Term project
TTK4550 Medical cybernetics

Implementation of a Metabolism Model for Insulin/Glucose Dynamics

Peter Martinius Stige

Supervisor:
Anders Lyngvi Fougner, ITK

Co-supervisor:
Steinar Sælid, Prediktor

Trondheim, December 18th, 2018

NTNU
Norwegian University of
Science and Technology

Faculty of Information Technology and Electrical Engineering
DEPARTMENT OF ENGINEERING CYBERNETICS

PROSJEKTOPPGAVE

Kandidatens navn: Peter Martinius Stige

Fag: Teknisk kybernetikk, fordypningsprosjekt (TTK4550)

Oppgavens tittel (norsk): Implementasjon av metabolismemodell for insulin/glukose-dynamikk

Oppgavens tittel (engelsk): Implementation of a metabolism model for insulin/glucose dynamics

Oppgavens tekst:

Based on Prediktor Medical's metabolism model for insulin/glucose dynamics, the goal of the project is to realize such a model for the following purposes:

- ❖ To be able to identify parameters in the model based on measured time series of glucose data over a period from a CGM such as the BioMKR, Freestyle Libre or others.
- ❖ To use this model as a predictive tool for the user to do "what if" analyses as well as to propose insulin amounts to apply based on planned meals.

The project shall focus on the following parts:

1. Describe the existing Prediktor Medical model and discuss the task of finding a suitable complexity level of a modified version for use and identification; the latter meaning which parameters to potentially identify for a given model type, assuming preadapted models for a person either healthy, having diabetes mellitus type 1 or 2. Modelling of different types of meals/food shall be a part of this.
2. Discuss the identifiability of a set of parameters for a given model using sensitivity analysis and singular value decomposition and/or the Fisher Information Matrix (FIM) approach for a given model. Discuss how excitations such as insulin injections, meals and exercise influences identifiability. Potential parameters to discuss/include are insulin sensitivity given as a parametrized function of time (cyclic on a 24-hour base) and meal class/type.
3. Based on an implementation of the sensitivity analysis and/or FIM approach, discuss the selection of associated excitations.
4. Make a Matlab simulation model and discuss/implement methods for parameter identification such as simplex based methods and Newton Raphson methods.
5. Test this on simulated measurement data.

Oppgaven gitt: 20. august 2018

Besvarelsen leveres innen: 18. desember 2018

Utført ved Institutt for teknisk kybernetikk

Veileder: Anders Lyngvi Fougner
Biveileder: Steinar Sælid, Prediktor Medical AS

Trondheim, 31. august 2018

Contents

List of Figures

List of Tables

Nomenclature

DM1   Diabetes Mellitus 1

DM2   Diabetes Mellitus 2

DS   Downhill Simplex

NR   Newton-Raphson

PB   Plackett-Burman

PM   Prediktor Medical

RSV   Right singular vector

SVD   Singular value decomposition

Abstract

For people diagnosed with diabetes mellitus, it is crucial to keep control of the plasma glucose concentration levels. One helpful tool can be to use a metabolism model to simulate and predict future glucose values based on meals, external insulin and exercise.

In this study, the identifiability of the GlucoPred metabolism model for glucose and insulin dynamics have been investigated. This model is developed by Prediktor Medical (PM), and used in the development of a non-invasive glucose measurement devise.

The existing GlucoPred model have been presented. That includes a description of the 14 state variables and their corresponding state transition equations. Also the 41 model parameters have been introduced.

The identifiability of the model have been investigated. This has been done using a sensitivity analysis approach. The influence of each parameter on the simulated output glucose measurements were used as an indication for identifiability of that parameter. Parameter sensitivity rank conditions were assessed by observing the sensitivity matrix singular values and right singular vectors, and used to suggest possible parameter sets to give identifiability. Different input combinations have been looked at for this, and differences between them discussed.

The estimation methods Downhill Simplex and Newton-Raphson with finite difference derivative approximations were tested on the parameters suggested through sensitivity analysis. Both the sensitivity analysis and the estimations were carried out for all three preadapted model types: healthy, DM1 and DM2.

1 Introduction

Diabetes Mellitus (DM) is a disease where defects in insulin action (Diabetes Mellitus type 2) or insulin secretion (Diabetes Mellitus type 1) leads to hyperglycemia. Worldwide, 425 million adults have diabetes, and out of those, 212 million are undiagnosed [1]. Precise insulin administration is key for humans with Diabetes Mellitus to control their glucose levels. Today the dosage of exogenous insulin is mainly done manually, with the patients injecting it subcutaneously, either with an insulin pen/syringe or via an insulin pump. Intra- and interday variability of physical properties of the patient [2; 3; 4; 5; 6], together with properties of different meals and foods, makes the task of dosing insulin a difficult exercise. Therefore, closed-loop systems for automatic control of blood glucose has been an area of extensive research for some years.

An important tool in these studies has been the use of insulin metabolism simulation models to study different aspects of the disease, without the use of more time consuming and expensive animal/human testing. These models consists of dynamic equations describing the insulin and blood glucose dynamics in the human body. It exists several different versions of these models. Examples are the UVa/Padova type 1 diabetes simulator [7], Bergman's minimal model [8; 9], Sorensens's model [10], Chase's model for critical care [11] and Cambridge's model [12]. The model to be analyzed in this article is the GlucoPred metabolism model, developed by Prediktor Medical.

To measure the blood glucose concentration, many people with diabetes use a continuous glucose monitor (CGM), while others also measure it by fingerprick measurements. The CGM has a small electrode placed under the skin that can measure the blood glucose continuously. The measurements are sent to another device, typically a smart phone or another devise with a screen, via a transmitter placed on the outside of the skin. This information is displayed to the user as a glucose curve on the screen, and is used by a person with diabetes to calculate/guess insulin dosages during the day.

CGM measurements can also be used to identify model parameters in insulin-glucose metabolism models. To do this, one must find the model parameters that makes the model simulation output look like the CGM measurement. This is a difficult task, if not impossible for most models, the reason being that one continuous measurement output is in most cases not enough to identify a bunch of model parameters. Therefore most models needs to be modified with a lower complexity level to become identifiable.

In this project, the GlucoPred model from PM will be described, and a suitable complexity level for parameter identification will be discussed. This means to discuss which parameters to potentially estimate for a given input. To do this, the metabolism model will be implemented, together with the sensitivity analysis calculations. Also, parameter estimations techniques will be presented, implemented and tested on the parameters from the sensitivity analysis. Sev-

eral different input combinations will be assessed and compared with respect to which parameters that can possibly be identified. This is done for all three model types, that is, healthy, DM2 and DM2.

In chapter 2, the basic theory needed for this project is presented, and chapter 3 presents the goal with this study. Chapter 4 contains the methodology used when doing the analysis and chapter 5 presents the results from it. In the chapters 6 and 7, discussion and conclusion is contained and suggestions for future work comes in chapter 8.

2 Theory

2.1 Metabolism models

There are a variety of different metabolism models that model the dynamics between glucose and insulin in the human body. They have different complexity levels, include different effects in their equations and are used differently dependent on what they are modelling. One example is a model mentioned earlier, the Bergman minimal model. It consists of three state variables, plasma glucose concentration $G(t)$, plasma insulin concentration $I(t)$ and insulin in remote compartment $X(t)$ [8]. It only models how glucose and insulin is injected into the blood, so for it to work as a complete glucose-insulin metabolism model, several things needs to be added, including for example a model of the digestive system and realistic exogenous insulin injection. This is one of the simplest models there is. A more complex model is the UVa/Padova type 1 diabetes simulator [7]. This model has 18 state variables, and is much more complex than the minimal model. This model is therefore used for more complex tasks than the minimal model.

A metabolism model contains a number of parameters representing different phenomena in the body. One example is the parameter S_i, usually representing insulin sensitivity in many metabolism models for glucose and insulin dynamics. These parameters have to be identified for the model to be able to represent a person realistically.

