

Model-based analysis and parameter estimation of a human blood glucose control system model



Domokos Mészéna

Pázmány Péter Catholic University

The Faculty of Information Technology and Bionics

Info-bionics Engineering

Scientific advisor:

Gábor Szederkényi, D.Sc.

Paper for the XXXII. National Conference of the Scientific Students' Association

2014

Abstract

A komplex biológiai folyamatok időbeli viselkedését legalább kvalitatív módon leíró matematikai modellek fontos szerepet játszanak a rendszer működésének megértésében és a szükséges külső beavatkozások (szabályozás) megtervezésében. Sajnos általánosságban elmondható, hogy biológiai rendszerek esetén a dinamikus mérések minősége és mennyisége elmarad a technológiai rendszereknél megszokott lehetőségektől. Emiatt a megfelelően pontos dinamikus modellalkotás lényegesen nehezebb. Kezdeti lépésként az ún. strukturális identifikálhatóság vizsgálatának feladata annak megállapítása, hogy az adott modellstruktúra esetén a modell ismeretlen paraméterei egyértelműen meghatározhatók-e elméleti szinten. A praktikus identifikálhatóság vizsgálata pedig arra irányul, hogy a megbecsülendő modell-paraméterek a gyakorlatban megfelelő minőségben kiszámíthatók-e a mérési adatokból. Msc hallgatói munkámban célkitűzésem volt tanulmányozni – az előzetes irodalomkutatásból választott – vércukor szabályozási rendszermodell (Blood Glucose Control System) strukturális és praktikus identifikálhatóságát, mivel e tulajdonságokat legjobb tudomásom szerint az irodalomban részletesen még nem tanulmányozták. Eredményeimben azt találtam, hogy a kiválasztott négy lineáris függőségű paraméter strukturális szempontból globálisan (tehát egyértelműen) identifikálható, amely biztató eredmény a paraméterbecslés gyakorlati kivitelezése szempontjából. Az identifikálhatósági analízis eredményeinek felhasználásával sikerült az eddigieknél pontosabban reprodukálni a szakirodalomban fellelhető bizonyos kísérleti adatokat. Feladatomat, mind az illesztést, és identifikációkat MATLAB szimulációs környezetben végeztem. Legutolsó eredményemként elvégeztem egy időbeli szenzitivitás analízist, megvizsgálva a becsült paraméterek érzékenységet a vizsgált állapotváltozókra nézve. Terveim között szerepel hozzáfogni egy, a klinikusok számára alkalmazható, optimális kísérlet-terv előkészítéséhez, a hatékonyabb paraméterbecslés érdekében. További céljaim között szerepel az eredményeim részletezése, publikálása, valamint nemlineáris szabályozási sémák megalkotása. Az identifikált modellen tesztelt szabályzó a jövőben hozzájárulhat egy biológiailag releváns, mérnökiileg megfelelően szabályozott vércukorszintmérő kifejlesztéséhez.

Contents

1	Introduction	1
1.1	Topic overview	1
1.2	Objectives	1
1.3	Organization of Blood Glucose Control System	2
1.4	Short description of diabetes	4
2	Methods	5
2.1	Dynamical model analysis for identification	5
2.1.1	Model identification processes	5
2.1.2	Structural identifiability analysis	6
2.1.3	Practical identifiability	8
2.1.4	Optimal experiment design	8
2.1.5	Parameter estimation techniques	9
2.1.6	Definition of identifiability <i>tableaus</i>	9
2.2	The applied mathematical description: <i>The Liu - model</i>	10
2.3	Brief description of the applied toolboxes	13
2.3.1	GenSSI Toolbox	13
2.3.2	AMIGO Toolbox	14
2.3.3	SUNDIALS Toolbox	14
3	Results	15
3.1	Building a MATLAB <i>SIMULINK</i> model	15
3.1.1	Results for structural identifiability, obtained <i>tableaus</i>	16
3.2	Fitting the model to experimental data	17
3.2.1	Iterative process of parameter estimation	20
3.3	Model validation using OGTT experimental data	21
3.4	Additional results for practical identifiability	22
3.4.1	Simulating the model in AMIGO Toolbox	22

3.4.2	Ranking of unknowns	24
3.5	Time-dependent sensitivity analysis	25
3.5.1	Improved parameter estimation technique	27
4	Discussion	29
4.1	Summary	29
4.2	Prospects for the future	29
	Acknowledgement	32
	Abbreviations	33
	List of Figures	33
	References	34

Chapter 1

Introduction

1.1 Topic overview

Mathematical models and simulations offer us a great possibility to obtain useful information about a given system. With these methods we are able to design *in silico* experiments, generate reliable predictions and hypotheses so as to better understand – for instance – complex biological processes. Nevertheless, the quality of our model highly depends on the model construction: namely the selection of states and parameters in the equations. This is what we might call the 'art of modelling' in quantitative natural sciences. Especially in biology, the model building problem is more complex because of the common nonlinearities. Even if we have experimental data, every measurement has a certain error introducing uncertainty to the system and complicating precise modelling. In this respect, parameter estimation by means of data fitting has become a critical step in the model building process. Nowadays, it is a common knowledge, that these main difficulties are often based on the poor or lack of identifiability, which is the difficulty or impossibility of choosing unique values for the unknown parameters [1]. Mathematically speaking, the mapping from the parameter space to the model output space is not injective. Therefore, as a first step we have to examine if our model is identifiable both in a structural and in practical sense, and based on these findings, we might be able to design optimal experimental set-ups to obtain more reliable models through improved parameter estimation. Moreover, we would like to obtain such a nonlinear model that will be suitable later for controller design.

1.2 Objectives

In my student research project, my aim is to analyse one of the existing models of Glucose Control System (in human blood), since as far as I know these identifiability properties have

1. INTRODUCTION

never been studied earlier in the literature. The main question is whether this system can be identified structurally and practically or not. The applied method follows a recent approach in the identification analysis of biological systems. In the present study I give an overview of general knowledge on the field, then demonstrate the mathematical formalism for Blood Glucose Control System and execute the identification processes to explore how the unknown parameters affect model fitting. During my investigation, I consider seven model parameters as unknowns. I develop the model by separating the parameters, taking in the initial values as nonlinear parameters and smoothing the intermittent experimental values with a cubic spline interpolation. With the results of the identifiability analysis (and with other methods) I have achieved a more accurate comparison with the experimental data found in literature [2]. Furthermore, I execute a time-dependent parameter sensitivity analysis to see the impact of the parameters respect to the state-variables. The sensitivity analysis is calculated for each of the model variables separately, as a function of specific, individual parameter perturbations. I make the simulation and the figures with MATLAB programming environment, I use the GenSSI Toolbox and the AMIGO Toolbox to perform the model identification procedures [3, 4] and I execute the time-dependent sensitivity analysis with the SUNDIALS Toolbox (using CVODES solvers) [5, 6].

1.3 Organization of Blood Glucose Control System

The regulation of blood glucose is one of the most fundamental phenomena in the human body. The system uses glucose as input (digested from food) and the absorption of glucose in the cells determines the actual concentration of blood glucose. All of this is controlled by rigorous hormonal and enzymatic processes. The following chart 1.1 shows the molecular control mechanism of blood glucose [7].

The system comprises of many complicated sub-processes, whose erroneous activity usually leads to common diabetic diseases. To understand diabetes, it is important to first understand the normal metabolic process of glucose (as a source of fuel for the body). Glucose intake is originated through food, while processes in the liver can also increase the glucose level in the bloodstream. Glucose can be taken up by cells in a way depending on the concentration of insulin or independent of it. The 'insulin-independent' way is characteristic for brain and nerve cells and uses GLUT3 transporter. The other, 'insulin-dependent' path is used by tissue cells and goes via the GLUT4 (in muscle, kidney and fat cells) and GLUT2 (liver) transporters. In the case of low blood glucose level, α -cells of the pancreas produce the hormone *glucagon*. Glucagon initiates a series of kinase activations, and finally leads to the activation of one phosphorylase enzyme, which catalyses the breakdown of *glycogen* into

1.3 Organization of Blood Glucose Control System

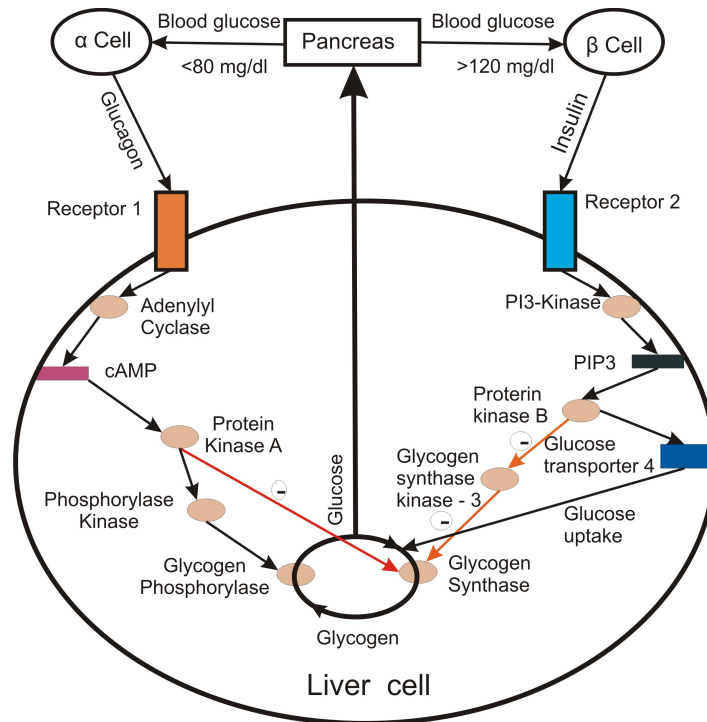


Figure 1.1: Glucose regulation - This is the schematic description of Blood Glucose Control System (BGCS) in humans.

glucose. On the contrary, when blood glucose level is too high, β -cells of the pancreas secrete *insulin*. Insulin triggers a series of reactions to activate the glycogen synthase enzyme, which catalyses the conversion of glucose into glycogen. Furthermore, insulin also initiates a series of activation for kinases in tissue cells to import glucose into the muscle or fat cell's intracellular storages [7, 8] (and see Figure 1.1). To summarize, glucose has an important role in the following metabolic processes:

- muscle and fat cells remove glucose from the blood,
- cells breakdown glucose via glycolysis and the citrate cycle, storing its energy in the form of adenosine triphosphate (ATP),
- liver and muscle store glucose as glycogen as a short-term energy reserve,
- adipose tissue stores glucose as fat for long-term energy reserve, and
- cells use glucose for protein synthesis.

When the body produces enough insulin, glucose levels in the bloodstream and cells are controlled automatically. The glucose in the bloodstream stays within a safe range, never getting too high or too low. When the supply of glucose is not maintained the body's glucose

1. INTRODUCTION

levels can either become too high (*hyperglycemia*) or too low (*hypoglycemia*). The second one is usually caused by an overdose of insulin.

