the framework to implement structural and practical identifiability analysis. A ranking of the parameters is also performed, describing the impact of each of the parameters to the model output as well as their estimability based on linear independence on the sensitivity coefficient. Correlation analysis is used to determine linear relationships between parameters, while non-linear functional relationships between the parameters are identified through the mean optimal transformation approach. Identifying these relationships reduces the number of non-identifiable parameters through their functional relationship [Cobelli 1980]. Removing non-identifiable parameters that are functionally linked leads to a more successful parameter estimation of the simplified model. The parameter estimation framework presented in this thesis incorporates all these function for the successful estimation of parameters.

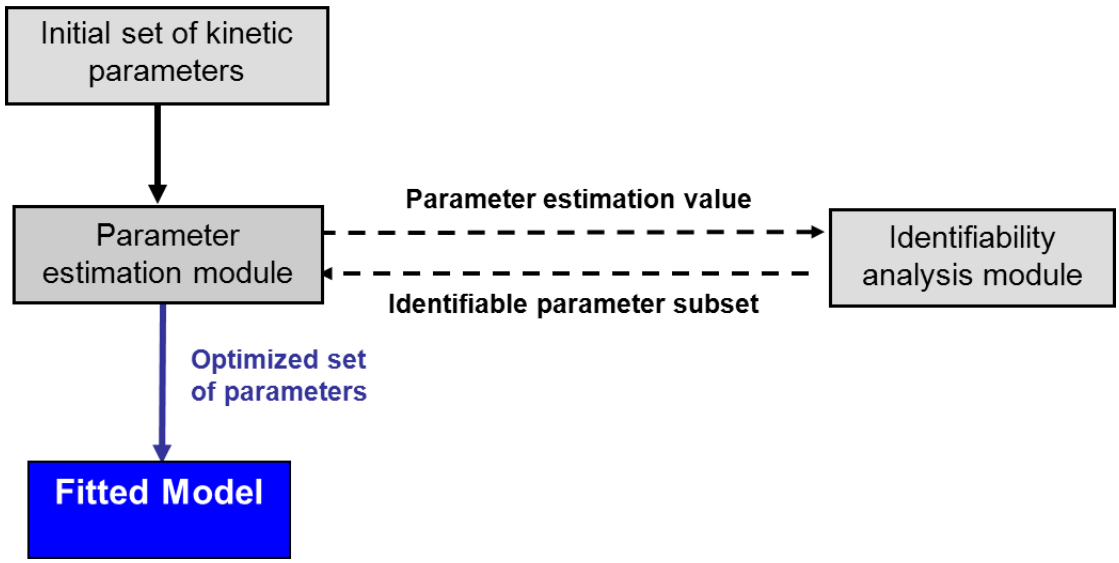


Figure 2.1: Complete parameter estimation framework. For different calculations the identifiability analysis module uses the optimized data of the parameter estimation module.

2.1 Overview of parameter estimation framework

The complete framework is depicted in Figure 2.1. It is divided into two modules, 1) the identifiability analysis module and 2) the parameter estimation module. The identifiability analysis in this framework uses a data driven method requiring an initial set of parameter values to start the analysis. The arrow from the parameter estimation module to the identifiability analysis module illustrates this initial set of parameter values computed by the CSUKF supplied to the identifiability analysis module to start the identifiability analysis. The basis of the identifiability analysis is the calculation of the profile likelihood of these parameters. This module attempts to resolve parameter non-identifiability through multiple steps based on ranking of the parameters, their correlation and functional relationship.

After resolving parameter non-identifiability, the arrow from the identifiability analysis module to the parameter estimation module returns the identifiable parameter subsets to the estimation module. The CSUKF then begins its basic operation of estimating the parameters starting with small random values. This estimation is repeatedly refined until the predefined stop criteria, such as the number of iterations, or in case the objective function reaches a stable value or numerical threshold. Once determined the optimized parameters are combined with the model yielding the optimized model.

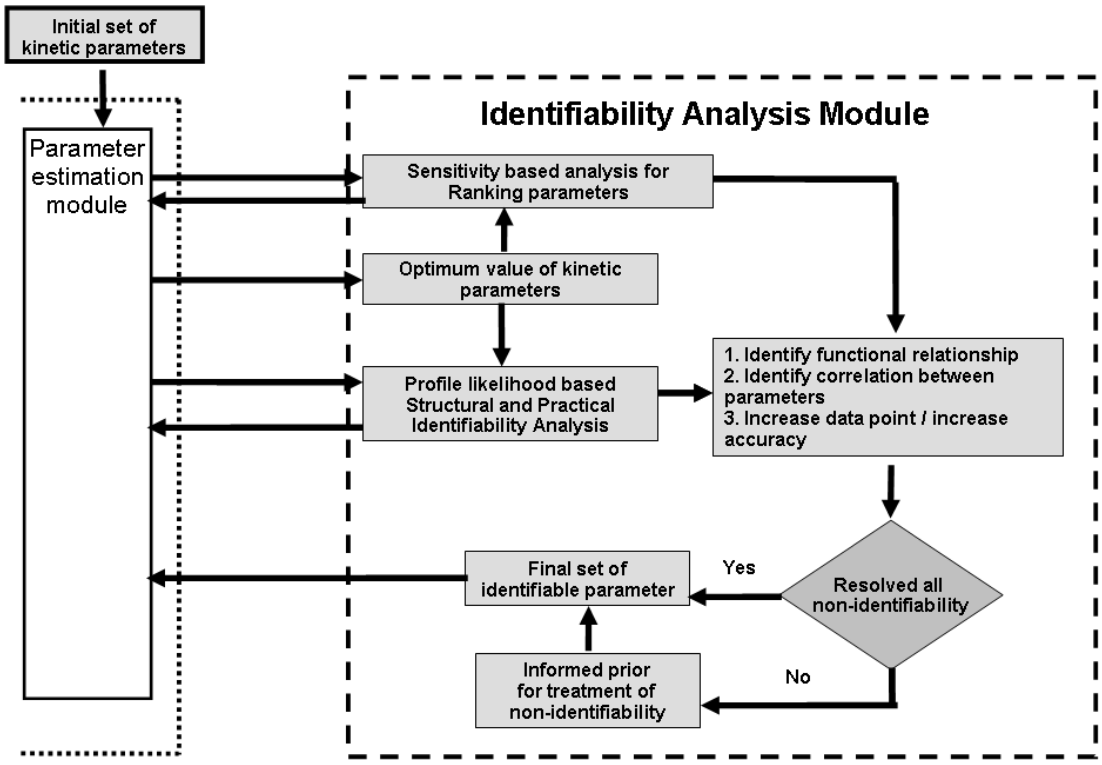


Figure 2.2: Identifiability analysis module. This module conducts the identifiability analysis of the parameters, including the ranking and identifies the correlation and the non-linear functional relationship between the parameters. The initial set of parameter values used in this module comes from the CSUKF in the parameter estimation module.

2.1.1 Identifiability analysis module

Figure 2.2 describes the identifiability analysis module in detail. The major functionality of the identifiability analysis are the profile likelihood based identifiability analysis, calculation of the ranking, determination of the correlation between parameters and identification of the non-linear functional relationships among the parameters. The identifiability analysis module starts with an initial set of optimized parameter values calculated by the parameter estimation module. In the identifiability analysis module, this data is first processed by the profile likelihood based structural and practical identifiability analysis sub-module to determine which, if any, parameters are non-identifiable. In addition to generating the initial optimal parameter set, this sub-module also makes use of CSUKF to generate the profile likelihood of the parameters. A sensitivity based analysis is then conducted for ranking the parameters according to their importance. This method uses a sensitivity matrix which is calculated locally at each iteration step of

the CSUKF. Next the correlation and functional relationship sub-modules try to identify the repetitive relationships between the parameters. The strategy is to first deduce the relationship between a high ranking and a low ranking non-identifiable parameter. Where possible, non-identifiable parameters with a high ranking should be replaced by values from wet-lab experiments. Then low ranking parameters are re-evaluated using these new values to determine if they are still non-identifiable. When additional wet lab data is not available, low ranking parameters are set to small nominal values as the lower sensitivity of these parameters have a minimal effect on the system output, then re-evaluate non-identifiability of the high ranking parameters [Yao 2003]. If high ranking parameters continue to be non-identifiable without determined functional relations, the model has to be reformulated so as to reduce the number of states and parameters as outlined in [Chis 2011b]. In this way the structural non-identifiability problem may be solved, and the model simplified. Solving the practical non-identifiability problem requires an increase in data accuracy or increase in the number of data points in a time series or both. In biological systems it is often the case that these solutions are not available for solving non-identifiability. Furthermore simplification of the model is not always feasible as it leads to a model without the required predictive behaviour. In these cases the Bayesian treatment of non-identifiability is applied. When the informative prior is available, Bayesian inference makes possible a unique estimation of parameter values which are non-identifiable in the frequentists approach [Samaniego 2010]. As CSUKF is a variation of the dynamic Bayesian inference, the CSUKF can supply the informative prior to make a unique prediction of parameters.

2.1.2 Parameter estimation module

Parameter estimation in kinetic models suffers from numerous difficulties due to the nature of such models. These models incorporate system noise due to the uncertainty in the model description during the mathematical formulation of the biological network. Noise hinders most conventional parameter estimation methods. Furthermore it is not always possible to directly measure all the variables in a kinetic model due to the limitation in the measurement method or device. Such variables are called hidden variables. For example, the activity level of regulatory protein in a gene regulatory network cannot be measured directly with currently available techniques [Nachman 2004] and thus they are considered as hidden variables. These hidden variables lead to rather complex cost functions that are to be minimized for the estimation of the parameters [Sitz 2002] and cause problems for the conventional estimation

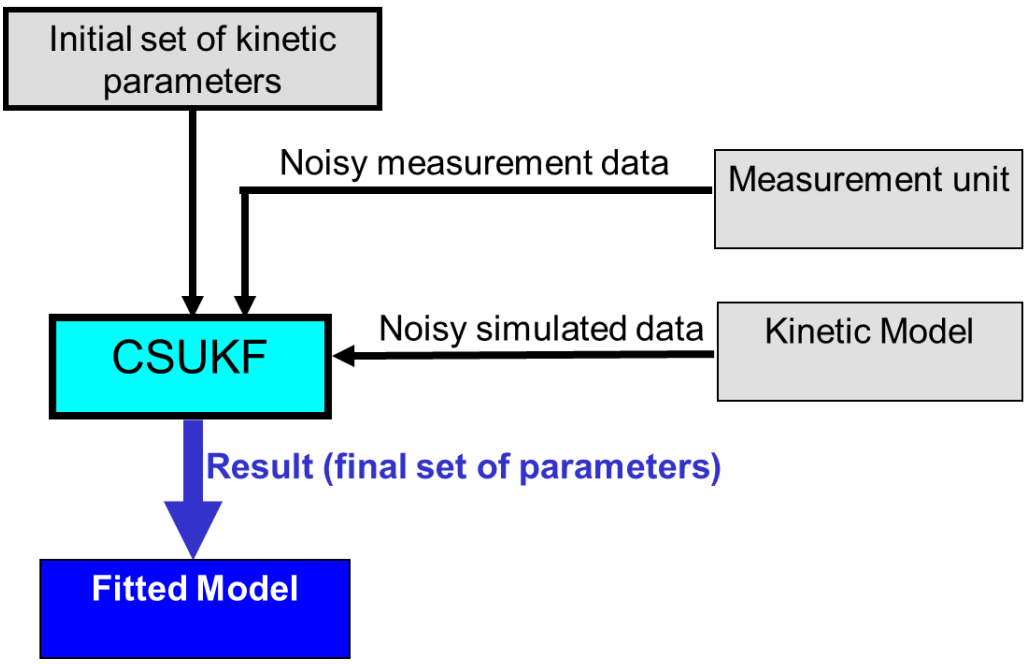


Figure 2.3: Parameter estimation module. CSUKF iteratively estimates parameter values by minimizing the difference between the noisy simulated data and the noisy measurement data

methods. Moreover, in such models only a small portion of noise corrupted measurement data is available from which parameters are to be estimated. This thesis proposes a constrained extension of square-root unscented Kalman filter, CSUKF, to be used as the parameter estimation module which lies at the core of the framework. This filtering technique tries to estimate the parameter values by minimizing the mean squared error. It takes into consideration both the system noise and measurement noise during the estimation of these parameters. Like any other filtering technique CSUKF has the ability to remove unwanted noise from the desired data. It utilizes the power of representing dynamic systems through state space models. It uses the general probability theory to estimate the parameter values of these biological models. Additionally, the CSUKF can also make prediction of hidden variables, which would have an influence on the calculation of the mean squared error. Figure 2.3 illustrates this module, centred on the CSUKF, which is discussed in detail in the next section.

2.2 Constrained Square Root Unscented Kalman Filter (CSUKF)

The Kalman filter (KF) performs state estimation by propagating the probability distribution function of the system states over time, taking into consideration the model uncertainties and measurement uncertainties. A major drawback of KF and its variants is that there is no general mechanism for incorporating constraints on the state space [Vachhani 2006, Kandepu 2008]. In biological systems the state constraints can be important to 1) include prior information about the state value ranges into the estimation process and 2) to make the state values biologically relevant, such as the metabolite concentrations are always positive or the binding constant of a Michaelis-Menten rate law cannot have negative values. Parameters in a biological model must be constrained to vary within a biophysically plausible range (e.g., within the range of diffusion limited rates). Therefore it is important to incorporate state constraints into the estimation process.

In control theory state constraints [Kandepu 2008] typically refer to the state space boundaries. A widely used technique to ensure state constraints is the moving horizon estimator (MHE) [Rao 2003]. MHE formulates the state estimation problem as a non-recursive constrained quadratic program [Teixeira 2008]. But as this process is non-recursive and solves a quadratic program at each step, it is in general computationally less efficient [Vachhani 2006, Kandepu 2008]. An interval constrained unscented transformation (ICUT) proposed by [Vachhani 2006] is considered to be an efficient way for introducing the state inequality constraints into the state estimation process through constrained sigma point selection. In this approach the sigma points that are outside the feasible region are projected onto the constrained state space boundary [Vachhani 2006].

In this thesis I propose a novel filtering technique, the constrained square-root UKF (CSUKF), that combines the interval constrained unscented transformation (ICUT) [Vachhani 2006] with the square-root UKF [Merwe 2001], to solve the problem of constrained state estimation. This ensures a numerically stable constrained estimation of the unknown parameter values utilizing the benefits of the unscented Kalman filter.

In the basic UKF the Choleskey decomposition of the covariance matrix (P) is calculated at each iteration. For stability P must be a positive semi-definite matrix, which means that its square root, which can be computed by Cholesky factorization, is a triangular matrix. When the process noise covariance (Q) is sufficiently small, meaning that the process is well known, then the rounding error may cause P to become

numerically unstable, e.g. by making it negative-definite or by having negative diagonal entries. Merwe *et al*, (2001) proposed a square-root implementation of the UKF, which instead of calculating the Cholesky decomposition at each iteration, propagates the triangular matrix form of the covariance matrix at each step of sigma point calculation [Merwe 2001]. By maintaining the covariance in the form of a triangular matrix the state covariance matrix remains positive semi-definite ensuring the numerical stability of the filtering algorithm. The algorithm proposed in this thesis modifies this efficient and numerically stable filtering algorithm to include the state constraints into the estimation. In CSUKF, the state function is denoted as f and the observation function is denoted as h. The states at time update step are the *a priori* estimate which is denoted with a superscript $^-$ and states on measurement update stage is the *a posteriori* estimate. The sigma point matrix $\mathcal{X}$ used in iteration k is based on the results of the previous iteration, $k-1$, and is denoted $\mathcal{X}(k|k-1)$. From the previous iteration the posterior mean $\hat{x}(k-1)$ and square-root of the covariance $P(k-1)$, of the estimated states are required for this calculation. These sigma points are propagated through the state function f to create the *a priori* transformed sigma point $\mathcal{X}^a(k|k-1)$ at iteration k. The *a priori* mean value estimation using $\mathcal{X}^a(k|k-1)$ is denoted as $\hat{x}^-(k)$ and the final or a posteriori mean value estimation is denoted as $\hat{x}(k)$.

2.2.1 Interval constrained unscented transformation

It is assumed that for an n dimensional state vector x the state constraints at iteration k for all $k \geq 0$ is represented by a box with the following equation

$$L(k) \leq x(k) \leq U(k) \tag{2.1}$$

where[1] $L \in \mathbb{R}^n$ is the vector of lower limits and $U \in \mathbb{R}^n$ is the vector of upper limits of the constrained boundary[2]. The values of these boundaries are assumed to be known. If any element of the state vector x is unbounded the corresponding boundary is set to $\pm\infty$ as appropriate. This constraint information is incorporated during the time-update step when the sigma points are calculated. If these sigma points fall outside the boundary, they are projected back to the boundary of the nearest feasible region. This method of selecting constrained sigma points for a two dimensional state variable, $x \in \mathbb{R}^2$, is depicted in Figure 2.4. The sigma points outside the boundary (the dotted lines) are

[1] For notational clarity the iteration index k has been removed

[2] Constraints can be varied at each iteration. Also different constraint boundaries can be set for different state variables

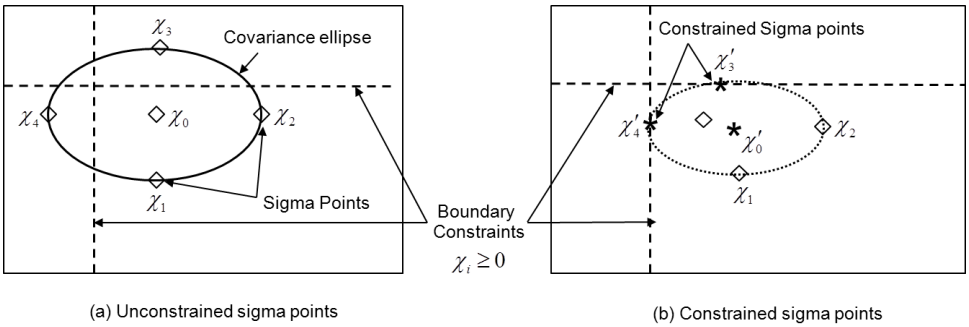


Figure 2.4: Unconstrained and constrained sigma points, constraints indicated by dashed lines. a) Illustrates the covariance ellipse (solid line) with unconstrained sigma points ($\diamond$) and mean x_0. Here two of the four unconstrained sigma points x_3 and x_4 lie outside of the the boundary. In b) The covariance ellipse (dotted line) calculated with the constrained sigma points ($\star$). The sigma points that were previously outside of the boundary in (a) are projected onto the boundary, resulting in the modified covariance ellipse as well as the shift in the constrained mean (x_0).

projected back onto the constraint boundary. The resulting selection of sigma points are non-symmetrically distributed on the ellipse, thus the weights need to be adjusted accordingly. The sigma points are then propagated through the non-linear transformation and the state estimation is made through a modified square-root UKF. This modification of the square-root UKF is made to ensure the adaptability of the algorithm in case of non-symmetric negative weights. Since the estimated state may still violate the constraints, in which case it would be projected onto the constrained boundary before starting the next iteration. The mean and covariance calculated with the constrained sigma points now include the information on the constraints, which increases accuracy of the filter estimate in a biological sense.

2.2.2 Sigma point selection method

Based on the present state covariance matrix, $2n+1$ sigma points are generated, satisfying the following equation

$$L \leq X(k|k-1) \leq U \tag{2.2}$$

For the sigma point selection method the same approach as outlined by [Vachhani 2006] is taken. In their approach the matrix $S = \left[\sqrt{P} \quad -\sqrt{P} \right]$ is defined as the direction of the sigma points where P is the state estimation covariance matrix. However since in CSUKF, the square-root of the state estimation covariance matrix is propagated at each iteration instead of the full form of the covariance matrix, it is not necessary to

calculate this square-root at each iteration. Instead this step direction matrix is defined as the square-root of the state covariance matrix, V, propagated at each iteration of the constrained square root UKF. At iteration k, the step size, ζ is selected as

$$\zeta_j \triangleq min(col_j(\Theta)), \tag{2.3}$$

$$\Theta_{(i,j)} \triangleq \begin{cases} \sqrt{n+\lambda} & \text{if } S_{(i,j)} = 0 \\ min\Big(\sqrt{n+\lambda}, \frac{U_i(k)-\hat{x}_i(k-1)}{S_{(i,j)}}\Big) & \text{if } S_{(i,j)} > 0 \\ min\Big(\sqrt{n+\lambda}, \frac{L_i(k)-\hat{x}_i(k-1)}{S_{(i,j)}}\Big) & \text{if } S_{(i,j)} < 0 \end{cases}$$

where n is the state dimension, $i = 1, \ldots, n$ and $j = 1, \ldots, 2n$ [Vachhani 2006]. The step size $\sqrt{n+\lambda}$ is a regular step size for the UT [Merwe 2001]. This step size is selected when it does not cause the sigma points to violate the constraint boundary. However if it does violate the constraint boundary then the step size is selected in a way so that the sigma points lie on the boundary when the step size is multiplied by the corresponding directional element and added to the state estimate $\hat{x}(k-1)$. The value of the scaling factor is $\lambda = \alpha^2(n+\kappa) - n$, as a scaled UT selection is being used. The value of α and κ are the same as described at the end of section 1.5.1.3 and [Julier 2002]. However it can also be chosen to any small random number. The sigma point matrix contains $2n+1$ sigma points (column) for the n state variables (rows). In this definition indexing starts at 0 for convenience.

$$\mathcal{X} = \begin{cases} \hat{x}(k-1) & j = 0 \\ \hat{x}(k-1) + \zeta_j col_j(S) & 1 \leq j \leq 2n \end{cases} \tag{2.4}$$

At iteration k these sigma points are calculated from the prior state estimation of $k-1$. The first sigma point is the state estimate $\hat{x}(k-1)$. Next $1, \cdots, n$ sigma points are selected by adding the state estimate $\hat{x}(k-1)$ with the multiplication of step size and direction. The next $n+1, \cdots, 2n$ sigma points are generated by subtracting the product of step size and direction from the state estimate $\hat{x}(k-1)$. Sigma points generated with this method may not be symmetric, that is they may not all be equidistant from $\hat{x}(k-1)$. Consequently the weights associated with these sigma points cannot be equally distributed. Instead they must be calculated in a way that if the state variables are sufficiently far from the constraint boundary, the weights should correspond to that regular form of UT as

mentioned in section 1.5.1.3. Otherwise the weights are chosen such that the weights vary linearly with the step size,

$$W_0 = b, \tag{2.5}$$

$$W_j = a\zeta_j + b, \qquad j = 1, \ldots, 2n \tag{2.6}$$

The equation outlined by [Vachhani 2006] is modified in order to solve for a and b. From the general weighting scheme of UT as described in section 1.5.1.3 and in [Julier 1996] it is known that the weights must sum to 1, therefore

$$\begin{aligned} \sum_{j=0}^{2n} W_j &= b + \sum_{j=1}^{2n}(a\zeta_j + b) \\ 1 &= aT_\zeta + (2n+1)b \\ \text{where } T_\zeta &= \sum_{j=1}^{2n} \zeta_j \text{and} \qquad \sum_{j=0}^{2n} W_j = 1 \\ a &= \frac{1-(2n+1)b}{T_\zeta} \end{aligned} \tag{2.7}$$

From Equation (2.5) it is known that b is independent of the step size ζ. Therefore, b will take the value of W_0 from the regular UKF,

$$b = \frac{\lambda}{n+\lambda} \tag{2.8}$$

Substituting for b into Equation (2.7) yields,

$$\begin{aligned} a &= \frac{1-(2n+1)\frac{\lambda}{(n+\lambda)}}{T_\zeta} \\ &= \frac{n+\lambda-2n\lambda-\lambda}{T_\zeta(n+\lambda)} \\ &= \frac{n(1-2\lambda)}{T_\zeta(n+\lambda)} \end{aligned} \tag{2.9}$$

So the final value of a and b are defined as

$$a = \frac{n(1-2\lambda)}{(n+\lambda)\left(\sum_{j=1}^{2n}\zeta_j\right)} \tag{2.10}$$

$$b = \frac{\lambda}{n+\lambda} \tag{2.11}$$

As the scaling factor for the sigma points selection have been used it is needed to add $1 - \alpha^2 + \beta$ with W_0^c to take the effect of the scaling factor into the weights. The weights are thus formulated as

$$W_j^m = W_j^c = \Big[b \qquad a \times \zeta_{1:2n} + b\Big], \quad j = 0 \ldots, 2n \tag{2.12}$$

$$W_0^c = W_0^c + (1 - \alpha^2 + \beta) \tag{2.13}$$

2.2.3 CSUKF formulation

In general UKF estimates the state variables by propagating the sigma points through the non-linear dynamic functions f and g. However UKF suffers from numerical instability when the process error covariance matrix Q is small, the square-root UKF (SR-UKF) tries to solve this problem by propagating the Cholesky factored square-root form of the covariance matrix P at each iteration. This UKF extension ensures equal or marginally better estimation accuracy with the added benefit of ensuring numerical stability [van der Merwe 2004]. Like SR-UKF, CSUKF make use of *QR decomposition*, *Cholesky factor updating* and *pivot-based least squares*. QR decomposition is used during the *time update* step of CSUKF to calculate the *prior* Cholesky factor of the compound matrix formed with the positively weighted sigma points and the matrix square-root of the additive process noise. For the negatively weighted sigma points the Cholesky factor is downdated using the *Cholesky factor update*. The same two approaches are used to update the prior of the square-root of the observation error covariance V_y. Finally the Kalman gain $K(k)$ is calculated using the *pivot-based least squares* method. Before going into the details of the filtering technique first a brief overview of these three technique used in CSUKF is given.

QR decomposition: For $A \in \mathbb{R}^{m \times n}$, QR decomposition produces an $m \times m$ orthogonal matrix O, whose columns are orthogonal unit vectors, and an invertible $m \times n$ upper triangular matrix G satisfying

$$A = OG \tag{2.14}$$

The upper triangular G, represents the transpose of the Cholesky factor of a matrix i.e. if A is the Cholesky factor of a matrix B, that is if $B = AA^T$ then G will represent the transpose of this Cholesky factor meaning $G^T G = AA^T$. In case of $m \geq n$, the bottom $(m - n)$ rows of the upper triangular matrix consist entirely of zeroes where it would be safe to take only the $n \times n$ elements.

Cholesky factor updating: The QR decomposition is applied to a compound matrix formed with the square-root of the sigma point weights and process noise covariance matrix. As the weights in the constrained sigma point selection can take negative values the square-root of this negative numbers cannot be taken directly as this will result in a complex number. To incorporate the effect of these weights it is needed to update the Cholesky factor with $'-'$. This is done by applying the *Cholesky factor updating* or *cholupdate* function. If for $P = SS^T$ where S is the Cholesky factor of matrix P then updating Cholesky factor S with the value of X with a $'-'$ would lead to

$$P = P - XX^T \tag{2.15}$$

This calculation is denoted as

$$\tilde{S} = cholupdate(S, X, '-') \tag{2.16}$$

If X is a matrix having m columns then the *cholupdate* function will make m consecutive updates of the Cholesky factor using m number of columns of matrix X.