2.2 GlucoPred metabolism model from PM

The GlucoPred model is a metabolism model for glucose and insulin dynamics, aiming for simulation and prediction of plasma glucose in a subject. The model consists of 14 state variables. These are:

$$
\mathbf{x} =
\begin{bmatrix}
G_p \\
G_t \\
I \\
X \\
S_{R1} \\
S_{R2} \\
S_{S1} \\
S_{S2} \\
M_{sto} \\
M_{gut} \\
H \\
M_{lg} \\
Y \\
Z
\end{bmatrix}
\begin{array}{l}
\text{- Glucose concentration in plasma [mg/dL]} \\
\text{- Glucose concentration at measurement site [mg/dL]} \\
\text{- Insulin concentration in central compartment [mU/L]} \\
\text{- Insulin concentration in remote compartment [mU/L]} \\
\text{- Rapid Insulin concentration in SC compartment [mU/L]} \\
\text{- Rapid Insulin concentration in second SC compartment [mU/L]} \\
\text{- Slow Insulin concentration in SC compartment [mU/L]} \\
\text{- Slow Insulin concentration in second SC compartment [mU/L]} \\
\text{- Glucose in stomach contents [g]} \\
\text{- Glucose in gut contents [g]} \\
\text{- Glucagon concentration in plasma [pg/ml]} \\
\text{- Muscle and liver glycogen store [g]} \\
\text{- Exercise input [unitless]} \\
\text{- Exercise memory [unitless]}
\end{array}
$$

The most important variable here is G_p, which is the glucose concentration in blood plasma. This is what a person with diabetes is seeking to control. G_t is the subcutaneous blood glucose concentration, that is, the glucose concentration that is measured with for example a CGM.

I is the insulin concentration in the blood, and X is the insulin in the remote compartment, where the insulin is used.

S_{R1}, S_{R2}, S_{S1} and S_{S2} models the dynamics for injecting external insulin subcutaneously for rapid and slow working insulin. Time constants in the model represent how fast the rapid insulin and slow insulin is moving from the first compartments (S_{R1} and S_{S1}) to the second compartments (S_{R2} and S_{S2}), and from the second compartments to the blood.

M_{sto} and M_{gut} represents the glucose quantity in the stomach and in the gut. H is the glucagon concentration in the blood and M_{lg} is the quantity of glycogen that is stored in the liver and muscles.

Y and Z is the two exercise variables. Y is a value representing the heart rate above basal heart rate during exercise. Z is a variable representing recovery after the exercise session. Right after the exercise is finished, Z is at its highest, and it decays with time. When Z is higher, more blood glucose is used for glycogen storage in the muscles and liver. This is because the glycogen storage is empty, as much of it has been used as energy during the exercise.

The possible inputs to the model is:

$$
\mathbf{u} =
\begin{bmatrix}
u_{Ir} \\
u_{Is} \\
u_{hr} \\
u_{meal}
\end{bmatrix}
\begin{array}{l}
\text{- Rapid insulin injection [U]} \\
\text{- Slow insulin injection [U]} \\
\text{- Heart rate [BPM]} \\
\text{- Meal intake, carbohydrates [g]}
\end{array}
$$

u_{Ir} and u_{Is} represents rapid and slow insulin injections, respectively. u_{hr} is the heart rate of the subject, and is used to indicate the heart rate during an exercise session. u_{meal} represents a meal, with a number of carbohydrates $[g]$ and the glycemic index $SlowFact$. The GlucoPred model includes 41 model parameters, and the list of all the parameters in the model is:

$$\mathbf{p} = [K_{mg}, \alpha, k_{Dia}, k_{IL}, \beta, G_{neo}, R_m, T_d, I_{Half}, R_i, U_b, n_h, n_i, T_Y, I_0, a_P,$$
$$M_{lMx}, HR_B, k_{abs}, T_{max}, T_{ds}, S_i, k_{gb}, k_{HL}, k_{Del}, E_{neo}, k_{gm}, G_I, G_H,$$
$$p_2, f, U_{ii}, V_i, R_{Hbas}, s_{Comp}, n, R_{Hmax}, r, k_{glg}, V_g, HR_M]$$

2.2.1 Model equations

In this section the model equations for the GlucoPred model is described. Most of this section is written by Odd Martin Staal [13], with some notational alterations from myself. In this section, the notation X^+ should be taken to mean that the term X is made non-negative, i.e. if $X < 0$, $X^+ = 0$, otherwise $X^+ = X$. Another notational convention is $X^{\in[a,b]}$ which should be taken to mean that X is limited to be in the range [a,b]. The state transition equations are:

$$\dot{G}_p = G_{prod}(H, M_{lg}, I) - G_{use}(G_p, X, Z, Y) + R_a(M_{gut}) \tag{2.1}$$

$$\dot{G}_t = k_{Del}(G_p - G_t) \tag{2.2}$$

$$\dot{I} = -nI + I_{endo}(G_p, \dot{G}_p) + I_{exo}(S_{R2}, S_{S2}) \tag{2.3}$$

$$\dot{X} = -p_2(X - I)^+ \tag{2.4}$$

$$\dot{S}_{R1} = -\frac{S_{R1}}{T_d} + u_{Ir} \tag{2.5}$$

$$\dot{S}_{R2} = \frac{S_{R1} - S_{R2}}{T_d} \tag{2.6}$$

$$\dot{S}_{S1} = -\frac{S_{S1}}{T_{ds}} + u_{Is} \tag{2.7}$$

$$\dot{S}_{S2} = \frac{S_{S1} - S_{S2}}{T_{ds}} \tag{2.8}$$

$$\dot{M}_{sto} = -f_{abs}(Y)M_{sto} + u_{meal} \tag{2.9}$$

$$\dot{M}_{gut} = -f_{abs}(Y)(M_{gut} - M_{sto}) \tag{2.10}$$

$$\dot{H} = n(S_H - H) \tag{2.11}$$

$$\dot{M}_{lg} = -10^{-3}Q_{liver}(H, M_{lg}, I) \tag{2.12}$$

$$\dot{Y} = \frac{1}{T_Y}(f_{hr}(u_{hr}) - Y) \tag{2.13}$$

$$\dot{Z} = -\frac{1}{T_{max}}Z + f_Y(Y)(1 - Z) \tag{2.14}$$

The functions embedded in the above equations are given in the following. The glucose production is given by

$$G_{prod} = \frac{1}{V_g} \left(E_{neo} + Q_{liver}(H, M_{lg}, I) + G_{neo} k_{HL} H \right) \tag{2.15}$$

where the liver net production is given by

$$Q_{liver} = (1 - G_{neo}) k_{HL} H \frac{M_{lg}}{M_{lMx}} - k_{IL} f_{kM}(M_{lg}) \frac{I}{I + I_{half}} \tag{2.16}$$

f_{kM} is a reduction factor when liver glucagon stores are full, given by

$$f_{kM} = \left(1 - s_{Comp} \left(\frac{M_{lg}}{M_{lMx}} - 1 \right) \right)^{+} \tag{2.17}$$

The glucose usage is given by

$$G_{use} = U_{ii} + k_{gb} G_p + f_{idep}(X, Y, Z) \frac{G_p}{K_{mg} + G_p} \tag{2.18}$$

where

$$f_{idep} = S_i X + k_{gm} + k_{kglg} + \alpha X Z + \beta Y \tag{2.19}$$

The rate of appearance of a meal is given by

$$R_a = 10^3 * f * f_{abs}(Y) * M_{gut} \tag{2.20}$$

The absorption function in the equation for R_a, x_{Sto} and x_{Gut} is given by

$$f_{abs} = (p_{mealGI} k_{abs}(1 - rY))^{\in[0,1]} \tag{2.21}$$

and contains a parameter p_{mealGI} between 0 and 1 that is related to the glycemic index of the last meal (set to 1 if unknown). Note that it also depends on recent exercise.

The endogenous insulin production is given by

$$I_{endo} = n \left(R_m \dot{G}_p^{+} + \frac{R_i}{1 + \left(\frac{G_I}{G_p} \right)^{n_i}} \right) \tag{2.22}$$

The contribution to plasma insulin from exogenous insulin input is

$$I_{exo} = \frac{10^3}{V_i} \left(\frac{S_{R2}}{T_d} + \frac{S_{S2}}{T_{ds}} \right) \tag{2.23}$$

The production of glucagon is given by

$$S_H = S_{H1} + S_{H2} \tag{2.24}$$

$$S_{H1} = \left(R_{Hbas} + (R_{Hmax} - R_{Hbas}) \left(1 - \frac{1}{1 + \left(\frac{G_H}{G_p} \right)^{n_h}} \right) \right)^{+} \tag{2.25}$$

$$S_{H2} = \left(\frac{k_{Dia}}{1 - a_P(I_0 - I)^{\in[0,k_{IHmax}]}} \left(\frac{G_p}{G_H} \right)^2 \right)^{\in[0,SHmax]} \tag{2.26}$$

In the exercise submodel the expressions are given by the following:

$$f_{hr} = \left(\frac{u_{hr} - HR_B}{HR_M - HR_B} \right)^{\in[0,1]} \tag{2.27}$$

$$f_Y = \frac{(Y k_a)^{k_{nY}}}{1 + (Y k_a)^{k_{nY}}} \tag{2.28}$$

2.2.2 Preadapted parameter sets

Prediktor Medical already have parameter sets adapted to healthy, DM1 and DM2 models. These sets represents a general healthy, DM1 or DM2 person, so that simulating the model will generate a typical output for a person with the disease (or absence of disease). These sets will not fit to every person in each category, but these can be used as a starting point when trying to estimate the parameters for a subject. These parameter sets can be found in Appendix A.1.