1.4 Short description of diabetes

Diabetes is a life-long disease with high levels of sugar in the blood. It can be caused by too low insulin production, resistance to insulin, or both. In the most frequent case, the pancreas does not produce enough insulin to press down the glucose level. In the other case, the muscle, fat, and liver cells do not respond to insulin normally. Type 1 diabetes is generally diagnosed in childhood. These patients have no or very small amount of insulin production in pancreas. To sustain their normal life conditions, daily insulin injections are required. Type 2 diabetes is more common than Type 1 (90 % of all cases) and it usually occurs in adulthood. Here, the pancreas cannot produce enough insulin to keep blood glucose levels normal, often because the body does not respond well to the insulin. Type 2 diabetes is becoming widespread due to the growing number of older people worldwide, to the percentage of obesity in the population, *etc.* There are several diagnostic tests to detect diabetes. Fasting Blood Glucose Test (FBGT) is one simple way, because it is easy to perform among any circumstances. In the most commonly applied Oral Glucose Tolerance Test (OGTT) person who has fasted earlier, drinks a glass of intense glucose solution. Drops of blood are extracted periodically and concentration of blood glucose is measured. Trends in glucose level variations over time, the maximum value, and several other features can help to decide whether the given person has diabetes or not [9].

Chapter 2

Methods

2.1 Dynamical model analysis for identification

2.1.1 Model identification processes

A dynamical model is a mathematical description focusing on selected features of the studied process, which can be built in several different forms depending on the objectives in mind. It has been shown that complex processes of constructing mathematical models for biological systems are challenging, but probably the most difficult task among them is the identification of the structure of the underlying biological network and its regulatory processes [10]. When we start building a mathematical model, we have to pay attention to (at least) three crucial aspects:

1. First of all, invalid hypotheses regarding variables and interactions to be included in the model may lead to incorrect interpretation of the results.
2. Second, overly complex model representation may provide very good fit to the observed time series data, but is rarely optimal against new datasets, and highly sensitive to the noise in measured data (due to over-fitting) [10].
3. Finally, the inclusion of too many components and interactions may eventually result in problems caused by computational 'explosion'. In such case, the system most probably will be non-identifiable for a couple of its generally nonlinear parameters.

Despite the improving quality of biological measurements, this model identification step still remains a mathematical and computational problem, since in many cases, no unique solution exists to the parameters. Therefore the first step is to examine whether the parameters are identifiable. But what could be the cause of this (earlier described) difficulty, namely the 'lack of identifiability'? We can distinguish between two types of identifiability: *structural*

2. METHODS

identifiability, which is an intrinsic, theoretical property of the model structure depending only on the system dynamics, the observation and the stimuli functions. It can be problematic even if we have perfect data. On the other hand, *practical identifiability* is related to the experimental design, sufficient excitation of the system dynamics and the measured artefacts and noise [1, 4] (and see Figure 2.1). We will apply the following general non-linear model form for describing the dynamics of the blood glucose control system:

$$\sum(\boldsymbol{\theta}) : \begin{cases} \dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) + \sum_{j=1}^n \mathbf{g}_j(\mathbf{x}, \boldsymbol{\theta}) \mathbf{u}_j \\ \mathbf{y} = \mathbf{h}(\mathbf{x}, \boldsymbol{\theta}), \mathbf{x}(t_0) = \mathbf{x}_0 \end{cases} \quad (2.1)$$

where $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbf{M} \subset \mathbb{R}^{n_x}$ is the state variable, with $\mathbf{M}$ a subset of $\mathbb{R}^{n_x}$ containing the initial state (which may depends on the parameters as well), $\mathbf{u} = (u_1, u_2, \dots, u_n) \in \mathbb{R}^{n_u}$ is an n_u - dimensional input (control) vector with $u_1, \dots, u_n$ smooth functions, and $\mathbf{y} = (y_1, y_2, \dots, y_n) \in \mathbb{R}^{n_y}$ is the n_y - dimensional output (experimental observables). The vector of unknown parameters is denoted by $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_n) \in \boldsymbol{\Theta}$, and in general is assumed to belong to an open and connected subset of $\mathbb{R}^{n_p}$. The entries of $\mathbf{f}$, $\mathbf{g} = (g_1, g_2, \dots, g_n)$ and $\mathbf{h}$ are analytic functions of their arguments. These functions and the initial conditions may depend on the parameter vector $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ [1].

2.1.2 Structural identifiability analysis

‘Per definitionem’, a given model will be structurally globally (or uniquely) identifiable , if

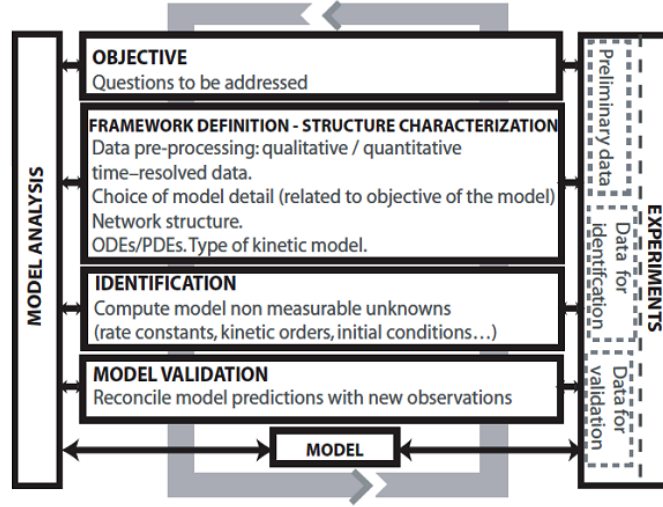
$$\tilde{\gamma}(t \mid \theta') \equiv \tilde{\gamma}(t \mid \theta'') \Rightarrow \theta' = \theta'' \quad (2.2)$$

where

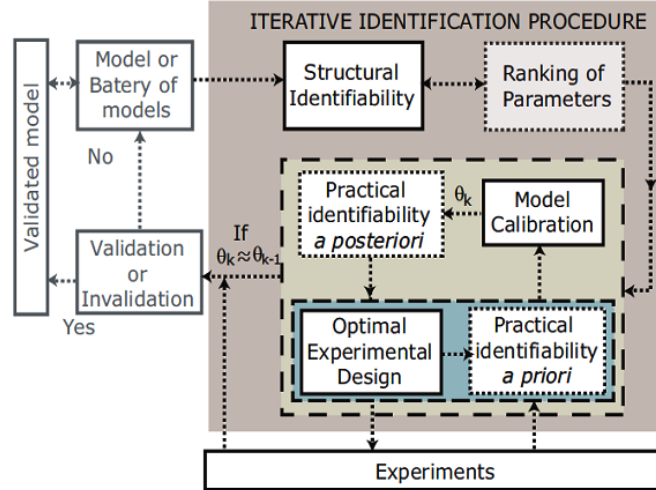
$$\tilde{\gamma}(t \mid \theta) = h(x(t, \theta), u(t), \theta) \quad (2.3)$$

and $x(t, \theta)$ denotes the solution of 2.1 with parameter vector θ . According to 2.2, a structurally non-identifiable model can produce exactly the same observed output with different parametrization. This is clearly a fundamental obstacle of determining the true model parameters from measurements even if the selected model structure is considered to be correct [11]. In other words, if the model is not uniquely identifiable, then there are several parameter vectors that correspond to exactly the same input-output behaviour [12]. When one cannot prove that the structure considered is globally identifiable, one might try to establish that it is identifiable at least locally (whether or not a neighbourhood exists, on which the earlier defined identifiability constraint is true). If one cannot uniquely identify θ_i neither globally

2.1 Dynamical model analysis for identification



(a) Model building loop



(b) Iterative identification process

Figure 2.1: The flow diagram of the place of the identification step in the model building procedure and the parts of identification method [4].

nor locally, one can say, that this parameter is structurally non-identifiable. In other words, structural identifiability regards the possibility of giving unique values to model unknown parameters from the available observables, assuming perfect experimental data (i.e. noise-free and continuous in time)[1]. If some parameters seem not to be identifiable, numerical approaches will not be able to find unique, reliable values for them. In those situations, the ways to overcome this problem will be to reformulate the model (for instance reducing the number of states and parameters) or to fix some parameter values (for instance those which can be determined experimentally in a reliable way).

2. METHODS

2.1.3 Practical identifiability

As we already mentioned in the introduction, practical identifiability analysis is able to evaluate the possibility of assigning unique values to parameters from a given set of experimental data or experimental scheme subject to experimental noise. We have to distinguish between practical identifiability a priori and a posteriori. The first one anticipates the quality of the selected experimental scheme, the expected uncertainty of the parameters. On the other hand, the latter determines the quality of the parameter estimation after model calibration with respect to the confidence regions. It is important to note that the major difference between the two analyses is that, a priori, we have to assume a maximum experimental error. However, a posteriori, the experimental error may be estimated either through experimental data manipulation (when experiments are available) or after model calibration using the residuals, – in other words – prediction errors, which are the differences among model and the experimental data [4, 13].

2.1.4 Optimal experiment design

A crucial aspect of experimental measurements is data quantity and data quality. As mentioned in the previous section, a given noisy data may cause problems in practical identifiability. This is why data generation and modelling have to be designed as a parallel process to avoid unsuited experimental and model output results. In addition, model-based, *in silico* experimentation can greatly reduce the cost of biological experiments and facilitate the understanding of complex biological systems. In the optimal experiment design (OED) we calculate the best scheme for measurements with the greatest precision and with uncorrelation in order to maximize the richness (quality and quantity) of information. The 'richness' of information may be quantified (e.g. with a defined matrix norm) by the *Fisher Information Matrix* (FIM) $\mathcal{F}$, which can be calculated as follows:

$$\mathcal{F} = \mathbf{E}_{y_m|\boldsymbol{\mu}} \left\{ \left[\frac{\partial J(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right] \left[\frac{\partial J(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right]^T \right\} \quad (2.4)$$

Where J is the objective function (*e.g.* a weighted quadratic least-squares function), E represents the expectation for a given value of the parameter μ close to the optimal solution θ^* . It is important to note that the Fisher Information Matrix will depend on the type of experimental noise. In optimal experimental design, we want to determine the time-varying stimuli profile, sampling times, experiment durations, initial conditions to maximize the norm of the Fisher Information Matrix with respect to the system dynamics and algebraic constraints of experimental limitations [13].

2.1.5 Parameter estimation techniques

In the following pages we concentrate on the identifiability, and on the parameter estimation (PE) methods. There are several possible ways to estimate model parameters. The three most frequently used general methods are: *Least Squares* (LS), *Maximum Likelihood* (ML) and *Bayesian Estimator*. However, here we do not describe them in details. Naturally, it is more difficult to determine nonlinearly depending parameters than linear ones. If some parameters seem not to be identifiable, numerical approaches will not be able to find unique, reliable values for them. In those situations, the ways to overcome this problem will be for instance the model reduction (or to merge non-identifiable parameters into a new and identifiable form).