Efficient least squares: To calculate the value of the Kalman gain K it is needed to solve an overdetermined least square problem $Ax = b$. This is solved with a QR decomposition method with pivoting. This method is roughly the same as $INV(A) * b$, except it is computed in a different way.

2.2.4 Constrained square-root state estimation

For the constrained square-root UKF implementation the state covariance matrix is first initialized as $\mathbb{E}[(x(0) - \hat{x}(0))(x(0) - \hat{x}(0))^T]$ and the matrix square root is once calculated via a Cholesky factorization, $V(0) = chol\Big[\mathbb{E}[(x(0) - \hat{x}(0))(x(0) - \hat{x}(0))^T]\Big]$. Then at each iteration of CSUKF, $k > 0$, $V(k)$ is propagated instead of the full covariance matrix $P(k)$ where $V(k)V(k)^T = P(k)$. This would avoid the need to refactorize $P(k)$ at each iteration. As the weights in the constrained sigma point calculation are asymmetric they may vary in magnitude and sign, i.e. they can be positive or negative. So the calculation of the square-root factor of the covariance matrix needs to be decomposed into two parts, one for all the weights W_j^c having positive value denoted as $V^{pos}(k)$ and one for all the weights W_j^c having a negative value denoted as $V^{neg}(k)$. This propagated Cholesky factor is then used directly to calculate the sigma points. To facilitate this decomposition two index

sets are defined as

$$\begin{aligned} I^+ &= \{j | W_j^C \geq 0\} \\ I^- &= \{j | W_j^C < 0\} \end{aligned} \tag{2.17}$$

The following decomposition of the covariance matrix by the UKF explains the decomposition of V

$$\begin{aligned} P(k) &= \left(\sum_{j=0}^{2n} W_j^c \Big(\mathcal{X}_j(k|k-1) - \hat{x}(k) \Big) \Big(\mathcal{X}_j(k|k-1) - \hat{x}(k) \Big)^T \right) + Q \\ &= \left(\sum_{j=0}^{2n} \sqrt{W_j^c} \Big(\mathcal{X}_j(k|k-1) - \hat{x}(k) \Big) \sqrt{W_j^c} \Big(\mathcal{X}_j(k|k-1) - \hat{x}(k) \Big)^T \right) + \sqrt{Q}\sqrt{Q}^T \\ &= \sum_{j \in I^+} \sqrt{W_j^c} \Big(\mathcal{X}_j(k|k-1) - \hat{x}(k) \Big) \sqrt{W_j^c} \Big(\mathcal{X}_j(k|k-1) - \hat{x}(k) \Big)^T + \sqrt{Q}\sqrt{Q}^T \\ &\quad - \sum_{j \in I^-} \sqrt{|W_j^c|} \Big(\mathcal{X}_j(k|k-1) - \hat{x}(k) \Big) \sqrt{|W_j^c|} \Big(\mathcal{X}_j(k|k-1) - \hat{x}(k) \Big)^T \end{aligned}$$

Any weights can have negative values. The decomposition then can be rearranged as

$$\begin{aligned} P(k) &= \left[\sqrt{W_j^c} \Big(\mathcal{X}_j(k|k-1) - \hat{x}(k) \Big) \quad \sqrt{Q} \right]_{j \in I^+} \begin{bmatrix} \sqrt{W_j^c} \Big(\mathcal{X}_j(k|k-1) - \hat{x}(k) \Big) \\ \sqrt{Q} \end{bmatrix}_{j \in I^+} \\ &\quad - \left[\sqrt{|W_j^c|} \Big(\mathcal{X}_j(k|k-1) - \hat{x}(k) \Big) \sqrt{|W_j^c|} \Big(\mathcal{X}_j(k|k-1) - \hat{x}(k) \Big)^T \right]_{j \in I^-} \\ &= V^{pos}(k) \Big(V^{pos}(k) \Big)^T - V^{neg}(k) \Big(V^{neg}(k) \Big)^T \end{aligned} \tag{2.18}$$

The *QR decomposition* is applied because $V^{pos}(k) \in \mathbb{R}^{n \times 3n}$ produces a computationally undesirable matrix, as it increases the number of columns [Terejanu 2008]. QR-decomposition is used to express $V^{pos}(k)$ in terms of an orthogonal matrix O and an upper triangular matrix $G^T \in \mathbb{R}^{n \times n}$. So for the positive W_j^c the equation can be written as

$$\left[\sqrt{W_j^c} \Big(\mathcal{X}_j(k|k-1) - \hat{x}(k) \Big) \quad \sqrt{Q} \right]^T = OG^T \tag{2.19}$$

It can be verified that at each iteration k the covariance matrix $P(k)$ can be calculated from the propagated $V^{pos}(k)$ and $V^{neg}(k)$ through this $O(G)^T$ according to the following

equation

$$\begin{aligned} P(k) &= GO^T OG^T - \Big(V^{neg}(k)\Big)V^{neg}(k)^T \\ P(k) &= GG^T - \Big(V^{neg}(k)\Big)V^{neg}(k)^T, \text{ as } O \text{ is orthogonal } O^T O = I \end{aligned} \tag{2.20}$$

From Equation (2.20) it can be seen that the full covariance matrix can be recovered from the upper triangular matrix generated during *QR decomposition*. This justifies the use of this method for updating the square-root factor. Therefore for the positive weights only the upper triangular matrix of the QR decomposition is taken as

$$G = qr(V^{pos}) \tag{2.21}$$

To include the effect of $V^{neg}(k)$ at each iteration in the square-root calculation a rank-1 downdate to Cholesky factorization is performed

$$V(k) = cholupdate(G, V^{neg}(k), '-') \tag{2.22}$$

This approach is applied to update the square-root factor in both the time-update and measurement-update step. The Kalman gain K is calculated as

$$K(k) = \frac{P^{xy}(k)}{\left(\Big(V_y(k)\Big)^T V_y(k)\right)} \tag{2.23}$$

where $P^{xy}(k)$ is the cross-correlation matrix and $V_y(k)$ is the measurement updated covariance matrix. Lastly with the measurement data $y(k)$, the state estimation and the square-root factor of the estimation covariance is calculated as

$$\hat{x}(k) = \hat{x}^-(k) + K(k)\Big(y(k) - \hat{y}^-(k)\Big) \tag{2.24}$$

$$V(k) = cholupdate(V^-(k), K(k)V_y(k), '-') \tag{2.25}$$

If individual values of the estimated state, $\hat{x}(k)$, violate the constraints, those values would be projected onto the constraint boundary. Despite this the covariance ellipse may still extend beyond the constraint boundary. However this does not pose a problem as the sigma points calculated from this covariance in the next iteration would again be projected onto the constraint boundary. The sigma points are propagated through the state function f and the weighted mean and covariance of the transformed sigma

points are then calculated. Sigma points are regenerated before transforming through the measurement function h to capture the effect of the additive noise. The weights are also recalculated at this step.

CSUKF uses the classical KF update during the measurement update step and does not assimilate the constraint information during the calculations of mean and covariance at this step. This will make CSUKF computationally faster than the algorithms where interval constraint is also imposed on the measurement update step. But this will not compromise the accuracy of CSUKF. This conclusion is based on the work of [Teixeira 2008] where they reviewed seven different algorithms with different approaches to constraint handling at the measurement-update stage. The approaches they considered are a) classical KF update where no constraint is incorporated, b) constrained Kalman update and, c) the sigma point constrained update. All of these algorithms have interval-constrained UT in the time-update stage. These algorithms were then used to estimate the state variable of a gas-phase reversible reaction. The work showed that, with a good initialization of the state variables, the algorithm without any constraint update on the measurement stage, interval constrained UKF (IUKF), has the best accuracy of estimating the state variables. Although with poor initialization, the performance slightly deteriorates, it still remains within the four best ones. The CPU processing time of this algorithm, is the best among all these seven algorithms. This showed that a computationally fast filtering technique with good accuracy can be achieved with constrained sigma point calculation only at the time update step and without making any constrained KF update at the measurement update step.

Initialization

The state-space vector and the square-root factor of the estimation covariance matrix are initialized with the expected value of the state vector as

$$\hat{x}(0) = \mathbb{E}[x(0)] \tag{2.26}$$

$$V(0) = chol\left\{\mathbb{E}\left[\Big(x(0)-\hat{x}(0)\Big)\Big(x(0)-\hat{x}(0)\Big)^T\right]\right\} \tag{2.27}$$

For $k \in \{1, \ldots, T\}$**:**
The sigma points satisfying $L \leq \mathcal{X}(k|k-1) \leq U$, are selected as

$$\mathcal{X} = \begin{cases} \hat{x}(k-1) & j = 0 \\ \hat{x}(k-1) + \zeta_j col_j(S) & 1 \leq j \leq 2n \end{cases} \tag{2.28}$$

where $\mathcal{X}$ is based on the direction, $S = [V(k-1) \quad -V(k-1)]$.
The step size, ζ is calculated as

$$\zeta_j \triangleq min(col_j(\Theta)), \tag{2.29}$$

$$\Theta_{(i,j)} \triangleq \begin{cases} \sqrt{n+\lambda} & \text{if } S_{(i,j)} = 0 \\ min\Big(\sqrt{n+\lambda}, \frac{U_i(k) - \hat{x}_i(k-1)}{S_{(i,j)}}\Big) & \text{if } S_{(i,j)} > 0 \\ min\Big(\sqrt{n+\lambda}, \frac{L_i(k) - \hat{x}_i(k-1)}{S_{(i,j)}}\Big) & \text{if } S_{(i,j)} < 0 \end{cases}$$

The weights are calculated as

$$W_0^m = b \tag{2.30}$$

$$W_0^c = b + (1 - \alpha^2 + \beta) \tag{2.31}$$

$$W_j^m = W_j^c a \zeta_j + b, \quad 1 \leq j \leq 2n \tag{2.32}$$

where

$$a = \frac{n(1-2\lambda)}{(n+\lambda)(\sum_{j=1}^{2n} \zeta_j)} \tag{2.33}$$

$$b = \frac{\lambda}{n+\lambda} \tag{2.34}$$

Time update

$$\mathcal{X}^a(k|k-1) = f[\mathcal{X}(k|k-1), u(k)] \tag{2.35}$$

$$\hat{x}^-(k) = \sum_{j=0}^{2n} W_j^m \mathcal{X}_j^a(k|k-1) \tag{2.36}$$

$$G^x(k) = qr\Big\{\Big[\sqrt{W_j^c}\Big(\mathcal{X}_j^a(k|k-1) - \hat{x}^-(k)\Big) \quad \sqrt{Q}\Big]_{j \in I^+}\Big\}, \; I^+ = \{j | W_j^c \geq 0\} \tag{2.37}$$

$$V_x^{neg}(k) = \Big[\sqrt{|W_j^c|}\,(\mathcal{X}_j^a(k|k-1) - \hat{x}^-(k))\Big]_{j \in I^-}, \quad I^- = \{j | W_j^c < 0\} \tag{2.38}$$

The prior cholesky factor is found by performing a downdate of the positive and negative square roots

$$V_x^-(k) = cholupdate(G^x(k), V_x^{neg}(k), '-') \tag{2.39}$$

Measurement update

To incorporate the additive process noise, in the measurement update stage, sigma points are redrawn and unconstrained weights are calculated

$$\begin{aligned}
\mathcal{X}^y(k|k-1) &= [\hat{x}^-(k) \quad \hat{x}^-(k) + \sqrt{(n+\lambda)}V_x^-(k) \\
&\qquad \hat{x}^-(k) - \sqrt{(n+\lambda)}V_x^-(k)] && (2.40)\\
W_0^m &= \frac{\lambda}{n+\lambda} && (2.41)\\
W_0^c &= \frac{\lambda}{n+\lambda} + (1-\alpha^2+\beta) && (2.42)\\
W_j^m &= W_j^c = \frac{\lambda}{2(n+\lambda)}, \quad j = 1\ldots,2n && (2.43)
\end{aligned}$$

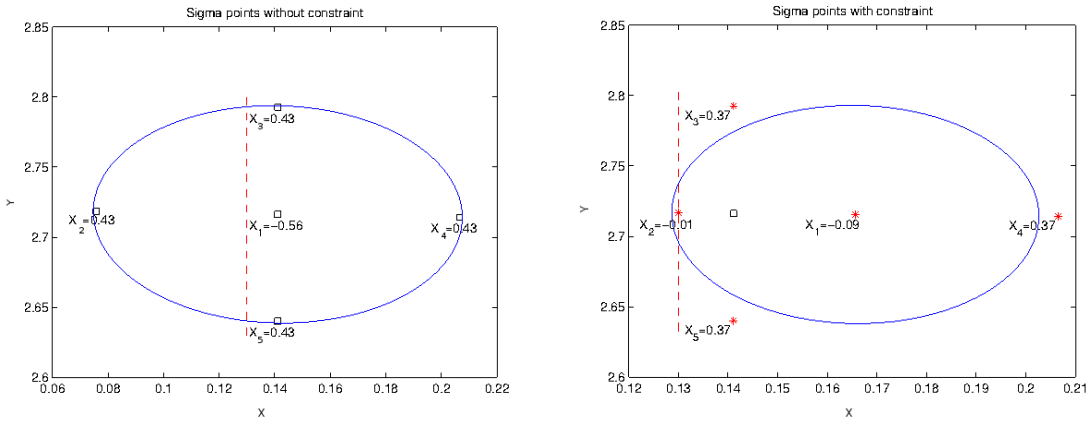
now the measurement update is performed as

$$\begin{aligned}
\mathcal{Y}(k|k-1) &= h[\mathcal{X}^y(k|k-1)] && (2.44)\\
\hat{y}^-(k) &= \sum_{j=0}^{2n} W_j^m \mathcal{Y}(k|k-1) && (2.45)\\
G_y(k) &= qr\Big\{\Big[\sqrt{W_j^c}(\mathcal{Y}_j(k|k-1) - \hat{y}^-(k)) \quad \sqrt{R}\Big]_{j\in I^+}\Big\},\ I^+ = \{j|W_j^c \geq 0\} && (2.46)\\
V_y^{neg}(k) &= \Big[\sqrt{|W_j^c|}\,(\mathcal{Y}_j(k|k-1) - \hat{y}^-(k))\Big]_{j\in I^-}, \quad I^- = \{j|W_j^c < 0\} && (2.47)\\
V_y(k) &= cholupdate(G_y(k), V_y^{neg}(k), '-') && (2.48)
\end{aligned}$$

For all index $j = 0,\ldots,2n$ for which the value of W_j^c is negative.

$$\begin{aligned}
P^{xy}(k) &= \sum_{j=0}^{2n} W_j^c[\mathcal{X}_j^a(k|k-1) - \hat{x}^-(k)][\mathcal{Y}_j(k|k-1) - \hat{y}^-(k)]^T && (2.49)\\
K(k) &= P^{xy}(k)/(V_y(k))^T V_y(k) && (2.50)\\
\hat{x}(k) &= \hat{x}^-(k) + K(k)\Big(y(k) - \hat{y}^-(k)\Big) && (2.51)
\end{aligned}$$

Here $y(k)$ is the measurement data. The square-root factor of the estimation covariance



(a) (b)

Figure 2.5: Selection of sigma points in the presence of constrain boundary. The dashed line represents the boundary. It shows how the sigma points are projected onto the boundary in case they violate it. The number on each of the sigma points represent their corresponding weights

matrix V_k is then updated as

$$V(k) \quad = \quad cholupdate\Big\{V^{-}(k), K(k)V_y(k),' -'\Big\} \tag{2.52}$$

2.2.5 State estimation with CSUKF

To verify the applicability and accuracy of CSUKF this algorithm is first applied for the state estimation of **the Higgins-Selḱov oscillator**. As described in section 1.1 this model has three parameters k_1, k_2, k_3 and two state variables X and Y where X is the substrate and Y is the product. The system is discretized with a sampling interval of 0.25 seconds with a total simulation time of 25 second. Process noise with a covariance of $10^{-3}I$ is added to the system states. This experiment assumes that only the additive time series data of the state variables, i.e. $X+Y$ is available as the measurement data. The covariance of the measurement noise is calculated from the statistics of the random noise which is added to the simulated data to generate the synthetic measurement data. The lower bound of the constraint is set ot 0.13 and the upper bound is set tos 2.9 for both states. In this experiment the state is initialized to some random numbers close to [0.1 4.5]. As these values fall outside the boundary they are projected on the boundary which makes the initial state value to be [0.13 2.9]. The estimation process is run 50 times to generate the estimation statistics presented in Table 2.1. Figure 2.5 illustrates how the sigma points progress when projected to the constraint boundary. In Figure 2.5a the sigma point X_2

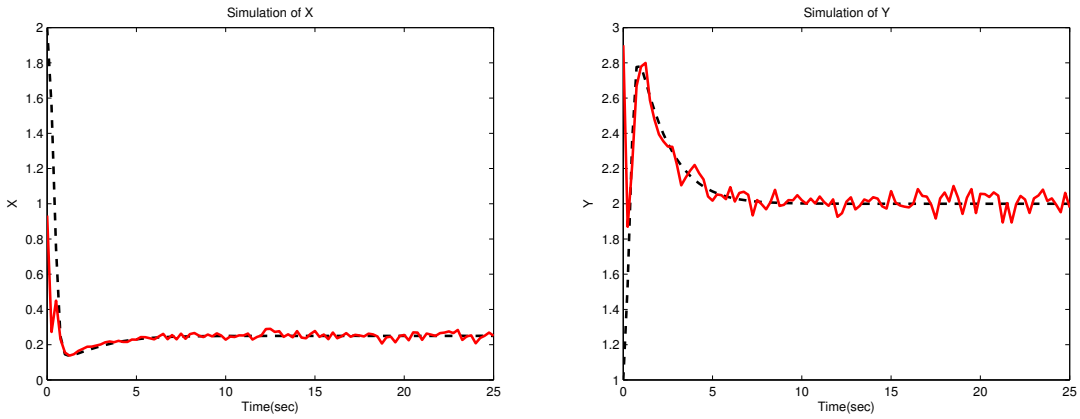
State variable	Mean value	Standard deviation	Actual value
X	0.25	0.008	0.25
Y	2.00	0.029	2.00

Table 2.1: The average estimated value along with the standard deviation obtained after running CSUKF 50 times to estimate X and Y

falls outside of the boundary (the dashed red line) and is consequently projected onto the boundary, in Figure 2.5b. The sigma points in Figure 2.5a are symmetric and therefore the weights are also symmetric. In Figure 2.5b the sigma points are not symmetric around the current mean value which is consistent with regular UKF [Teixeira 2008] in the presence of constraints. Due to this non-symmetric distribution of the sigma points, the weights also have to be adjusted accordingly. From Figure 2.5b it can be seen that the weighted mean of the constrained sigma points ($\star$) deviates from the mean of the unconstrained sigma points ($\square$). This variation has also been noticed in [Teixeira 2010]. Moreover as the probability distribution function is not truncated, part of the covariance ellipse calculated from the projected sigma points might also fall outside of the constrain boundary. This is not a problem when the actual state value is far from the boundary. However if the mean of the state value is close to the constrained boundary then this might create a ping-pong effect, meaning that if the state estimation value is moving towards the boundary, this effect would push it away again. [Teixeira 2008] reported that for constrained UKF, the state estimation will always stay within the constrain boundary. This is also the case for CSUKF, where the estimated state values are always within the boundary, due to the introduction of this constrained selection scheme of sigma points. This should make the CSUKF estimation more biologically meaningful.

Figure 2.6a and 2.6b show the true and estimated states trajectory of X and Y using CSUKF for a single run. For the sake of clarity in the estimation performance the simulated trajectory is plotted along with the actual state trajectory instead of $(X + Y)$. In both figures, the black dotted line represents the actual state trajectory and the red straight line represents the state estimation.

The CSUKF estimate converges to the true state value as the time progresses, despite selecting the initial state vector outside of the constraint boundary. This convergence is true for all 50 runs of CSUKF for this state estimation, as summarized in Table 2.1, giving a clear indication of high accuracy of CSUKF. In this algorithm the square-root form of the covariance matrix was propagated at each iteration which ensures numerical stability of the algorithm [van der Merwe 2004]. One of the limitations of the CSUKF is



(a) (b)

Figure 2.6: Estimated state trajectory for **the Higgins-Selkov oscillator** using CSUKF with an initial state estimate far from the constrain boundary.

that when the state estimate is very close to a boundary, then the weighted sample mean might not be equal to the current estimate which might give rise to a biased estimate. This is also mentioned for constraints with regular UKF [Vachhani 2006]. A careful selection of constraints may solve this problem.

2.3 Orthogonal-based parameter ranking

In a model not all parameters have the same level of importance. Some parameters have high sensitivity, for which a small change will have a large impact on the model output. Other parameters have low sensitivity meaning they can be varied without having much impact on the model output. Consequently, parameters with high sensitivity need to be estimated accurately while parameters with low sensitivity can, if necessary, be fixed to a small arbitrary number.

The proposed framework utilizes the orthogonal based parameter ranking method introduced by Yao *et al.* [Yao 2003, McAuley 2010]. This method was chosen to calculate the ranking in a more accurate and robust way. This is a data-based method since the ranking is dependent on the calculation of the parameter estimations. Specifically the ranking of parameters is made by analysing their sensitivity towards the system output. A sensitivity matrix is formed to measure the effect of a change in the parameter on the change in the system output. This effect is measured through the sensitivity coefficient. The sensitivity coefficient is an element in the sensitivity matrix which is calculated by taking the partial derivative of the system state output with respect

to each of the model parameters. Based on the sensitivity coefficient the parameters are ranked from most estimable to least estimable. In addition to the sensitivity of the parameters, the orthogonal method also considers linear independence of the parameters during the ranking. Parameters that are linearly correlated in their sensitivity coefficient are difficult to estimate.

2.3.1 Sensitivity matrix calculation

The sensitivity matrix Z^a has p columns, one for each of the model parameters and n rows, for the number of data points available for parameter estimation. If d different output variables are available, measured m times, then the number of rows are $n = d \times m$. The sensitivity matrix Z^a is defined as,

$$Z^a = \frac{\partial X}{\partial \theta} = \begin{pmatrix} z^a_{11} & z^a_{12} \cdots & z^a_{1p} \\ z^a_{21} & z^a_{22} \cdots & z^a_{2p} \\ \vdots & \ddots & \vdots \\ z^a_{n1} & z_{n2} \cdots & z^a_{np} \end{pmatrix} \tag{2.53}$$

where X denotes the vector of output elements, θ represents the parameter vector and $z^a_{i,j} = \frac{\partial x_i}{\partial \theta_j}$. In order to effectively compare the coefficients a scaled sensitivity coefficient matrix Z is defined as

$$z_{i,j} = \frac{\partial x_i}{\partial \theta_j} . \frac{\hat{\theta}_j}{\hat{x}_i} \tag{2.54}$$

where $\hat{\theta}_j$ represents the current estimate of parameter θ_j and $\hat{x}_i$ is the estimated value of the i^{th} state variable.

A conventional solution of Z can be obtained if the analytical solution of Equation (2.54) exists [Yue 2006]. However this is a rarity in non-linear biological systems [Yue 2006]. As a result, the Z has to be solved numerically for each iteration. Two of the most commonly used numerical methods are finite difference method (FDM) and decoupled direct method (DDM) [Yue 2006]. DDM suffers from numerical stability in some form of its implementation and needs to be used carefully [Rabitz 1983]. Therefore, this work applies the central difference (CD) method, a form of FDM, to calculate $z_{i,j}$ from the difference of the forward and reverse perturbation. This method is the most widely used [Hakami 2003] and straightforward in its implementation as only the calculation of the state variables are needed [Yue 2006]. Applying CD yields,

$$z_{i,j} = \frac{x_i(\theta_j + \Delta\theta_j) - x_i(\theta_j - \Delta\theta_j)}{2\Delta\theta_j} \tag{2.55}$$

In this approach, the choice of the step size $\Delta\theta_j$ plays an important role as the numerical value obtained through CD is highly dependent on the value of this step size. A variable step size results in a calculation where the sensitivity matrix captures more detailed information. In this implementation, the square root of the corresponding diagonal element of the estimation error covariance matrix is chosen as the step size, which gives $\Delta\theta_j = \sqrt{P_{j,j}}$ or $\Delta\theta_j = V_{j,j}$ for CSUKF. This ensures that the step size is reset for each iteration and remains within the feasible parameter range of the perturbed system. It has been shown that the accuracy error of this numerical calculation reduces linearly with reduction in step size [Brennan 2011]. As the CSUKF estimation error decreases at each iteration with the incorporation of new information, it is expected that choosing the step size in this manner increases the accuracy of the numerical calculation for CD. Choosing the parameter values within one standard deviation, ensures a relatively small step sizes as well. This makes the standard deviation a feasible choice for the step size.

2.3.2 Orthogonal based ranking

Orthogonal based ranking is a method that ranks the parameters by analysing how strong a parameter influences one or more of the measured responses. It iterates over each of the columns of Z to select the column with the highest sum of squared value. As each of the columns corresponds to individual parameters, this represents the parameter having the highest impact on the model output. The selected column is then added as a column to the matrix E_L (L denoting the iteration number). At each iteration new columns are added to E_L in order from the highest to lowest sensitivity of the parameters. For example at iteration four the matrix E will have four columns sorted according to the order of sensitivity. As the selected column of matrix Z might be correlated with another, some form of orthogonalization of the columns in matrix Z is needed to rank the influences of the individual parameters on the responses. After a parameter is chosen, the net influence of each of the remaining parameters on the selected parameter is adjusted by regressing the original columns of Z on to the column associated with the most estimable parameter (denoted as $\hat{Z}_L$). The residual matrix, Rs_L, is calculated by measuring the orthogonal distance between Z and the regression matrix $\hat{Z}_L$. The column with the highest sum of squared value in the residual matrix, denoted Cs_L, is chosen as the next most estimable parameter. It considers the parameters that are linearly independent in the sensitivity coefficient to be estimable. The reason is that if the sensitivity of one parameter is linearly dependent on some or all of the other parameters then the change of this parameter can be represented by the change of some or all of those parameters. The steps are repeated

until all iterations are finished or a specific cutoff value of Cs_L is reached.

Algorithm

1. Calculate the sensitivity matrix Z as discussed in section 2.3.1.
2. Calculate the sum of squared values of each column of Z and choose the column with the highest value as the most estimable.
3. Add the chosen column to E_L where $L = 1$ for the first iteration. With increasing number of iterations the size of E_L matrix grows as columns from Z are added to E_L.
4. Calculate an orthogonal projection $\hat{Z}_L$ for the column that exhibits the highest independence to the vector space spanned by E_L

$$\hat{Z}_L = E_L(E_L^T E_L)^{-1} E_L^T Z \tag{2.56}$$

5. As a measure of independence the residual matrix $Rs_L = Z - \hat{Z}_L$ is calculated.
6. The sum of squares value for each column of the Rs_L matrix is calculated, resulting in the vector Cs_L. The column corresponding to the largest sum of squares is chosen for the next estimable parameter.
7. Select the corresponding column in Z and augment it with the matrix E_L by marking the new column.
8. Iterate steps 4-7 until the cut-off value is reached or until all the parameters have been ranked and selected to be identifiable.