2.2.3 Food modelling

The task of modelling different types of food and their effect on the blood glucose concentration is a complex one. The main parameter when it comes to this is the amount of carbohydrates the meal contains. Carbohydrate comes in different forms, and it is the part of a meal that affects the blood glucose the most. Carbohydrates can be digested both quickly and slowly, dependent on what type of carbohydrate it is. This will of course be shown in the blood glucose, as slower carbohydrates affects the glucose much slower than the faster carbohydrates will. Other parts that affect the blood glucose are fat and protein. These elements can also be transformed to glucose in the blood for energy utilization. Fat and protein can also affect the blood glucose in a different way. When eating a meal with much fat and protein, the body will take longer to digest the main energy source, carbohydrates, because it has to deal with the fats and proteins at the same time.

In the GlucoPred model, the meal is modelled simply by two parameters. The first one is the amount of carbohydrates $[g]$. The other one is a parameter
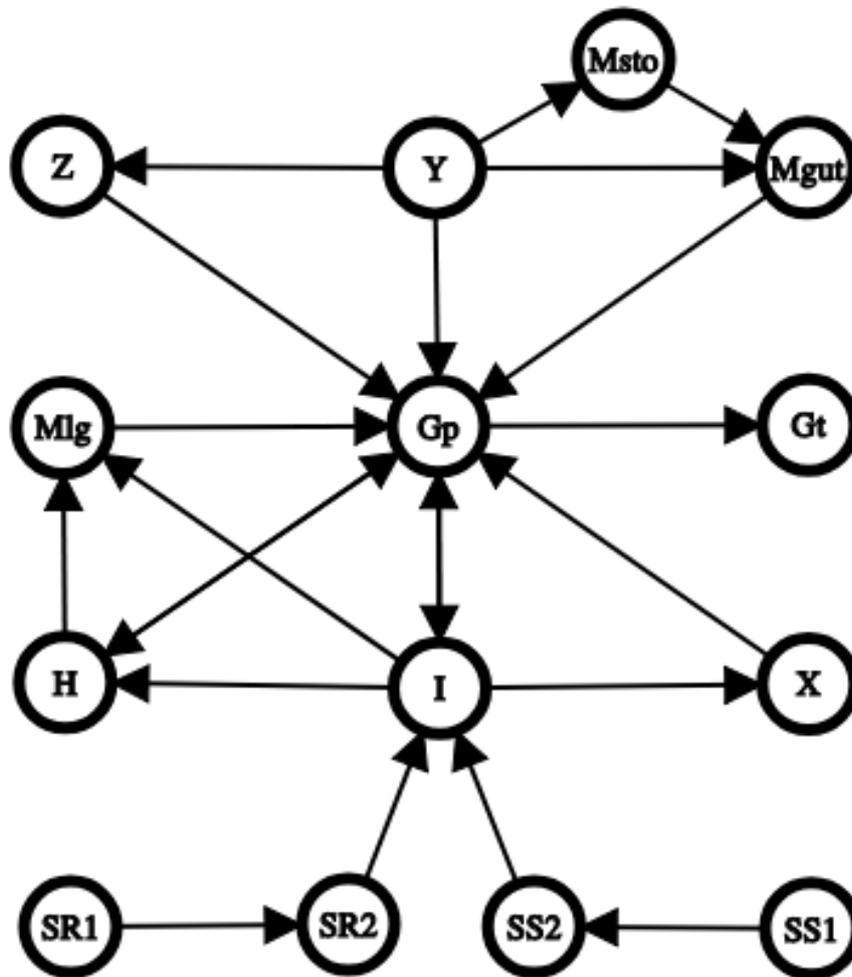
Figure 2.1: Graph overview of the GlucoPred model.

Every node represents a state variable. An arrow from node A to node B means that the state represented by node A is featured in the differential equation for the state represented by node B.

representing how "fast" these carbohydrates will be digested and turned into plasma glucose, called $SlowFact$. This parameter can also be seen as a type of glycemic index. A simplification that is made in the GlucoPred model is that fats and proteins in the meal are not modelled. A real meal would of course contain carbohydrates, fats and proteins, but only the former one is taken into consideration here. Another simplification is that the whole meal has the same $SlowFact$, as would not be the case in real life, where different parts of the meal would be digested with different rates.

2.3 Identifiability

Observability is a property a system can have which implies that the values of the state variables can be determined from the output of the system. This is a desired property of a system because unknown state variables can make control of the system harder.

Identifiability considers whether the parameters in a system of differential equations can be uniquely determined from the input and the output of the system. If this is the case, the system is called identifiable. If a system is non-observable, it implies that it is also non-identifiable. Many types of identifiability have been introduced ([14; 15]), but structural and practical identifiability are the ones talked about here, together with conditional identifiability.

2.3.1 Structural identifiability

A structural identifiability analysis investigates whether a system is identifiable given perfect and noise-free measurement data from it. In other words, a structurally identifiable system has a model structure which yields identifiability in theory. Perfect and noise-free measurements are of course not possible in the real world, but this type of analysis can give some insight into a system. This type of analysis should be done before you analyze a model more closely, but is not always carried out because of the computational costs it represents.

2.3.2 Practical identifiability

A practical identifiability analysis looks at the identifiability when we do not have perfect and noise-free measurements. Lack of practical identifiability can be caused by several things. one of them is that the model structure makes it not structurally identifiable, and therefore also not practically identifiable. Even though a parameter set for a model is structurally identifiable, it is not always identifiable in practice, and this may be caused by two things. The first is that the model is not sensitive to one or more parameters in the set. The other is that two or more parameters correlate, that is, their contribution to the model output cannot be distinguished from each other [16]. A kind of practical identifiability analysis is sensitivity analysis, described in section 2.4.

2.3.3 Conditional identifiability

For many models, the parameter vector is generally unidentifiable. If setting some of the parameters constant, the rest of the parameter vector is identifiable, this modified model can be called conditionally identifiable [15].

2.3.4 Local identifiability

A model is locally identifiable if the model parameters can be identified in a given region of the parameters space around the true parameter values. This means that if $f(\mathbf{x}, \mathbf{p_1}) = f(\mathbf{x}, \mathbf{p_2})$, where f represents the model equations, the parameter values $\mathbf{p_1}$ and $\mathbf{p_2}$ are either $\mathbf{p_1} = \mathbf{p_2}$ or one of them are outside the given region in which the local identifiability is applied.

2.3.5 Varying conditions that affect identifiability

When performing an identifiability analysis, there are several conditions which can influence the result. For the GlucoPred model, it is of course important which one of the preadapted model parameter sets that is used. These can be healthy, DM1 or DM2 parameter sets. Other factors that affect the sensitivities are:

- **How often the glucose values are measured.** Normally, for a CGM, a glucose measurement is displayed every fifth minute. This is an average value of measurements taken those last five minutes. Longer time steps between the measurements will lead to less chance of identifiability for a model. In this study the simulation is done with time steps of 30 seconds, and so is the simulated measurements. A change in the frequency of the simulated measurements (for example every 2nd minute, every 5th minute or every 10th minute) could have affected the results, but was not performed here.

- **The length of the experiment.** The longer the simulation time, the higher the likelihood of getting a more correct estimate of the parameter set. In this study a simulation time of 500 minutes is used in the initial analysis, and later it is adapted to the length of the input.

- **What is measured.** In a clinical environment, it is possible to measure both glucose and insulin concentrations in plasma. In addition, a pretty good estimate of the injected subcutaneous insulin is available. In this study, it is assumed that only plasma glucose measurements, G_t, are available, when doing the sensitivity and identifiability analysis.