2.1.6 Definition of identifiability *tableaus*

In the process of identification, there are many different types of approaches. Naturally, to successfully tackle identification problems, the applied method must be carefully selected taking into account (among other factors) the model structure and the availability/quality of measurements and the expected complexity of computation. Numerous approaches are available, such as: *Taylor Series*, *Generating Series* (using the Lie-derivatives), *Differential Algebra* methods, *Similarity Transformation*, *Direct tests*, etc. The recent results in literature reveal that the generating series approach (calculating the Lie-derivatives), in combination with the so called identifiability 'tableaus' formalism offers an advantageous compromise among range of applicability, computational complexity and information provided [1, 3, 11]. We define the Lie-derivates of $\mathbf{g}$ along a vector field $\mathbf{f}$ as follows [13]:

$$L_{\mathbf{f}}g(\mathbf{x}, \boldsymbol{\theta}, t) = \sum_{j=1}^n \frac{\partial g(\mathbf{x}, \boldsymbol{\theta}, t)}{\partial x_j} f_j(\mathbf{x}, \boldsymbol{\theta}, t) \quad (2.5)$$

with f_j the j^{th} component of $\mathbf{f}$.

The identifiability tableau is a visualisation of a $K \times n$ (n is the number of parameters, K is the non-zero coefficients of the generating series) matrix representing non-zero elements of the Jacobian computed in series expansion with respect to the parameters. Each column of the table corresponds to a parameter, while all rows represent non-zero coefficients of the series (default value is infinite). The most significant properties of the tableaus are the following:

2. METHODS

- The corresponding parameters may be non-identifiable if the Jacobian of the series coefficients is structurally rank deficient, in other words, the tableau presents zero columns.
- If the rank of the Jacobian is complete (such as it equals to the number of parameters), then it will be possible to, at least, locally identify the parameters.

An identifiability tableau is a way of expressing the dependencies between parameters, with the help of the so called *exhaustive summary*. A vector-valued function $s(\theta)$ is an exhaustive summary if it contains only information about the parameters θ that can be extracted from knowledge of the control $u(t)$ and the measured quantities $y(t, \theta)$. In the case of generating series approach, $s(\theta)$ equals to the series coefficients, evaluated at the initial conditions (or initial states) [3].

2.2 The applied mathematical description: *The Liu - model*

As it has been pointed out in the previous sessions, a model is a mathematical description of the chosen important features of the studied process, which can be built in many different forms depending on the objectives in mind. Our aim is to simply yet realistically follow the dynamics of blood glucose and through refining the model understand the system and the possibilities of controlling it.

The ODE and PDE modelling of blood glucose was started with the so called *minimal model* of Bergman (1979) which contains only 2 ODEs [14]. There is also a more sophisticated form of the minimal model, which includes 3 ordinary, nonlinear differential equations. However, these Bergman models describe only the main insulin-glucose dynamical properties, and a much more complicated description of blood glucose behaviour is presented later by Sørensen (1985) with 19 state variables [15] including almost everything what we know about the system's governing factors. But Sørensen's model is very hardly understandable because of its complexity. In contrast to these previous models, like Bergman and Sørensen, the recently published model of Liu and Tang (2008) applies a more straightforward approach: it describes the aspects of the blood glucose system at the level of molecular processes, taking into account some biochemical considerations but not incorporating all individual molecular interactions responsible for important cellular functions [7, 8]. With this approach, the consequences are more plausible, and different biological processes can be separated. Its complexity is somewhere half-way between the minimal model of Bergman and the most complex Sørensen model. This model has been widely used for model analysis and as a basis of different applications in the literature, for this reason we consider the Liu-model (in fact, the firstly

2.2 The applied mathematical description: *The Liu - model*

published simplified version:[7]) for our further analysis. The dynamics of the state variables and corresponding parameters using the above notation are as follows:

$$\frac{dp_1}{dt} = -(a_1 + a_2)p_1 + u_1(g_2), \quad (2.6)$$

$$\frac{dp_2}{dt} = -(b_1 + b_2)p_2 + u_2(g_2), \quad (2.7)$$

$$\frac{dh_1}{dt} = -a_4h_1(R_1^0 - r_1) - a_3h_1 + a_1p_1 \frac{V_p}{V} \quad (2.8)$$

$$\frac{dh_2}{dt} = -b_4h_2(R_2^0 - r_2) - b_3h_2 + b_1p_2 \frac{V_p}{V} \quad (2.9)$$

$$\frac{dr_1}{dt} = a_4h_1(R_1^0 - r_1) - a_5r_1, \quad (2.10)$$

$$\frac{dr_2}{dt} = b_4h_2(R_2^0 - r_2) - b_5r_2, \quad (2.11)$$

$$\frac{dg_1}{dt} = \frac{k_1r_2}{1 + k_2r_1} \frac{V_{max}^{gs}g_2}{K_m^{gs} + g_2} - k_3r_1 \frac{V_{max}^{gp}g_1}{K_m^{gp} + g_1} \quad (2.12)$$

$$\begin{aligned} \frac{dg_2}{dt} &= -\frac{k_1r_2}{1 + k_2r_1} \frac{V_{max}^{gs}g_2}{K_m^{gs} + g_2} + k_3r_1 \frac{V_{max}^{gp}g_1}{K_m^{gp} + g_1} \\ &- f_u(g_2, h_2) + G_{in}, \end{aligned} \quad (2.13)$$

where

$$\begin{aligned} f_u(g_2, h_2) &= U_b \left(1 - \exp \left(\frac{-g_2}{C_2} \right) \right) \\ &+ \frac{g_2}{C_3} \left(U_0 + \frac{(U_m - U_0) \left(\frac{h_2}{C_4} \right)^\beta}{1 + \left(\frac{h_2}{C_4} \right)^\beta} \right), \end{aligned} \quad (2.14)$$

We consider the following nonlinear feedback rates:

$$u_1 = \frac{G_m}{1 + b_1 \exp(a_1(g_2 - 1000))}, \quad (2.15)$$

$$u_2 = \frac{R_m}{1 + b_2 \exp(a_2(C_1 - g_2))} \quad (2.16)$$

The state variables of this simplified Liu-model are the following: p_i , the plasma hormones, h_i the cellular hormones and r_i the hormone-bound receptor concentrations, where $i = 1, 2$ stand for glucagon and insulin, respectively. The variable g_1 represents blood glycogen and g_2 blood glucose levels, the latter being the measured output. The model contains (in a simplified form) the hormone dynamics, glycogen-glucose transition in the liver, furthermore, the insulin-independent and insulin dependent utilization ($f_u(g_2; h_2)$) of glucose. Feedback is incorporated in the glucose-dependent hormone infusion rates, u_i . The parameters a_i and b_i denote the reaction rates in glucagon and insulin dynamics. For the model parameter

2. METHODS

State variables	Description	Focused parameters	Description
h_1	Cellular glucagon concentration	$\mathbf{a}_1$	Glucagon transitional rate
h_2	Cellular insulin concentration	$\mathbf{b}_1$	Insulin transitional rate
p_1	Plasma glucagon concentration	$\mathbf{k}_1$	Feedback gains for glycogen-glucose transition
p_2	Plasma insulin concentration	$\mathbf{k}_2$	
r_1	Hormone-bound glucagon receptor concentration	$\mathbf{k}_3$	
r_2	Hormone-bound insulin receptor concentration	$\mathbf{C}_2$	Crucial parameters of glucose utilization
g_1	Blood glycogen level	β	
g_2	Blood glucose level		

Table 2.1: State variables and parameters of the simplified Liu-model.
Parameters with linear (a_1, b_1, k_1, k_3) , and nonlinear (k_2, C_2, β) dependencies

Known parameters	Description	Known parameters	Description
a_2, a_3, a_4, a_5	Glucagon transitional, degradation and association rates	K_m^{gp}	Michaelis-Menten constant of glycogen phosphorylase
b_2, b_3, b_4, b_5	Insulin transitional, degradation and association rates	K_m^{gs}	Michaelis-Menten constant of glycogen synthase
R_1^0	Total concentration of glucagon receptors	V, V_p	Volume of cellular and plasma insulin
R_2^0	Total concentration of insulin receptors	U_b, U_0, U_m	Max. velocity of the different glucose utilizations
V_{max}^{gp}	Max. velocity of glycogen phosphorylase	G_m	Max. glucagon infusion rates
V_{max}^{gs}	Max. velocity of glycogen synthase	R_m	Max. insulin infusion rates

Table 2.2: Known parameters of the simplified Liu-model.

These parameters remain fixed during the model fitting

estimation we choose seven of the parameters of the model based on both their uncertainty and biological importance. These are the following: the hormone transitional rates, a_1 and b_1 the feedback gains for glycogen-glucose transition, k_i ($i = 1, 2, 3$) and the two most crucial parameters of glucose utilization, C_2 and β . Three from these coefficients, namely k_2 , C_2 and β cause a nonlinear dependence, the other four, such as a_1 , b_1 , k_1 and k_3 are linearly depending.

2.3 Brief description of the applied toolboxes

2.3.1 GenSSI Toolbox

Both of the Toolboxes were developed by the (Bio-) Process Engineering Group of the IIM-CSIC Marine Research Institute, Vigo, Spain. The GenSSI received its name from 'Generating Series' and 'Structural Identifiability' expressions. It offers an easy to use technique for studying structural identifiability, which is done by computing the generating series using symbolic calculations, through iteratively computing the Lie derivatives of the analytic output. The derivatives can be calculated up to an arbitrary degree defined by the user as one of the script inputs. The necessary number of successive differentiations heavily depends on the structure of the investigated system, but typically four or five is sufficient. The result, so the complete and reduced identifiability tableaus are 0-1 matrices (also plotted within the MATLAB environment as black and white block-images representing of the Jacobian of the non-zero generating series coefficients. Some of the useful features of GenSSI Toolbox [3]:

- The Toolbox is applicable for a whole class of non-linear models.
- Computational information is displayed at each step, and not 'all or nothing' response.
- The use of *tableaus* is a very efficient way to summarize the information about the parameters.
- Problems of low memory can be handled by the user, through the manipulation of the Lie derivation order.
- Testing of structural local identifiability is also incorporated in the toolbox for the cases when the model is not globally identifiable.

2. METHODS

2.3.2 AMIGO Toolbox

AMIGO is a toolbox which covers most steps of the identification procedure: sensitivity analysis, ranking of parameters, parameter estimation, identifiability analysis and optimal experimental design. So the structure and available functions of the toolbox are also much more diverse than of GenSSI. In this regard, it was more difficult to implement a selected model and to execute the given subtasks. The main beneficial properties of AMIGO are the following [4]:

- Maximum flexibility for the definition of models and observation functions. (It is able to handle models specified in Fortran or Matlab, as well as in the widely used SBML.)
- Multiple types of experimental noise conditions and different types of cost functions for parameter estimation and experimental design are available.
- Use of the Fisher Information Matrix (FIM) to asymptotic analyses and to calculate OED.
- AMIGO includes the state of art initial value problem (IVP) and non-linear optimization (NLP) methods so as to handle a large variety of problems.

2.3.3 SUNDIALS Toolbox

This software received its name from the acronym of "*SU*ite of *Nonlinear and DI*fferential/*AL*gebraic equation Solvers", thus it is a family of software tools for integration of ODE and DAE initial value problems and for the solution of nonlinear systems of equations. It consists of CVODE, IDA, and KINSOL solvers, and variants of these with sensitivity analysis capabilities. The Toolbox (called *SundialsTB*) is a collection of matlab functions which provide interfaces to the sundials solvers [5, 6]. We use SundialsTB to calculate the time-dependent sensitivities of the model state variables using direct differential methods.