This orthogonalization based ranking method is very similar to the multiple regression method forward selection. The forward selection method starts with no variables in the model and includes them one by one if they are statistically significant. This method uses a threshold value as the stopping criteria for the addition of variables. The cut-off value in the orthogonal algorithm is analogous to this threshold value used by forward selection [Yao 2003]. The choice of the cut-off value of the stop criteria is somewhat arbitrary and depends on individual applications. In their work, [Yao 2003] choose the cut-off value of 0.04. However they concluded that this cut-off value depends on the level of noise in the experiments and for an imperfect model structure a higher cut-off value would lead to a more appropriate result. Taking their conclusion into consideration the cut-off value are

also varied here depending on the model complexity.

This sensitivity based analysis is similar to the structural identifiability analysis in the sense that it tries to identify parameters which do not change linearly with the perturbation (the sensitivity coefficients are not linearly dependent). It is also close to practical identifiability analysis as it requires pre-specified parameter values to start which could be either a nominal value or an actual estimate and it also requires the number and location of measurement time points [Miao 2011]. Thus to carry out the ranking of parameters, it is necessary to first have the estimate of the pre-specified parameter values. This sensitivity based ranking method explains, if a parameter has a high sensitivity then it is much important for the system behaviour and therefore has a high rank. A well known fact in systems biology is that not all of the parameters have equal sensitivity and only a small subset of parameters have a higher influence in the model behaviour [Ryan 2007, Schenkendorf 2011]. Once identified this small subset needs to be estimated accurately while the remaining parameters can be set to small random values.

2.4 Profile likelihood based identifiability analysis using CSUKF

Parameter identifiability has so far been addressed from the perspective of observability with the Kalman filter and its non-linear extensions. However since the computational complexity of these observability based methods increase with non-linearity and model size, this analysis is not well suited for non-linear biological models. To better suit biological models, profile likelihood based parameter identifiability analysis is integrated into CSUKF, instead of using the more common observability based method.

[Raue 2009] proposed a method for identifiability analysis based on likelihood confidence intervals generated from the profile likelihood trajectory. This profile likelihood trajectory is calculated for each parameter, θ_i, along the minimum of the $\chi^2(\theta)$ with respect to all other parameters $\theta_{j\neq i}$. In this manner it explores the parameter space of each of the parameters in the direction of the least increase of $\chi^2(\theta)$. In order to calculate this confidence interval it is necessary to have the asymptotic χ^2 distribution of the profile likelihood which follows a χ^2 distribution with df degrees of freedom. Justification for the integration of this method with CSUKF lies with the fact that the Kalman filter can have a likelihood interpretation with equations derived from a chi-square merit function. The Kalman filter is known as a recursive least squares (RLS) filter whereas the χ^2 merit function is a maximum likelihood function and the fitting

criteria for a χ^2 process is known as least squares fitting. According to [Thacker 1998], the Kalman filter equations can also be derived from χ^2 merit function. The Kalman filter is used to extend the likelihood estimation for cases where some of the state variables are not directly observed [Holmes 2004]. Thus the profile likelihood based identifiability analysis can also be applied using CSUKF.

2.4.1 The Profile Likelihood

In profile likelihood the unknown parameter vector $\theta \in \mathbb{R}^n$ is partitioned as $\theta = (\psi, \eta)$ where ψ is the 1-dimensional parameters of interest and η is (n-1)-dimensional nuisance parameter. Nuisance parameters are those parameters which are not of direct interest but must be accounted for in order to have a successful analysis of the parameter of interest. The parameter of interest is kept at a fixed value and the nuisance parameters are varied to have a maximum likelihood estimation (MLE). Then

$$pl_\psi = \max_\eta l_n(\psi, \eta) \tag{2.57}$$

is called the profile likelihood of ψ.

2.4.2 Asymptotic Confidence Interval

The confidence interval (CI) σ is an interval estimate determined by the position of its upper and lower endpoints to indicate the reliability of that estimate. A (1-α)CI of a parameter estimate consists of all those values that are to be considered valid by an α-level hypothesis analysis. In other words the true value of a parameter estimate will fall within the (1-α)CI with probability greater than or equal to (1-α) [Neale 1997].

The most widely used confidence interval calculated in maximum likelihood estimation (MLE) is the asymptotic confidence intervals. In an asymptotic confidence interval the CI of a estimated parameter θ_i is calculated as

$$\sigma_i = \hat{\theta}_i \pm \sqrt{\chi^2(df, 1-\alpha).C_{i,i}} \tag{2.58}$$

where $\chi^2(df, 1-\alpha)$ represents the $1-\alpha$ quantile χ^2 distribution with df degrees of freedom. $C_{i,i}$ is the variance from the covariance matrix calculated using the information matrix of MLE $l()$ [Raue 2011]. The information matrix is evaluated at the maximum of the likelihood surface. This information matrix reflects the curvature of the likelihood function in infinitely large samples [Reid 2007, Neale 1997] which is calculated by calculating the Hessian ($\frac{\partial^2 logl(\theta)}{\partial\theta_i \partial\theta_j}$) matrix at the maximum. Although this approach of

calculating the CI enjoys simplicity and speedy computation, it is less accurate when the sample sizes are small. The information matrix assumes the likelihood to be quadratic in form, which poses a constraint when calculating the CI if the likelihood is not symmetric. The information matrix fits a parabola (a quadratic function) with the log-likelihood and assumes the fitted parabola to be the log likelihood itself, which may overestimate the confidence interval and give erroneous results [Schaber 2011, Schenkendorf 2009].

In an asymptotic confidence interval, the mean and variance of the result need to be calculated. Even if the mean and variance are calculated accurately, for example using the sigma point method of CSUKF, to calculate the CI the probability distribution is assumed to be normal. This assumption of a normal distribution of the maximum likelihood estimator works well when the sample size is large. However there is no precise definition of how large a sample needs to be in order to consider it to be sufficient, as this is dependent on the model [Fisher 1922]. For small sample sizes the normality of the probability distribution might not be enough.

2.4.3 Likelihood based confidence interval

Graphically a smooth likelihood function is a parabola whose top represents the Maximum Likelihood Estimator (MLE). The curvature of the graph provides an estimate of the inverse of the variance-covariance of the MLE [Murphy 2000]. Instead of calculating the CI from the curvature of the ML, likelihood based confidence intervals are calculated from the profile likelihood. Here the approximated confidence interval of parameter θ is defined as

$$\{\theta|2[l(\hat{\theta}) - l(\theta)] < \chi^2(df, 1-\alpha)\} \tag{2.59}$$

In Equation (2.59) $l(\hat{\theta})$ is the MLE of θ and $l(\theta)$ is the log-likelihood defined for values of θ in the parameter space [Venzon 1988]. The difference mentioned in Equation (2.59) between the objective function evaluated at the optimal point and the objective function evaluated keeping one parameter fixed follows a a χ^2 distribution with df degrees of freedom according to the following equation

$$\{\theta \in: \chi^2(\theta) - \chi^2(\hat{\theta}) < \chi^2(df, 1-\alpha)\} \tag{2.60}$$

The border of Equation (2.60) represents the likelihood based confidence interval [Raue 2009, Raue 2011]. Depending on the value of the degree of freedom (df), two types of confidence intervals emerge. If df is considered to be one, then it results in

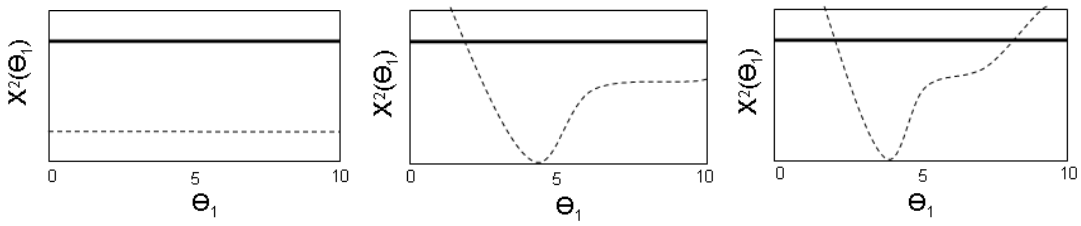


Figure 2.7: Parameter Identifiability Analysis (adopted from [Raue 2009]). The dotted line represents the profile likelihood $\chi^2(\theta_1)$ value and the straight line is the threshold at $\chi^2(df, 1-\alpha)$. (a) Defines the structural non-identifiability as the profile likelihood line is flat. (b) Defines the practical non-identifiability where the profile likelihood goes to ∞ for either $\theta \to \infty$ or $\theta \to -\infty$. (c) An identifiable parameter, where θ has a finite confidence interval for both high and low values from the global minimum

point-wise confidence intervals and when df is equal to the number of parameters then it is simultaneous confidence interval.

2.4.4 Structural identifiability analysis

A parameter θ_i is structurally identifiable if it can be estimated uniquely from the model structure. If not then it is structurally non-identifiable. The profile likelihood approach manifests a flat profile likelihood trajectory for a structurally non-identifiable parameter (Figure 2.7a). A structural non-identifiability is dependent only on the model structure and its parameterization, and is independent of the amount or accuracy of the experimental data. Structural non-identifiability is specifically caused by the insufficient mapping of the states with the measurement data that arises due to redundant parametrization in the solution of observables [Raue 2009]. This mapping causes the χ^2 value to remain constant as any change in the parameters does not have an impact on the value of the objective function. This redundant parametrization results in functionally related parameters. For example, in the equation $\theta_1\theta_2 - 20 = 0$, parameter θ_1 and θ_2 are functionally related, that is $g(\theta_{sub}) = 0$ and they can be varied within themselves without having any impact on the observables $y(t_i, \theta)$. As a result no unique solution is possible for these parameters. The profile likelihood of such parameters are flat and their likelihood based confidence interval are infinite meaning that they are structurally non-identifiable as can be seen in Figure 2.7a .

2.4.5 Practical identifiability analysis

In contrast to structural identifiability, practical identifiability analysis answers whether it is possible to uniquely identify a parameter from a set of experimental data considering the amount and quality of the data [Raue 2011, Quaiser 2009]. For a practical non-identifiable parameter, the profile likelihood flattens out and would not cross the χ^2 threshold of the likelihood confidence interval (discussed in section 2.4.2) for either increasing or decreasing or both values of parameter θ (Figure 2.7b). A practical non-identifiability might arise due to insufficient measurement data or data with too much noise [Balsa-Canto 2010, Raue 2011]. In the case of a practical non-identifiable parameter, the likelihood based confidence interval of an estimate $\hat{\theta}_i$ is infinitely extended in the increasing or decreasing direction of θ_i as been described in Figure 2.7b. As a result the χ^2 stays below the $\chi^2(df, 1-\alpha)$ for a df degrees of freedom. For a parameter to be practical non-identifiable it is not necessary for the confidence interval to be infinite in both the directions. It might happen that either σ^- or σ^+ is infinite, but not both.

2.4.6 Implementation of identifiability analysis with CSUKF

Using the representation of χ^2 in vector form and the notations from the CSUKF derivation, the χ^2 merit function for the k^{th} iteration of the CSUKF can be written as

$$\chi^2(k) = (y(k) - \hat{y}^-(k))R^{-1}(y(k) - \hat{y}^-(k))^T \tag{2.61}$$

The χ^2 value of a complete CSUKF run is then

$$\chi^2 = \sum_{k=1}^{n}(y(k) - \hat{y}^-(k))R^{-1}(y(k) - \hat{y}^-(k))^T \tag{2.62}$$

where n is the number of data points, R is the observation error covariance matrix, $y(k)$ is the vector of observation data and $\hat{y}^-(k)$ is the current estimate of the observed state variables.

The identifiability analysis begins with the best set of parameter values found after the first complete run of the CSUKF. In this first run, CSUKF tries to estimate all the unknown parameters after initializing them to a small random number. This estimation also includes those parameters that turn out afterwards to be non-identifiable. During the calculation of the profile likelihood it is possible that a parameter set with a lower objective function is found. In this case the profile likelihood calculation is restarted

with the newly found best set of values. A specific step size (which is initially 10% of the current value) is then added iteratively with the global minimum point of a specific parameter θ_i to create the positive profile likelihood trajectory. The same approach is taken for the calculation of the profile likelihood trajectory for the negative values where the step size is deducted iteratively.

During the calculation of the profile likelihood, one parameter is always kept fixed and the maximization is performed over the nuisance parameters. In CSUKF the system noise covariance matrix Q and measurement noise covariance matrix R are generated considering all parameters to be free and to be estimated. It is necessary to remove the effect of the fixed parameter variance from these covariance matrices during the calculation of the profile likelihood value. This can be done by striking out the rows and columns corresponding to the fixed value from the inverse of the covariance matrix. The following steps are performed to generate that matrix

1. Take the inverse of the covariance matrix.

2. Delete the j^{th} row and column of the inverted matrix, where j is the index of the fixed parameter.

3. Take the inverse of the modified matrix for the covariance matrix without the effect of the variance of the fixed parameter.

This removes the effect of the fixed parameter from the covariance matrix [Baker 2001]. The step size is used to increase or decrease the parameter values iteratively to calculate the objective function while keeping the j^{th} parameter fixed at this point and running the estimation over the remaining parameters. This calculation is performed until a pre specified stop criteria is met. In the experiments mentioned in this thesis, the stop criteria is 100 iterative runs on both the increasing and decreasing direction of the parameter values. These χ^2 values are then plotted against the logarithmic value of each of the parameters in the direction matrix. Analysis of these plots yields the structural and practical non-identifiable parameters. These plots also help to calculate the likelihood confidence interval for the identifiable parameters. If a global minimum of a parameter θ_i is represented with θ_{i0} then the upper bound of the confidence interval of this parameter is the value at point j where $\theta_{ij} > \theta_{i0}$ and where the profile likelihood trajectory crosses the χ^2 threshold with df degrees of freedom. Similarly the lower bound of this parameter is where the value at point j is $\theta_{ij} < \theta_{i0}$, and where the profile likelihood trajectory θ_{ij} crosses the χ^2 threshold.

2.5 Determining correlation and functional relationship between parameters

2.5.1 Correlation between parameters

Parameters that are highly correlated are not identifiable. Therefore finding this correlation helps to determine the non-identifiable parameters. Fernandez *et. al.* calculated the correlation matrix of the parameter estimates from the Fisher information matrix (FIM) [Rodriguez-Fernandez 2006b]. The inverse of the FIM gives an estimation of the lower bound of the covariance matrix according to the Cramèr−Rao inequality [Kay 1993]. From this covariance matrix the correlation coefficients can be calculated. However if the models are non-linear with respect to the parameters then FIM may lead to a poor approximation of this covariance matrix [Schenkendorf 2009]. In this framework, the correlation coefficient is calculated from the square-root of the covariance matrix generated by CSUKF during the parameter estimation process. If V represents the square-root of the estimation covariance matrix then the covariance matrix can be easily obtained as $P_\theta = VV^T$. The covariance matrix calculated using this sigma point method is highly accurate and does not need to calculate any gradients or the Jacobian [Schenkendorf 2009]. The correlation coefficient between θ_i and θ_j can then be calculated as

$$corr(i,j) = \frac{P_{(i,j)}}{\sqrt{P_{(i,i)}P_{(j,j)}}} \tag{2.63}$$

2.5.2 Determining functional relationship between parameters

In this thesis the mean optimal transformations [Hengl 2007], a non-parametric bootstrap-based algorithm, is used to determine the functional relationship between the parameters. This method is able to determine the non-linear relationship between parameters. It is based on the optimal transformation of the dependent (response) variable and a set of independent (predictor) variables. This transformation is estimated by the non-parametric regression method alternating conditional expectation (ACE) [Breiman 1985]. This transformation uncovers the non-linear relationship between the variables. The strength of ACE lies in the fact that it can recover functional form of the relationship even for complicated ones [Wang 2004]. As it is non-parametric an initial guess concerning the underlying functional form of the relations is not required.

In ACE the response variables Y^{ace} are replaced by functions $\xi(Y^{ace})$ and the predictor variables $X_1^{ace}, \cdots, X_n^{ace}$ are replaced by $\phi(X_1^{ace}), \cdots, \phi(X_n^{ace})$ [Breiman 1985].

The general form of an ACE regression model is

$$\xi(Y^{ace}) = \alpha + \sum_{i=1}^{n} \phi(X_i^{ace}) + \varepsilon \tag{2.64}$$

An optimization is then performed iteratively to estimate functions, ϕ_i and ξ by minimizing the error variance that is not explained by the transformed dependent variable on the sum of transformed independent variables defined as

$$\varepsilon^2(\xi, \phi_1, \cdots, \phi_p) = E\{[\xi(Y^{ace}) - \sum_{i=1}^{p} \phi_i(X_i^{ace})]\}^2 \tag{2.65}$$

The final $\phi(X_i^{ace})$ and $\xi(Y^{ace})$ after the minimization are estimates of the optimal transformations.

In the mean optimal transformation approach, [Hengl 2007] proposed a test function H_k^{ace} that interprets the optimal transformation as a functional relationship between the variables. This test function states that if $H_{i1}^{ace} \leq T_1^{thr}$ the response parameter θ_{i1} on the left hand side does not have any functional relationship with any other parameter $\theta_k, k \in \{i_1, \cdots, i_n\}$. To test the whole group of parameters, the mean of H_k^{ace}, $\bar{H}^{ace}(\theta_i, \theta_k)$ is calculated which is then used to identify whether a given set of parameter has enough information to establish a relationship. If $\bar{H}^{ace}(\theta_i, \theta_k) > T_2^{thr}$ then a strong relationship exists between parameter θ_i and the parameters in vector θ_k where T_1^{thr} and T_2^{thr} are thresholds defined as $T_1^{thr} = 0.01$ and $T_2^{thr} = 0.07$. Finally it returns a $p \times p$ matrix where p is the number of parameter with either a zero or one as an entry in the cell. If there is one in a cell then there is a functional relationship between the two parameters mentioned in the row and the column corresponding to the cell.

2.6 Treatment of non-identifiability with informed prior

Situations might arise in biological models where it is not possible to perform additional measurement in order to solve the non-identifiability of parameters. Furthermore simplification of the model may also not be possible when it will significantly reduce the model's desired predictive capability. In these scenarios neither the frequestists nor any classical statistical methods may be used as they are incapable of estimation in the presence of non-identifiable parameters [Neath 1997, Rannala 2002, Samaniego 2010]. Classical approaches like least squares estimation, likelihood require parameters to be identifiable before proceedings into estimation [Neath 1997]. Because of the absence of identifiability in these approaches, they are not able to distinguish between two or more

possible sets of parameter values on the basis of observed data. In contrast to the classical approach, Bayesian methods can provide point estimates of the parameters without solving this non-identifiability if the informative prior distribution can be provided.

The Bayesian paradigm using proper priors has no difficulty in treating non-identifiable parameters. Identifiability is more of a problem of model specification rather than one of inference and therefore Bayesian inference can produce a unique estimate of parameters even in the presence of non-identifiability [Samaniego 2010]. Bayesian inference for a parameter starts by stipulating a prior distribution on the parameters of interest. This prior distribution is then updated on the basis of available observation data to generate a posterior distribution on which the inference is finally based. The data available to estimate a non-identifiable parameter are defective in the sense that they do not provide enough information to uniquely estimate these parameters. However the data may still be informative for Bayesian inference to determine the posterior distribution of these parameters. The Kalman filter and its non-linear variants can be derived within a Bayesian framework [Chen 2003] as it is considered as one of the simplest dynamic Bayesian networks. Therefore this Bayesian treatment of non-identifiability can also be adapted to CSUKF for making point estimation of the parameters without solving the non-identifiability of the parameters, an approach that has no classical counterpart. The next section discusses how identifiability of parameters are addressed using the probability distribution function (PDF) and how a proper prior can solve parameter non-identifiability.

2.6.1 Parameter identifiability with PDF

To describe parameter identifiability in the view of probability distribution, the variables are first defined, X as the vector of observed random variable, θ as the set of parameter values and p to be the conditional probability distribution of X given parameter θ. Now if there exists some $\theta_1 \neq \theta_2$ satisfying

$$p(X|\theta_1) = p(X|\theta_2) \tag{2.66}$$

then the parameters of the model are adjudged to be not identifiable. Equation (2.66) says that at least any two sets of parameter values θ_1 and θ_2 produces the exact same probability distribution of the observable, i.e. all possible sets of observations have the same probability for any two different sets of parameter values [Rannala 2002]. Having this probability distribution, it is not possible to uniquely distinguish these two parameter sets from the observation data. The result is that the parameters sets are non-identifiable.

2.6.2 Informative prior for Bayesian inference

An informative prior can be used to perform a legitimate Bayesian inference for the parameters even if they are non-identifiable. The following example is used to illustrate this principle. Given a parameter vector θ with two elements $\beta^{(1)}, \beta^{(2)}$ is considered available. Now the case with two sets of θ with different parameter values, where $\theta_1 = \{\beta_1^{(1)}, \beta_1^{(2)}\}$ and $\theta_2 = \{\beta_2^{(1)}, \beta_2^{(2)}\}$ is considered. For this parameter vector if the likelihood is a function of $\beta^{(1)} + \beta^{(2)}$ only, then it is impossible to separately identify β values. Now if an informed prior for $\beta^{(1)}$ assigns $\beta^{(1)} = y$ with probability one then $\theta_1 = \theta_2$ is possible if and only if $\beta_1^{(2)} = \beta_2^{(2)}$. This makes the model identifiable. Thus Bayesian inference is possible even for the models which are non-identifiable from the perspective of likelihood, if an informative prior is available. However this specific fact is not by itself sufficient enough to trust the solution from the Bayesian inference. Bayesian inference of a parameter might not represent its true value. If not dealt with care, Bayesian inference will not converge to an acceptable close to the true value of a parameter θ even as the sample size n goes to infinity [Samaniego 2010].

2.6.3 Informed prior in CSUKF

The CSUKF is a non-linear extension of Kalman filter, as such it can be derived within the framework of Bayesian inference. As a result the Bayesian treatment of non-identifiability can also be applied when using CSUKF. For this treatment it is needed to assign the informed prior to the parameters before starting the actual estimation process. There are a number of parameters in CSUKF that can be selected for *a priori* knowledge. They are the initial state, $x(0)$, the initial estimation covariance matrix, $P(0)$, the process noise covariance matrix, Q, the measurement noise covariance matrix, R, and the UT scaling parameters, α, β, κ. The scaling parameters α, β and κ only affects the higher order terms and have little effect on the estimation accuracy. The two parameters of CSUKF that can be used to introduce the informed prior, is the initial state assumption $x(0)$ and the initial estimation covariance matrix $P(0)$. The other two parameters that can have a higher impact are the state noise covariance matrix Q and the measurement noise covariance matrix R. R is generally known from the noise model of the measurement data. Therefore Q can be initialized in order to introduce the prior information to CSUKF. In CSUKF this proper prior is being formulated by informatively initializing the state variable $x(0)$, the estimation error covariance matrix $P(0)$ and the state noise covariance matrix Q. The information from the parameter ranking can be used to formulate this informative prior for the parameters. As the high rank parameters have a much higher

influence on the system output, they are considered to be estimated with higher precision in the model. Therefore during the initialization of the $P(0)$ and $Q(0)$ the high rank parameters are initialized with low variances. For the low rank parameters $P(0)$ and $Q(0)$ is initialized with high variances. Variance for the parameters which are deemed to be non-identifiable through the orthogonal method are also initialized with high values. This variance initialization is done chronologically with the ranking. In this case as there is no information on the state variables, they can be initialized to small random number. If two sets of data are available, one from the mutant and one from the wildtype, then the mutant data can be used to formulate the initial estimation on the state variables and their covariance matrix which can then be used to formulate the informative prior for the estimation with the wildtype data.

Chapter 3

Application of the framework

The primary focus of the proposed framework is the successful estimation of the unknown parameter values of biological kinetic models. To verify the accuracy and applicability of the proposed framework, it is implemented and used to estimate parameters in a number of previously published biological models. At first the applicability and accuracy of the estimation algorithm, CSUKF, is confirmed by estimating the parameters in a very simple kinetic model composed of the reactions of the upper part of glycolysis in yeast [Hynne 2001, Klipp 2005]. The statistics generated from the results of this estimation is then compared with some widely used global optimization algorithms as well as the two widely used non-linear extensions of the Kalman filter. The applicability of the algorithm is then tested using a fairly large kinetic model of sucrose accumulation in the sugar cane culm tissue developed by Rohwer *et al.* [Rohwer 2001, Uys 2007]. During this test the estimation runs into the problem of parameter non-identifiability which are then addressed with the proposed framework. Using this framework the non-identifiable parameters are first determined and then solved before going for the final estimation. The framework also proposes an alternative way of estimating the parameters without solving the non-identifiability. This alternative solution uses the proper prior distribution of the state variables to have a unique estimation of the parameter values even in the presence of non-identifiability. This approach is verified with the sugar cane culm model as well as with a Gene regulatory network supplied by the DREAM6 Estimation of Model Parameters Challenge [Prill 2010]. Each of these biological models are represented using the state-space model described in Equation (1.17). The problem of parameter estimation in these models are formulated as a state estimation problem, described in Equation (1.22), then discretized to the final state space representation shown in Equation (1.23).

To evaluate the framework synthetic data is obtained from simulating the model and are used as the measurement data. This synthetic data is generated by first running the model simulation with a known parameter set and then a white Gaussian noise is added with the simulated data. At first, it may seem unusual to use such data as it seems that all the parameters are known beforehand. However this is not the case, as the information

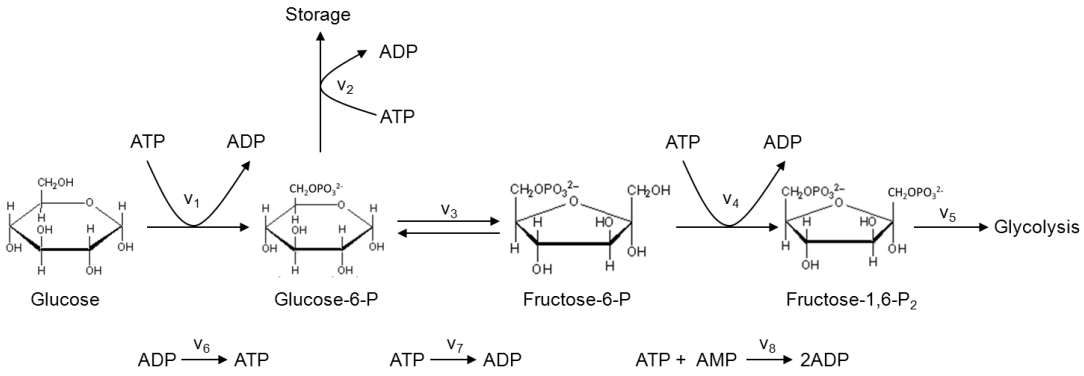


Figure 3.1: Schematic diagram of the simplified glycolysis model (adapted from [Klipp 2005]). Abbreviations are as follows Gluc6P: glucose-6-phosphate; Fruc6P: fructose-6-phosphate; Fruc1,6P_2 : fructose-1,6-bisphosphate; ATP: adenosine-triphosphate; ADP: adenosine-diphosphate; AMP:adenosine-monophosphate. The flux, v_i denotes the i^{th} reaction represented by rate laws depending on their enzymatic reactions. Details of the rate laws can be found in the A.2.

is lost between the movement of the parameter values to simulate the synthetic data and then returning to parameters via estimation [Chen 2010]. This kind of synthetic data plays a critical role in the development and the validation of most numerical algorithms. The numerical tool-kit MATLAB is used for the implementation of the framework. The biological models are also coded in MATLAB as a system of ODEs. For comparison with four widely used global optimization algorithm, the upper part of glycolysis of yeast is also modelled in Copasi [Hoops 2006] in addition to MATLAB.