- **Model input.** The model input has of course a major impact on the identifiability of the system. To identify a set of system parameters, the input must be "rich" enough so that the parameters are possible to determine individually. The concept of persistently exciting input has been introduced on this matter. A signal is persistently exciting if the input generates enough information to the output so that, in finite time, the estimates of the system parameters converges to their true value [17]. The inputs in

this study is a set of meal, exercise and external insulin. These are varied to see which inputs gives rise to identifiability in which parameters, but the concept of persistently exciting input is not taken into account.

2.4 Sensitivity analysis

Sensitivity analysis is a kind of practical identifiability analysis and a way to obtain information about how much the different model parameters influences the model output. The sensitivity of a parameter can be seen as

$$\mathbf{S}_{p_i} = \frac{\partial \mathbf{y}}{\partial p_i}$$

That is, the derivative of the output $\mathbf{y}$ with respect to the parameter p_i. This only gives the sensitivity at one time instant. To get the parameter sensitivity over the course of a simulation, a system of ODEs must be developed, based on the original system equations, like in Stigter et al. (2017) [18]. These equations will model the sensitivity dynamics between the model and its parameters, and can be given as:

$$\dot{\mathbf{x}}_p(t) = \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \mathbf{x}_p(t) + \frac{\partial \mathbf{f}}{\partial \mathbf{p}} \tag{2.29}$$

$$\dot{\mathbf{y}}_p(t) = \frac{\partial \mathbf{h}}{\partial \mathbf{x}} \mathbf{x}_p \tag{2.30}$$

Here $\mathbf{f}$ is the vector of the original model equation expressions, $\mathbf{h}$ is the output equations, $\mathbf{x}$ is the state vector and $\mathbf{p}$ is the parameter vector. $\mathbf{y}_p(t)$ is a vector with the parameter sensitivity at time t. These vectors can be put together to a sensitivity matrix $\mathbf{S}$. In this matrix each row is the sensitivity vector of one output at one time instant. Each column corresponds to the sensitivity for a single parameter over the course of the simulation. This way, by plotting one column over the simulation time, we can observe in which parts of the simulation a parameter has high sensitivity, and for which it has lower or no sensitivity.

2.5 Model parameter estimation methods

To identify model parameters for the metabolism model, parameter estimation algorithms must be used. These try to minimize an objective function with respect to the parameters. The objective function is based on a criteria for which minimization identifies the parameters, if the model is identifiable.

2.5.1 Objective function for parameter estimation

One way objective function minimization can be done is to use the model and vary the parameters so that a simulation of it is fitted to the output measurement. The two sided desirability function [19; 20; 21] can be used as an objective function for this purpose. This function is on the form

$$d_{tot}(\mathbf{x}, \mathbf{y}) = -d_y \prod_{i=1}^{n_{par}} d_{x,i} \qquad (2.31)$$

where d_y is given by

$$d_y = d(y, n_y) \qquad (2.32)$$

$$y = \sqrt{\frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} (y_{sim} - y_{ref})^2} \qquad (2.33)$$

$d_{x,i}$ is given by

$$d_{x,i} = d(x_i, n_x) \qquad (2.34)$$

and $d(x, n)$ is given by

$$d(x, n) = \exp\left[-(\tilde{x}(x))^n\right] \qquad (2.35)$$

$$\tilde{x}(x) = \frac{2x - (x_{max} + x_{min})}{x_{max} - x_{min}} \qquad (2.36)$$

Here x_{min} and x_{max} are the minimum and maximum values of the parameters estimated, given for all parameters in Appendix A.2.

d_y is a measure of the difference between $\mathbf{y}_{sim}$ and $\mathbf{y}_{ref}$. $\mathbf{y}_{sim}$ is the glucose values of a simulation with parameter values from $\mathbf{x}$, and $\mathbf{y}_{ref}$ the simulated measurement glucose values that we want to obtain with the model parameter estimation. n_{sim} is the number of time steps in the simulation.

$d_{x,i}$ gives a value near 1 if the parameter value is in the desired region, and lower values the further away it is from that region. The higher n is, the wider this region becomes, so that parameters moving away from the middle does not decrease as fast. When $n \to \infty$, the function will approach the boxcar function. The boxcar function is a function which evaluates to zero outside the region $[x_{min}, x_{max}]$, and 1 inside it.

2.5.2 Downhill simplex method (Nelder-Mead)

The downhill simplex method is a derivative-free method for solving nonlinear optimization problems. Instead of derivatives it uses a simplex in the parameter space to narrow the search space until it finds a minimum point of the

function. A simplex is a polytope with n+1 vertices, in n dimensions. Each vertex of the simplex is a set of parameter values in the parameter space, with a corresponding objective function value.

The fundamental idea behind the simplex method is that the vertex in the current simplex with the "worst" objective function value is mirrored through the line between the two vertices with the "best" objective function values. Then the objective function is evaluated at this new point. If it is good, it replaces the worst vertex in the simplex. If it is not so good, a new point closer to the worst point along the mirroring line is evaluated. This way, the simplex will always try to move towards the parts of the parameter space with better objective function values.

2.5.3   Newton-Raphson method

The Newton-Raphson method is an algorithm for finding the roots of nonlinear functions. In optimization, the Newton-Raphson method is used on the derivative of the objective function. By finding the root of the derivative, or gradient, of the objective function, hopefully an extremum point of the objective function is found.

For parameter estimation in our case, we want to find the roots of the gradient of the objective function $d_{tot}(\mathbf{x}, \mathbf{y})$, and by that find an extremum point of it. This method makes use of both the gradient of the objective function, and the Jacobian of this gradient.

2.6   Finite difference approximation

Some mathematical functions or expressions are so complex that it is too hard or time consuming to find its derivative analytically. Then it is appropriate to apply a derivative approximation technique. One of the simplest ones of these is called finite difference. With this technique, you evaluate the function value $f(x)$ for two different x values close to each other to approximate the derivative in this area of the variable space. If the point for which one wants to find the approximated derivative is x^*, one approach is to evaluate the function at x^* and another point $x^* + dx$. This is showed in (2.37), and the expression is called the difference quotient. $L_{dq}(x^*)$ then is an approximation of the derivative $f'(x^*)$. Another approach is to evaluate the function at $x^* - dx$ and $x^* + dx$, showed in equation (2.38), which is called the symmetric difference quotient.

$$L_{dq}(x^*) = \frac{f(x^* + dx) - f(x^*)}{dx} \tag{2.37}$$

$$L_{sdq}(x^*) = \frac{f(x^* + dx) - f(x^* - dx)}{2dx} \tag{2.38}$$

3   Aim of the study

The aim of this study was to investigate the parameter identifiability of the GlucoPred metabolism model, developed by Prediktor Medical. The model's intended use is for simulation and prediction of future plasma glucose values, but also as a helping tool during the development of the non-invasive glucose measurement device from Prediktor Medical. The model parameters must be individualized so that the model can represent a specific person, for the model to be used for simulation and prediction. This identifiability study is meant as a step in the direction of being able to identify the GlucoPred model parameters for individuals.

The study will present:

- A presentation of the GlucoPred model, with the dynamic equations modelling the glucose and insulin dynamics in the body.

- A sensitivity analysis discussing which input gives rise to identifiability in which parameters.

- Parameter estimation with two different estimation methods, for the parameters from the sensitivity analysis.

4 Method description and implementation

Optimally, both a structural and practical identifiability analysis should have been carried out when investigating the GlucoPred model. First the structural analysis should be used to investigate which restrictions the model equations themselves put on the identifiability of the system. Then the practical analysis should be used to look at what type of input the different parameters are sensitive to, and in what order the input is needed for them to be identifiable. In this study however, a structural identifiability analysis is not carried out, mostly because of the high computational cost of the calculations. In addition, a structural identifiability analysis would only give us a yes/no answer on whether the model structure gives an opportunity to identify the model parameters, when we have perfect infinite measurements.

4.1 Simulation model and sensitivity implementation

To be able to implement both the sensitivity analysis and estimation algorithms, the GlucoPred model itself would have to be implemented. This was done in Matlab with a simulation time step of 0.5 minutes. To the model simulation, input of meal and exercise can be given, in addition to rapid or slow external insulin doses.