Chapter 3

Results

3.1 Building a MATLAB *SIMULINK* model

At first, our goal is to implement the system in MATLAB (the Simulink model in Figure 3.2) In the next step, we examine the interactions between the system's states drawing the structure graph of the variables of the equations. In the graph, each node represents a state variable; each graph edge corresponds to an 'influence' of variables among the differential equations.

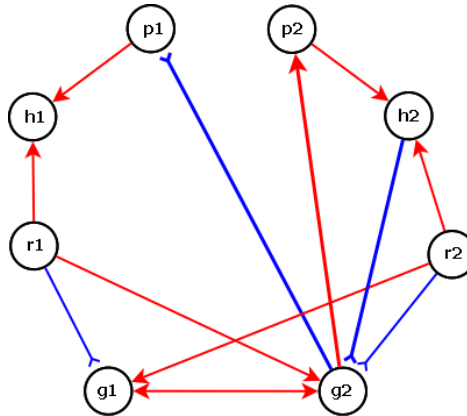


Figure 3.1: Structure graph - The structure graph represents the interactions between the system's states. Red and blue arrows denote excitatory (+) and inhibitory (-) influences between the state variables, respectively.

For experimental input we use data from the same article [2]. Unfortunately, the quality of used measurements is not perfect - and not sufficiently informative -, because it is executed as a Chinese nutritional experiment, namely: the investigation of whether parboiled rice or normal rice is better for digestion. In addition, the sampling of the blood glucose level contains only 11 averaged data points. In the near future I want to explore a 'true to nature', diabetic experiment designed directly for our purposes and compare with my simulations.

3. RESULTS

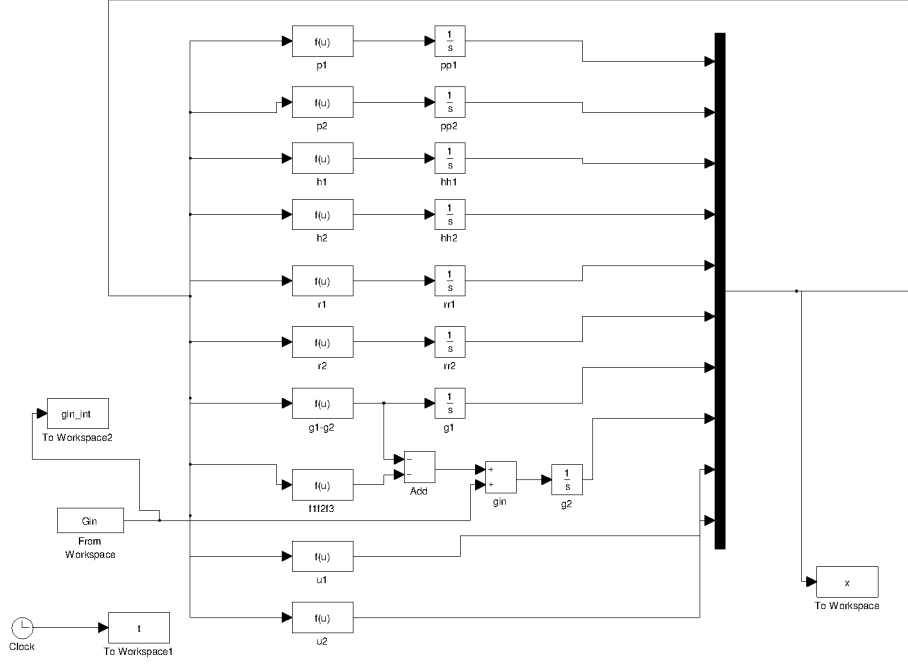


Figure 3.2: SIMULINK model - Simplified mathematical model of blood glucose control system is constructed by using the Matlab *SIMULINK* Toolbox

3.1.1 Results for structural identifiability, obtained *tableaus*

After all, we examine the (earlier described) identifiability of this improved and simplified model. Our selected method is the *Generating Series* technique, for computing we are going to use the functions in the GenSSI Toolbox [3]. In the following figures we can see the results for the linearly depending subset of the selected parameters (a_1, b_1, k_1, k_3) . A black square at the coordinates (i, k) indicates that the corresponding non-zero generating series coefficient i depends on the parameter θ_k . Eventually, it is surprising that each of these coefficients were structurally globally identifiable. In Figure 3.3, we can see firstly, the complete identifiability tableau for these four parameters with linear dependencies. Now the rank is complete, the program will not consider another derivative and at least structural local identifiability is guaranteed. Then the Figure 3.4 shows first order reduced identifiability tableau, which helps to compute the corresponding parameters, until the remaining identifiability tableau cannot be reduced anymore. The remaining parameters are computed, if possible. Step by step, when a row has just one non-zero element in the reduced identifiability tableau, we eliminate it from the tableau, and the corresponding parameter is structurally globally identifiable. In our case, all of these four parameters are such (See Figure 3.3 and Figure 3.4).

3.2 Fitting the model to experimental data

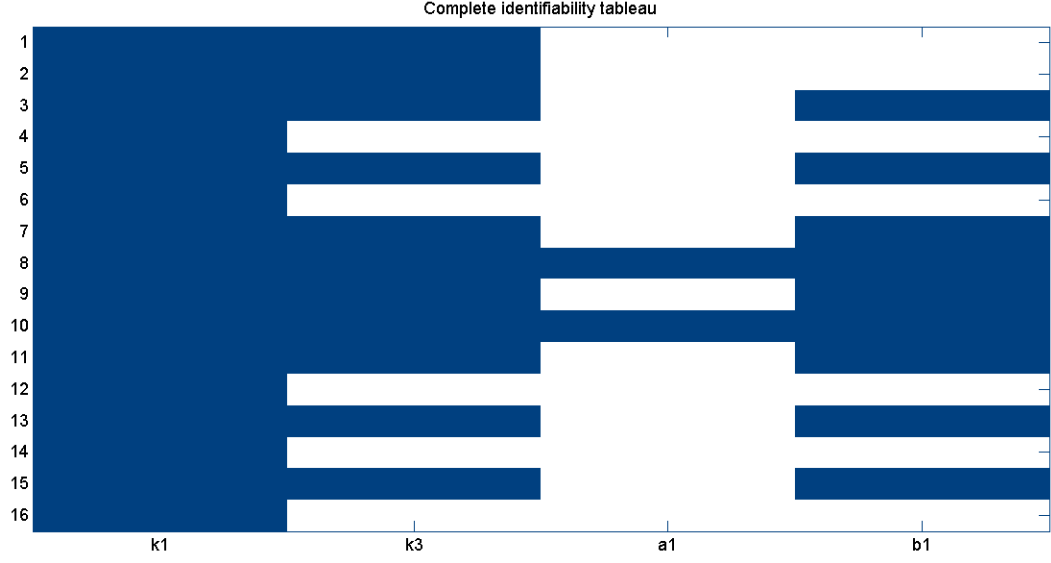


Figure 3.3: First (complete) identifiability tableau - The complete rank means at least structural local identifiability is guaranteed.

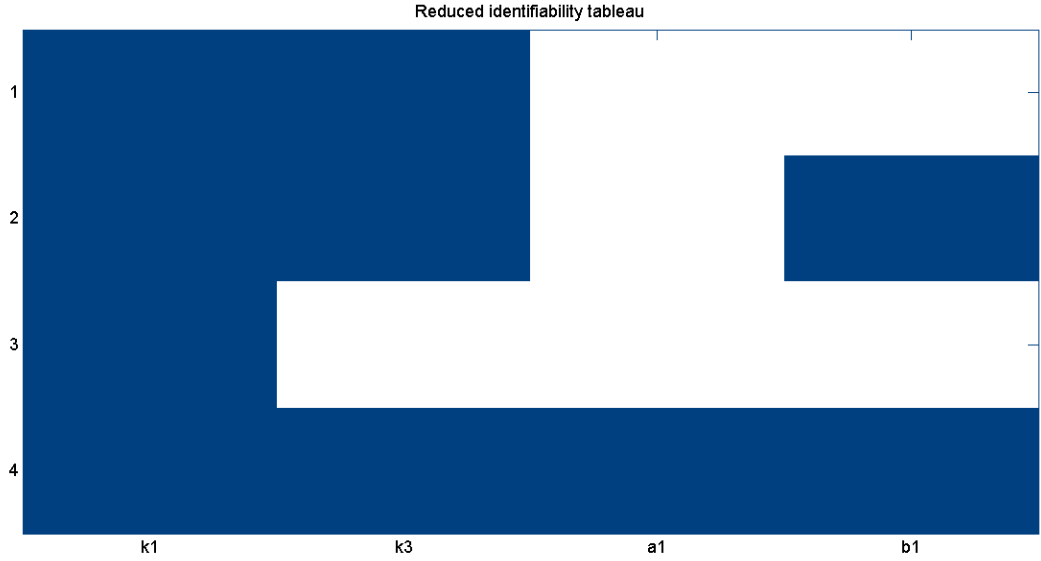


Figure 3.4: Reduced identifiability tableau - All of the four linearly depending parameters are structurally globally identifiable.

3.2 Fitting the model to experimental data

During the investigation of the model we choose seven of the model parameters based on both their uncertainty and biological importance. These seven have no exact data or reliable measurement in literature, but they have remarkably high influence on the output, since their biological relevance. These are the following: the hormone transitional rates, a_1 and b_1 the

3. RESULTS

feedback gains for glycogen-glucose transition, k_i where ($i = 1, 2, 3$); and the two most crucial parameters of glucose utilization, C_2 and β . Three from these coefficients, namely k_2 , C_2 and β cause a nonlinear dependence, the other four are linearly depending. Because of this, the starting task is to separate the two different behavioural types of parameters in the model, to fix the actual type and to estimate the corresponding others. Since the initial values of the state variables have also very important role in the model output, we also may consider the initial values as nonlinear parameters to be estimated. The so called objective functions are able to map the parameters onto fit indices: for each combination of parameter values, the predictions are computed, and the fit to the data is visualized. Our selected parameter estimation cost function is the standard normed quadratic error ratio (alias the ration of the two *Euclidean* norms) between the experimental data taken from literature and the simulated output:

$$\|x_i\|_2 = \sqrt{\sum_{j=1}^n |x_i|^2} \quad (3.1)$$

The value of the estimation objective function is finally 3.7 % lower (i.e. the goodness of fit is improved by 3.7 %) compared to the original fit taken from literature. In Figure 3.6, we receive the following fit with only linearly depending parameters considered for parameter estimation (For the exogenous glucose input, see Figure 3.5).