3.1 Glycolysis model from yeast

The first application of CSUKF is to estimate parameters in a relatively simple kinetic model of the upper part of glycolysis in yeast. [Klipp 2005] simplified the glycolysis model originally developed by [Hynne 2001]. This model is restricted to just the first four reactions of glycolysis, as shown in Figure 3.1 accounting for the adenylates ATP, ADP and AMP. This model explains the degradation of glucose in the process of yielding energy and building blocks for cellular processes. In the model there are 15 parameters. In this model, if a reaction $A \rightarrow B$ is determined by flux v_i then the parameters, maximal velocity V_{max} and enzyme affinity for reaction K_m are defined as $V_{max,i}$ and $K_{m,i}$ in the rate laws, where i represents the reaction number. Among these 15 parameters four parameters are chosen for estimation using the synthetic time series data of the metabolite concentrations. These four parameters are k_2, $V^f_{max,3}$, $V_{max,4}$ and k_{8r}. Their

actual values are 2.2600, 140.2820, 44.7287 and 133.3300 $mmol/ml$ respectively. The glucose concentration are kept fixed at 12.8174$mmol/ml$ while all the other metabolite concentrations are free to vary throughout the simulation to generate a set of time-series data. The metabolite concentrations are initially set to zero and the energy currency as $ATP = 2.1$, $ADP = 1.4$, $AMP = 0.1$ $mmol/ml$. This experiment has assumed that the measurement data of all metabolites, whose concentration are changing over time, are available.

Experimental setup The experiment begins by integrating the ODEs over a time interval from zero to 25 seconds with all the parameters considered to be known. The ODE45 function (a numerical Runge-Kutta method for numerical integration) of MATLAB is used to solve the ODEs. The simulated time-series data y of the metabolite concentrations are used to generate synthetic measurement data y_{noisy} by adding uncorrelated white noise ε^0 with it, $y_{noisy} = max[0, y + 0.2 * r * y]$ where y is the simulated measurement data and ε^0 is defined as $\varepsilon^0 = 0.2 * r * y$ where r is a random variable having normal distribution with zero mean and one standard deviation. y_{noisy} has a random variation of 20% of the actual measurement data. Finally the measurement data is clipped to zero if it goes below zero. Sampling at an interval of $\Delta t = 0.25$ second gives a time series with 101 data points.

3.1.1 Comparison of global optimization algorithm

Before going into a detailed analysis of CSUKF, the first experiment performs a comparison between four of the most widely used global optimization algorithms to estimate the parameters of this model. This analysis is made in order to have a better understanding on the performance and working principles of these algorithms, which in turn aid in the implementation of CSUKF. The four algorithms are evolutionary programming (EP), genetic algorithm (GA), simulated annealing (SA) and particle swarm optimization (PSO). All of these algorithms have some form of stochasticity. For real problems stochastic global optimization methods have been shown to arrive at relatively good solutions in moderate computational time, whereas deterministic methods have proven to be too expensive computationally [Moles 2003, Rodriguez-Fernandez 2006b]. Copasi is used for this comparison as these algorithms are already been implemented in Copasi and building the model in this software is straight forward. The optimization parameters chosen as the default values in Copasi are given in Table 3.1. The experiment chooses these default parameters of copasi as they are mostly used by the experimentalists. This also ensures equal starting points for all four algorithms.

Algorithm	Number of Generations	Population Size		Random Number Generator	Seed
EP	200	20		1	0
GA	200	20		1	0
	Start Temperature	**Cooling Factor**	**Tolerance**		
SA	1	0.85	1E-06	1	0
	Iteration Limit	**Swarm Size**	**Std. Dev.**		
PSO	2000	50	1E-06	1	0

Table 3.1: Setup of the optimization parameters used for the comparison.

Algorithm	Parameter Name	Mean	Std. Dev.	Median	CPU Time (second)
EP	k_2	2.675	1.685	2.247	1,104.62
	$V^f_{max,3}$	130.276	44.631	140.571	
	$V_{max,4}$	163.501	1,032.808	44.970	
	k_{8r}	112.375	49.446	133.316	
GA	k_2	7.493	0.014	7.495	640.76
	$V^f_{max,3}$	0.025	0.026	0.019	
	$V_{max,4}$	0.008	0.007	0.006	
	k_{8r}	0.020	0.037	0.016	
SA	k_2	2.165	0.334	2.249	84,540.14
	$V^f_{max,3}$	115.746	44.715	140.616	
	$V_{max,4}$	$2.72E+23$	$7.94E+23$	44.890	
	k_{8r}	122.010	50.265	133.500	
PSO	k_2	2.256	0.048	2.249	16,359.82
	$V^f_{max,3}$	137.744	20.201	140.616	
	$V_{max,4}$	45.525	4.469	44.890	
	k_{8r}	133.460	0.278	133.500	

Table 3.2: Results obtained by repeating the computation of the four algorithms in the case study model 100 times (CPU time is for total 100 runs). Statistics of the data are calculated from this 100 runs. CPU time is the total time that it took to run each of the methods 100 times.

Each algorithm run 100 times for statistical analysis of the optimization results. These results are summarized in Table 3.2. From Table 3.2 it can be seen that among the two EC algorithms, the performance of EP is better than that of GA. Although GA has much smaller standard deviations, its estimated mean values are far from the original values. The higher accuracy in EP could be due to the reproduction operator which carries more information and efficient than GA [Chiong 2007]. [Fogel 1995] also reported that EP outperforms GA on obtaining the mean solution from an optimization problem by minimizing the objective function. However the mean value of EP is still far from the actual value, this may be due to a known limitation of such algorithms, that they can get stuck in local minima if proper algorithmic parameters are not chosen, yielding solutions that are often not optimal [Hewlett 2007]. In contrast the median values from EP closely approximates the actual value, which makes this algorithm still applicable to the parameter estimation. The average objective function value of EP, 12.35, also suggested a better result compare to the average objective function value of GA, 206.18. The performance of SA is similar to that of EP except for parameter $V_{max,4}$ where it has a very large mean and standard deviation. This makes SA even worse than EP. When results from each of the 100 runs for SA are analysed individually, it is found that about quarter of those runs have an extreme outlier where the parameter value for $V_{max,4}$ is more than $E + 20$. The objective function value with these outliers are also very high. If statistics are calculated after removing these outliers, the estimation accuracy dramatically improves for all the parameters with mean value of 140.62 and standard deviation of 0.006 for $V_{max,4}$. The accuracy for SA also improves, if the estimation is introduced with noise free synthetic measurement data as was shown in [Baker 2010]. This gives an indication that the noise in the measurement data is causing SA to get stuck in local optimums and causing this big outlier values. Similarly Hart *et al.* showed that the evolutionary algorithms have higher probability of success than simulated annealing after sufficiently large number of functional evaluations [Hart 1996]. However the median value for all the four parameters is similar to that of EP. But the time taken by SA to reach this accuracy is very high which makes this algorithm infeasible for large parameter estimation algorithms. This is a known problem of SA that it can reach the vicinity of the global optima quite quickly but then moves slowly when approaching the optima [Banchs 1997]. Compared to the other three global optimization algorithms PSO performs the best. The average objective function value, 0.606, also indicates the improved performance of PSO. The reason is that PSO utilizes a significantly different information sharing mechanism than the other algorithms. In PSO, only the information

from the best particle is distributed and the evolution only targets the best solution [Jones 2006]. Furthermore PSO has the advantage that it can escape from local minima [Wachowiak 2010]. But the time taken by PSO, while not as long as SA, is still very high compared to GA and EP. This could be due to the high number of functional evaluations required for PSO to converge to a solution as reported in [Wachowiak 2010] or it could be due to the settings of the control parameters of the algorithm for which appropriate settings is key to success [Jones 2006]. While the control parameters are kept unchanged throughout for proper comparability, their performance may vary significantly if non-default options are used [Elizabeth 2004]. The problem remains that the selection of the values for the control parameters, that would ensure the algorithms best performance is not intuitive. Therefore it would be much easier for the typical users to go with the default options in which case this comparison result would give a vital clue on the performance of the algorithms for these users.

3.1.2 Applying CSUKF to the glycolysis model

Having illustrated the applicability and accuracy of CSUKF for state estimation in chapter 2.2.5, the focus here is on applying CSUKF to the parameter estimation of the upper part of the glycolysis model. CSUKF is completely implemented within MATLAB. For this parameter estimation experiment the model is also written in MATALB. As a first model for testing parameter estimation with the CSUKF, this model is chosen for its relative simplicity. The same four parameters are selected that were used for the algorithm comparison problem in the previous section. These four parameters are selected purposefully for the large spreads of their magnitudes, from $V^f_{max,3} = 140.282$, the largest parameter, to $k_2 = 2.26$, the smallest. As described in the experimental setup section, synthetic measurement data is generated for this experiment. The unknown parameters are initialized with a small random number between zero and one. The simulated time-series data with the estimated parameters are generated by integrating the ODEs of the model at a step size of 0.25 seconds over 25 seconds, for a total of 101 data points. The constraints on the metabolite concentration and the parameter values are defined to be zero for the lower bound and 500 for the upper bound. The parameters are converted to time invariant states, changing the parameter estimation problem into a state estimation problem as mentioned in Equation (1.22). The process noise covariance matrix has also to be initialized with augmented noises of the parameters and the state variables. For the parameter values, the diagonal elements of the process noise covariance matrix, Q, is initialized with 20% variation of the initial parameter

values $\theta(0)$ i.e. $diag(Q) = 0.2 * \theta(0)$. Similarly for the state variables, a 20% variation from the initial state values (which is already randomly initialized), $x(0)$, is incorporated, $diag(Q) = 0.2 * x(0)$. The measurement noise covariance matrix R is initialized with the same ε^0 used to generate the synthetic measurement data, i.e. with a random variation of 20% of the actual measurement data. The CSUKF adjusts the parameter values by adjusting the state covariance matrix to minimize the mean-squared error between the simulation data and the measurement data. This process continues in an iterative manner until the system reaches a steady state (i.e. where the metabolite concentration no longer changes over time).

Results To obtain representative statistics of the estimation, the experiment is repeated 100 times. The results as summarized in Table 3.3, show that the estimated value of the parameters closely and consistently approximate the original value. A representative example (run 3) of the parameter estimation is illustrated in Figure 3.2. The dashed lines represent the true parameter value and the solid lines correspond to estimation trajectory. As can be seen in the figure, the estimation trajectories of all the four parameters closely approach to their actual values. This figure illustrates how even when the estimation is started far from the true value, the algorithm quickly zeros in on the value range and then gradually converges to the actual values, for parameters with both high and low values. This confirms the flexibility of CSUKF with regards to the magnitude of the parameters.

Parameter Name	**Actual Value**	**Estimation**	
		Average	**Std. Dev.**
k_2	2.260	2.260	0.076
$V^f_{max,3}$	140.282	140.235	0.592
$V_{max,4}$	44.729	44.742	0.266
k_{8r}	133.330	133.330	0.387

Table 3.3: Summary statistics of the parameter estimation values obtained from CSUKF. For each estimated parameter, the mean and standard deviation are calculated from 100 runs. In each run the parameters were initialized with small random numbers between 0 and 1.

A closer view of one of the parameter trajectories, $V^f_{max,3}$, is shown in Figure 3.3 (corresponding to run 5). In this close up it is possible to see the standard deviations

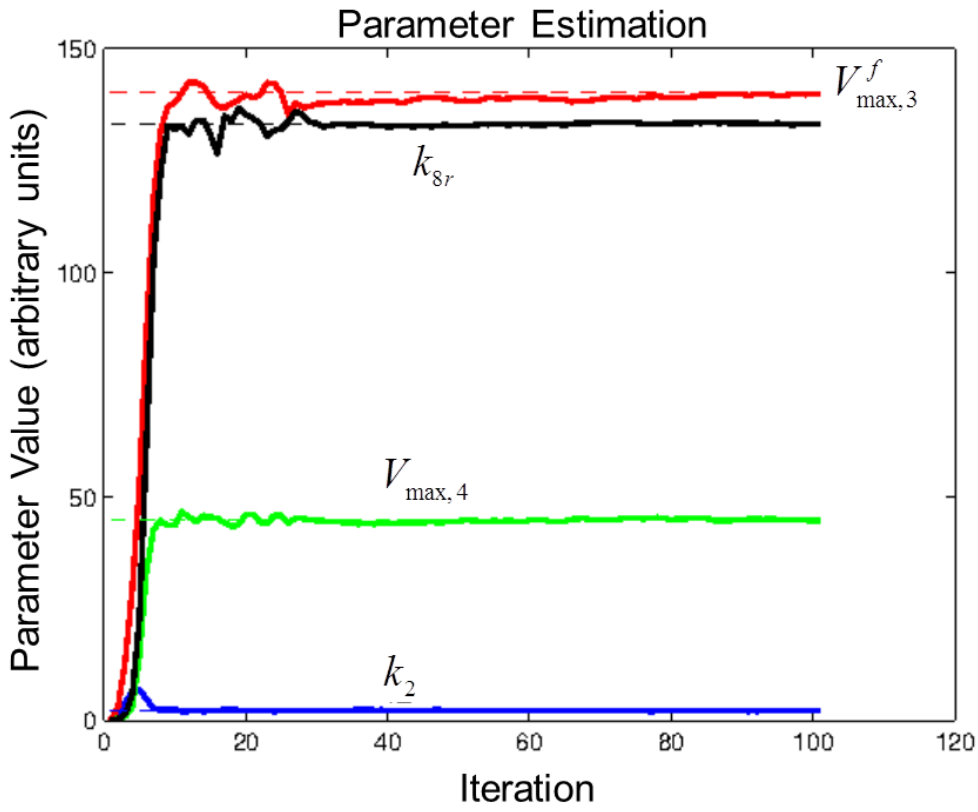


Figure 3.2: Parameter estimation trajectory from a representative sample run (run 3) of the recursive estimation of the parameter values. The dashed lines represent the actual parameter values and the solid lines represent the corresponding estimation trajectory.

throughout the refinement of the estimation. Like before the dashed line represents the actual parameter value and the solid line represents the estimation trajectory. The vertical bar on the trajectory shows the estimation standard deviation calculated from the diagonal element of the square-root estimation covariance matrix V at each 10^{th} time step. As can be seen from the figure the estimation standard deviation decreases with each iteration step. It starts with much high value at the beginning and then decreases as the estimation reaches the actual value. Similar behaviour has also been observed in case of UKF as described in [Quach 2007]. In order to test whether the selection of the initial value of the parameters has any impact on the estimation accuracy, the CSUKF is repeatedly run 30 times, initializing the parameters with different random values between 0.01 to 20 including the endpoints. In all the runs CSUKF estimated the parameters with similar accuracy mentioned in Table 3.3.

This experiment chooses flat priors (uninformed priors) for state distribution. How-

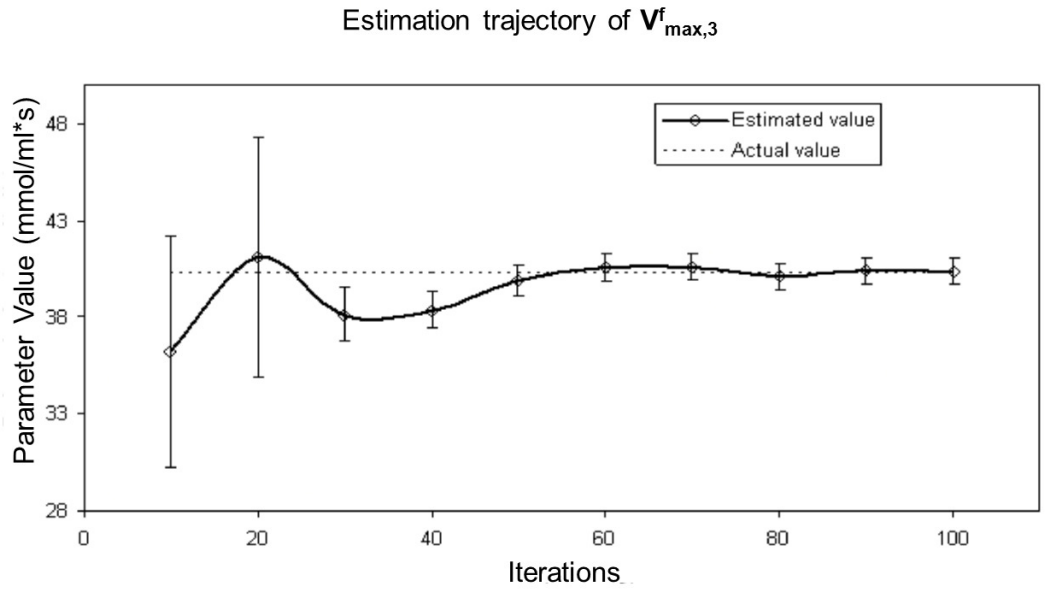


Figure 3.3: Estimation trajectory of $V^f_{max,3}$ from a representative sample run (run 5) with standard deviation. The dashed line represents the actual value and the solid line represents the estimation trajectory. The vertical bar represents the standard deviation. For visibility and ease in analysis plotting begins with the 10^{th} sampling point and the standard deviation is given for every 10^{th} sampling point.

ever a more elaborate prior distribution could be used with the CSUKF in order to favour meaningful regions of the parameter and state space. The impact of such informative priors has been demonstrated for two different models in sections 3.4 and 3.5. Moreover as constraints can be added into CSUKF, this algorithm may yield results which emphasize biological relevance. One other advantage of CSUKF is its simultaneous computation of the square-root of the variance of the estimator where it might not otherwise be straightforward to compute due to the complexity of the model.

A comparison was made between the constrained square-root unscented Kalman filter and the two other widely used non-linear extensions of the Kalman filter, the extended Kalman filter and, the (regular) unscented Kalman filter. Table 3.4 quantifies the average of the estimation results obtained from this comparison together with the previous results from the four global optimization algorithms, EP, GA, SA, PSO. Each of these algorithms is run 50 times. In each run the parameters are initialized with small random number between 0 and 1. As can be seen from the table, CSUKF outperforms the four global optimization algorithms either in accuracy, in time or both.

Algorithm	Parameter Name	Actual Value	Mean	Std. Dev.	Total time (second)
EP	k_2	2.26	3.043	2.123	697.94
	$V^f_{max,3}$	140.282	130.923	53.152	
	$V_{max,4}$	44.7287	252.103	1,457.136	
	k_{8r}	133.33	117.811	47.018	
GA	k_2	2.26	7.495	0.014	327.28
	$V^f_{max,3}$	140.282	0.026	0.024	
	$V_{max,4}$	44.7287	0.007	0.005	
	k_{8r}	133.33	0.024	0.042	
SA	k_2	2.26	2.148	0.412	43,311.26
	$V^f_{max,3}$	140.282	119.213	43.958	
	$V_{max,4}$	44.7287	$2.29E+23$	$7.01E+23$	
	k_{8r}	133.33	119.730	55.072	
PSO	k_2	2.26	2.250	$3.78E-05$	8,044.39
	$V^f_{max,3}$	140.282	140.616	0.003	
	$V_{max,4}$	44.7287	44.890	0.0004	
	k_{8r}	133.33	133.500	0.003	
EKF	k_2	2.26	3.579	0.028	169,699.07
	$V^f_{max,3}$	140.282	92.645	0.209	
	$V_{max,4}$	44.7287	36.129	0.144	
	k_{8r}	133.33	120.048	0.343	
UKF	k_2	2.26	2.264	0.030	4,363.57
	$V^f_{max,3}$	140.282	140.310	0.454	
	$V_{max,4}$	44.7287	44.709	0.160	
	k_{8r}	133.33	133.330	0.324	
CSUKF	k_2	2.26	2.269	0.075	4,361.32
	$V^f_{max,3}$	140.282	140.239	0.679	
	$V_{max,4}$	44.7287	44.730	0.279	
	k_{8r}	133.33	133.327	0.383	

Table 3.4: Comparison of CSUKF to EKF, UKF, EP, GA, SA and PSO. Results are the average from running each algorithm 50 times, however the time is the total time for each algorithm to complete all 50 runs. The same noisy data was used for all algorithms. The unknown parameters were initialized to a small random value between 0 and 1 for each run of each algorithm.

Although the two EC algorithms run much faster than CSUKF, its estimation accuracy

is far below CSUKF. The performance of SA is quite low considering accuracy because of the presence of outliers. This has been discussed in section 3.1.1. On the other hand PSO has similar accuracy to CSUKF, but requires longer time. The accuracy of UKF is similar to that of CSUKF. This is consistent with the work of Rudolph van der Merwe where they confirmed that the square-root UKF has equal to marginally better estimation accuracy than UKF [van der Merwe 2004]. But with CSUKF constraints can be introduced in to the estimation, which is not possible with UKF [Kandepu 2007]. The performance of both CSUKF and UKF exceeds that of the EKF. This agrees with the results of other similar works like [Merwe 2001, Kandepu 2007]. To test whether the performance of both EKF and UKF are dependent on the starting values these algorithms are run with multiple starting values chosen in the similar fashion as mentioned for CSUKF, ranging from 0.01 to 20. It has been observed in the experiment that the estimation performance of EKF is dependent on the starting values of the parameters. When the parameters are initialized with values less than one they do not reach the actual value for any of the four parameters. However, EKF does converge to the actual values when the parameters are initialized to any value greater than one. Although the UKF generally performs similar to CSUKF when the parameters are initialized with small random numbers between 0.01 and one. If the parameters are initialized very close to zero, the estimation is inconsistent, even resulting in negative values for the parameters. Furthermore, UKF suffers from numerical instability if parameters are initialized with high values. To accurately determine the numerical stable range for UKF, additional experiments are performed with a direct search of the initial values, and it is found that UKF is stable when initialized with all parameter values between 0.01 and 5. The average running time of CSUKF is also marginally better than the UKF. This is due to the propagation of the square-root of the estimation covariance matrix, P, eliminating the Cholesky decomposition of P at each iteration. Despite this difference, both CSUKF and UKF have a similar run time. In terms of the run time, EKF is the computationally most expensive method of the three. As the profiling in Matlab explains, this is due to the calculation of linearization or the Jacobian in the state space equation. This conforms with the literature, that the requirement to explicitly calculate Jacobians is a major bottleneck for complex functions in EKF [Madhumita 2010]. This run time is expected to rise as the model grows bigger. Taken together, the advantages, flexibility, speed and most importantly accuracy of CSUKF makes it the clear choice for parameter estimation in biological models.

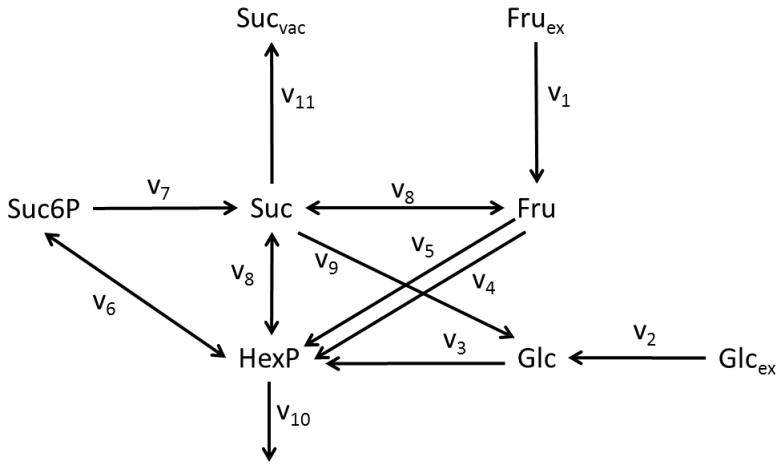


Figure 3.4: Schematic diagram of the sucrose accumulation model of sugar cane culm tissue. Abbreviations are as follows Suc: sucrose; Suc6P: sucrose-6-phosphate; HexP (fructose 6-phosphate and UDP-glucose); Fru: fructose; Glc: glucose. The subscript 'ex' stands for extracelullar and the subscript 'vac' stands for vacuolar. The Hexose phosphate pool was considered as an equilibrium block. The numbered v's denote the reactions which are represented by rate laws, details of which can be found in A.1

3.2 Sucrose accumulation model in the sugar cane culm tissue

To verify that the algorithm works for relatively large models, the CSUKF was applied to parameter estimation of the second case model, the kinetic model for sucrose accumulation in the sugar cane culm tissue [Rohwer 2001, Uys 2007]. This model helps to assess the biochemical control of sucrose accumulation and futile cycling in sugarcane. This model allows for the potential of different manipulation strategies to enhance sucrose accumulation and to select the most promising ones. This is a typical medium scale model in systems biology. The schematic diagram of the model is given in Figure 3.4.

Experimental setup In this model the structure of the network is mathematically formulated as a set of ODEs. The model consists of 54 parameters from which 12 parameters are selected to be estimated corresponding to the 12 parameters that Rohwer estimated in his work [Rohwer 2001]. List of these 12 parameters can be found in Table 3.5. Each of the 12 parameters is initialized to a small random number between zero and one. The remaining 42 parameters are known and kept fixed during the estimation. The five metabolites that are considered to be explicit variables, i.e. whose concentrations

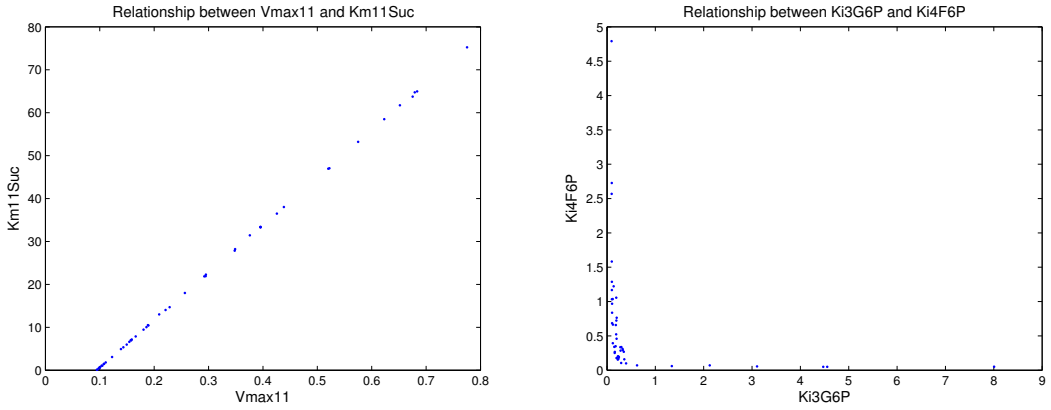
are changing over the time are Fru, Glc, HexP, Suc6P and Suc. The rest of the metabolite concentrations are considered constant. These five metabolites have an initial concentration of $Fru = Glc = HexP = Suc6P = Suc = 1$. As with the other models, this experiment also assumes that the measurement data for all the changing metabolites are available. Synthetic data, corrupted with noise, as described in section 3.1, is again generated as the measurement data. This model requires a much longer time to reach steady state, so the time interval used to generate the time series data is adjusted accordingly, zero to 2340 seconds with a step size of $\Delta t = 10$ second. The process noise covariance matrix, Q, and the measurement noise covariance matrix R is setup using the same method described in section 3.1.2. In this experiment a sensitivity based ranking of the parameters is coupled with CSUKF to find the most estimable parameters. This method performs identifiability analysis based on the orthogonality of the sensitivity matrix [Yao 2003]. Global identifiability cannot be generally guaranteed for non-linear models, thus this sensitivity based analysis is used to determine the probability that the parameters are identifiable [Berit 2008]. This sensitivity analysis considers two features that govern the estimability of the parameters. First, it considers the strength of the influence of a parameter on one or more of the measured responses. Second, it considers the correlation between the effect of the parameters on the model predictions.