The creation of the sensitivity matrix for a simulation with a parameter set $\mathbf{p}$ and a given input was implemented with equations (2.29) and (2.30). These were calculated for every time step in parallel with the original system equation simulation. Because of the complexity of the model equations, the derivatives $\dfrac{\partial \mathbf{F}}{\partial \mathbf{x}}$ and $\dfrac{\partial \mathbf{F}}{\partial \mathbf{p}}$ needed for this were approximated with finite difference (section 2.6), instead of calculating them analytically. The perturbation dx was 1% of the normal range of the state variable or parameter.

4.2 Initial values

Initial values for the model parameters were chosen based on a preadapted set of initial values used for GlucoPred earlier, for persons either healthy or having diabetes mellitus type 1 or 2. The only parameter changed in this were T_d, the time constant for rapid insulin dynamics. It was set to 30 instead of 15 for DM1 to get more stable glucose values for the given input.

To make the initial state variable values closer to a steady-state of the model from the start, the initial values for the two glucose variables G_p and G_t were both changed from 100 mg/dL to 85.25 mg/dL. The initial glucagon concentration H was also changed from 100 pg/ml to 160 pg/ml. This was done after observing the initial transient period of the model simulation, and that these

state variables tended to approach these values with no input. Then the initial values for the state variables became:

$$\mathbf{x}_0 = \begin{bmatrix} G_p \\ G_t \\ I \\ X \\ S_R1 \\ S_R2 \\ S_S1 \\ S_S2 \\ M_{sto} \\ M_{gut} \\ H \\ M_{lg} \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 85.25 \text{ mg/dL} \\ 85.25 \text{ mg/dL} \\ 10 \text{ mU/L} \\ 10 \text{ mU/L} \\ 0 \text{ U/L} \\ 0 \text{ U/L} \\ 0 \text{ U/L} \\ 0 \text{ U/L} \\ 0 \text{ g} \\ 0 \text{ g} \\ 160 \text{ pg/ml} \\ 50 \text{ g} \\ 0 \\ 0 \end{bmatrix}$$

4.3 Initial parameter sensitivity analysis

To start with, the sensitivity of each parameter for itself was investigated and plotted for the timespan of the experiment. An example is showed in figure 4.1, for the parameter S_i. This was done using the input showed in table 4.1. The interpretation of this example is that the parameter S_i has high sensitivity during the meal and when the blood glucose is high. This can be seen by the larger negative value of the sensitivity during the period of high positive glucose values. This means that, during or right after the meal, if the parameter S_i is increased a little, the blood glucose will decrease a little. The higher the absolute value of the sensitivity, the more a perturbation in the parameter will affect the output glucose.

This plot provided an intuitive and easy way to see what type of input, and where during the simulation, the different parameters had the biggest influence on the output glucose value.

An assumption made in this part was that a parameter with very low sensitivity for the initial experiment input from table 4.1, generally would have pretty low sensitivity also for the same input but in different order and quantity. It was on the basis of these sensitivities that the initial parameter sets $P_{h,init}$, $P_{dm1,init}$ and $P_{dm2,init}$ were chosen before the start of the sensitivity matrix analysis in section 4.4.
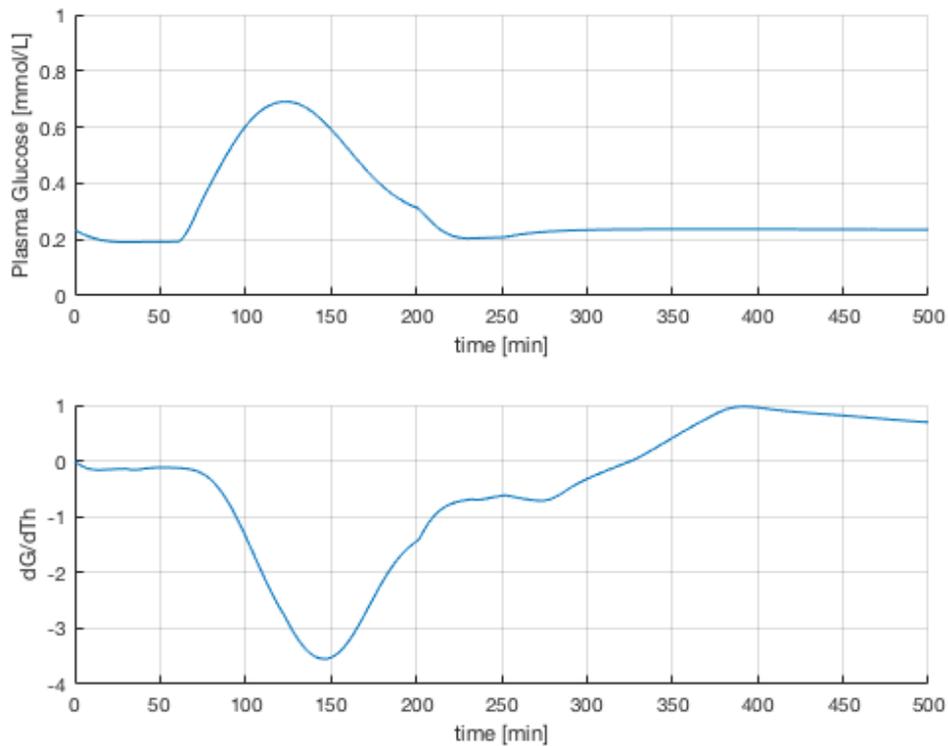
Figure 4.1: Sensitivity plot example.

For parameter S_i, dm2 preadapted model, with meal and exercise input. The upper plot is plasma glucose, and the lower plot is the sensitivity values for S_i.

4.4   Sensitivity matrix analysis

Parameter sensitivity for preadapted models for healthy, DM1 and DM2 subjects was investigated in this part of the study. Different input is of course a major factor in identifiability and for which model parameters can be estimated correctly. A single meal (together with any needed insulin doses), a single exercise session, and combinations of these, were assessed as input to the model.

In both the sensitivity matrix analysis and parameter estimation, described in the current section and section 4.5, the amount of the different inputs were the following, unless otherwise stated:

- Rapid insulin: 6.5 Units

- Slow insulin: 20 Units

- Exercise: 150 BPM for 50 minutes

- Meal: 80 grams of carbohydrate and $SlowFact = 0.6$

Subject	Input	At time
Healthy	- Meal: 80 g carbs, 0.6 GI	60 min
	- Exercise: 150 BPM	200 - 250 min
DM1	- Slow insulin dose: 20 Units	0 min
	- Meal: 65 g carbs, 0.6 GI	60 min
	- Rapid insulin dose: 5 Units	70 min
	- Exercise: 150 BPM	200 - 250 min
DM2	- Meal: 65 g carbs, 0.6 GI	60 min
	- Rapid insulin dose: 5 Units	70 min
	- Exercise: 150 BPM	200 - 250 min

Table 4.1: Input to initial parameter sensitivity analysis

As an indication for which parameters that could be estimated, a sensitivity matrix S_{ens} was calculated based on the input combinations. A singular value decomposition (SVD) was performed on S_{ens}. From this, the three matrices $\mathbf{U}$, $\mathbf{S}$ and $\mathbf{V}$ were obtained. The diagonal matrix $\mathbf{S}$, contained singular values and the right-singular vectors (RSV) corresponding to them were the columns of $\mathbf{V}$. Like in Stigter et al. (2017) [18], the singular values and RSVs were plotted, to easily observe which parameters contributed to which singular values. An example of this plot is shown in figure 4.2. Low singular values are indicating low identifiability, and higher singular values indicates that parameters may be identifiable.

The RSV corresponding to a singular value significantly lower than the rest was plotted to find out which parameters contributed to this. One of these was removed from the set of possibly identifiable parameters, or another action was taken to get rid of the low singular value. To figure out the most sensitive/identifiable parameters, the RSV to the three (or more) highest singular values was plotted. The parameters contributing most to these would have a good chance of being identifiable. In addition, a sum of RSV-contribution weighted by the corresponding singular values for each parameter was computed and plotted. This measure, called V_{sum} here, makes sure the highest singular values contribute most when selecting identifiable parameters based on it, and that lower singular values do not.

$$V_{sum,p} = \frac{1}{S_{max}} \sum_{i=1}^{n_{Pte}} |V_{p,i}|\mathbf{s}_j \qquad (4.1)$$

For a given parameter p, this measure is computed with equation (4.1), where n_{Pte} is the number of parameters analyzed, V is the RSV matrix, $\mathbf{s}$ is the vector of singular values (where values below 0.01 is set to zero) and S_{max} is the highest value in $\mathbf{s}$. This gives a value for each parameter for how much it contributes

in the RSV for the highest singular values, and is easily interpreted when plotted. A parameter with a high V_{sum} is an interesting parameter because it contributes much to the RSV corresponding to the highest singular values of the sensitivity matrix.