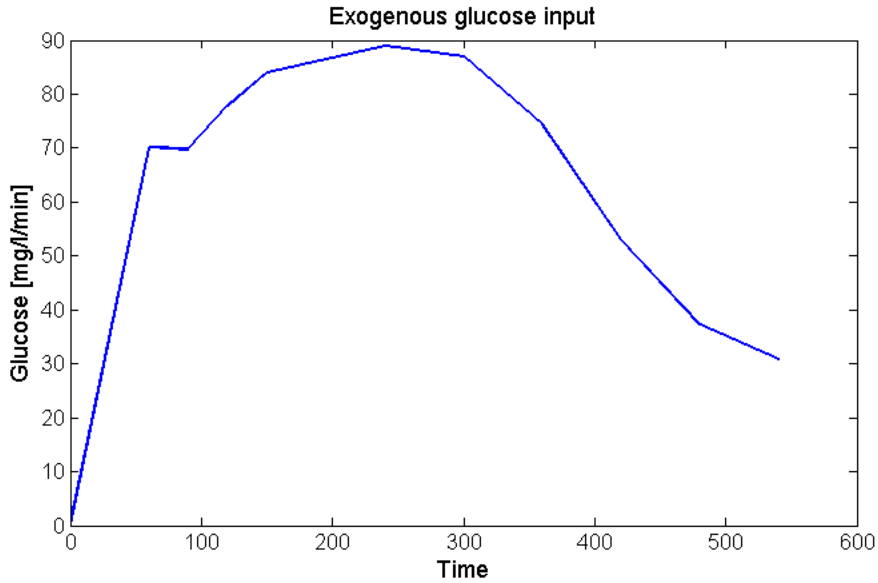


Figure 3.5: Glucose input - The exogenous glucose input for the model fitting (mg/l/min) (based on the experimental data of [2].)

3.2 Fitting the model to experimental data

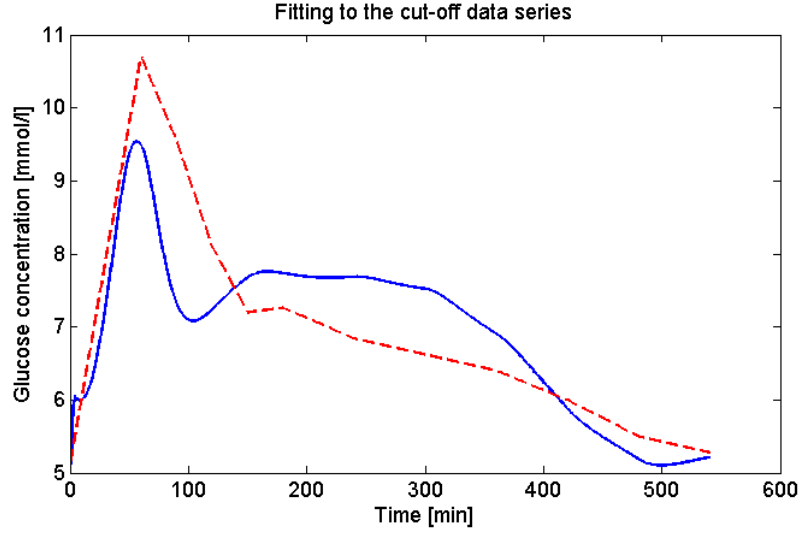


Figure 3.6: Fitted model to the experimental values - Red curve: Interpolated experimental values, in blue: simulated model output (the concentration of the blood glucose level is measured in mmol/l).

It is easy to see, that the quality of fit with sparse data values (from 11 data points) is worse, than the fit with interpolated values (See in Figure 3.7). Moreover, as our output is a smooth differentiable function in contrast to sparse discrete data points connected by linear lines, fitting to a smooth experimental output seemed more realistic. Therefore we apply so called spline interpolations (using cubic splines) to approximate the unknown measurements.

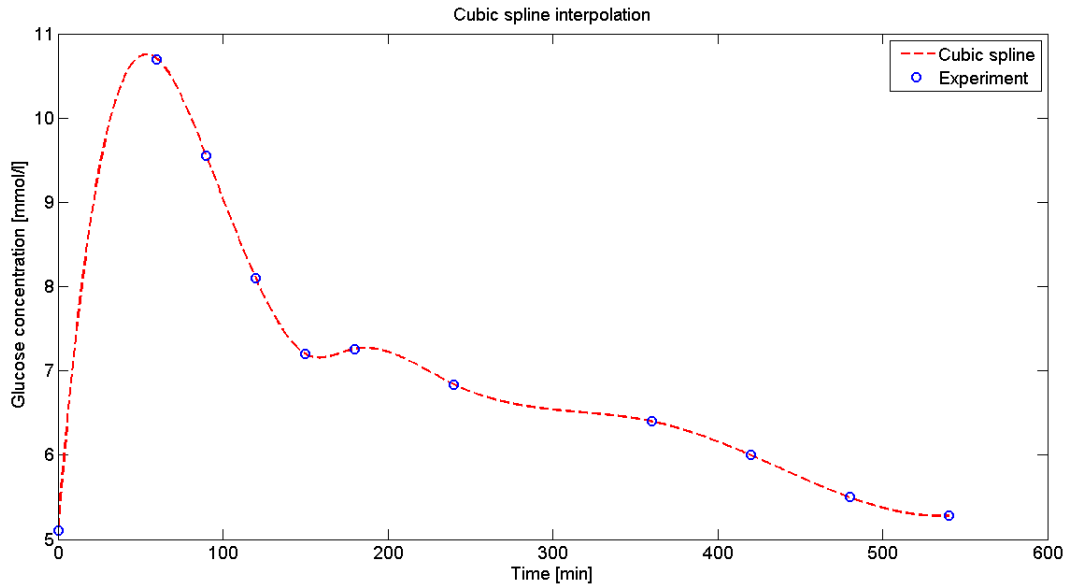


Figure 3.7: Spline - Cubic spline interpolation to smooth the intermittent or linearly interpolated experimental data.

3. RESULTS

Another issue which has to be considered is that there might be errors in the experimental data. We can see in Figure 3.7 (around 200 minutes) a local maximum. Since the attenuation of glucose level is thought to be a monotonically decreasing function of time, that upstroke might be an experimental artefact. However, having a slight down-stroke (at $t = 150$) after the global maximum (at $t = 80$) is explained by the sudden increase in insulin production (causing glucose uptake from blood), and the following upstroke can be seen as getting back to a decreasing trajectory starting at $t = 80$. Moreover, the data points are actually obtained as an average of 8 independent measurements, all showing the aforementioned features.

3.2.1 Iterative process of parameter estimation

To summarise, we managed to improve the model with small modification to make a better fit than the previously published result. Furthermore, the parameter estimation procedure is an iterative process, where θ_1 (linear dependencies) is estimated using a least squares procedure, while θ_2 (nonlinear dependencies) is estimated by the pattern search minimization method.

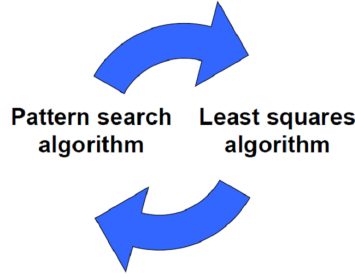


Figure 3.8: Parameter estimation - The estimation of the model parameters is implemented as an iterative process by fixing and estimating the linearly and nonlinearly depending subsets, alternately.

Pattern search algorithm (PSA) is a family of numerical optimization methods that do not require the gradient of the problem to be optimized. This fact is very important, first of all in the case of nonlinear systems, hence PSA can be used on functions that are not continuous or differentiable. Such optimization methods are also known as direct-search, derivative-free optimization, *etc.* The pseudo-code of the PSA is shown in the next Figure (3.9) [16]:

```

Let  $x_0 \in \mathbb{R}^n$  and  $\Delta_0 > 0$  be given.
For  $k = 0, 1, \dots$ , until convergence do:
  1. Compute  $f(x_k)$ .
  2. Determine a step  $s_k$  using an unconstrained exploratory moves algorithm.
  3. If  $f(x_k + s_k) < f(x_k)$ , then  $x_{k+1} = x_k + s_k$ . Otherwise  $x_{k+1} = x_k$ .
  4. Update  $C_k$  and  $\Delta_k$ .
  
```

Figure 3.9: Pattern search algorithm - Generalized case for unconstrained minimalization

3.3 Model validation using OGTT experimental data

In order to check our model estimation, we have to distinguish between two different types of methods, namely the verification and validation techniques. In the context of computer simulation verification of a model is the process of confirming it is correctly implemented with respect to the conceptual model, in other words, it matches specifications and assumptions acceptable for the given purpose of application. In our case, it means that the build-up of the estimated model is correct, and the parameters are structurally identifiable. In contrast, validation checks the accuracy of the model's representation of the real system. Validation is usually achieved through the calibration of the model, an iterative process of comparing the model to actual measurements and using the differences (residuals) between them. After that, we can perform a hypotheses testing (in general, t -tests are used) to confirm our results within a given confidence level.

Our improved model is validated using a new input based on the widely used oral glucose tolerance test [17]. The glucose tolerance test also referred to as either the OGT test or OGTT, is a method which can help to diagnose instances of diabetes mellitus or insulin resistance. The test is used to determine whether the body has difficulty metabolising intake of sugar/carbohydrate. The patient is asked to take a glucose drink and their blood glucose level is measured before and at intervals after the sugary drink is taken.

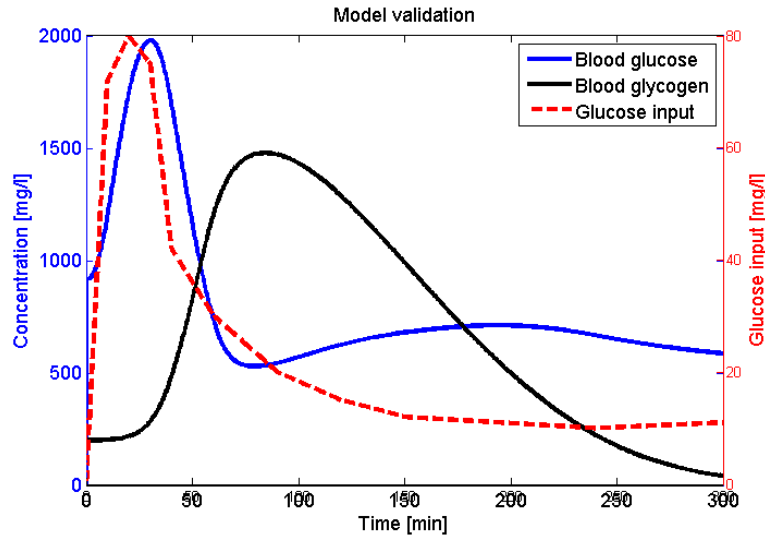


Figure 3.10: OGTT levels - Glucose input, glycogen (g_1 state variable) and blood glucose levels (g_2 state variable).

The model output reproduces characteristic features of measurements on healthy subjects, such as a down-stroke in glucose level (g_2 state variable) due to the temporary increase of insulin (which can lead to a hypoglycaemic state in patients with reactive hypoglycaemia).