Results Table 3.5 summarizes these results with the estimation from 50 runs of CSUKF along with the ranking of the parameters chosen from the most common ranking of those 50 runs. The full algorithm of this ranking method is described in section 2.3. The selection of the stop criteria of this algorithm is crucial. In their work [Yao 2003] used heuristics for their stop criteria of the selection of the identifiable parameters. We also used heuristics to find that $Max(C_L) < 0.004$ was a reasonable stopping criteria. This stopping criteria is used throughout these experiments. The standard deviation of seven parameters in the estimation was higher than 100% of their mean values for this model as described in Table 3.5. Furthermore, six parameters have a mean value that is more than one standard deviation apart from the actual value. All these indicates a poor performance of the CSUKF for parameter estimation in this model. This performance is due to several of the parameters being non-identifiable, thus allowing their values to vary within a wide range. As can be seen in Table 3.5 V_{max6r} has the highest rank, and correspondingly this parameter also has the highest magnitude in the sensitivity coefficient matrix. However one point to note is that this sensitivity depends on the initial value of the parameters and in some cases has high sensitivity at the beginning stages of the estimation, which then decreases as the estimation approaches the actual value. This dependence of the

Parameter Name	Actual value	CSUKF		Ranking
		Mean	Std. Dev.	
k_{i1Fru}	1	1.00	0.01	4
k_{i2Glc}	1	1.00	0.009	9
k_{i3G6P}	0.1	0.67	1.46	5
k_{i4F6P}	10	0.63	0.85	NI
$k_{i6Suc6P}$	0.07	0.45	0.77	8
$k_{i6UDPGlc}$	1.4	0.32	0.40	3
V_{max6r}	0.2	0.34	0.67	1
k_{m6UDP}	0.3	4.73	3.45	6
$k_{m6Suc6P}$	0.1	5.97	4.58	2
k_{i6F6P}	0.4	0.65	1.06	NI
V_{max11}	1	0.28	0.19	7
k_{m11Suc}	100	21.43	21.82	NI

Table 3.5: Parameter estimation results and parameter ranking from the application of CSUKF to the sugarcane model. The mean and standard deviation of the estimated parameters is calculated from 50 repetitions. The ranking is chosen to be the most commonly occurring rankings from the 50 runs. The NI stands for Non-identifiable. In each repetition the parameters are randomly initialized to values between zero and one.

sensitivity coefficients on the initial parameter values has been previously described in [Yao 2003]. Additionally the sensitivity is time varying with the dynamics of the model as previously reported in the work of [Yue 2006]. The three parameters found to be non-identifiable, i.e. those for which a unique solution cannot be found are, k_{i4F6P}, k_{i6F6P} and k_{m11Suc}. This could be due to a functional relationship of these parameters between themselves or among the others. An exhaustive functional analysis of k_{m11Suc} with each of the other parameters individually reveals that k_{m11Suc} has a strong linear relationship with parameter $V_{max1}1$ as shown in Figure 3.5a. A hyperbolic relationship can also be seen between k_{i3G6P} and k_{i4F6P}, illustrated in Figure 3.5b. However, similar analysis was not able to find a simple relationship directly between k_{i6F6P} and any one of the other identifiable parameters. But, this parameter was found to have a very low sensitivity towards the state variables which kept this parameter from being selected as estimable. As explained by [Yao 2003, McAuley 2010], changing the stop criteria to a different value may cause this parameter to become identifiable. This proved to be the case when the value of the stop criteria was reduced to 0.0004. Having directly dealt



(a) (b)

Figure 3.5: Functional relationships between parameters. a) Linear; the relationship between V_{max11} and k_{m11Suc}. b) near hyperbolic; the relationship between k_{i3G6P} and k_{i4F6P}.

with these three parameters the actual value of k_{i4F6P}, k_{i6F6P} and k_{m11Suc} are considered to be known through biological experiments.The estimation process is thus repeated after removing these previously non-identifiable parameters from the estimation procedure. The CSUKF is again executed 50 times. To make a comparison on the estimation performance for the remaining nine parameters, UKF, genetic algorithm (GA) and non-linear least squares (NLSQ) are also applied, each running 50 times. The SimBiology toolkit (in MATLAB) implementation of the later two algorithms with the default settings are used for this comparison. The results are summarized in Table 3.6. As can be seen from the table, CSUKF estimates some of the parameter values with low standard deviation. However five high rank parameters still have large standard deviation (more than 50% of their mean values), specifically the two parameters k_{m6UDP} (150% of the mean value) and $k_{m6Suc6P}$ (88% of the mean value). These two parameters also suffered from high standard deviation when they were previously estimated with the full set of 12 parameters. Although the sensitivity method determines these parameters to be identifiable, the high variation is contradictory to this. An explanation is that although these parameters have no linear dependency, a higher order dependency might still exist which this method does not consider. In [Yao 2003] it was shown that this sensitivity method would identify parameters as estimable even when they are highly functionally correlated, unless they have an exact linear relationship in the sensitivity coefficient. In other words, parameters selected using this method are biased by linear independence. Parameters that have non-linear functional relationship with each other might still be considered to be identifiable [Yue 2006]. A similar conclusion is reached by Floor *et*

Parameter Name	CSUKF		UKF		GA		NLSQ	
	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
k_{i1Fru}	1.00	0.01	1.00	0.01	0.97	0.15	0.99	0.007
k_{i2Glc}	1.00	0.009	1.00	0.008	1.00	0.09	0.99	0.001
k_{i3G6P}	0.14	0.04	0.12	0.02	0.85	0.69	0.1	0.01
$k_{i6Suc6P}$	0.85	0.98	0.88	1.25	0.94	0.72	1.35	2.135
$k_{i6UDPGlc}$	0.55	0.51	1.45	0.12	0.97	0.74	1.29	0.305
V_{max6r}	0.60	1.01	1.11	1.36	0.86	0.56	3.27	4.932
k_{m6UDP}	5.32	8.03	0.03	0.56	0.9	0.55	0.89	1.747
$k_{m6Suc6P}$	8.71	7.67	-2.36	2.53	0.88	0.62	0.78	1.77
V_{max11}	1.00	0.005	0.99	0.001	1.04	0.29	0.99	0.001

Table 3.6: Comparison of parameter estimation methods. The mean and standard deviation from 50 repetitions for each of the estimation methods, CSUKF, UKF, GA and NLSQ are presented for the nine identifiable parameters. The values of the three non-identifiable parameters are considered to be known. In each repetition of each method, the identifiable parameters are randomly initialized to small values between zero and one.

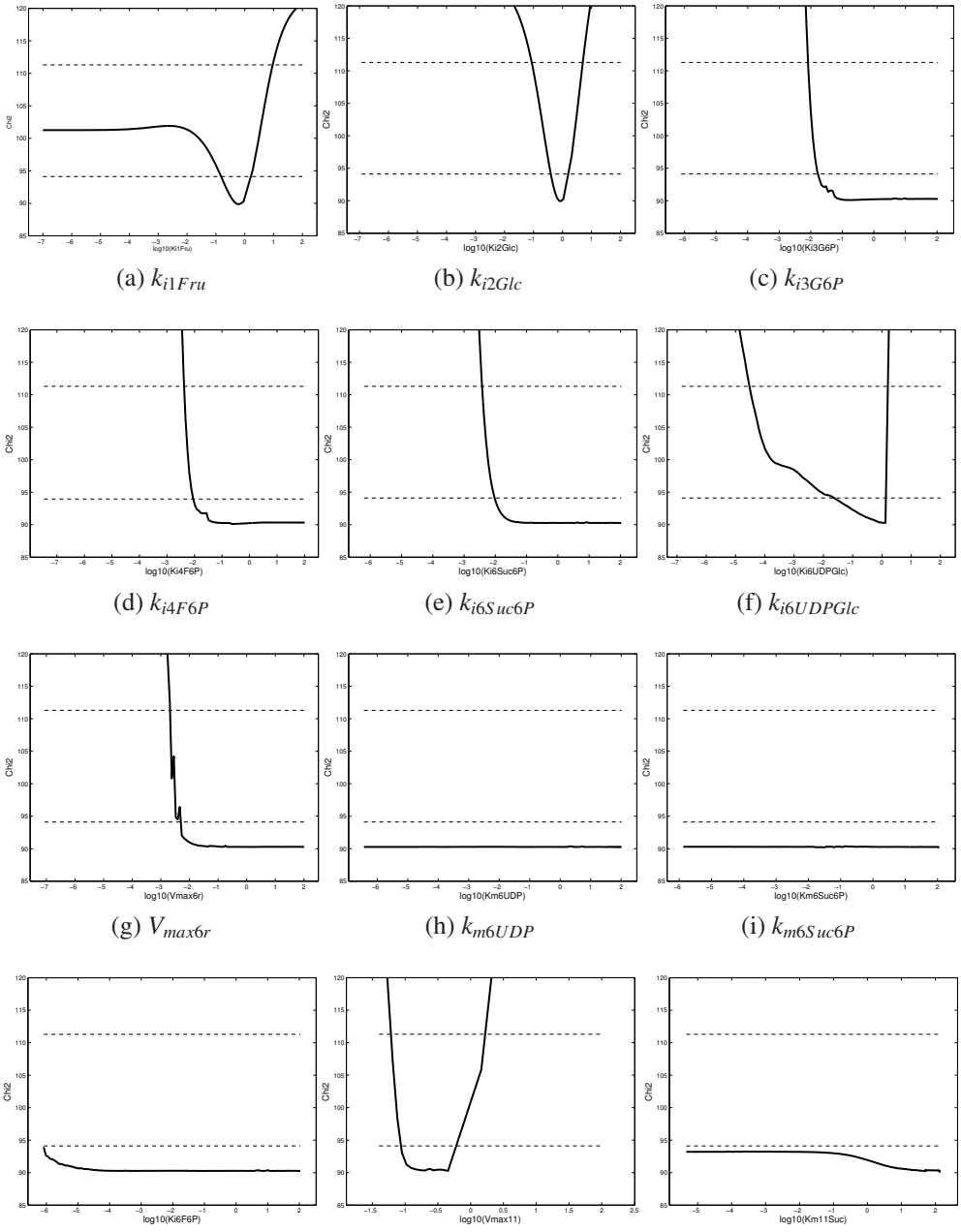
al. in their work on a non-linear dynamic model of submerged arc silicon [Berit 2005]. An alternative reason for this high variance could be that as this method only considers the identifiability within a finite set of points in the parameter space, the individual parameters within this set might have a very large domain, which subsequently causes a large variation within the individual parameters. This means that although the parameters are indeed identifiable, they are poorly resolved. From the table it can be seen that parameter k_{i1Fru}, k_{i2Glc}, k_{i3G6P} and V_{max11} are identified equally well in all the algorithms except GA. Compared to the other algorithms GA has higher standard deviation for these four parameters agreeing to the conclusion in section 3.1.1. From the above discussion it can be concluded that sensitivity analysis alone is not complete enough to make a conclusion regarding the nature of identifiability of the model. There is a clear indication that among the nine parameters found to be identifiable by the sensitivity method, some are still non-identifiable. Thus to make a more complete and accurate analysis, profile likelihood based identifiability are additionally considered.

3.3 Application of the complete framework

In this section, the complete parameter estimation framework is used to develop a unique estimated parameter set for the sugarcane model. As discussed in the previous section, orthogonal based sensitivity analysis cannot fully determine the non-identifiable

parameters. To overcome these limits and to generate a complete *structural* and *practical* parameter identifiability analysis, a profile-likelihood based identifiability analysis is incorporated. In this approach parameter identifiability is investigated by calculating the profile likelihood of the parameters. In this framework the CSUKF is used to calculate this profile likelihood. This approach can be considered as an alternative to the more computationally expensive observability based analysis, the most commonly used identifiability analysis in control theory [Farina 2006, Lillacci 2010]. A variation on the synthetic measurement data is made by taking a time interval of [0 2340] seconds with step size of 10 seconds to generate the measurement data. To begin the profile likelihood analysis it is necessary to first find the current optimum value of the parameters. The CSUKF is applied to find this set of global optimum values. If a new, better, set of parameter values are found at any time during the profile likelihood calculation; the program is restarted with the new optimum value. A good model to data agreement is found with an objective function value of $\chi^2 = 90.271$ for 12 parameters with 234 data points. The maximal and minimal step size is adjusted depending on the parameters and their profile likelihood values. If the profile likelihood is not reasonable smooth, a smaller step size is chosen. A larger step size is chosen if the iteration stops prematurely because of reaching the maximal number of iterations. Figure 3.6 depicts the result of the profile likelihood analysis for these 12 parameters using a confidence interval of 95%. Defining $\Delta_\alpha = \chi^2(df, 1-\alpha)$, the point-wise confidence interval threshold for 95% confidence level is thus $\Delta_\alpha = 3.84$ and simultaneous confidence interval threshold is $\Delta_\alpha = 21.03$.

As depicted in Figure 3.6, the parameters k_{i1Fru}, k_{i2Glc}, $k_{i6UDPGlc}$ and V_{max11} are the identifiable parameters because they have a finite likelihood based confidence interval in both increasing and decreasing direction of the parameter values. This basic principle of identifiability has been described by Raue *et al.* in his work [Raue 2009, Raue 2011]. Parameters k_{m6UDP} and $k_{m6Suc6P}$ are *structurally* non-identifiable because they each have a flat profile likelihood. This explains the reason for the high standard deviation in the estimation result previously reported in both Table 3.5 and Table 3.6. High standard deviation for *structural* non-identifiable parameters have already been reported in other works, such as Hengl *et al.*, where they observed high standard deviation for *structurally* non-identifiable parameters in parameter estimation of a non-linear dynamical model for the endocytosis of the erythropoietin receptor [Hengl 2007]. This is due to the fact that these parameters can take any value within a wide and possibly infinite range, and yet still reach the same global optimum. *structural* non-identifiability is mostly due to the over-parametrization of the model, that is the model has more parameters than



(a) k_{i1Fru} (b) k_{i2Glc} (c) k_{i3G6P}

(d) k_{i4F6P} (e) $k_{i6Suc6P}$ (f) $k_{i6UDPGlc}$

(g) V_{max6r} (h) k_{m6UDP} (i) $k_{m6Suc6P}$

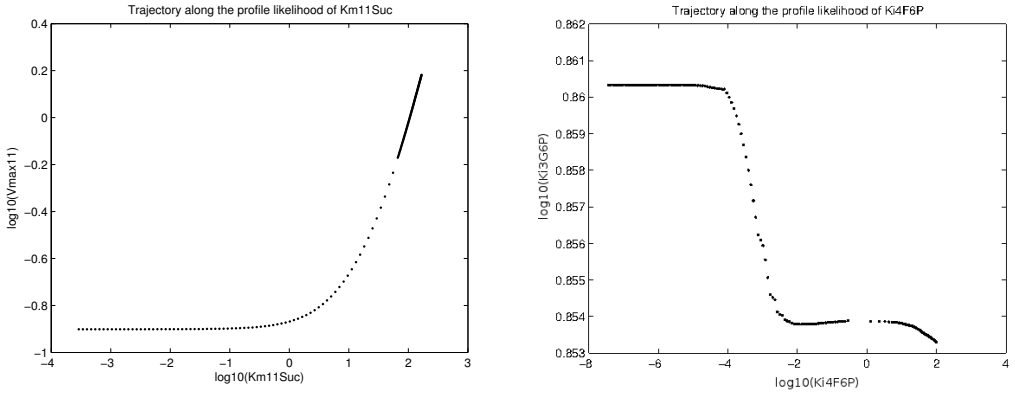
(j) k_{i6F6P} (k) V_{max11} (l) k_{m11Suc}

Figure 3.6: Profile likelihood based parameter identifiability analysis. The solid line represents the profile likelihood trajectory versus parameter, with the parameter values in log scale. In each plot the dotted lines represent the two thresholds. The lower threshold is the 95% point wise confidence interval and the upper threshold is the 95% simultaneous confidence interval. If the profile likelihood of a parameter crosses the threshold line for both high and low values then the parameter is identifiable. A horizontal (i.e. flat) profile likelihood indicates *structural* non identifiability, while crossing the threshold(s) on only one side indicates *practical* non identifiability.

can be estimated from the data [Chis 2011a]. As this model has two *structurally* non-identifiable parameters it indicates that the model is somehow over-parametrized. Over-parametrization might be due to a functional relationship between one parameter with any other parameters of the model [Rannala 2002]. To make these parameters *structurally* identifiable additional measurement data is needed through different mapping of the observation function for the hidden variables.

In Figure 3.6, the parameters k_{i3G6P}, k_{i4F6P}, $k_{i6Suc6P}$, V_{max6r}, k_{i6F6P} and k_{m11Suc} are shown to be *practically* not identifiable, as their likelihood-based confidence region is infinitely extended in either increasing or decreasing parameter values. This means that they cannot be reliably estimated with acceptable accuracy from the available noisy measurement data [Raue 2009, Raue 2010, Miao 2011]. Among these six *practically* non-identifiable parameters, three were previously determined to be non-identifiable using the orthogonal method. As the sensitivity based method is similar to the *practical* analysis approach [Miao 2011], this result also illustrates the accuracy of the profile likelihood based method for identifying *practically* non-identifiable parameters. As per definition parameters having a functional relationship are *structurally* non-identifiable [Hengl 2007, Raue 2009]. By that definition k_{m11Suc} was supposed to be *structurally* non-identifiable because it is linearly dependent on V_{max11} as shown in Figure 3.5a. But profile likelihood analysis found it to be *practically* non-identifiable but *structurally* identifiable. The reason is depicted in Figure 3.7a where the change of parameters V_{max11} is plotted for different values of k_{m11Suc} used in the calculation of the profile likelihood of k_{m11Suc} in a log-log plot. From Figure 3.7a it can be seen that while for higher values of k_{m11Suc} ($log_{10}(k_{m11Suc}) > 0$) they form a linear relationship, this does not hold for smaller values of k_{m11Suc}. Figure 3.7b plots the change of parameter k_{3G6P} along the value of k_{i4F6P} used for profile likelihood calculation of k_{i4F6P}. They show a linear relationship between -2.4 and -4 for $log_{10}(k_{i4F6P})$, the trajectory remains almost constant for the other values illustrating that there is no obvious relationship between the two parameters at these values.

After identifying all non-identifiable parameters it is necessary to solve this non-identifiability, before proceeding with the identification procedure. The first approach to solve *structural* non identifiability is to have qualitatively more measurement data to change the observable mapping function as shown by Raue *et al.* [Raue 2009, Raue 2011] in their example model. The second approach is applied to functionally related parameters. In this approach, where possible, the higher ranked parameter needs to be measured and the lower ranked parameter(s) can be estimated. If the measurement of



(a) Relationship between V_{max11} and k_{m11Suc} (b) Relationship between k_{i4F6P} and k_{3G6P}

Figure 3.7: Relationship between the change of parameters in respect to the different values of the other parameters calculated during profile likelihood calculation. The plot is log-log scale. a) The plot of V_{max11} versus k_{m11Suc}. b) The plot of k_{i4F6P} versus k_{3G6P}.

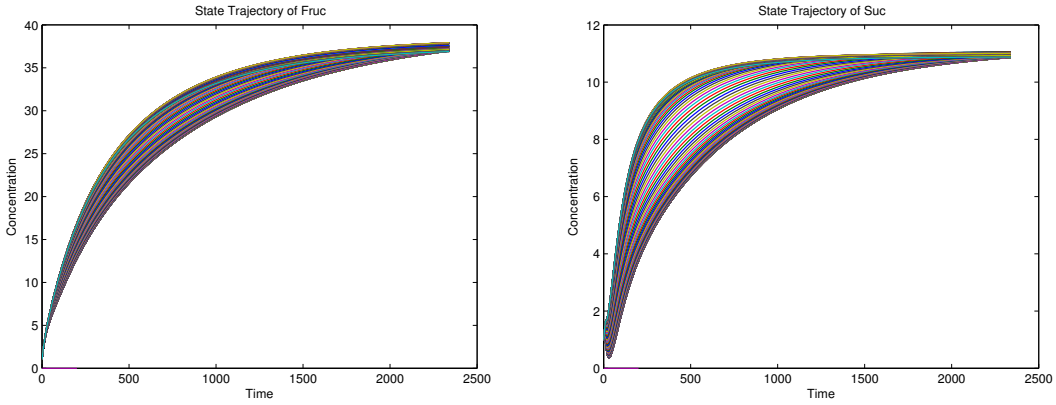
this high rank parameter is not possible due to wet lab constraints, then the alternative approach would be to estimate the high rank parameter while keeping the low rank parameter(s) fixed to a nominal value as suggested by [Yao 2003]. Since low rank parameters have lower sensitivity, they would not effect the model trajectories as much as the high rank parameter. In this example, the second approach is applied here. The identifiability analysis identified k_{m6UDP} and $k_{m6Suc6P}$ to be *structurally* non-identifiable. This gives a hint that these two parameters might have a functional relationship with the other parameters. However, the method does not explicitly state which parameters those might be. In order to identify the functional relationship between these two parameters and any other parameters, the mean optimal transformations approach (MOTA) was applied using the profile likelihood estimation data of k_{m6UDP} and $k_{m6Suc6P}$ individually [Hengl 2007]. MOTA identified a functional relationship between k_{m6UDP}, k_{i3G6P} and V_{max6r}. When applied to $k_{m6Suc6P}$, a functional relationship was found with V_{max6r}. To solve *structural* non identifiability of both k_{m6UDP} and $k_{m6Suc6P}$, direct measurement of V_{max6r} is considered here as it has a functional relation with both k_{m6DUP} and $k_{m6Suc6P}$ and also it has a very high rank (ranking 1).

As *practical* non-identifiability occurs mostly due to insufficient amount and quality of experimental data, this can be solved by increasing the amount and quality of measured data [Raue 2009, Raue 2011]. The observables along a *practical* non-identifiability changes only negligible, but the model behaviour in terms of internal states might change strongly. High variation on the model trajectories of these states along the profile

likelihood of the *practically* non-identifiable parameters would give the points where the uncertainty in a specific parameter has the largest impact on the model uncertainty [Raue 2010]. Improving measurement data at these points would solve the *practical* non-identifiability [Raue 2009].

Another reason for *practical* non identifiability would be a correlation between parameters [Faller 2003, Rodriguez-Fernandez 2006b]. In this case the flattening out of the likelihood of a *practically* non-identifiable parameter could continue along the correlated parameters. If measurement data for any of the highly correlated parameters is available, the non-identifiability of the related parameters would be solved. A similar approach has also been described in the work of Guedj *et al.* where they analysed the *practical* identifiability of a dynamic model of HIV through the correlation of the parameters [Guedj 2007]. Recall that CSUKF estimates both the mean and the square-root of the covariance at each iteration. The square-root of the covariance matrix can then be used to calculate the correlation coefficient matrix. Using the correlation matrix and the built in MATLAB function *'corrcoef'* any significant correlations among the parameters are found. This analysis found a strong correlation between k_{i3G6P} and k_{i4F6P}. An accurate value for either of these two parameters would solve the *practical* non identifiability of the second. As k_{i4F6P} has already been shown to be non-identifiable during the orthogonal based ranking method, measurement of this parameter would help to resolve the non-identifiability of both these parameters. The correlation method also found k_{i6F6P} to be significantly correlated with V_{max6r} and $k_{i6UDPGlc}$. Among these three parameters V_{max6r} has already been selected for measurement. We select k_{i6F6P} for measurement considering the same reason for which k_{i4F6P} is measured.

To solve the *practical* non-identifiability of k_{m11Suc}, the state trajectories of Fruc and Suc are plotted against the profile likelihood values of k_{m11Suc} in Figure 3.8a and 3.8b respectively. This trajectory reveals spots where the uncertainty of θ has the largest impact on the model. Thus it suggests points of measurement data at which enhancement in the precision or increase in the data points would improve parameter identification. It can be seen from both figures that there is a large variation between state trajectories, for $Fruc$ between time point 50 to 2000 and for Suc between 30 and 2000. This suggests that having more measurement data of the states may solve the *practical* non-identifiability of k_{m11Suc} because this measurements might bring the parameter out of the flat χ^2 value. Therefore new synthetic measurement data over the same time interval, but with a time step of 0.25 seconds is generated. This approach was successfully used by Raue *et al.* to solve the *practical* non-identifiability in their example application of the JAK-STAT

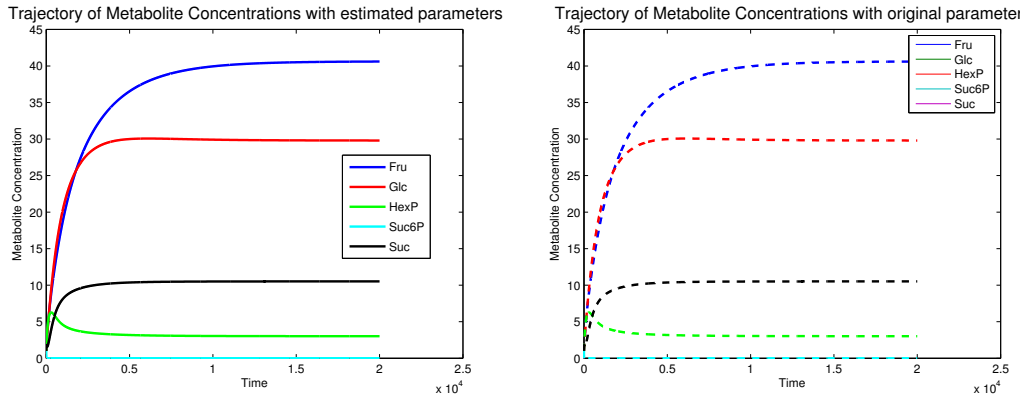


(a) (b)

Figure 3.8: Trajectory of a) Fruc and b) Suc along the values of k_{m11Suc} used during the calculation of the profile likelihood. Places of larger variability denotes points where measurement of a species would efficiently estimate the parameter

Parameter Name	**Obtained**	**Value**	σ^+	σ^-	**Original Value**
k_{i1Fru}	Estimated	0.999	1.191	0.181	1
k_{i2Glc}	Estimated	1.001	2.07	0.4	1
k_{i3G6P}	Estimated	0.1	0.111	0.099	0.1
k_{i4F6P}	Measured	10.00	-	-	10
$k_{i6Suc6P}$	Estimated	0.05	0.094	0.01	0.07
$k_{i6UDPGlc}$	Estimated	1.16	2.32	0.05	1.4
V_{max6r}	Measured	0.2	-	-	0.2
k_{m6UDP}	Estimated	0.4	0.63	0.18	0.3
$k_{m6Suc6P}$	Estimated	0.16	0.56	0.06	0.1
k_{i6F6P}	Measured	0.4	-	-	0.4
V_{max11}	Estimated	0.99	1.451	0.089	1
k_{m11Suc}	Estimated	99.59	102.48	96.70	100

Table 3.7: Final parameter values with confidence intervals after solving all non-identifiability problems. To achieve this, three non-identifiable parameters (k_{i4F6P}, V_{max6r} and k_{i6F6P}) were explicitly measured and the rest were estimated.