The effects on the glucose from parameters with similar sensitivity curves will be hard to distinguish. The identifiability of the system with parameters that contribute similarly to the measured output is low. This can be shown from a simple example. The four parameters K_{mg}, E_{neo}, k_{gm} and k_{glg} have very similar sensitivity plots, as shown in the bottom plot of figure 4.2. This reflects itself in the singular values of the sensitivity matrix, shown in the upper plot in the same figure. The logarithm of these are all around zero or below, and the smallest one is much smaller than the others. One can observe from the middle plot that k_{gm} and k_{glg} contributes to the RSV corresponding to the lowest singular value. Therefore, those two are most likely not in an identifiable set together, and one of them has to be removed.
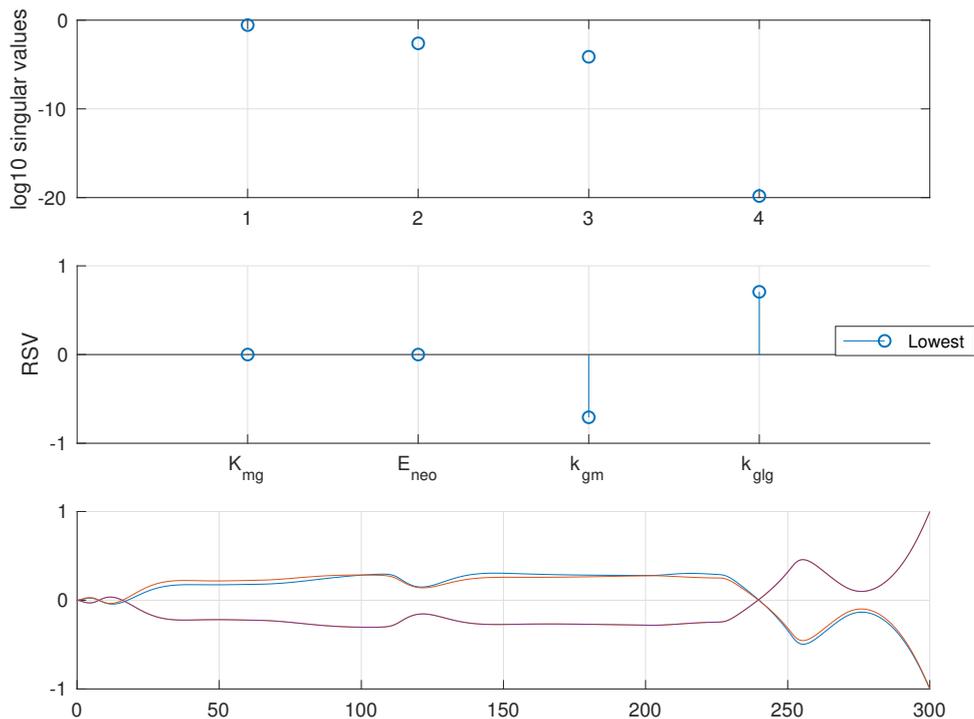

Figure 4.2: singular values, RSV and sensitivity plot example.

For parameters K_{mg}, E_{neo}, k_{gm} and k_{glg}, healthy preadapted model, with meal and exercise input. RSV corresponding to the lowest singular value.

4.5 Parameter Estimation

After the sensitivity analysis was done, estimation techniques were implemented to try to estimate the parameters. Instead of the $d_{x,i}$ described in equation 2.34, the boxcar function was used. The reason for this was that in the initial testing of the estimation methods, it was observed that the parameter bias $d_{x,i}$ introduced, did disturb the estimations. The estimations were dragged towards the middle of their range, where $d_{x,i}$ were at its highest, despite the glucose curve fitting better elsewhere in the parameter space. The boxcar function got rid of this bias by outputting 1 whenever the parameter where inside its range, and 0 elsewhere.

The two parameter estimation techniques implemented were Downhill Simplex (Nelder-Mead) and Newton-Raphson optimization. The parameter estimation took place in the same order as the sensitivity analysis, first healthy, then DM1 and DM2 after that. In this part it was important to observe whether the parameters that looked possible to estimate in the sensitivity analysis were actually possible to estimate.

Table 4.2 shows a small extract from the full result table written down during the parameter estimation. This is for one meal input and one exercise input for the healthy model. The table also displays which parameters that were estimated, the estimation results in **DS result** and **NR result**, and also the absolute value of the desirability function from (2.31), in **DS score** and **NR score**. A score value of 1 is the maximum and indicates that the curve of a simulation using the new parameters, fitted to the simulated measurement is done perfectly, and that the parameters are inside their normal range. It does not, however, say anything about how precise the parameter estimations themselves were. For that, one has to compare the results with the actual parameter values used in the simulated measurements (Appendix A.1). The parameter estimation accuracy value introduced in section 4.5.3 does this.

Input	Parameters	DS result	DS score	NR result	NR score
Meal	k_{abs}, k_{HL}	0.0500, 3.60	1.000	0.0500, 3.66	0.9994
Meal	k_{abs}, k_{HL}, S_i	0.0513, 2.97, 0.160	0.9959	0.0498, 3.88, 0.217	0.9987
Meal	k_{abs}, k_{HL}, p_2	0.0580, 3.59, 0.0580	0.9943	0.0502, 3.67, 0.0208	0.9993
Exercise	β, k_{HL}	1.00, 3.60	1.000	1.40, 3.73	0.9991
Exercise	β, k_{HL}, G_H	3.21, 2.93, 65.9	0.9975	2.00, 4.37, 48.2	0.9975

Table 4.2: Excerpt from the parameter estimation result.

The reason DS was chosen as one of the parameter estimation algorithms is that the derivatives of the objective function $d_{tot}(\mathbf{x}, \mathbf{y})$ with respect to the parameters is hard to obtain analytically. The derivatives would have become complicated expressions that it would have taken some time to arrive at. The DS algorithm is a derivative free method, which means that only evaluations of the objective function itself are taken into account, not the derivatives.

The NR algorithm was chosen to compare the two fundamentally different methods. DS which is a derivative free method against NR, which is a quasi-Newton method. Also, both of these methods were mentioned in the problem formulation text.

You want to make sure that all areas of the possible parameter space is explored during estimation. In this study this area is the area inside the normal parameter ranges. This can be done by trying different combinations of initial starting points for every parameter to estimate. If we define that there are only two different initial values each parameter should have, a full factorial experiment would require 2^{n_p} different estimations, where n_p is the number of parameters in the estimation set. This quickly becomes infeasible with a growing n_p. An experimental design called Plackett-Burman (PB) [22] can be used to limit the number of estimations to $n_p + 1$. At the same time it makes sure that a small as possible area of the parameter space is left unexplored. The PB design was used in this project. For both the Downhill Simplex and Newton-Raphson, $n_p + 1$ estimations were carried out for each combination of model input and parameter set, where the best result was used. An example design is shown in table 4.3. This would have been for a parameter set consisting of three parameters and four algorithm runs. A $+$ means that the parameter will be assigned the higher value as initial value, and a $-$ means that it will be assigned the lower value.

Run	p_1	p_2	p_3
1	$+$	$+$	$+$
2	$+$	$-$	$-$
3	$-$	$+$	$-$
4	$-$	$-$	$+$

Table 4.3: Plackett-Burman experimental design for $N = 4$ and $n_p = 3$.

4.5.1 Downhill Simplex

For the Downhill Simplex algorithm, the initial simplex in the parameter space was made using PB design. For estimation run r_i, where $i = 1, 2 \ldots N$, the ith row of the PB design (Example in table 4.3) was used to decide new low and high values for the different parameters. For parameter p_j, if $PB_{i,j} = +$, the values were:

$$p_{j,low} = 0.57p_{j,max} + 0.43p_{j,min}$$
$$p_{j,high} = 0.78p_{j,max} + 0.22p_{j,min}$$

If $PB_{i,j} = -$, the values were

$$p_{j,low} = 0.22p_{j,max} + 0.78p_{j,min}$$
$$p_{j,high} = 0.43p_{j,max} + 0.57p_{j,min}$$

Where $p_{j,max}$ and $p_{j,min}$ were the maximum and minimum values of the parameter ranges in Appendix A.2. This way, the possible initial parameter values were shifted closer to max or min of the parameter range, based on whether $PB_{i,j}$ was $+$ or $-$, respectively.