3. RESULTS

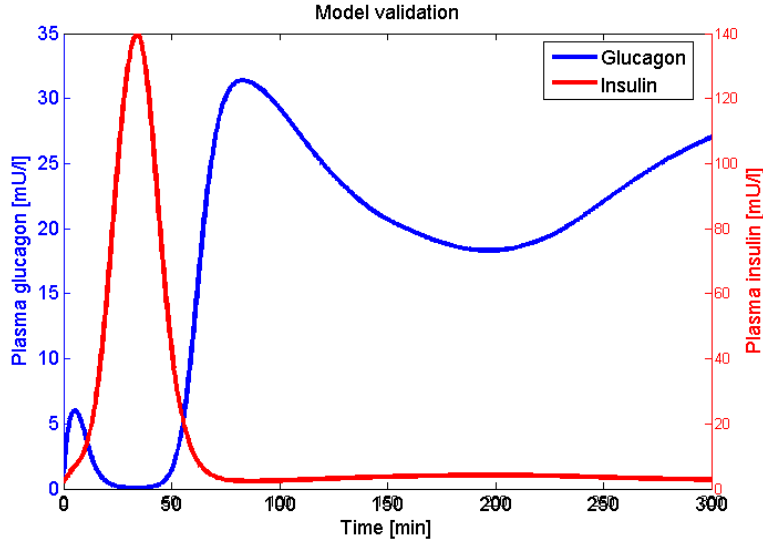


Figure 3.11: OGTT levels - Glucagon (h_1 state variable) and insulin hormone levels (h_2 state variable).

3.4 Additional results for practical identifiability

3.4.1 Simulating the model in AMIGO Toolbox

After that we have presented the structural properties of the model, we continue the further analysis implementing the BGCS to the AMIGO Toolbox. The toolbox offers several dynamic simulation functions to solve the system dynamics under given values of model unknowns and given experimental schemes. In our latest work, we concentrate on the 'sensitivity analysis' and 'rank of parameters' tasks. The rankings of parameters can assess their influence in the observables (*e.g.* in the blood glucose concentration levels). Results may be analysed for each experiment or for the whole experimental scheme. In our case we have only one real experimental time series, which was adopted from the literature [12]. In the next two figures we see the model outputs for two different inputs, first was 'sustained', the second is 'pulse-down' type of stimulation (first subplots). All of the state variables are presented on the following eight subplots (g_2 is shown on the last subplot). The second experimental set-up has very similar output dynamics as the earlier described OGTT behaviour taken from the literature [17]. These *in silico* simulations are useful as they help to understand what type of input is supposed to produce a distinguishing output most informative of the model. Experimentally implementing such inputs then can lead to measurements based on which the model will become more easily estimated.

3.4 Additional results for practical identifiability

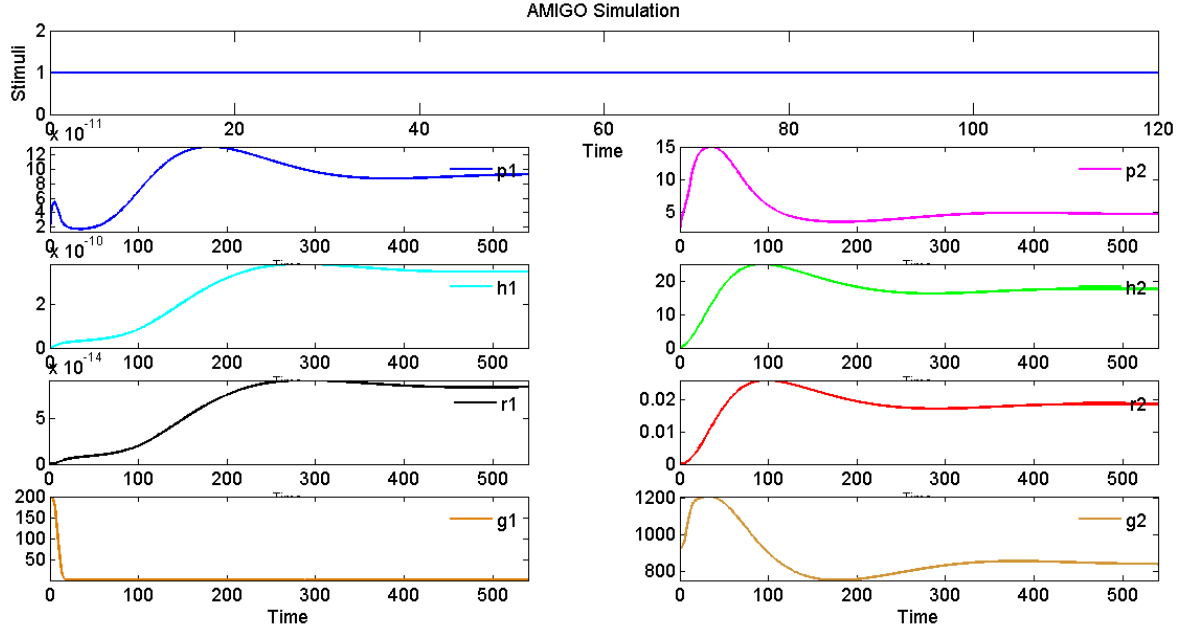


Figure 3.12: AMIGO Simulations - Changes of state variables with 'sustained' type of input. Blood glucose level changes (g_2) is indicated with brown colour on the last subplot (this behaviour is similar to what we have seen in the literature [7, 8]).

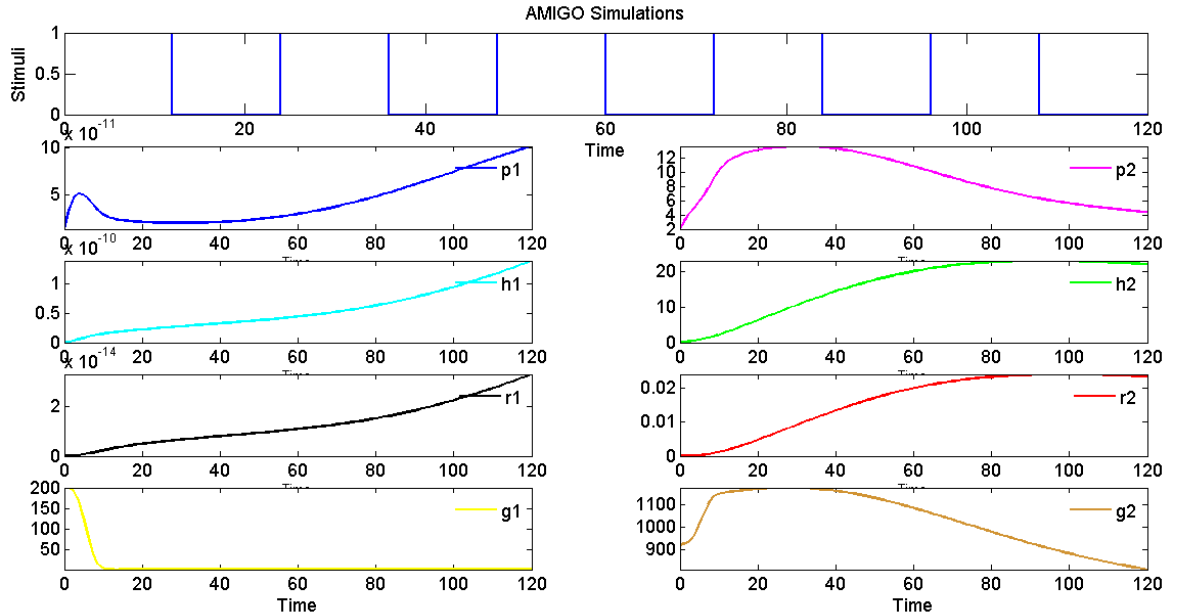


Figure 3.13: AMIGO Simulations - Model outputs for 'pulse-down' type of input. Blood glucose level changes (g_2) are indicated with brown colour on the last subplot. Now we can see something new in its behaviour.

3. RESULTS

3.4.2 Ranking of unknowns

Observables depend differently on each parameter and this can be used to rank the parameters in order of their relative influence on model predictions. There are several indexes to rank the parameters. However different criteria may lead to different results, but in the successful case, all criteria lead to the same conclusions even though the relative order of the most relevant parameters may slightly vary from one criterion to the other (Figure 3.14). In practice the so called δ_{MSQR} is probably the most widely used, which uses local parametric rankings. Of course, the values of the parameters are not known a priori, and even when optimally computed, optimal values are subject to uncertainty, which depends on the type of experiments and the properties of experimental noise. We note that in the case of 'lack of structural identifiability', global ranking may be used to make decisions as to reformulate the model. If we fix the less relevant parameters, we can improve either practical or structural identifiability [13].

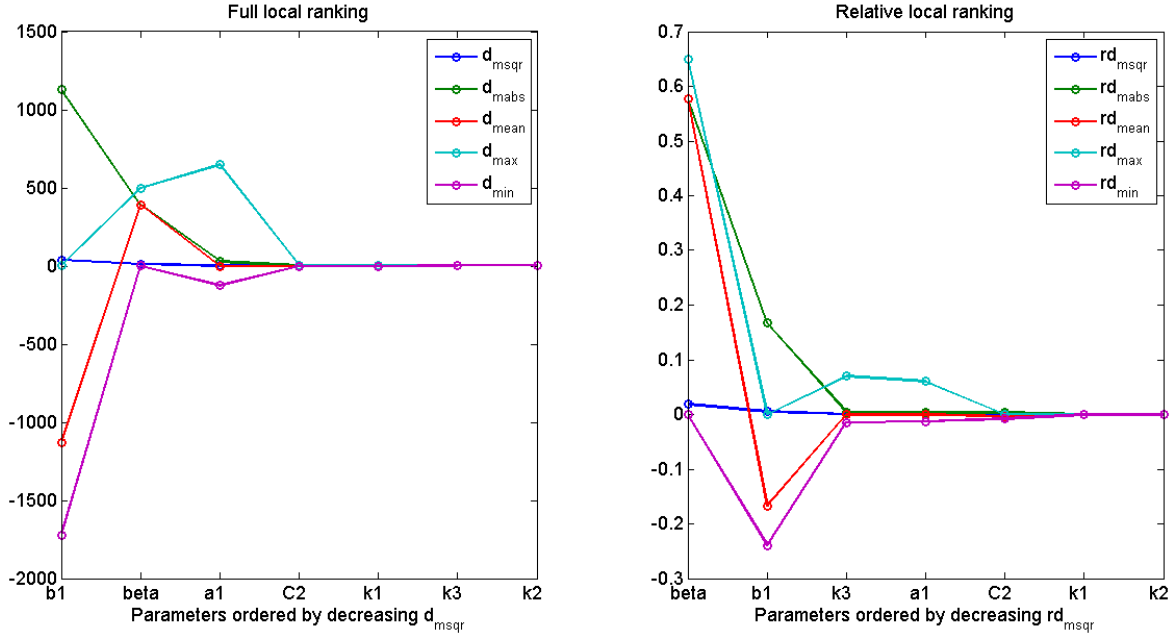


Figure 3.14: Ranking of unknowns - Local relative rank of parameters. Results obtained for the nominal value of parameters and the given experimental scheme. On the 'x' axis we can see the ordered parameters respect to their relevance (firstly, β and b_1 have governing role).

The next figure also reveals that there are some parameters (namely: β , a_1 and b_1) which are more clearly influencing the observables. We show the results of the so-called relative sensitivity analysis methods, whereas we chose our experimental output (blood glucose level, g_2) as the reference point (termed obsV) for the analysis. It is clear, that β and b_1 are the most relevant parameters for the Liu-model.