(a) (b)

Figure 3.9: Simulation of the dynamics states in sugarcane culm model. a) Simulation based on the estimated parameter values. b) Simulation based on the actual parameter values.

signalling pathway [Raue 2009]. In that application they analysed the model states x along the profile likelihood of the *practical* non-identifiable parameter p_3 to determine the measurement data needed to solve this *practical* non-identifiability. As species x_2 and species x_3 had large variations on their trajectory they measured the fraction of dimerized pSTAT(x_3) relative to total phosphorylated STAT($x_2 + x_3$) in cytoplasm.

Table 3.7 summarizes the final estimation results for all parameters together with their confidence intervals. The estimated values all closely approach the original value and all the original values lie within the confidence intervals of the estimation. This illustrates that after resolving parameter non-identifiability, CSUKF can estimate parameters with both high and low magnitude equally well. With the estimated parameter values in hand the dynamics of the sugarcane model states were simulated. The results of this simulation are illustrated in Figure 3.9a, while Figure 3.9b shows the same state dynamics generated using the original parameter values. As can be seen the dynamics in both cases show the same behaviour. They are presented separately as when plotted together the dynamics of the two cases cannot be discerned. Therefore it can be concluded that the framework estimated the parameters with high accuracy and the estimated parameter values represent the real dynamics of the model.

3.4 Using the informed prior

Despite the most sophisticated measurement techniques and powerful measurement devices developed to measure in-vivo biological data, efficient methods to measure

biochemical parameters are still limited [Maerkl 2007]. Even after immense development in the measurement devices of time series data, these datasets are most often noisy and incomplete due to the model complexity and the limitation of measurement techniques [Jia 2011]. Thus it may not be possible to always measure parameter values directly nor to have more data points in the time-series data, in order to solve parameter non-identifiability. To overcome this limitation, the Bayesian treatment is applied to non-identifiability. As the CSUKF is an extension of dynamic Bayesian inference, the CSUKF is used to provide this solution. Using this approach the CSUKF can uniquely estimate the parameter values even in the case of non-identifiability, provided that an informed prior distribution for the parameters has been defined. The CSUKF is based on a Gaussian distribution for both the model and prior pdf, which belongs to the exponential family, thus the conjugate prior distribution can be used to define this prior. This leads to the same form of the transformed pdf [Suzdaleva 2007]. An example of a similar treatment of non-identifiable parameters can be found in the work of Lindley and El-Sayyad where they used Bayesian inference to estimate parameters subject to a linear functional constraint [Lindley 1968]. In this thesis I first applied this approach to the previous sugarcane model and then to a gene regulatory network with 29 parameters to be estimated.

In this experiment the prior information into the distribution is introduced through the square-root of the covariance matrix for the initial state estimation matrix, V and the state noise covariance matrix, Q. Results from the ranking of the parameters are used to formulate this informed prior. Both the V and Q matrices are first initialized on the basis of the rank of the parameters. For V, the high ranking parameters are initialized with low standard deviations as they are more sensitive towards the model and can be determined more confidently, while the low ranking parameters were initialized with high standard deviations. Similarly the matrix Q was initialized with the square of these standard deviation values. The state variables are again initialized with small random numbers, between zero and one. Table 3.8 summarizes the parameter estimation results for all the 12 parameters of the sugarcane model using the informed prior distribution. All parameters were estimated with unique values where the highest standard deviation was for k_{i4F6P} having the value 1.16 which is 18% of its estimated mean value. However the estimation accuracy is not equally good for all the parameters as six parameters have relative percent error more than 100%. This result could be due to the fact that although Bayesian estimation can uniquely estimate parameter values, it does not guarantee the accuracy of the estimation, as explained by [Samaniego 2010]. However successful estimation with generally low standard deviations conceptually proves the hypothesis

Parameter Name	Original Value	Estimation	Std. Dev.
k_{i1Fru}	1.00	1.00	0.0100
k_{i2Glc}	1.00	1.00	0.0100
k_{i3G6P}	0.10	0.16	0.0080
k_{i4F6P}	10.00	6.26	1.1600
$k_{i6Suc6P}$	0.07	0.25	0.0010
$k_{i6UDPGlc}$	1.40	0.14	0.0005
V_{max6r}	2.00	0.07	0.0003
k_{m6UDP}	0.30	4.69	0.0550
$k_{m6Suc6P}$	0.10	3.49	0.0100
k_{i6F6P}	0.40	0.93	0.0050
V_{max11}	1.00	1.03	0.0200
k_{m11Suc}	100.00	104.64	2.1200

Table 3.8: Result of all the 12 parameter estimation using informed prior. 100 runs of the estimation was made to calculate the statistics.

that the framework can uniquely estimate parameters even in the presence of non-identifiability.

3.5 Gene regulatory network

To verify the applicability of the proposed framework to a wide variety of biological models, in particular where it is not possible to acquire additional measurements, the framework is applied to estimate the parameters of a gene regulatory network. The first model in Dream6 estimation of model parameters challenge is selected for this experiment [Prill 2010, Schneider 2011]. The purpose of this challenge was twofold, first to develop, improve and apply optimization methods for parameter estimation. Second to propose an efficient experimental design that is specifically suited for complex models with limited and noisy data. The schematic diagram of the model is given in Figure 3.10. The model assumes linear kinetics for mRNA degradation and protein synthesis and degradation. mRNA synthesis is typically modelled using Hill-type kinetics with one or two regulatory inputs. Each regulatory input works as an inhibitory or an activating input. A constant rate of transcription is assumed when there is no regulatory input to a gene. In the network depicted in Figure 3.10, lines connecting protein coding sequences with proteins represent a protein production process. For simplicity the two steps in protein production process, the transcription and translation, are not explicitly shown

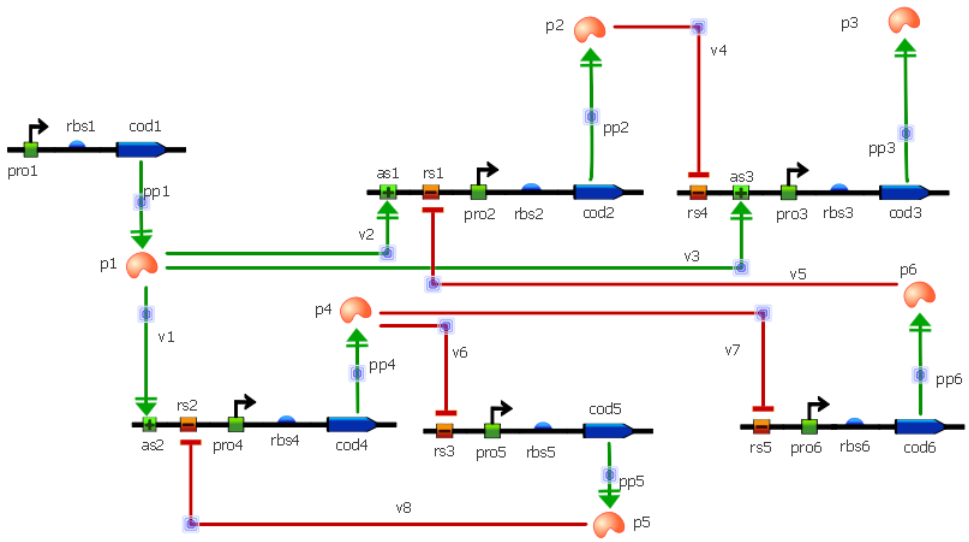


Figure 3.10: Schematic diagram of the gene regulatory network. Abbreviations are as follows, as: activator binding site; rs: repressor binding site; rbs: ribosomal binding site; prom: promoter; cod: protein coding sequence; prot: protein. The *v*'s denote the activation and repression reactions. Details of the rate laws can be found in the A.3

in the figure but they are included in the mathematical formulation. The transcription process is modelled using a single rate equation which is expressed as a sum of the transcription activity of all activators (as) in a specific promoter region multiplied by the product of the transcriptional activity of all the repressor (rs) binding sites in the same region. The rate of production of protein is given by linear rate equations multiplying the ribosomal strength with the transcription rate.

The contest provides a limited amount of microarray data which gives time-courses of all mRNAs as a start up data. In addition it also provides the network topologies and mathematical descriptions of the models at the start up. To reflect the actual scientific practice, additional experimental can be bought later on. Those are the time-course data for mRNA and proteins in response to different network perturbations namely gene deletion, siRNA-mediated knock-down and change of RBS activity.

The model had a total of 30 parameters 29 of which are to be estimated. Only the mRNA degradation rate constant is kept fixed to a value of one. Though the values of all protein degradation rate constants are identical they are unknown and consequently must be estimated. The remaining parameters, the two regulation process parameters, activation K_d and repression h, the promoter strength and the ribosomal binding site strength, have to be estimated.

The gene regulatory network involves a number of biological mechanisms for which measurement values are rarely available, such as chromatin, transcription and so on. Parameter estimation in the presence of such hidden variables is a difficult task due to the complex formulation of the objective function. CSUKF has the ability to jointly estimate parameters and the hidden variables or unobserved states in a gene regulatory network. In this experiment the only data available is the time-series data for all mRNAs and protein abundance for wild type and mutant with increase of RBS4 activity by 100%. The main focus in this experiment is to determine a unique solution utilizing the proper prior of the probability distribution, despite parameters found to be non-identifiable in the conventional sense.

Experimental setup The measurement data is provided over the interval from zero to 20 seconds with a step size of one second. The Dream6 organizers used a noise model $y_{noisy} = max[0, y + 0.1 \times r_1 + C \times r_2 \times y]$ to simulate this synthetic noisy measurement data, where y is the simulated value, r_1 and r_2 are Gaussian random variables with standard deviation of one and $C = 0.2$. The lower bound of the constraint was set at 10^{-8} and the upper bound was set at 100. The parameters are initialized to small random numbers between zero and one. Experiments with two phases are conducted for the estimation with informed prior. First the CSUKF is used to estimate parameters with the mutant data of high RBS4 activity, specifying the prior distribution of the model parameters based on their ranking. The point estimate and covariance matrix from this experiment is then used to form the informative prior for the final phase of estimation with the wild type data.

3.5.1 First phase experiment

With the RBS4 mutant data, the first phase of experiment is conduced which is again divided into two stages. In the first stage the ranking of the parameters are calculated using the method described in section 2.3.2. With no information of the pdf at the start of the experiment, the diagonal of both the state-estimation covariance matrix, P, and process noise covariance matrix, Q, are initialized with small random numbers between 0.001 to 0.1. The measurement noise covariance matrix R is initialized according to the noise model discussed in section 3.5. In the second stage this ranking is used to formulate the initial state estimation covariance matrix P and the process noise covariance matrix Q. Table 3.9 lists the ranking of the parameters and the corresponding standard deviations used to formulate the informed prior. In both the estimations the parameters are initialized to small random numbers as no information on the parameter values were available.

Parameter Name	Ranking	Initialized Std. Dev.
rbs4_strength	1	0.001
pro6_strength	2	0.002
v8_h	3	0.003
pro1_strength	4	0.004
pro5_strength	5	0.005
pro3_strength	6	0.006
pro2_strength	7	0.007
pro4_strength	8	0.008
rbs6_strength	9	0.009
rbs5_strength	10	0.01
v5_h	11	0.011
v7_Kd	12	0.012
v6_Kd	13	0.013
p_degradation_rate	14	0.014
v4_h	15	0.015
v7_h	16	0.016
rbs1_strength	17	0.017
rbs3_strength	18	0.018
v6_h	19	0.019
v8_Kd	20	0.02
v3_h	21	0.021
rbs2_strength	22	0.022
v5_Kd	23	0.023
v2_h	24	0.024
v1_Kd	25	0.025
v2_Kd	26	0.026
v3_Kd	27	0.027
v1_h	28	0.028
v4_Kd	29	0.029

Table 3.9: Ranking and standard deviation of the parameters used for formulating the informed prior. The standard deviations are used to initialize the diagonal entry of the matrix V and Q.

3.5.2 Second phase experiment

In the second phase of the experiment the informative prior is formulated from the information gathered in the first phase. The parameter values and the estimation covariance matrix calculated in the first phase of the parameter estimation is used to form the informative prior pdf for the second phase of the estimation. In this experiment the wildtype data is used as the measurement data. The parameters other than the ones involved in the increase of RBS4 activity (initialized to half of the mean value) are initialized to the values with small random perturbation around the mean values of the first phase experiment. The square root of the P matrix is initialized to the values of the final V matrix of the first phase estimation. The state error covariance matrix Q is initialized with the same matrix formulated with ranking used in the first phase. The measurement noise covariance matrix R is initialized according to the noise model as described in section 3.5. The experiment is repeated 50 times.

Table 3.10 summarizes the results of the parameter estimation performed both with and without the informed prior. The results present the mean and standard deviation from performing each estimation 50 times. It can be seen from this table that the estimated values reach closer to the actual value even after random initialization, indicating a higher estimation accuracy in the experiment when using the informed prior compared to the estimation accuracy without informed prior. This accuracy could be due to the use of two sets of data, the wild type and the mutant. However, this increase in accuracy is still not satisfactory enough as six parameters have percent relative error more than 100%. Huang *et al.* also reported such unsatisfactory fitting for some of the subjects in their Bayesian inference procedure with informed prior to study HIV dynamics [Huang 2006]. The estimation result conducted with informed priors could estimate parameters more concisely with low standard deviations (maximum standard deviation is of $pro4\_strength$ with 60% of the mean value) compared to the estimation with no informed priors (maximum standard deviation for three parameters go beyond 100% of the mean value). This low standard deviation indicates that the framework can uniquely estimate the parameter values of the model in the presence of informative prior. Huang *et al.* also had similar result of low standard deviation in the estimation [Huang 2006]. Therefore it can be summarized that the informed prior helps to uniquely estimate the parameter values of this model with higher accuracy. This illustrates the applicability of the proposed method of using informed prior for CSUFK to uniquely estimate parameters of a kinetic model. Additional work still needs to be done to increase the estimation accuracy, that is get as close as to the actual value which will decrease the

relative percent error. This would be one of the main foci of future work.

		With Informed Prior		Without Informed Prior	
Parameter Name	Actual value	Estimate	Std. Dev.	Estimate	Std. Dev.
p_degradation_rate	0.8	0.85	0.05	0.72	0.27
pro1_strength	3.0	3.04	0.05	2.94	0.15
pro2_strength	8.0	5.85	0.47	6.66	2.58
pro3_strength	6.0	7.12	0.68	9.14	4.43
pro4_strength	8.0	2.93	1.78	1.50	1.87
pro5_strength	3.0	3.03	0.07	3.46	0.80
pro6_strength	3.0	3.27	0.03	3.27	0.54
rbs1_strength	3.9	3.98	0.23	3.33	1.41
rbs2_strength	5.0	5.94	0.33	4.82	2.18
rbs3_strength	5.0	5.13	0.32	4.31	1.56
rbs4_strength	1.0	1.46	0.29	1.29	0.81
rbs5_strength	5.0	5.23	0.31	3.77	1.56
rbs6_strength	5.0	5.03	0.28	4.55	1.64
v1_Kd	1.0	1.54	0.18	1.40	1.62
v1_h	4.0	2.54	0.92	2.98	2.36
v2_Kd	1.0	1.87	0.15	1.17	0.74
v2_h	2.0	3.74	1.28	3.32	2.09
v3_Kd	0.1	0.56	0.18	0.61	0.31
v3_h	2.0	4.05	0.34	2.99	2.20
v4_Kd	10.0	8.04	1.12	7.17	3.10
v4_h	4.0	2.49	0.42	2.97	2.06
v5_Kd	1.0	2.22	0.41	2.16	1.52
v5_h	1.0	1.20	0.08	1.27	0.29
v6_Kd	0.1	0.28	0.02	0.64	0.57
v6_h	2.0	3.20	0.39	5.55	3.07
v7_Kd	0.1	0.26	0.02	0.48	0.28
v7_h	2.0	2.78	0.35	5.34	3.03
v8_Kd	0.2	0.41	0.30	2.14	2.40
v8_h	4.0	1.77	0.33	1.12	0.50

Table 3.10: Estimation results with and without the informed prior. The results are presented as the mean and standard deviation for each estimated parameters, calculated with 50 repetitions.

Chapter 4

Conclusion and outlook

Mathematical models of biological networks help to understand the functional and physical interactions between different parts of a cell. Although different kinds of mathematical models have been developed, detailed kinetic models have been the most effective, due to their ability to describe a more detailed and fine-grained picture of the mechanism of the network. These kinetic models heavily depend on accurate parameter values to correctly simulate the behaviour of the original system. However a lack of information on the parameter values from wet lab experiments derails the successful use of kinetics for biological models. This necessitates the use of efficient computational methods to estimate the missing parameter values. Numerous computational methods have been developed in recent times to successfully estimate these parameter values. Most of these estimation methods are based on minimizing the difference between the observed and the simulated quantities. Due to the non-convex nature of the parameter estimation problem for most biological models, the majority of conventional parameter estimation methods do not guarantee that an optimal solution is ever found and often fail to arrive at satisfactory solutions. The partially observable nature of biological systems further complicates the use of conventional methods for the estimation problem. Finally the non-identifiability issue of the parameters due to the structure of the model and complexity in achieving good measurement data, with respect to both quantity and quality, further inhibits the applicability of conventional methods. This thesis focuses on the development of a framework that can handle all of these complexities of parameter estimation, more efficiently than the conventional methods. This framework proposes a novel approach to parameter identification combining a newly proposed filtering technique with existing approaches for successful parameter estimation of the kinetic models. It considers both the system and measurement noises during the estimation which is an important feature for all biological models. As reasoning under uncertainty is essential in biological models, the proposed framework has the strength of addressing the parameter estimation problem through general probability theory. It also uses the power of representing partially observable dynamic system through state space models to facilitate the estimation of these parameter values. This framework consists of two

modules, one, the parameter estimation module and the other, the identifiability analysis module. The parameter estimation module consists of a novel filtering technique based on the extension of SQ-UKF, the constrained square-root unscented Kalman filter (CSUKF). The CSUKF provides an efficient, numerically stable filtering technique capable of making the estimation within a constrained parameter space. This method considers the noise in the system due to the uncertainty in the model and the noise in the measurement data due to the inaccuracy in the measurement method or device. The CSUKF applies a widely used concept in engineering for parameter estimation, to jointly estimate the states of the non-linear dynamic systems and the parameters. The efficiency of this method has been demonstrated with the state estimation of the Higgins-Sel'kov oscillator and parameter estimation of the upper part of glycolysis model. When applied to the state estimation of the Higgins-Sel'kov oscillator from the combined measurement data, the estimation converges quickly to the actual state value. During the estimation of the four parameter values of the glycolysis model CSUKF performs with an high accuracy with the extra benefit of ensuring numerical stability of the algorithm. CSUKF also showed independence on the initial starting values when tested with different random values between 0.01 and 20 in this experiment. The second module is the parameter identifiability analysis module. This module consists of a profile likelihood based identifiability analysis used to identify both the structural and practical non-identifiable parameters. This module also makes use of an orthogonal ranking method to identify the parameters having the highest influence on the model states. This sensitivity based ranking helps to find the parameters that need to be estimated with higher accuracy. It also includes methods for correlation and functional relationship identification in order to identify dependency relationships between parameters. To uniquely estimate these parameters these relationship must first be resolved. The integration of all these information helps in solving the non-identifiability of the parameters. This has been demonstrated for the parameters of the sugarcane culm model. Initially CSUKF was applied in this model to find out the parameters that are identifiable based on orthogonality and also to rank the parameters depending on their importance. These analysis found 3 out of 12 parameters to be non-identifiable. The thorough identifiability analysis using the profile likelihood based method confirmed this result by determining these three parameters to be practically non-identifiable. This profile likelihood method also identified three more parameters to be practically non-identifiable as well as two other parameters to be structurally non-identifiable. The measurement data of three parameters k_{i4F6P}, V_{max6r} and k_{i6F6P} as well as more data points in the time series of

Suc and $Fruc$ were considered to be available in order to solve these non-identifiability. However, the general ability to solve parameter non-identifiability is hampered if it is not possible to collect extra measurement data, which is often the case in biological models. In this case, an informed prior from Bayesian inference theory can be used to uniquely estimate the parameter values, even in the presence of non-identifiability. As CSUKF is considered to be a variant of Bayesian inference, this treatment of non-identifiability is used in conjunction with CSUKF for the unique estimation of the parameters in this framework. This informed prior is formulated with the help of the information from different parts of the identifiability analysis module. The performance of this estimation has been demonstrated by estimating the parameters from the sugarcane model and the gene regulatory model. When applied to the sugarcane model, it estimates ten parameters with low standard deviation ≤ 0.01 and two parameters with relatively high standard deviation (for k_{i4F6P}, 18% of its mean value and for k_{m11Suc}, 2% of its mean value). When applied to the gene regulatory model all the parameters were estimated with considerably low standard deviation. Having lower standard deviation for the parameters estimated using the CSUKF with informed prior indicates less variance in their estimation and thus conceptually validating the hypothesis that the parameters can be uniquely estimated using CSUKF when informed prior is used. However in both the experiments the estimation result using the informed prior does not reach equally well to the actual value for all the parameters. This indicates that further improvement needs to be made in this approach to make it more efficient.

The framework developed in this thesis would help to estimate the unknown parameter values of a model in a more efficient and robust way by solving the non-identifiability of the parameters through different approaches included in the framework. This work makes it possible to have more accurate biological models through identifying the parameter values within a biologically significant range at the same time considering the uncertainty in the model and data. It also includes an approach that uniquely estimates the parameters in cases where it is not possible to solve the non-identifiability. All these features make the framework robust enough to be used in large and sparse biological models.

While conducting extensive research in the field of parameter identifiability and estimation, a number of issues related to this research arose. Although not central to the research conducted in this thesis, answering these open questions might provide a breakthrough in their respective fields. A brief summary of the most interesting questions are presented.

- **Flux calculation with dynamic labelling:** The framework proposed in this thesis could be applied to other fields of biological models such as determining fluxes from ^{13}C non-stationary labelling data. Dynamic ^{13}C metabolic flux analysis, represents biological systems with ODEs for isotopic transient phases and produces time-series data. This framework fits perfectly with the flux calculation in these systems. At the same time this framework would also help to solve one of the critical bottlenecks of such isotopic non-stationary MFA which is the the flux non-identifiability problem. It can also be used to determine accurate flux confidence intervals which would make it easier to interpret the flux results for physiological significance. More generally the framework can be used for any estimation purposes where time-series data is available.

- **Adaptive CSUKF:** One of the aspects not directly addressed in this thesis, is the initialization of the process and measurement noise distribution. In the thesis they are mostly initialized from knowledge of the model or calculated from the noise model of the measurement data. However precise knowledge on this noise distributions specially for the process noise Q would help in the accuracy of the estimation, specifically formulating the informed prior for uniquely estimating the parameters. Adaptive methods can estimate noise covariance at the same time as estimating the parameter values, which would automatically tune the filter parameters to match the original statistics. This adaptive method might be formulated in fuzzy logic, by modifying the maximum-likelihood principle to approximate the error covariance matrix or on a master-slave basis, where the master works as a regular UKF to estimate states or parameters and the slave estimates the diagonal of the noise covariance matrix for the master.

- **CSUKF with non-Gaussian posteriors:** Extending this framework for modelling the posterior distribution with heavily tailed density functions, such as the student-t distribution, would make the framework applicable to a large number of applications. Currently the CSUKF only propagates the mean and covariance of the Gaussian posterior and updates on the Kalman filter. Propagating heavy tailed distribution is not straight forward in CSUKF as only the mean and covariance would not propagate this distribution. It would also need to propagate the fourth order moments, which would require a modifications in the sigma points.

- **Scaling parameters:** In this thesis the CSUKF was not tested with different values of the scaling parameters. There are some guidelines provided in the literature for

the values of the scaling parameter which were used in this thesis but no definitive criterion was set to judge their values. These scaling parameters can also be made adaptive to adjust in conjunction with the parameter estimation.

- **Enhancing accuracy when used with informed prior:** When it is not possible to have more measurement data either qualitatively or quantitatively then the framework tries to determine a unique estimation of parameter values by using the informed prior. However, when using the informed prior, the framework focuses on providing a unique estimation but not on the accuracy of the estimation. As a result, in some cases the estimation accuracy drops for some of the parameters as shown in Table 3.8 and 3.10. Extensive research on improving the accuracy while uniquely estimating parameters in the presence of non-identifiability is needed to make this framework more efficient.

- **Use in synthetic biology:** Synthetic biology aims at engineering biological systems for specific biotechnological or biomedical purposes. As synthetic biology shares some of the common problems of systems biology, the proposed framework can be used to engineer synthetic biological systems by inferring both structural and parametric information.

The above lists provide a portion of the most promising research opportunities on this parameter estimation framework in terms of their potential benefit to the research community. Part of it is already planned to carry out in the existing group, specifically to apply this framework in the dynamic tracer experiments. With small modifications the complete framework can be applied to various fields solving a large paradigm of problems in and out of biology.