Based on these new low and high values, the initial simplex was created, also with the PB design. For the ith vertex of the simplex and the jth parameter, if $PB_{i,j} = +$, the initial parameter value was $p_{j,high}$, and if $PB_{i,j} = -$, it was $p_{j,low}$.

The stopping criterion was either when 100 iterations were done, or before that if the difference between the objective function values of the best and worst vertices in the simplex were small enough.

4.5.2   Newton-Raphson

For the Newton-Raphson algorithm, the PB design was also used to find the initial values. Now, for the ith estimation run and the jth parameter, if $PB_{i,j} = +$, the initial value were:

$$p_{j,init} = 0.75p_{j,max} + 0.25p_{j,min}$$

If $PB_{i,j} = -$, the value were:

$$p_{j,init} = 0.25p_{j,max} + 0.75p_{j,min}$$

In the initial iteration, the Jacobian $\mathbf{J}_{\nabla d}$ of the gradient of the desirability function $d_{tot}(\mathbf{x}, \mathbf{y})$ (The objective function) were approximated by first approximating the gradient $\nabla_p d_{tot}(\mathbf{x}, \mathbf{y})$ with respect to the parameter vector investigated. Both approximations were carried out using finite differences (Section 2.6).

In the version of the algorithm used in this study, the approximation of the Jacobian matrix $\mathbf{J}_{\nabla d}$ was not calculated for every iteration. Instead it was updated based on the Jacobian from the previous iteration, the gradient at the current search point and the direction of the previous iteration step. This was done via the update equation:

$$\mathbf{J}_{\nabla d}^{(n+1)} = \mathbf{J}_{\nabla d}^{(n)} + \frac{\nabla_p d_{tot} \cdot \mathbf{d}^{(n)}}{\mathbf{d}^{(n)} \cdot \mathbf{d}^{(n)}} \tag{4.2}$$

$$\tag{4.3}$$

where

$$\mathbf{d}^{(n)} = \mathbf{p}^{(n)} - \mathbf{p}^{(n-1)} \tag{4.4}$$

was the direction vector of the parameter estimation update in iteration n. The reason for this Jacobian update for each iteration was that the approximation of the Jacobian was computationally harder and therefore took longer time than the update.

4.5.3 Measure of estimation accuracy

During the estimations, a measure of estimation accuracy was needed. Even though the values **DS score** and **NR score** give information about the curve fit of the estimation, that will not be a guarantee for good parameter estimations, as explained earlier. A measure to observe whether the parameter estimates were close to the original parameter values (Parameter values that were known, because the glucose measurements were simulated measurements from the GlucoPred model itself), relative to the parameter range. Therefore a value called ee_i was used for the estimation error of parameter p_i, defined by:

$$ee_i = 100 \frac{|p_i - p_{i,est}|}{p_{i,max} - p_{i,min}}\%$$

where $p_{i,est}$ is the estimate provided by the estimation method, and $p_{i,max}$ and $p_{i,min}$ are the maximum and minimum values of parameter p_i. This gave an intuitive value representing the parameter estimation error as a percentage of the normal range of the parameter.

4.6 Method procedure

To sum up, for each input set that was investigated, the procedure was the following.

1. Removing parameters (this means setting them constant when estimating), from the set of all parameters, that were not interesting. This included non relevant, zero or low sensitivity parameters for subjects of that condition (healthy, DM1 or DM2), parameters clearly not relevant for that input and parameters which could be measured with reasonable precision in real life. This was done using the initial parameter sensitivity analysis described in section 4.3.

2. Observe sensitivity for the remaining parameter set. If one or more of the singular values were significantly lower than the rest, the parameter(s) responsible for this were revealed in the RSV plot for these singular

values. It could be that a parameter had very low or zero sensitivity, or it could be that two or more parameters correlated in their sensitivities, making them hard to distinguish when estimating. A choice was taken for which parameter to take out of the set, based on the RSV and the general sensitivities for the parameters involved.

3. When none of the singular values of the sensitivity matrix for the remaining parameters were much smaller than the others, the RSV for the highest singular values were investigated. The plots for RSV and V_{sum} together gave a good picture of which parameters to investigate further. A set consisting of between three and seven parameters were chosen for a final analysis.

4. The sensitivities, the RSV and V_{sum} for the chosen parameters in part 3 were examined just like in point 2 and 3 to make a suggestion for parameter combinations of these that could be identifiable. Parameters with the highest V_{sum} rankings and high contributions to the RSVs corresponding to the highest singular values were the most interesting ones. If the contribution to these RSV were similar between two parameters, this could suggest some correlation, so these two parameters would not be picked together in an identifiable set. Based on this, a set of one to three parameters were chosen as a most likely identifiable set. In addition to this, one or two parameters were chosen to be added to the first set, in the case where the first set was easily estimated correctly. This corresponds to the two columns "Most identifiable" and "Together with" in table 5.1 introduced later.

5. At the end, the parameters chosen were estimated with the parameter estimation methods Downhill Simplex and Newton-Raphson. This gave indications of whether the sensitivity analysis gave the right impression of the identifiability or not.

5 Results and observations

The results from the parameter sensitivity analysis and the parameter estimation are shown here. The analysis is divided into three separate section for the healthy, DM1 and DM2 models. A summary of the sensitivity analysis results are shown in table 5.1.

5.1 Healthy subject

After removing the not interesting parameters, the resulting parameter set to investigate was:

$$P_{h,init} = [K_{mg}, \alpha, k_{IL}, \beta, G_{neo}, R_m, I_{Half}, R_i, n_h, T_Y, M_{lMx}, k_{abs}, S_i, k_{gb}, k_{HL}, G_I,$$
$$G_H, p_2, f, U_{ii}, R_{Hbas}, R_{Hmax}]$$

5.1.1 One meal

First for the healthy model, the effect of only one meal was examined. After removing the parameters only related to exercise (α, β and T_Y) from the initial parameter set $P_{h,init}$, the parameter sensitivities were assessed for the one meal input. In figure 5.1a it can be seen that the last singular value drops lower than the others. From the plot of the RSV corresponding to this singular value (the middle plot), it was observed that the parameter R_{Hmax} was the reason for it being so low. Removing this parameter, the plot in figure 5.1b was displayed. Here the most interesting thing was to look at the highest singular values, and one could observe that the parameters that contributed most to the RSV of the highest singular values were the ones with the highest V_{sum}-values. The parameters with the five highest V_{sum}-values, k_{abs}, S_i, k_{HL}, p_2 and f, were investigated further.

From figure 5.1c it can be observed that for these five parameters, the lowest singular value has a "drop" from the second lowest one. This may indicate non identifiability. The corresponding RSV is plotted in the same figure, and one can see that it is the parameters f and S_i that contributes the most in the corresponding RSV. This indicates that a covariance between these two parameters, and possibly also p_2, caused that "drop" to the lowest singular value. Removing parameter f gave a lowest logarithmic singular value of 0.6, which was much better than before. When plotting RSVs for the three highest singular values, it could be seen that the parameters k_{abs} and k_{HL} were the most likely to be identifiable, together with either S_i or p_2 which had some correlation. This can be seen from the RSV plot in figure 5.1d.