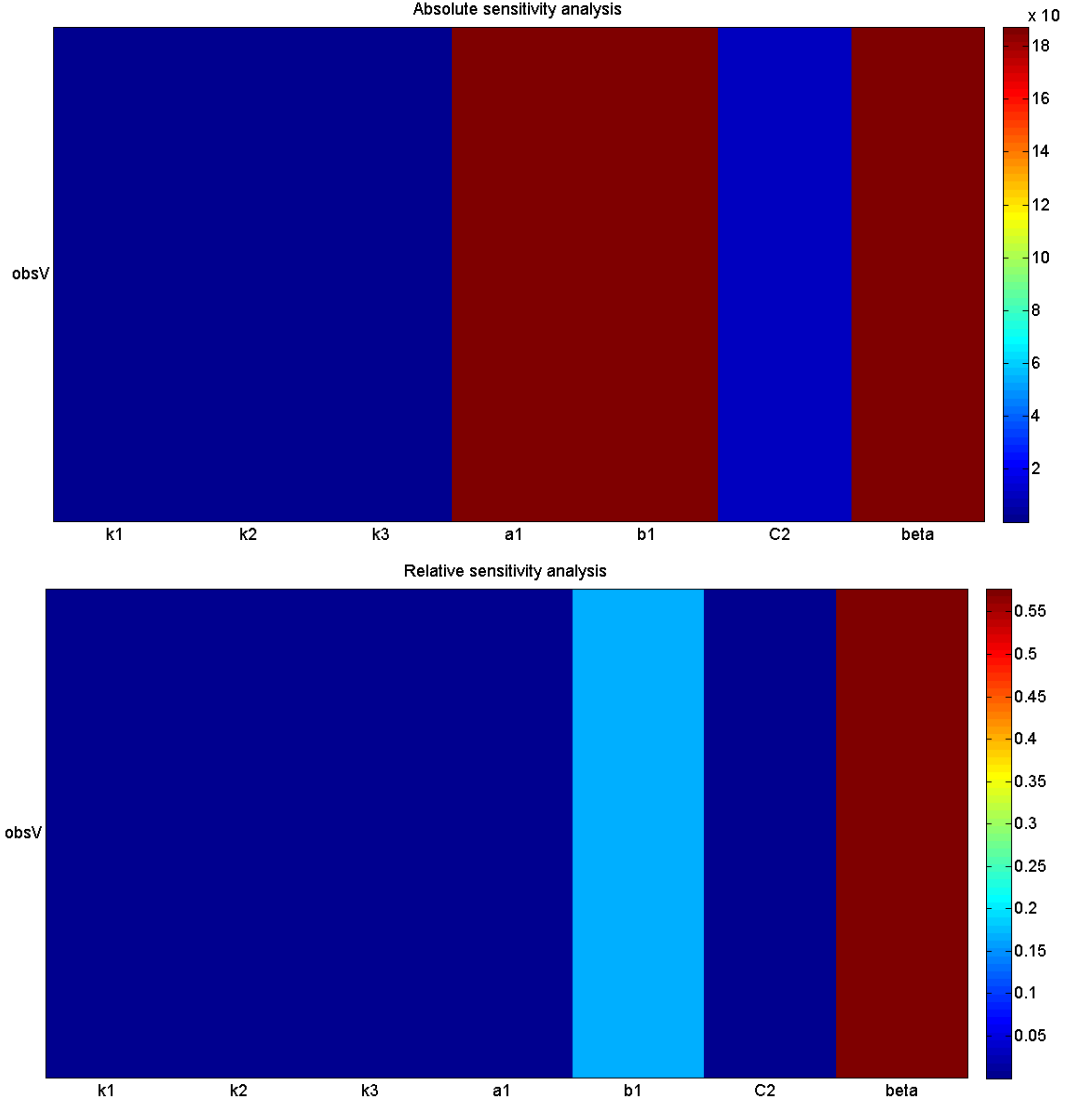


Figure 3.15: Absolute and relative sensitivities calculated with AMIGO Toolbox (obsV is adjusted for the g_2 variable).

3.5 Time-dependent sensitivity analysis

In the case of systems biological models, many parameters are difficult and sometimes even impossible to measure accurately with experiments. Some of the parameter values usually have large variations using different experimental conditions. Thus, our confidence on the model predictions is limited due to the uncertainties of the parameters [18]. As we mentioned earlier, sensitivity analysis is a technique to determine how the fluctuations in mathematical model outputs belong to the variations in the model inputs (namely the parameters and initial conditions). In other words, sensitivity analysis is used to quantify the parameter impacts on the experimental observations. After these calculations we are able to refine those parameters

3. RESULTS

which contribute most to the variation, and in the same time, we might reduce a lot of experimental effort and increase the predictive accuracy [19]. This information extracted from sensitivity analysis can be useful in both an 'understanding' context, suggesting hypotheses about mechanisms in a biological system, and a 'design' or 'control' context, suggesting how we may modify the system to produce certain behaviours or to hold the output in a certain tolerance scheme. And finally, we will see in the following section, that sensitivity analysis is valuable for model reduction and parameter estimation as well [19].

Among several types of sensitivity calculations we chose the so called direct differential method to calculate the time-dependent sensitivities for all of the variables respect to each parameter (resulting fifty six plots). The theory of the derivation is the following [18]:

$$S = \frac{dx}{d\theta} \quad (3.2)$$

$$\frac{d}{dt} \frac{dx}{d\theta} = \frac{d}{d\theta} \frac{dx}{dt} = \frac{d}{d\theta} f(x, \theta, u) \quad (3.3)$$

Note that $\dot{x} = \frac{dx}{dt} = f(x, \theta, u)$ is the right side of the model ODE system, using the chain rule of differentiation we get the following expression:

$$\frac{d}{d\theta} f(x, \theta, u) = \frac{df}{dx} \frac{dx}{d\theta} + \frac{df}{d\theta} \quad (3.4)$$

And if we introduce a new variable $\frac{dx}{d\theta} = x_\theta$, we receive a new ODE system for x_θ , where $\frac{df}{dx} = J$ is in fact the Jacobian, and $\frac{dx}{d\theta} = S$ is the sensitivity coefficient:

$$\dot{x}_\theta = \frac{df}{d\theta} + J \cdot S \quad (3.5)$$

This ODE system was implemented in MATLAB and the sensitivities were calculated with SUNDIALS Toolbox using the CVODES solvers [5, 6].

The advantage of these solvers is that they can calculate the sensitivities of different variables with respect to one parameter simultaneously. In contrast, the disadvantage of the solvers (working with direct differential method) is that the Jacobian matrix (J) needs to be defined and this step can be time-consuming. But in our case, the computational time was surprisingly small (less than one second).

In Figure 3.16 it seems interesting even at first sight that the decays of the parameters are different but in some cases can be similar by pairs. This consideration helps us to build new sets of parameters based on their similar sensitivity dynamics. The time-scale naturally divides the parameters into three groups: parameters with transient (k_3, a_1), middle-range

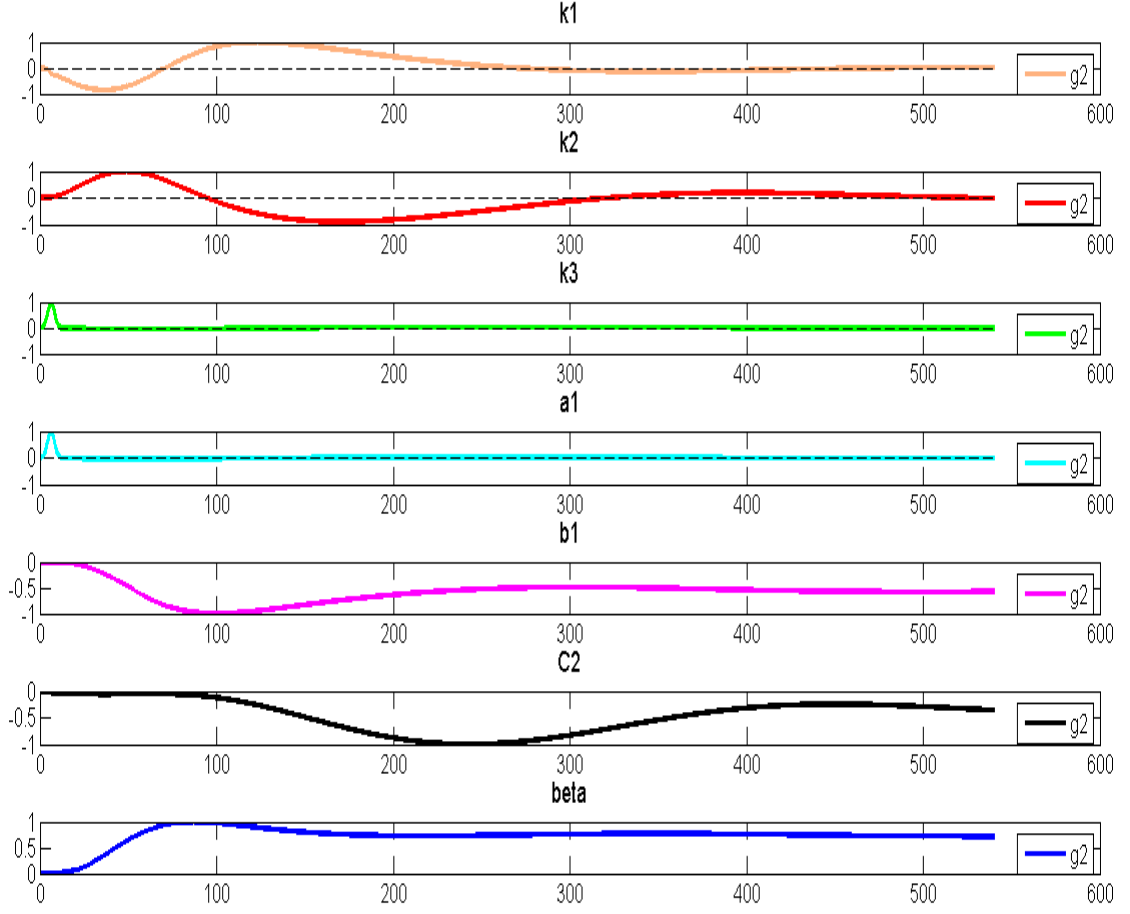


Figure 3.16: Sensitivities are calculated for the g_2 (blood glucose level) state variable, respect to the estimated parameters over time.

(k_1 , k_2) and long-term (b_1 , C_2 , β) impact. This is consistent with the biological assumption that the overall response consists of subprocesses where different biological parameters have temporary significance. Strictly glucagon connected parameters have non-zero sensitivities only up to 10 minutes, corresponding to an initial glucagon release, which is quickly shut down by the increased blood glucose level. On the other hand, parameters related to insulin-dependent glucose uptake have negligible impact on this time-scale but exert influence on long-term dynamics and steady-state concentrations.

3.5.1 Improved parameter estimation technique

The information extracted from sensitivity analysis is valuable for model reduction and parameter estimation. In the parameter estimation step we focus on the separation of the

3. RESULTS

parameters based on their sensitivities, and not on their linear or nonlinear dependencies. We suppose that if we have insensitive parameters at some time intervals, these parameters might be neglected to avoid problems arising from lack of identifiability. This consideration can build a tight and useful connection between sensitivity and identifiability.