Appendix A

Appendix

A.1 Rate law of the Sugarcane culm model

Rate laws used in the Sugarcane calm model, developed as Rohwer *et al.* [Rohwer 2001]

$$v_1 = V_{max1}\frac{[Fru_{ex}]}{k_{m1Fru,ex}\left(1+\frac{[Fru]}{k_{i1Fru}}\right)+[Fru_{ex}]}$$

$$v_2 = V_{max2}\frac{[Glc_{ex}]}{k_{m2Glc,ex}\left(1+\frac{[Glc]}{k_{i2Glc}}\right)+[Glc_{ex}]}$$

$$v_3 = V_{max3}\frac{\frac{[Glc]}{k_{m3Glc}}\frac{[ATP]}{k_{m3ATP}}}{\left(1+\frac{[ATP]}{k_{m3ATP}}\right)\left(1+\frac{[Glc]}{k_{m3Glc}}+\frac{[Fru]}{k_{m4Fru}}+\frac{0.113\times[HexP]}{k_{i3Glc6P}}+\frac{0.0575\times[HexP]}{k_{i4Fru6P}}\right)}$$

$$v_4 = V_{max4}\frac{\frac{[Fru]}{k_{m4Fru}}\frac{[ATP]}{k_{m4ATP}}}{\left(1+\frac{[ATP]}{k_{m4ATP}}\right)\left(1+\frac{[Glc]}{k_{m3Glc}}+\frac{[Fru]}{k_{m4Fru}}+\frac{0.113\times[HexP]}{k_{i3Glc6P}}+\frac{0.0575\times[HexP]}{k_{i4Fru6P}}\right)}$$

$$v_5 = \left(\frac{V_{max5}}{1 + \frac{[Fru]}{k_{i5Fru}}}\right)\left(\frac{\frac{[Fru]}{k_{m5Fru}}\frac{[ATP]}{k_{m5ATP}}}{1 + \frac{[Fru]}{k_{m5Fru}} + \frac{[ATP]}{k_{m5ATP}} + \frac{[Fru][ATP]}{k_{m5Fru}K_{m5ATP}} + \frac{[ADP]}{k_{i5ADP}}}\right)$$

$$v_6 = V_6^f \frac{0.0575 \times [HexP] \times 0.8231 \times [HexP] - \frac{[Suc][UDP]}{k_{eq6}}}{\begin{array}{l} 0.0575 \times [HexP] \times 0.8231 \times [HexP]\left(1 + \frac{[Suc6P]}{k_{i6Suc6P}}\right) + k_{m6Fru6P}\left(1 + \frac{[P]}{k_{i6P}}\right)(0.8231 \times [HexP] + k_{i6UDPGlc}) + k_{m6UDPGlc} \times 0.0575 \times [HexP] + \frac{V_6^f}{V_6^r k_{eq6}} \\ \left([Suc6P] \times k_{m6UDP}\left(1 + \frac{0.8231 \times [HexP]}{k_{i6UDPGlc}}\right) + [UDP] \times k_{m6Suc6P}\left(1 + \frac{k_{m6UDPGlc} \times 0.0575 \times [HexP]}{k_{i6UDPGlc} \times k_{m6Fru6P}\left(1 + \frac{[P]}{k_{iP}}\right)}\right) + [Suc6P]\left(1 + \frac{0.0575 \times [HexP]}{k_{i6Fru6P}}\right)\right) \end{array}}$$

$$v_7 = V_{max7}\frac{[Suc6P]}{k_{m7Suc6P} + [Suc6P]}$$

$$v_8 = -V_8^f \frac{[Suc][UDP] - \frac{[Fru] \times 0.8231 \times [HexP]}{k_{eq8}}}{[Suc][UDP]\left(1 + \frac{[Fru]}{k_{i8Fru}}\right) + k_{m8Suc}\left([UDP] + k_{i8UDP}\right) + k_{m8UDP}[Suc] + \frac{V_8^f}{V_8^r k_{eq8}}\left([Fru]k_{m8UDPGlc}(1 + \frac{[UDP]}{k_{i8UDP}})\right) + 0.8231 \times [HexP]\left(k_{m8Fru}(1 + \frac{k_{m8UDP}[Suc]}{k_{i7UDP}}k_{m8Suc})\right) + [Fru]\left(1 + \frac{[Suc]}{k_{i8Suc}}\right)}$$

$$v_9 = \left(\frac{V_{max9}}{1 + \frac{[Glc]}{k_{i9Glc}}}\right)\left(\frac{[Suc]}{k_{m9Suc}\left(1 + \frac{[Fur]}{k_{i9Fru}}\right) + [Suc]}\right)$$

$$v_{10} = V_{max10}\frac{0.0575 \times [HexP]}{k_{m10Fru6P} + 0.0575 \times [HexP]}$$

$$v_{11} = V_{max11} \frac{[Suc]}{k_{m11Suc} + [Suc]}$$

The ODEs are:

$$\frac{d[Glc]}{dt} = v_2 - v_3 + v_9$$

$$\frac{d[Fru]}{dt} = v_1 - v_4 - v_5 - v_8 + v_9$$

$$\frac{d[HexP]}{dt} = v_3 + v_4 + v_5 - 2v_6 - v_8 - v_{10}$$

$$\frac{d[Suc6P]}{dt} = v_6 - v_7$$

$$\frac{d[Suc]}{dt} = v_7 + v_8 - v_9 - v_{11}$$

A.2 Rate law of the Glycolysis model

Rate laws used in Glycolysis model, proposed in [Klipp 2005]

$$v_1 = \frac{V_{max,1}[ATP]}{K_{1ATP}}$$

$$v_2 = k_2[ATP][Gluc6P]$$

$$v_3 = \frac{\frac{V^f_{max,3}}{k_{Gluc6P,3}}[Gluc6P] - \frac{V^r_{max,3}}{k_{Fruc6P,3}}[Fruc6P]}{1 + \frac{[Gluc6P]}{k_{Gluc6P,3}} + \frac{[Fruc6P]}{k_{Fruc6P,3}}}$$

$$v_4 = \frac{V_{max,4}([Fruc6P])^2}{k_{Fruc6P,4}\left(1 + k\left(\frac{[ATP]}{[ADP]}\right)\right) + ([Fruc6P])^2}$$

$$v_5 = k_5[Fruc1,6P_2]$$

$$v_8 = k_{8f}[ATP][AMP] - k_{8r}([ADP])^2$$

The ODEs are:

$$\frac{d[Gluc6P]}{dt} = v_1 - v_2 - v_3$$

$$\frac{d[Fruc6P]}{dt} = v_3 - v_4$$

$$\frac{d[Fruc1,6P_2]}{dt} = v_4 - v_5$$

$$\frac{d[ATP]}{dt} = -v_1 - v_2 - v_4 + v_6 - v_7 - v_8$$

$$\frac{d[ADP]}{dt} = v_1 + v_2 + v_4 - v_6 + v_7 + 2v_8$$

$$\frac{d[AMP]}{dt} = -v_8$$

A.3 Rate law of the Gene regulatory network

$$as_1 = \frac{(\frac{p1}{v2\_kd})^{v2\_h}}{(1 + \frac{p1}{v2\_kd})^{v2\_h}}$$

$$as_2 = \frac{(\frac{p1}{v1\_kd})^{v1\_h}}{(1 + \frac{p1}{v1\_kd})^{v1\_h}}$$

$$as_3 = \frac{(\frac{p1}{v3\_kd})^{v3\_h}}{(1 + \frac{p1}{v3\_kd})^{v3\_h}}$$

$$rs_1 = \frac{1}{1 + (\frac{p6}{v5\_kd})^{v5\_h}}$$

$$rs_2 = \frac{1}{1 + (\frac{p5}{v8\_kd})^{v8\_h}}$$

$$rs_3 = \frac{1}{1 + (\frac{p4}{v6\_kd})^{v6\_h}}$$

$$rs_4 = \frac{1}{1 + (\frac{p2}{v4\_kd})^{v4\_h}}$$

$$rs_5 = \frac{1}{1 + (\frac{p4}{v7\_kd})^{v7\_h}}$$

$$r_1 = pro1\_strength$$

$$r_2 = pp1\_mrna\_degradation\_rate * pp1\_mrna$$

$$r_3 = rbs1\_strength * pp1\_mrna$$

$$r_4 = p1\_degradation\_rate * p1$$

$$r_5 = pro2\_strength * ((as_1) * (rs_1))$$

$$r_6 = pp2\_mrna\_degradation\_rate * pp2\_mrna$$

$$r_7 = rbs2\_strength * pp2\_mrna$$

$$r_8 = p2\_degradation\_rate * p2$$

$$r_9 = pro3\_strength * ((as_3) * (rs_4))$$

$r_{10} = pp3\_mrna\_degradation\_rate * pp3\_mrna$

$r_{11} = rbs3\_strength * pp3\_mrna$

$r_{12} = p3\_degradation\_rate * p3$

$r_{13} = pro4\_strength * ((as_2) * (rs_2))$

$r_{14} = pp4\_mrna\_degradation\_rate * pp4\_mrna$

$r_{15} = rbs4\_strength * pp4\_mrna$

$r_{16} = p4\_degradation\_rate * p4$

$r_{17} = pro5\_strength * (rs3)$

$r_{18} = pp5\_mrna\_degradation\_rate * pp5\_mrna$

$r_{19} = rbs5\_strength * pp5\_mrna$

$r_{20} = p5\_degradation\_rate * p5$

$r_{21} = pro6\_strength * (rs5)$

$r_{22} = pp6\_mrna\_degradation\_rate * pp6\_mrna$

$r_{23} = rbs6\_strength * pp6\_mrna$

$r_{24} = p6\_degradation\_rate * p6$

The naming of the variables are, as: activator binding site, rs: repression binding site, rbs: ribosomal binding site strength, pro: promoter strength, p: protein concentration, pp: protein production process. The ODEs of the model are:

$$\frac{d(pp1\_mrna)}{dt} = r_1 - r_2$$

$$\frac{(dp1)}{dt} = r_3 - r_4$$

$$\frac{d(pp2\_mrna)}{dt} = r_5 - r_6$$

$$\frac{d(p2)}{dt} = r_7 - r_8$$

$$\frac{d(pp3\_mrna)}{dt} = r_9 - r_{10}$$

$$\frac{d(p3)}{dt} = r_{11} - r_{12}$$

$$\frac{d(pp4\_mrna)}{dt} = r_{13} - r_{14}$$

$$\frac{d(p4)}{dt} = r_{15} - r_{16}$$

$$\frac{d(pp5\_mrna)}{dt} = r_{17} - r_{18}$$

$$\frac{d(p5)}{dt} = r_{19} - r_{20}$$

$$\frac{d(pp6\_mrna)}{dt} = r_{21} - r_{22}$$

$$\frac{d(p6)}{dt} = r_{23} - r_{24}$$

Bibliography

[Ashyraliyev 2009] Maksat Ashyraliyev, Yves Fomekong-Nanfack, Jaap A. Kaandorp and Joke G. Blom. *Systems biology: parameter estimation for biochemical models*. FEBS Journal, vol. 276, no. 4, pages 886–902, 2009. (Cited on pages 16, 17, 18 and 19.)

[Baker 2001] Rose D. Baker. *A methodology for sensitivity analysis of models fitted to data using statistical methods*. IMA Journal of Management Mathematics, vol. 12, no. 1, pages 23–39, 2001. (Cited on page 65.)

[Baker 2010] Syed Murtuza Baker, Kai Schallau and Björn H. Junker. *Comparison of different algorithms for simultaneous estimation of multiple parameters in kinetic metabolic models*. Journal of Integrative Bioinformatics, vol. 7, no. 3, page e133, 2010. (Cited on pages 18 and 75.)

[Balsa-Canto 2008] Eva Balsa-Canto, Martin Peifer, Julio Banga, Jens Timmer and Christian Fleck. *Hybrid optimization method with general switching strategy for parameter estimation*. BMC Systems Biology, vol. 2, no. 1, page 26, 2008. (Cited on pages 16, 17 and 22.)

[Balsa-Canto 2010] Eva Balsa-Canto, Antonio Alonso and Julio Banga. *An iterative identification procedure for dynamic modeling of biochemical networks*. BMC Systems Biology, vol. 4, no. 1, page 11, 2010. (Cited on pages 31 and 64.)

[Banchs 1997] Rafael E. Banchs. *Simulated Annealing*. Rapport technique, Time Harmonic Field Electric Logging, The University of Texas at Austin, 1997. (Cited on page 75.)

[Banga 2003] J Banga, C Moles and A Alonso. *Global optimization of bioprocesses using stochastic and hybrid methods*. Nonconvex Optimization and Its Applications. Frontiers In Global Optimization, vol. 74, pages 45–70, 2003. (Cited on pages 20 and 22.)

[Barabasi 2004] Albert-Laszlo Barabasi and Zoltan N. Oltvai. *Network biology: understanding the cell's functional organization*. Nature Reviews Genetics, vol. 5, no. 2, pages 101–113, February 2004. (Cited on page 8.)

[Berit 2005] Lund Berit Floor, E. Berntsen Hans and A. Foss Bjarne. *Methods for parameter ranking in nonlinear, mechanistic models*. In IFAC World Congress, 2005. (Cited on page 86.)

[Berit 2008] Lund Berit Floor and A. Foss Bjarne. *Parameter ranking by orthogonalization Applied to nonlinear mechanistic models*. Automatica, vol. 44, pages 278–281, 2008. (Cited on page 83.)

[Bray 2003] Dennis Bray. *Molecular Networks: The Top-Down View*. Science, vol. 301, no. 5641, pages 1864–1865, 2003. (Cited on page 1.)

[Breiman 1985] Leo Breiman and Jerome H. Friedman. *Estimating Optimal Transformations for Multiple Regression and Correlation*. Journal of the American Statistical Association, vol. 80, no. 391, pages 580–598, 1985. (Cited on page 66.)

[Brennan 2011] Conor Brennan. *Section One: Numerical Differentiation*. University Lecture, 2011. (Cited on page 58.)

[Brockmann 2008] D. Brockmann, K.-H. Rosenwinkel and E. Morgenroth. *Practical identifiability of biokinetic parameters of a model describing two-step nitrification in biofilms*. Biotechnology and Bioengineering, vol. 101, no. 3, pages 497–514, 2008. (Cited on page 33.)

[Chen 2003] Zhe Chen. *Bayesian Filtering: From Kalman Filters to Particle Filters, and Beyond*. Rapport technique, McMaster University, 2003. (Cited on page 68.)

[Chen 2010] William W. Chen, Mario Niepel and Peter K. Sorger. *Classic and contemporary approaches to modeling biochemical reactions*. Genes & Development, vol. 24, no. 17, pages 1861–1875, 2010. (Cited on page 72.)

[Chiong 2007] Raymond Chiong and Ooi Koon Beng. *A Comparison between Genetic Algorithms and Evolutionary Programming based on Cutting Stock Problem*. Engineering Letters, pages 72–77, 2007. (Cited on page 75.)

[Chis 2011a] Oana-Teodora Chis, Julio R. Banga and Eva Balsa-Canto. *GenSSI: a software toolbox for structural identifiability analysis of biological models*. Bioinformatics, vol. 27, no. 18, pages 2610–2611, 2011. (Cited on page 89.)

[Chis 2011b] Oana-Teodora Chis, Julio R. Banga and Eva Balsa-Canto. *Structural Identifiability of Systems Biology Models: A Critical Comparison of Methods*. PLoS ONE, vol. 6, no. 11, page e27755, 11 2011. (Cited on pages 33 and 40.)

[Chou 2008] I-Chun Chou. *Parameter Estimation and Network Identification in Metabolic Pathway Systems*. PhD thesis, Georgia Institute of Technology, 2008. (Cited on pages 7, 8, 9 and 17.)

[Cobelli 1980] C. Cobelli and J. J. DiStefano 3rd. *Parameter and structural identifiability concepts and ambiguities: a critical review and analysis*. AJP - Regu Physiol, vol. 239, no. 1, pages R7–R24, 1980. (Cited on pages 31 and 37.)

[dÁlchè Buc 2010] Florence dÁlchè Buc and Nicolas Brunel. Computational systems biology: Data driven inference of parameters and network structure. MIT Press, 2010. (Cited on pages 15 and 17.)

[Denis-Vidal 2001] Lilianne Denis-Vidal, Ghislaine Joly-Blanchard and Céline Noiret. *Some effective approaches to check the identifiability of uncontrolled nonlinear systems*. Mathematics and Computers in Simulation, vol. 57, no. 1-2, pages 35 – 44, 2001. (Cited on page 31.)

[Denis 2003] Dochain Denis. *State and parameter estimation in chemical and biochemical processes: a tutorial*. Journal of Process Control, vol. 13, no. 8, pages 801 – 818, 2003. (Cited on page 23.)

[Diego 2010] Diego, Suárez Cecilia, Cecchi Guillermo A., Marshall Guillermo and Stolovitzky Gustavo Fernández Slezak. *When the Optimal Is Not the Best: Parameter Estimation in Complex Biological Models*. PLoS ONE, vol. 5, no. 10, page e13283, 10 2010. (Cited on pages 16 and 17.)

[Eberhart 1995] R. Eberhart and J. Kennedy. *A new optimizer using particle swarm theory*. pages 39–43, 1995. (Cited on page 22.)

[Eiben 2008] A. E. Eiben and J. E. Smith. Introduction to evolutionary computing (natural computing series). Springer, October 2008. (Cited on page 21.)

[Elizabeth 2004] Dolan Elizabeth D, Moré Jorge J. and Munson Todd S. *Benchmarking Optimization Software with COPS 3.0*. Technical report, Office of Science, U.S. Department of Energy, 2004. (Cited on page 76.)

[Faller 2003] D. Faller, U. Klingmüller and J. Timmer. *Simulation Methods for Optimal Experimental Design in Systems Biology*. SIMULATION, vol. 79, no. 12, pages 717–725, 2003. (Cited on page 91.)

[Farina 2006] Marcello Farina, Rolf Findeisen, Eric Bullinger, Sergio Bittanti, Frank Allgöwer and Peter Wellstead. *Results Towards Identifiability Properties of Biochemical Reaction Networks*. In Proceedings of the 46th IEEE Conference on Decision and Control, pages 2104–2109, CA, USA, December 2006. (Cited on page 87.)

[Fisher 1922] R.A. Fisher. *On the Mathematical Foundations of Theoretical Statistics*. Philosophical Transactions of the Royal Society of London. Series A, vol. 222, no. 2, pages 309–368, 1922. (Cited on page 62.)

[Fogel 1995] David B. Fogel. *A Comparison of Evolutionary Programming and Genetic Algorithms on Selected Constrained Optimization Problems*. SIMULATION, vol. 64, no. 6, pages 397–404, 1995. (Cited on page 75.)

[Geffen 2008] Dara Geffen. Parameter identifiability of biochemical reaction networks in systems biology. Master's thesis, Queen's University, Kingston, Ontario, Canada, 2008. (Cited on page 31.)

[Guedj 2007] Jérémie Guedj, Rodolphe Thiébaut1 and Daniel Commenges. *Practical identifiability of HIV dynamics models*. Bulletin of Mathematical Biology, vol. 69, no. 8, pages 2493–2513, 2007. (Cited on pages 33 and 91.)

[Hakami 2003] A Hakami, MT Odman and AG Russell. *High-order, direct sensitivity analysis of multidimensional air quality models*. Environmental science and technology, vol. 37, no. 11, pages 2442–2452, 2003. (Cited on page 57.)

[Hart 1996] William E. Hart. *A Theoretical Comparison of Evolutionary Algorithms and Simulated Annealing*. In 5^{th} Annual Conference on Evolutionary Programming, pages 147–154. MIT press, 1996. (Cited on page 75.)

[Hedar 2002] A. R. Hedar and M. Fukushima. *Hybrid Simulated Annealing and Direct Search Method for Nonlinear Unconstrained Global Optimization*. Optimization Methods and Software, vol. 17, no. 5, pages 891–912, 2002. (Cited on page 22.)

[Heinrich 1996] Reinhart Heinrich and Stefan Schuster. The regulation of cellular systems. Springer, August 1996. (Cited on page 8.)

[Hengl 2007] S. Hengl, C. Kreutz, J. Timmer and T. Maiwald. *Data-based identifiability analysis of non-linear dynamical models*. Bioinformatics, vol. 23, no. 19, pages 2612–2618, 2007. (Cited on pages 31, 32, 66, 67, 87, 89 and 90.)

[Hess 1979] Benno Hess. *The Glycolytic Oscillator*. Journal of Experimental Biology, vol. 81, pages 7–14, june 1979. (Cited on page 3.)

[Hewlett 2007] Joel Hewlett, Bogdan Wilamowski and Günhan Dündar. *Merge of Evolutionary Computation with Gradient Based Method for Optimization Problems*. IEEE International Symposium on Industrial Electronics, pages 3304–3309, 2007. (Cited on page 75.)

[Holmes 2004] Elizabeth E. Holmes. *Beyond theory to application and evaluation: diffusion approximations for population viability analysis*. Ecological Applications, vol. 14, pages 1272 – 1292, 2004. (Cited on page 61.)

[Hoops 2006] Stefan Hoops, Sven Sahle, Ralph Gauges, Christine Lee, Jürgen Pahle, Natalia Simus, Mudita Singhal, Liang Xu, Pedro Mendes and Ursula Kummer. *COPASI – a COmplex PAthway SImulator*. Bioinformatics, vol. 22, no. 24, pages 3067–3074, 2006. (Cited on page 72.)

[Hoos 2004] Holger H. Hoos and Thomas Stützle. Stochastic local search foundations and applications. Morgan Kaufmann, 1 édition, 2004. (Cited on page 20.)

[Huang 2006] Yangxin Huang, Dacheng Liu and Hulin Wu. *Hierarchical Bayesian Methods for Estimation of Parameters in a Longitudinal HIV Dynamic System*. Biometrics, vol. 62, no. 2, pages 413–423, 2006. (Cited on page 99.)

[Hynne 2001] F. Hynne, S. Danø and P. G. Sørensen. *Full-scale model of glycolysis in Saccharomyces cerevisiae*. Biophysical Chemistry, vol. 94, no. 1-2, pages 121–163, 2001. (Cited on pages 71 and 72.)

[Ingram 2006] Piers Ingram, Michael Stumpf and Jaroslav Stark. *Network motifs: structure does not determine function*. BMC Genomics, vol. 7, no. 1, page 108, 2006. (Cited on page 3.)

[Jamshidi 2009] Neema Jamshidi and Bernhard Palsson. *Using in silico models to simulate dual perturbation experiments: procedure development and interpretation of outcomes*. BMC Systems Biology, vol. 3, no. 1, page 44, 2009. (Cited on page 11.)

[Jazwinski 1970] Andrew H. Jazwinski. Stochastic processes and filtering theory. Academic Press, April 1970. (Cited on page 23.)

[Jia 2011] Gengjie Jia, Gregory N. Stephanopoulos and Rudiyanto Gunawan. *Parameter estimation of kinetic models from metabolic profiles: two-phase dynamic decoupling method*. Bioinformatics, vol. 27, no. 14, pages 1964–1970, 2011. (Cited on pages 17 and 94.)

[Jones 2006] Karl O. Jones. *Comparison of Genetic Algorithms and Particle Swarm Optimization for fermentation feed profile determination*. In International Conference on Computer Systems and Technologies - CompSysTech'2006, pages 2–7, 2006. (Cited on page 76.)

[Julier 1995] S. J. Julier, J. K. Uhlmann and H. F. Durrant-Whyte. *A new approach for filtering nonlinear systems*. Proceedings of the American Control Conference, vol. 3, pages 1628–1632, 1995. (Cited on page 28.)

[Julier 1996] Simon Julier and Jeffrey K. Uhlmann. *A General Method for Approximating Nonlinear Transformations of Probability Distributions*. Rapport technique, Robotics Research Group, Dept of Engineering Science, University of Oxford, 1996. (Cited on pages 28 and 46.)

[Julier 1997] Simon J. Julier and Jeffrey K. Uhlmann. *A New Extension of the Kalman Filter to Nonlinear Systems*. In International Symposium on Aerospace/Defense Sensing, Simulation and Controls, pages 182–193, FL, USA, 1997. (Cited on pages 23, 25, 26, 27 and 28.)

[Julier 2000] Simon Julier, Jeffrey Uhlmann and Hugh F. Durrant-Whyte. *A New Method for the Nonlinear Transformation of Means and Covariances in Filters and Estimators*. IEEE transactions on Automatic Control, vol. 45, no. 3, 2000. (Cited on page 28.)

[Julier 2002] Simon J. Julier and Idak Industries. *The scaled unscented transformation*. In Proceedings of IEEE American Control Conference, pages 4555–4559, 2002. (Cited on pages 29 and 45.)

[Julier 2004] S. J. Julier and J. K. Uhlmann. *Unscented filtering and nonlinear estimation*. Proceedings of the IEEE, vol. 92, no. 3, pages 401–422, 2004. (Cited on pages 4, 25, 26 and 28.)

[Kaj 2004] Madsen Kaj, Nielsen Hans Bruun and Tingleff Ole. *Methods for non-linear least squares problems*. Lecture note, Technical University of Denmark, 2004. (Cited on page 19.)

[Kalman 1960] R. E. Kalman. *A New Approach to Linear Filtering and Prediction Problems*. Transactions of the ASME - Journal of Basic Engineering, no. 82 (Series D), pages 35–45, 1960. (Cited on page 26.)

[Kandepu 2007] Rambabu Kandepu, Bjarne A. Foss and Lars Imsland. *Applying the unscented Kalman filter for nonlinear state estimation*. Journal of Process Control, vol. 18, no. 7-8, pages 753–768, 2007. (Cited on pages 25 and 81.)

[Kandepu 2008] Rambabu Kandepu, Lars Imsland and Bjarne A. Foss. *Constrained State Estimation Using the Unscented Kalman Filter*. In 16th Mediterranean Conference on Control and Automation, pages 1453 – 1458, Congress Centre, Ajaccio, France, June 2008. (Cited on page 42.)

[Karl 2010] Hedrick Karl. *Control of Nonlinear Dynamic Systems: Theory and Applications*. University Lecture, 2010. (Cited on page 33.)

[Kay 1993] Steven M. Kay. Fundamentals of Statistical Signal Processing: Estimation Theory. Prentice Hall, Englewood Cliffs, NJ, 1993. (Cited on page 66.)

[Kirkpatrick 1983] S. Kirkpatrick, C. D. Gelatt and M. P. Vecchi. *Optimization by Simulated Annealing*. Science, vol. 220, no. 4598, pages 671–680, 1983. (Cited on page 21.)

[Kitano 2001] Hiroaki Kitano. Foundations of systems biology. MIT Press, October 2001. (Cited on page 1.)

[Kitano 2002a] Hiroaki Kitano. *Computational systems biology*. Nature, vol. 420, no. 6912, pages 206–210, November 2002. (Cited on page 37.)

[Kitano 2002b] Hiroaki Kitano. *Systems Biology: A Brief Overview*. Science, vol. 295, no. 5560, pages 1662–1664, 2002. (Cited on page 1.)

[Klipp 2005] Edda Klipp, Ralf Herwig, Axel Kowald, Christoph Wierling and Hans Lehrach. Systems biology in practice: Concepts, implementation and application. Wiley-VCH, May 2005. (Cited on pages 1, 7, 71, 72 and 108.)

[Koyama 2010] Shinsuke Koyama, Lucia Castellanos Pèrez-Bolde, Cosma Rohilla Shalizi and Robert E. Kass. *Approximate Methods for State-Space Models*. Journal of the American Statistical Association, vol. 105, no. 489, pages 170–180, 2010. (Cited on page 25.)

[Leven 2004] William F. Leven and Aaron D. Lanterman. *Multiple Target Tracking with Symmetric Measurement Equations Using Unscented Kalman and Particle Filters*. In Proceedings of the 36^{th} Southeastern Symposium on System Theory, pages 195–199, 2004. (Cited on page 4.)

[Li 2004] Peihua Li, Tianwen Zhang and Bo Ma. *Unscented Kalman filter for visual curve tracking*. Image and Vision Computing, vol. 22, no. 2, pages 157 – 164, 2004. (Cited on page 34.)

[Liebermeister 2006] Wolfram Liebermeister and Edda Klipp. *Bringing metabolic networks to life: integration of kinetic, metabolic, and proteomic data*. Theoretical Biology and Medical Modelling, vol. 3, no. 1, page 42, 2006. (Cited on page 16.)

[Lillacci 2010] Gabriele Lillacci and Mustafa Khammash. *Parameter Estimation and Model Selection in Computational Biology*. PLoS Comput Biol, vol. 6, no. 3, page e1000696, 03 2010. (Cited on pages 14, 17, 22, 23, 32 and 87.)