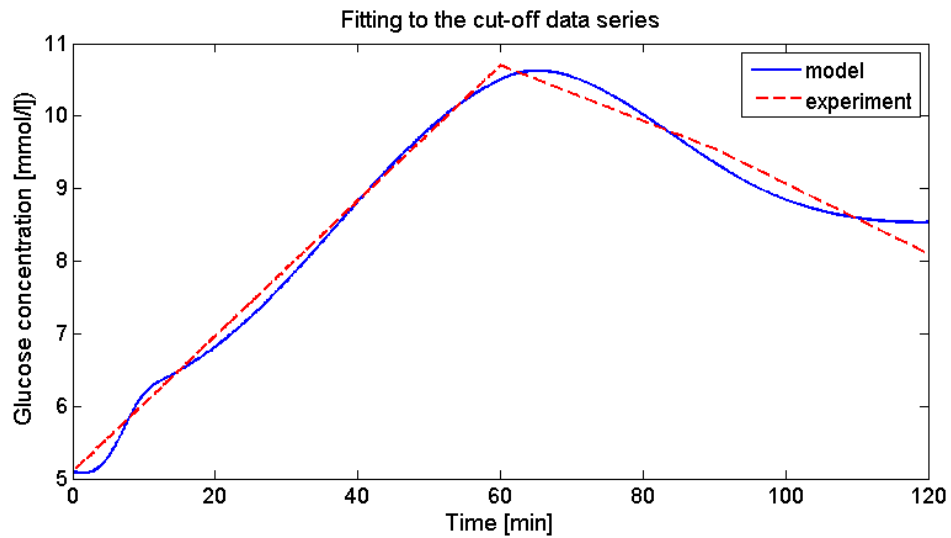


Figure 3.17: 'Cut-off' data fitting - 'Cut-off' data fitting to improve the estimation of the parameters with transient effect.

So for instances, if we fix all of the parameters except a_1 and k_3 (as in Figure 3.17), we may end up with a much better fit using the first part of the measurement and the transient parameters compared with the earlier estimated whole parameter set using the complete experiment.

Chapter 4

Discussion

4.1 Summary

In my student research project, I study the model identification theory of biological systems and after that, I implement in practice a model selected from literature. My results demonstrate that in the case of structural identifiability, the four considered linear parameters are globally identifiable, which is an interesting phenomenon in simulations of biological system. For the three nonlinearly depending parameters, locally identifiability is guaranteed at least. I also investigate some aspects of practical identifiability and optimal experimental design. Furthermore, I improve the model through several small modifications and made a better fit to published experimental data. Hereupon, I execute an extensive time-dependent sensitivity analysis with a direct differential method. The results show that we can separate different time scales among the effects of the parameters. These time scales have biological meaning as well. After all, we re-estimate the parameter sets using this newly known grouping principle and we utilise our findings for even better model fitting as in the literature previously shown. At the end of my work, I feel that I enquired some good and useful engineering knowledge and skills. In the recent past, we have presented two posters about our results of the structural identifiability (at the 11th Conference on Computational Methods in Systems Biology in Klosterneuburg, Austria [20]) and about the time-dependent sensitivity analysis (at the 11th International Workshop on Computational Systems Biology, Lisbon, Portugal [21]).

4.2 Prospects for the future

My future plans include the development of a more precise model and the creation of several different, even hypothetical input data sets to further examine the behaviour of the system. I would like to use not only healthy but pathological datasets as well. Besides this, I also plan

4. DISCUSSION

to determine whether any of the three nonlinear parameters are at least locally structurally identifiable (according to my preliminary results unfortunately all of them seem to be non-identifiable). Lastly, I also attempted to implement a parallel-sequential experimental scheme to improve identifiability, calculating the new, optimal experimental set-up and to implement the Monte-Carlo based robust identifiability analysis, both of them for the purpose of efficient parameter estimation. I had some – currently unsolved – difficulties with the too big computational costs and running times and I did not receive appreciable results so far. But in the forthcoming days, I plan to ask for an access on a high-capacity server and continue my student research work on it. Finally, in the near future we attempt to design a nonlinear controller to the system, which later – implemented in a glucose monitoring and insulin pump device – could prove useful in clinical practice and research. I hope that my further results (from the motivated continuation of this project) might contribute to developments in the theory and practice of blood glucose modelling and controlling.

Acknowledgements

I express my deep gratitude to my supervisor, Dr. Gábor Szederkényi. I thank him that he has fully promoted my student research work. I wish to express my thanks to my 'colleagues' and friends, in particular to Eszter Lakatos and to Zoltán Tuza for their excellent technical assistances and their encouragements.

Abbreviations

AMIGO	- Advanced Model Identification using Global Optimization
ATP	- Adenosine Triphosphate molecule
BGCS	- Blood Glucose Control System
DAE	- Differential algebraic equation
FBGT	- Fasting Blood Glucose Test
FIM	- Fisher Information Matrix
GenSSI	- Generating Series approach for testing Structural Identifiability
IVP	- Initial Value Problem
LS	- Least Squares
ML	- Maximum Likelihood
NLP	- Nonlinear Programming
ODE	- Ordinary Differential Equation
OED	- Optimal Experiment Design
OGTT	- Oral Glucose Tolerance Test
PDE	- Partial Differential Equation
PE	- Parameter Estimation
PSA	- Pattern Search Algorithm
SA	- Sensitivity analysis
SBML	- Systems Biology Mark-up Language

List of Figures

1.1	Glucose regulation	3
2.1	The flow diagram of the place of the identification step in the model building procedure and the parts of identification method [4].	7
3.1	Structure graph	15
3.2	SIMULINK model	16
3.3	First (complete) identifiability tableau	17
3.4	Reduced identifiability tableau	17
3.5	Glucose input	18
3.6	Fitted model to the experimental values	19
3.7	Spline	19
3.8	Parameter estimation	20
3.9	Pattern search algorithm	20
3.10	OGTT levels	21
3.11	OGTT levels	22
3.12	AMIGO Simulations	23
3.13	AMIGO Simulations	23
3.14	Ranking of unknowns	24
3.15	Absolute and relative sensitivities calculated with AMIGO Toolbox (obsV is adjusted for the g_2 variable)	25
3.16	Sensitivities are calculated for the g_2 (blood glucose level) state variable, respect to the estimated parameters over time.	27
3.17	'Cut-off' data fitting	28

References

- [1] O.-T. Chis, J. R. Banga, and E. Balsa-Canto, “Structural identifiability of systems biology models: a critical comparison of methods,” *PloS one*, vol. 6, no. 11, p. e27755, 2011. 1, 6, 7, 9
- [2] M. Korach-André, H. Roth, D. Barnoud, M. Péan, F. Péronnet, and X. Leverve, “Glucose appearance in the peripheral circulation and liver glucose output in men after a large 13c starch meal,” *The American journal of clinical nutrition*, vol. 80, no. 4, pp. 881–886, 2004. 2, 15, 18
- [3] O. Chiş, J. R. Banga, and E. Balsa-Canto, “Genssi: a software toolbox for structural identifiability analysis of biological models,” *Bioinformatics*, vol. 27, no. 18, pp. 2610–2611, 2011. 2, 9, 10, 13, 16
- [4] E. Balsa-Canto and J. R. Banga, “Amigo, a toolbox for advanced model identification in systems biology using global optimization,” *Bioinformatics*, vol. 27, no. 16, pp. 2311–2313, 2011. 2, 6, 7, 8, 14, 33
- [5] A. C. Hindmarsh, P. N. Brown, K. E. Grant, S. L. Lee, R. Serban, D. E. Shumaker, and C. S. Woodward, “Sundials: Suite of nonlinear and differential/algebraic equation solvers,” *ACM Transactions on Mathematical Software (TOMS)*, vol. 31, no. 3, pp. 363–396, 2005. 2, 14, 26
- [6] R. Serban and A. C. Hindmarsh, “Cvodes: the sensitivity-enabled ode solver in sundials,” in *ASME 2005 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. American Society of Mechanical Engineers, 2005, pp. 257–269. 2, 14, 26
- [7] W. Liu and F. Tang, “Modeling a simplified regulatory system of blood glucose at molecular levels,” *Journal of theoretical biology*, vol. 252, no. 4, pp. 608–620, 2008. 2, 3, 10, 11, 23

-
- [8] W. Liu, *Introduction to Modeling Biological Cellular Control Systems*. Springer, 2012, vol. 6. 3, 10, 23
- [9] B. Agar, M. Eren, and A. Cinar, “Glucosim: educational software for virtual experiments with patients with type 1 diabetes,” *27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 845–848, 2006. 4
- [10] I. Chou, E. O. Voit *et al.*, “Recent developments in parameter estimation and structure identification of biochemical and genomic systems,” *Mathematical biosciences*, vol. 219, no. 2, pp. 57–83, 2009. 5
- [11] G. Szederkényi, J. R. Banga, and A. A. Alonso, “Inference of complex biological networks: distinguishability issues and optimization-based solutions,” *BMC systems biology*, vol. 5, no. 1, p. 177, 2011. 6, 9
- [12] E. Walter and L. Pronzato, “Identification of parametric models,” *Communications and Control Engineering*, 1997. 6, 22
- [13] E. Balsa-Canto, A. A. Alonso, and J. R. Banga, “An iterative identification procedure for dynamic modeling of biochemical networks,” *BMC systems biology*, vol. 4, no. 1, p. 11, 2010. 8, 9, 24
- [14] R. N. Bergman, Y. Z. Ider, C. R. Bowden, and C. Cobelli, “Quantitative estimation of insulin sensitivity,” *American Journal of Physiology-Endocrinology And Metabolism*, vol. 236, no. 6, p. e667, 1979. 10
- [15] J. T. Sorensen, “A physiologic model of glucose metabolism in man and its use to design and assess improved insulin therapies for diabetes,” Ph.D. dissertation, Massachusetts Institute of Technology, 1985. 10
- [16] E. D. Dolan, R. M. Lewis, and V. Torczon, “On the local convergence of pattern search,” *SIAM Journal on Optimization*, vol. 14, no. 2, pp. 567–583, 2003. 20
- [17] W. Consultation, *Definition, diagnosis and classification of diabetes mellitus and its complications*. Part, 1999, vol. 1. 21, 22
- [18] Z. Zi, “Sensitivity analysis approaches applied to systems biology models,” *Systems Biology, IET*, vol. 5, no. 6, pp. 336–346, 2011. 25, 26
- [19] K. Rateitschak, F. Winter, F. Lange, R. Jaster, and O. Wolkenhauer, “Parameter identifiability and sensitivity analysis predict targets for enhancement of stat1 activity in

REFERENCES

- pancreatic cancer and stellate cells,” *PLoS computational biology*, vol. 8, no. 12, p. e1002815, 2012. 26
- [20] E. Lakatos, D. Mészéna, and G. Szederkényi, “Identifiability analysis and improved parameter estimation of a human blood glucose control system model,” in *LECTURE NOTES IN COMPUTER SCIENCE*, A. Gupta and T.A. Henzinger (Eds.): *CMSB 2013, LNBI 8130 Springer*, 2013, p. 248–249. 29
- [21] D. Mészéna, E. Lakatos, and G. Szederkényi, “Sensitivity analysis and parameter estimation of a human blood glucose regulatory system model,” in *11th International Workshop on Computational Systems Biology, Lisbon, Portugal (accepted)*, 2014. 29