[Lindley 1968] V. Lindley D. and M. El-Sayyad G. *The Bayesian Estimation of a Linear Functional Relationships*. Journal of the Royal Statistical Society, vol. 30, no. 1, pages 190–202, 1968. (Cited on page 94.)

[Ljung 1994] Lennart Ljung and Torkel Glad. *On global identifiability for arbitrary model parametrizations*. Automatica, vol. 30, no. 2, pages 265 – 276, 1994. (Cited on pages 31 and 32.)

[Madhumita 2010] Mamata Madhumita. Study of kalman, extended kalman and unscented kalman filter (an approach to design a power system harmonic estimator). Master's thesis, Dept. of Electrical Engineering, National Institute of Technology, Rourkela, 2010. (Cited on page 81.)

[Maerkl 2007] Sebastian J. Maerkl and Stephen R. Quake. *A Systems Approach to Measuring the Binding Energy Landscapes of Transcription Factors*. Science, vol. 315, no. 5809, pages 233–237, 2007. (Cited on page 94.)

[Malcolm 2009] James G. Malcolm, Martha E. Shenton and Yogesh Rathi. *Y.: Neural tractography using an unscented Kalman filter*. In Information Processing in Medical Imaging (IPMI), pages 126–138, 2009. (Cited on page 34.)

[Marquardt 1963] Donald W. Marquardt. *An algorithm for least-squares estimation of nonlinear parameters*. SIAM Journal on Applied Mathematics, vol. 11, no. 2, pages 431–441, 1963. (Cited on page 20.)

[McAuley 2010] Kim B. McAuley, Shaohua Wu and Thomas J. Harris. *Selecting Parameters to Estimate to Obtain the Best Model Predictions*. In Proceedings of The 2010 International Conference on Modelling, Identification and Control (ICMIC), pages 161–166, Okayama, Japan, August 2010. (Cited on pages 56 and 84.)

[Mehmood 2011] Tahir Mehmood, Harald Martens, Solve Saebo, Jonas Warringer and Lars Snipen. *Mining for genotype-phenotype relations in Saccharomyces using partial least squares*. BMC Bioinformatics, vol. 12, no. 1, page 318, 2011. (Cited on page 11.)

[Merwe 2001] Rudolph Van Der Merwe and Eric A. Wan. *The Square-Root Unscented Kalman Filter for State and Parameter-Estimation*. In International Conference on Acoustics, Speech, and Signal Processing, pages 3461–3464, 2001. (Cited on pages 4, 28, 42, 43, 45 and 81.)

[Miao 2011] Hongyu Miao, Xiaohua Xia, Alan S Perelson and Hulin Wu. *On Identifiability of Nonlinear ODE Models and Applications in Viral Dynamics*. SIAM Review, vol. 53, no. 1, pages 3–39, 2011. (Cited on pages 60 and 89.)

[Miko 2008] Ilona Miko. *Genetic dominance: genotype-phenotype relationships*. Nature Education, vol. 11, no. 1, 2008. (Cited on page 11.)

[Moles 2003] Carmen G. Moles, Pedro Mendes and Julio R. Banga. *Parameter Estimation in Biochemical Pathways: A Comparison of Global Optimization Methods*. Genome Research, vol. 13, no. 11, pages 2467–2474, 2003. (Cited on pages 17, 18 and 73.)

[Müller 2002] T. G. Müller, N. Noykova, M. Gyllenberg and J. Timmer. *Parameter identification in dynamical models of anaerobic waste water treatment*. Mathematical Biosciences, vol. 177-178, pages 147 – 160, 2002. (Cited on page 31.)

[Murphy 2000] S. A. Murphy and A.W. van Der Vaart. *On Profile Likelihood*. Journal of the American Statistical Association, vol. 95, no. 450, pages 449–465, 2000. (Cited on page 62.)

[Nachman 2004] I. Nachman, A. Regev and N. Friedman. *Inferring quantitative models of regulatory networks from expression data*. Bioinformatics, vol. 20, no. suppl 1, pages i248–i256, 2004. (Cited on page 40.)

[Neale 1997] Michael C. Neale and Michael B. Miller. *The Use of Likelihood-Based Confidence Intervals in Genetic Models*. Behavior Genetics, vol. 27, no. 2, pages 113–120, 1997. (Cited on page 61.)

[Neath 1997] Andrew A Neath and Francisco J Samaniego. *On the Efficacy of Bayesian Inference for Nonidentifiable Models*. The American Statistician, vol. 51, no. 3, pages 225 – 232, 1997. (Cited on page 67.)

[Nikolaev 2006] Evgeni V Nikolaev, Jordan C Atlas and Michael L Shuler. *Computer models of bacterial cells: from generalized coarsegrained to genome-specific modular models*. Journal of Physics: Conference Series, vol. 46, no. 1, page 322, 2006. (Cited on page 2.)

[Orderud 2006] Fredrik Orderud. *Comparison of Kalman Filter Estimation Approaches for State Space Models with Nonlinear Measurements*. In proceedings of IEEE Computers in Cardiology, pages 125–128, 2006. (Cited on page 31.)

[Pedamallu 2008] Chandra Sekhar Pedamallu and Linet Özdamar. *Investigating a hybrid simulated annealing and local search algorithm for constrained optimization.* European Journal of Operational Research, vol. 185, no. 3, pages 1230–1245, 2008. (Cited on page 22.)

[Pohjanpalo 1978] H. Pohjanpalo. *System identifiability based on the power series expansion of the solution*. Mathematical Biosciences, vol. 41, no. 1-2, pages 21 – 33, 1978. (Cited on pages 31 and 32.)

[Press 1992] William H. Press, Brian P. Flannery, Saul A. Teukolsky and William T. Vetterling. Numerical Recipes in Fortran 77: The Art of Scientific Computing. Cambridge University Press, 2 édition, September 1992. (Cited on page 25.)

[Prill 2010] Robert J. Prill, Marbach Daniel, Saez-Rodriguez Julio, Sorger Peter K., Alexopoulos Leonidas G., Xue Xiaowei, Clarke Neil D., Altan-Bonnet Gregoire

and Stolovitzky Gustavo Prill. *Towards a Rigorous Assessment of Systems Biology Models: The DREAM3 Challenges*. PLoS ONE, vol. 5, no. 2, page e9202, 02 2010. (Cited on pages 71 and 95.)

[Qi 2011] Correa-Victorino Qi Cheng, A. Charara and A. Heudiasyc Lab. *A new nonlinear observer using unscented Kalman filter to estimate sideslip angle, lateral tire road forces and tire road friction coefficient*. In Proceedings of Intelligent Vehicles Symposium (IV), pages 709–714, 2011. (Cited on page 34.)

[Quach 2007] Minh Quach, Nicolas Brunel and Florence dÁlchè Buc. *Estimating parameters and hidden variables in non-linear state-space models based on ODEs for biological networks inference*. Bioinformatics, vol. 23, no. 23, pages 3209–3216, 2007. (Cited on pages 15, 16, 17, 34 and 78.)

[Quaiser 2009] Tom Quaiser and Martin Monnigmann. *Systematic identifiability testing for unambiguous mechanistic modeling - application to JAK-STAT, MAP kinase, and NF-kappaB signaling pathway models*. BMC Systems Biology, vol. 3, no. 1, page 50, 2009. (Cited on pages 31, 32 and 64.)

[Rabitz 1983] H Rabitz, M Kramer and D Dacol. *Sensitivity Analysis in Chemical Kinetics*. Annual Review of Physical Chemistry, vol. 34, no. 1, pages 419–461, 1983. (Cited on page 57.)

[Rajeswari 2009] K. Rajeswari, K. Lingamurty and S. Rao. *Unscented Kalman Filter with Application to Bearings- Only Target Tracking*. IETE Journal of Research, vol. 55, no. 2, pages 63–67, 2009. (Cited on page 34.)

[Rannala 2002] Bruce Rannala. *Identifiability of parameters in MCMC Bayesian inference of phylogeny.* Systematic Biology, vol. 51, no. 5, pages 754–760, 2002. (Cited on pages 6, 35, 67, 68 and 89.)

[Rao 2003] Christopher V. Rao, James B. Rawlings and David Q. Mayne. *Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations*. IEEE Transactions on Automatic Control, vol. 48, pages 246–258, 2003. (Cited on page 42.)

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

Bioinformatics, vol. 25, no. 15, pages 1923–1929, 2009. (Cited on pages 33, 60, 62, 63, 87, 89, 90, 91 and 93.)

[Raue 2010] A Raue, V Becker, U Klingmüller and J Timmer. *Identifiability and observability analysis for experimental design in nonlinear dynamical models*. Chaos Woodbury Ny, vol. 20, no. 4, page 045105, 2010. (Cited on pages 89 and 91.)

[Raue 2011] A. Raue, C. Kreutz, T. Maiwald, U. Klingmuller and J. Timmer. *Addressing parameter identifiability by model-based experimentation*. IET Systems Biology, vol. 5, no. 2, pages 120–130, 2011. (Cited on pages 61, 62, 64, 87, 89 and 90.)

[Reid 2007] N. Reid and D.A.S. Fraser. *Efficient Selection of the Most Similar Image in a Database for Critical Structures Segmentation*. In PHYSTAT2003, SLAC, volume 4792 of *LNCS*, pages 203–210. Springer Verlag, 2007. (Cited on page 61.)

[Ribeiro 2004] Maria Isabel Ribeiro. *Kalman and Extended Kalman Filters: Concept, Derivation and Properties*. Institute for Systems and Robotics Lisboa Portugal, 2004. (Cited on page 27.)

[Rodriguez-Fernandez 2006a] Maria Rodriguez-Fernandez, Jose Egea and Julio Banga. *Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems*. BMC Bioinformatics, vol. 7, no. 1, page 483, 2006. (Cited on pages 17, 18 and 20.)

[Rodriguez-Fernandez 2006b] Maria Rodriguez-Fernandez, Pedro Mendes and Julio R. Banga. *A hybrid approach for efficient and robust parameter estimation in biochemical pathways*. Biosystems, vol. 83, no. 2-3, pages 248 – 265, 2006. (Cited on pages 17, 66, 73 and 91.)

[Rohwer 2001] Johann M. Rohwer and Frederik C. Botha. *Analysis of sucrose accumulation in the sugar cane culm on the basis of in vitro kinetic data*. Biochemical Journal, vol. 358, no. 2, pages 437–445, 2001. (Cited on pages 71, 82 and 106.)

[Ryan 2007] N Ryan, Waterfall Joshua J, Casey Fergal P, Brown Kevin S, Myers Christopher R and Sethna James P Gutenkunst. *Universally Sloppy Parameter Sensitivities in Systems Biology Models*. PLoS Computation Biology, vol. 3, page e189, 10 2007. (Cited on page 60.)

[Samaniego 2010] Francisco J Samaniego. A comparison of the bayesian and frequentist approaches to estimation. Springer Series in Statistics, 2010. (Cited on pages 6, 35, 40, 67, 68, 69 and 94.)

[Sanders 2009] Mathijs Sanders. *Metabolic network modeling: Rate Laws and Parameter Estimation*. Tehcnical note, Information and Communication Theory Group, TU Delft, 2009. (Cited on page 16.)

[Schaber 2011] Jörg Schaber and Edda Klipp. *Model-based inference of biochemical parameters and dynamic properties of microbial signal transduction networks*. Current Opinion in Biotechnology, vol. 22, no. 1, pages 109–116, 2011. (Cited on pages 17 and 62.)

[Schallau 2010] Kai Schallau and Björn H. Junker. *Simulating Plant Metabolic Pathways with Enzyme-Kinetic Models*. Plant Physiology, vol. 152, no. 4, pages 1763–1771, April 2010. (Cited on page 9.)

[Schenkendorf 2009] R. Schenkendorf, A. Kremling and M. Mangold. *Optimal experimental design with the sigma point method.* IET systems biology, vol. 3, no. 1, pages 10–23, 2009. (Cited on pages 62 and 66.)

[Schenkendorf 2011] Renè Schenkendorf and Michael Mangold. *Qualitative and Quantitative Optimal Experimental Design for Parameter Identification of a MAP Kinase Model*. In Proceedings of the 18th IFAC World Congress, pages 11666–11671, 2011. (Cited on pages 17 and 60.)

[Schneider 2011] Jean Schneider. *DREAM6 Estimation of Model Parameters Challenge*. http://www.the-dream-project.org/challanges/dream6-estimation-model-parameters-challenge, 2011. (Cited on page 95.)

[Schön 2006] Thomas Schön. *Estimation of Nonlinear Dynamic Systems - Theory and Applications*. Linköping studies in science and technology. dissertations. no. 998, Linköping, Sweden, Feb 2006. (Cited on page 24.)

[Sel'kov 1968] E. E. Sel'kov. *Self-Oscillations in Glycolysis 1.A Simple Kinetic Model*. European Journal of Biochemistry, vol. 4, pages 79–86, jun 1968. (Cited on page 3.)

[Simon 2006] Dan Simon. Optimal state estimation: Kalman, h infinity, and nonlinear approaches. Wiley Interscience, 2006. (Cited on page 32.)

[Singh 2006] Gulshan Singh and Kalyanmoy Deb. *Comparison of multi-modal optimization algorithms based on evolutionary algorithms*. In Proceedings of the 8th annual conference on Genetic and evolutionary computation, GECCO '06, pages 1305–1312, New York, NY, USA, 2006. ACM. (Cited on page 11.)

[Sitz 2002] A. Sitz, U. Schwarz, J. Kurths and H. U. Voss. *Estimation of parameters and unobserved components for nonlinear systems from noisy time series*. Phys. Rev. E, vol. 66, no. 1, page 016210, 2002. (Cited on page 40.)

[Stelling 2004] Jörg Stelling. *Mathematical models in microbial systems biology*. Current Opinion in Microbiology, vol. 7, no. 5, pages 513 – 518, 2004. (Cited on pages 7 and 8.)

[Stephanopoulos 1998] Gregory N. Stephanopoulos, Aristos A. Aristidou and Jens Nielsen. Metabolic engineering: Principles and methodologies. Academic Press, 1 édition, October 1998. (Cited on page 8.)

[Steuer 2006] Ralf Steuer, Thilo Gross, Joachim Selbig and Bernd Blasius. *Structural kinetic modeling of metabolic networks*. Proceedings of the National Academy of Sciences, vol. 103, no. 32, pages 11868–11873, 2006. (Cited on pages 7 and 9.)

[Steuer 2008] Ralf Steuer and Björn H. Junker. *Computational Models of Metabolism: Stability and Regulation in Metabolic Networks*. In Advances in Chemical Physics, pages 105–251. John Wiley & Sons, Inc., 2008. (Cited on page 9.)

[Stitt 2010] Mark Stitt, Ronan Sulpice and Joost Keurentjes. *Metabolic Networks: How to Identify Key Components in the Regulation of Metabolism and Growth*. Plant Physiology, vol. 152, no. 2, pages 428–444, 2010. (Cited on page 7.)

[Sun 2008] Xiaodian Sun, Li Jin and Momiao Xiong. *Extended Kalman Filter for Estimation of Parameters in Nonlinear State-Space Models of Biochemical Networks*. PLoS ONE, vol. 3, no. 11, page e3758, 11 2008. (Cited on pages 4 and 17.)

[Suzdaleva 2007] Evgenia Suzdaleva. *Initial conditions for Kalman filtering: prior knowledge specification*. In Proceedings of the 7th Conference on 7th WSEAS

International Conference on Systems Theory and Scientific Computation - Volume 7, pages 45–49, 2007. (Cited on page 94.)

[Teixeira 2008] Bruno Otávio Soares Teixeira, Leonardo A. B. Tôrres, Luis Antonio Aguirre and Dennis S. Bernstein. *Unscented filtering for interval-constrained nonlinear systems.* In 47^{th} IEEE Conference on Decision and Control, pages 5116–5121, 2008. (Cited on pages 42, 51 and 55.)

[Teixeira 2010] Bruno O. S. Teixeira, Leonardo A. B. Tôrres, Luis A. Aguirre and Dennis S. Bernstein. *On unscented Kalman filtering with state interval constraints*. Journal of Process Control, vol. 20, no. 1, pages 45–57, 2010. (Cited on page 55.)

[Terejanu 2008] Gabriel A. Terejanu. *Unscented Kalman Filter Tutorial*. Web-tutorial, Department of Computer Science and Engineering, University at Buffalo, Buffalo, New York, 2008. (Cited on page 49.)

[Thacker 1998] N.A. Thacker and A.J. Lacey. *Tutorial: The Kalman Filter*. Imaging Science and Biomedical Engineering Division, Medical School, University of Manchester, 1998. (Cited on page 61.)

[Uys 2007] Lafras Uys, Frederik C. Botha, Jan-Hendrik S. Hofmeyr and Johann M. Rohwer. *Kinetic model of sucrose accumulation in maturing sugarcane culm tissue*. Phytochemistry, vol. 68, no. 16-18, pages 2375–2392, 2007. (Cited on pages 71 and 82.)

[Vachhani 2006] Pramod Vachhani, Shankar Narasimhan and Raghunathan Rengaswamy. *Robust and reliable estimation via Unscented Recursive Nonlinear Dynamic Data Reconciliation*. Journal of Process Control, vol. 16, no. 10, pages 1075 – 1086, 2006. (Cited on pages 42, 44, 45, 46 and 56.)

[Vajda 1989] Sandor Vajda, Keith R. Godfrey and Herschel Rabitz. *Similarity transformation approach to identifiability analysis of nonlinear compartmental models*. Mathematical Biosciences, vol. 93, no. 2, pages 217 – 248, 1989. (Cited on page 31.)

[van der Merwe 2004] Rudolph van der Merwe. *Sigma-Point Kalman Filters for Probabilistic Inference in Dynamic State-Space Models*. PhD thesis, OGI School of Science and Engineering at Oregon Health and Science University, 2004. (Cited on pages 4, 14, 15, 24, 25, 28, 29, 31, 47, 55 and 81.)

[Venzon 1988] D. J. Venzon and S. H. Moolgavkar. *A Method for Computing Profile-Likelihood Based Confidence Intervals*. Applied Statistics, vol. 37, no. 1, pages 87–94, 1988. (Cited on page 62.)

[Wachowiak 2010] M. P. Wachowiak, Smolikova Wachowiak R. and Smolik D. *Parameter Estimation of Nonlinear Econometric Models Using Particle Swarm Optimization*. Central European Revue of Economic Issues, vol. 13, no. 4, pages 193–199, 2010. (Cited on page 76.)

[Wan 2001] Eric Wan and Rudolph Van Der Merwe. *Chapter 7 The Unscented Kalman Filter*. In Kalman Filtering and Neural Networks, pages 221–280. Wiley, 2001. (Cited on page 29.)

[Wang 2004] Duolao Wang and Michael Murphy. *Estimating optimal transformations for multiple regression using the ACE algorithm*. Journal of Data Science, vol. 2, pages 329–346, 2004. (Cited on page 66.)

[Wang 2008] Xu Wang. *Method of Steepest Descent and its Applications*. Lecture note, University of Tennessee, 2008. (Cited on page 19.)

[Welch 2006] Greg Welch and Gary Bishop. *An Introduction to the Kalman Filter*. TR 95-041, Department of Computer Science, University of North Carolina at Chapel Hill, 2006. (Cited on pages 4, 13 and 26.)

[Wiechert 1996] Wolfgang Wiechert and de Graaf AA. *In vivo stationary flux analysis by 13C labeling experiments*. Adv Biochem Eng Biotechnol, vol. 54, pages 109–154, 1996. (Cited on page 9.)

[Wiechert 2001] Wolfgang Wiechert. *13C Metabolic Flux Analysis*. Metabolic Engineering, vol. 3, no. 3, pages 195 – 206, 2001. (Cited on page 9.)

[Wu 2005] Yuanxin Wu, Dewen Hu, Meiping Wu and Xiaoping Hu. *Unscented Kalman filtering for additive noise case: augmented versus nonaugmented*. IEEE Signal Processing Letters, vol. 12, no. 5, pages 357–360, 2005. (Cited on page 29.)

[Xia 2003] X. Xia and C. H. Moog. *Identifiability of nonlinear systems with application to HIV/AIDS models*. IEEE Transactions on Automatic Control, vol. 48, no. 2, pages 330 – 336, 2003. (Cited on page 31.)

[Yao 2003] K. Zhen Yao, Benjamin M. Shaw, Bo Kou, Kim B. McAuley and D. W. Bacon. *Modeling Ethylene/Butene Copolymerization with Multi−site Catalysts: Parameter Estimability and Experimental Design*. Polymer Reaction Engineering, vol. 11, no. 3, pages 563–588, 2003. (Cited on pages 40, 56, 59, 83, 84, 85 and 90.)

[Yue 2006] Hong Yue, Martin Brown, Joshua Knowles, Hong Wang, David S. Broomhead and Douglas B. Kell. *Insights into the behaviour of systems biology models from dynamic sensitivity and identifiability analysis a case study of an NF-kb signalling pathway*. Molecular BioSystems, vol. 2, pages 640–649, 2006. (Cited on pages 57, 84 and 85.)

[Zheng 2009] Li Zheng, O'Doherty Joseph E., Hanson Timothy L., Lebedev Mikhail A., Henriquez Craig S. and Nicolelis Miguel A. L. Li. *Unscented Kalman Filter for Brain-Machine Interfaces*. PLoS ONE, vol. 4, no. 7, page e6243, 07 2009. (Cited on page 34.)

[Zhu 2008] Tianjun Zhu and Hongyan Zheng. *Application of Unscented Kalman Filter to Vehicle State Estimation*. In Proceedings of the 2008 ISECS International Colloquium on Computing, Communication, Control, and Management - Volume 02, pages 135–139, Washington, DC, USA, 2008. IEEE Computer Society. (Cited on page 34.)

Curriculumn Vita

Personal Data

Name:	Syed Murtuza Baker
Academic degree:	Masters in Information Technology
Gender:	Male
Date of birth:	11^{th} of September, 1981
Place of birth:	Dhaka, Bangladesh
Nationality:	Bangladeshi
Current address:	Correnstrasse 3
	06466 Gatersleben
E-mail:	baker@ipk-gatersleben.de

Education

01/2009 - present	IPK Gatersleben, Ph.D. Studies on computational systems biology supervised by Dr. Björn Junker and Prof. Falk Schreiber
07/2004 - 08/2005	University of Dhaka Masters in Information Technology CGPA 3.92 / 4.0
09/1999 - 04/2003	East West University Bachelor of Science in Computer Science CGPA 3.86 / 4.0 (Magna cum laude)

Research & Work Experience

01/2009 - present	Research associate at IPK Gatersleben Research group Systems Biology
05/2008 - present (On study leave)	Senior lecturer, Dept. Of Computer Science and Engineering East West University
09/2005 - 04/2008	Lecturer, Dept. Of Computer Science and Engineering East West University
05/2003 - 06/2004	Software developer Big apple computer solution center

Publication list:

Journal publications:

1 Syed Murtuza Baker, Hart Poskar, Björn Junker, **"Unscented Kalman filter with parameter identifiability analysis for the estimation of multiple parameters in kinetic models"**, In EURASIP Journal on Bioinformatics and Systems Biology, vol. 2011, no. 1, pp. 7, 2011.

2 Syed Murtuza Baker, Kai Schallau, Björn H. Junker, **"Comparison of different algorithms for simultaneous estimation of multiple parameters in kinetic metabolic models"**, In Journal of Integrative Bioinformatics, vol. 7, no. 3, pp. e133, 2010.

3 Mohammad Shoyaib, M. Abdullah-Al-Wadud, Syed Murtuza Baker, Nurul Mohammad Islam, Oksam Chae, **"Predicting Protein-protein Interaction Using Amino Acid Sequence Information: A Computational Approach"**, In Plant Tissue Cult. & Biotech., vol. 20, no. 1, pp. 37-45, 2010.

4 Firoz Anwar, Syed Murtuza Baker, Taskeed Jabid, Md Mehedi Hasan, Mohammad Shoyaib, Haseena Khan, Ray Walshe, **"Pol II promoter prediction using characteristic 4-mer motifs: a machine learning approach"**, In BMC Bioinformatics, vol. 9, no. 1, pp. 414, 2008.

Conference publications:

1 Syed Murtuza Baker, Björn Junker, **"Adaptive Unscented Kalman Filter for Estimation of Parameters in Kinetic Metabolic Models"**, In The Fourth International Conference on Computational Systems Biology, Suzhou, China, pp. 185-186, 2010.

2 Syed Murtuza Baker, Hart Poskar, Björn Junker, **"Unscented Kalman Filter for estimation of multiple parameters in kinetic models"**, In Proceedings of the Seventh International Workshop on Computational Systems Biology, Luxembourg, pp. 19-22, 2010.

3 Firoz Anwar, Syed Murtuza Baker, Mohammad Shoyaib, Taskeed Jabid, Mehedi Md. Hasan, **"Identification of Gene using Machine Learning Technique"**, In

International Conference on Bioinformatics Computational Biology, BIOCOMP 2008, Las vegas, USA, pp. 1012-1014, 2008.

4 Taskeed Jabid, Mohammad Shoyaib, Syed Murtuza Baker, Firoz Anwar, Md. Mehedi Hasan, **"Protein-Protein Interaction Detection from Primary Structure using Support Vector Machine"**, In International Conference on Bioinformatics Computational Biology, BIOCOMP 2008, Las vegas, USA, pp. 965-969, 2008.

5 Taskeed Jabid, Firoz Anwar, Syed Murtuza Baker, Mohammad Shoyaib, **"Identification of promoter through stochastic approach"**, In 10^{th} international conference on computer and information technology, ICCIT 2007, Dhaka, Bangladesh, pp. 1-4, 2007.

Publications in preparation:

1 Syed Murtuza Baker, C. Hart Poskar, Falk Schreiber, Björn H. Junker, **Constrained Unscented Square-Root Kalman Filter for Parameter Estimation in Non-linear Kinetic Models of Biochemical Network**, in preparation.

2 Syed Murtuza Baker, C. Hart Poskar, Falk Schreiber, Björn H. Junker, **A parameter estimation framework in kinetic models of biological system**, in preparation.

3 Syed Murtuza Baker, C. Hart Poskar, Björn H. Junker, **Identifiability analysis of metabolic fluxes in steady state ^{13}C stable isotope labelling experiments**, in preparation.

Declaration

Herewith I declare that I independently wrote the following doctoral thesis using no other than the sources those which are listed. The principles ”Verantwortung in der Wissenschaft” (Responsibility in Science), recommended by the Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, were observed.

Gatersleben, November 26, 2012 Syed Murtuza Baker

Erklärung

Hiermit erkläre ich, dass ich diese Arbeit selbständig und ohne fremde Hilfe verfasst habe. Ich habe keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt. Die den benutzten Werken anderer Autoren wörtlich oder inhaltlich entnommenen Stellen sind als Solche kenntlich gemacht worden. Bisher habe ich mich noch nicht um einen Doktorgrad beworben.
Desweiteren erkläre ich, daß keine Strafverfahren gegen mich anhängig sind.

Gatersleben, den November 26, 2012 Syed Murtuza Baker