The inclusion of k_{abs} in the identifiable set were not so surprising with meal input, as this parameter describes the gut absorption time constant. k_{HL} were more surprising as it describes the rate of glycogenolysis, that is how much
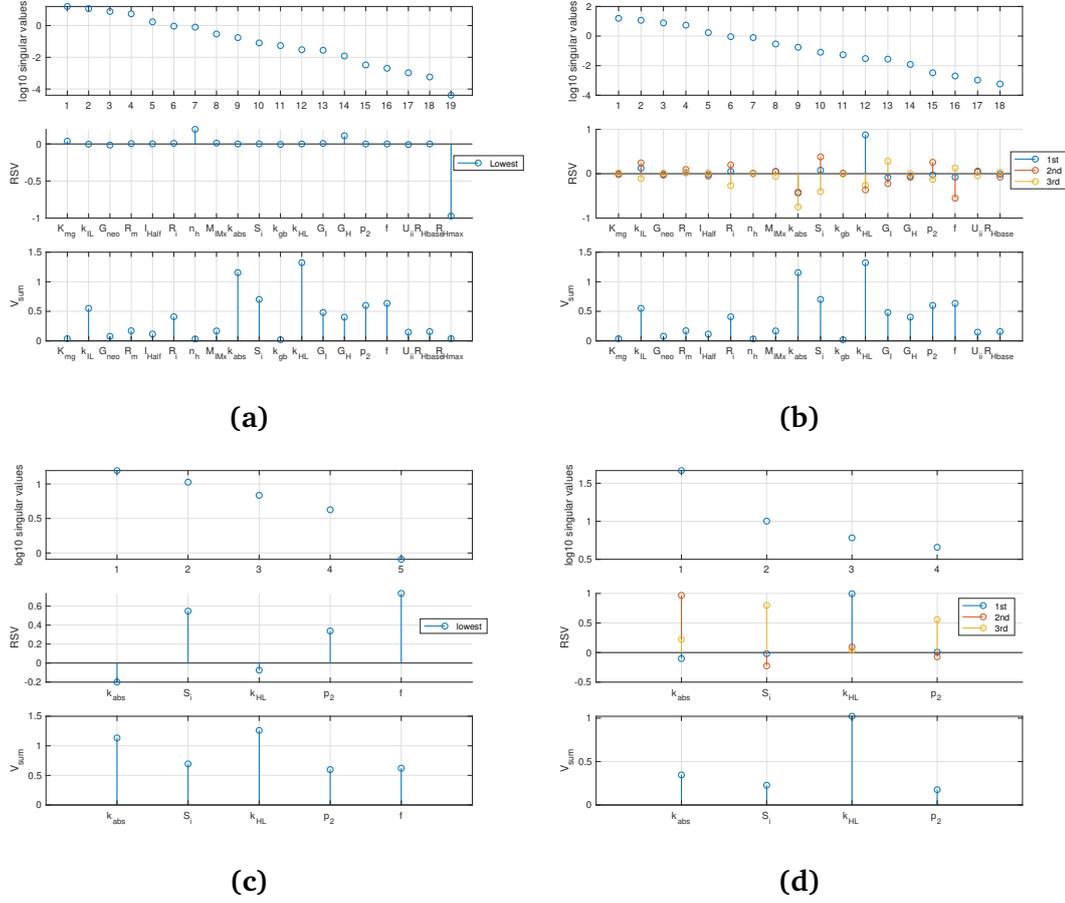
Figure 5.1: Singular values, RSV and Vsum, for 1 meal input and healthy subject. (a) For all parameters and RSV for lowest singular value. (b) For all parameters except R_{Hmax}, and RSV for three highest singular values. (c) For the 5 most sensitive parameters and RSV to lowest singular value. (d) For the 4 highest parameters and RSV to the three highest singular values

glucose is produced from glycogen in the liver. This should not be a dominating parameter during a meal, but as one can see during the rest of this chapter describing the results, k_{HL} is a parameter which has high sensitivity for almost all input.

For only one meal input k_{abs} and k_{HL} were possible to estimate together almost perfect with both estimation methods. Also those two parameters together with either S_i or p_2 could be estimated pretty good with NR, where $ee_{k_{HL}} = 4\%$ was the worst estimation.

5.1.2 One exercise session

Next, the effect of only on exercise session was examined. The parameters related to meal intake was removed (k_{abs} and f). Again, R_{Hmax}, contributed the most to a singular value that was much lower than the others, so it could be removed from the set. The parameters β, k_{HL} and G_H contributed much to one separate of the three highest singular values each. These three parameters were assessed further with another sensitivity plot. In figure 5.2 one can see that they still do not look like they correlate, because they each "own" one of the three RSV. This means that they could be identifiable. The parameter k_{HL} is again the one with highest V_{sum} and contributes most to the highest singular value.

It is not surprising that β, being an exercise parameter, has high sensitivity for exercise input. G_H and k_{HL} were not surprising either, as glucagon secretion (which rate is determined by G_H) and the liver glucose production would be higher during exercise on an empty stomach.
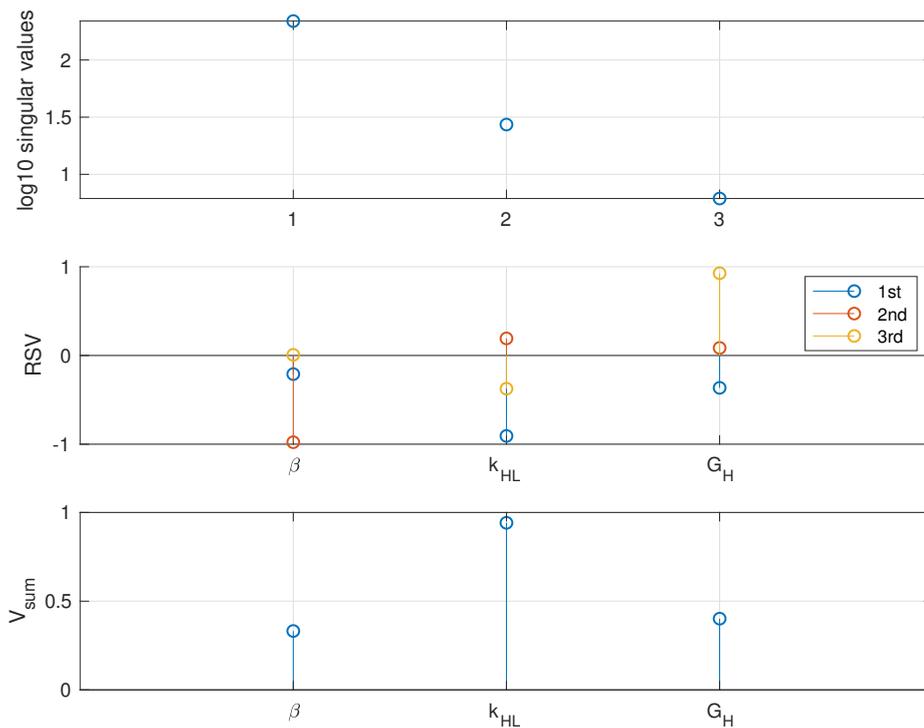
Figure 5.2: 1 exercise input sensitivity healthy

Singular values, RSV corresponding to the three highest singular values and Vsum, for 1 exercise input and healthy subject.

With only exercise input, β and k_{HL} could be estimated perfectly with DS.

Together with G_H the estimates were not so good, with $ee_\beta = 22.5\%$ for DS.

In the following sections, there will be less plots and shorter explanations, but the same procedure has been followed also for these sensitivity analyzes.

5.1.3   One meal and one exercise session afterwards

Starting with $P_{h,init}$ again, the parameters to investigate further based on the highest RSV and V_{sum} was α, β, k_{abs}, S_i, k_{HL}, G_H p_2 and f. These are a combination of the parameters that seemed maybe identifiable for 1 meal and the parameters for 1 exercise, in addition to α and f.

α had very low sensitivity for exercise alone, and that is not surprising when looking at the system equations. This parameter is dependent both on insulin in remote compartment, X, and the exercise memory state, Z. Therefore, when both food and exercise were inputs to the model, α got a higher sensitivity. Another observation here was that the sensitivity for α got lower as the time between meal and exercise got longer. In addition, it appeared as though the total identifiability for this set was higher when the time between the meal and exercise was higher, since the lowest singular value was higher. This might be because it is easier to distinguish the effects of the meal and exercise with longer time between them.

The RSV to the lowest singular value changed dependent on the time between the meal and the exercise, but the insulin sensitivity parameter S_i was always the biggest contributor to it. This suggests that it may not be possible to estimate S_i for this input. With S_i removed, G_H and p_2 would always be among the least sensitive parameters. Also removing these from the set left five parameters to maybe identify: α, β, k_{abs}, k_{HL} and f.

With the exercise right after the meal (Not very likely in reality) α and β could be separated and form an identifiable set with k_{HL}. When estimating, it showed that these three together were almost perfectly estimated with DS, but poorly with NR, where $ee_\beta = 65\%$ was the worst.

With the exercise 60 minutes after the meal, k_{abs} and k_{HL} was most identifiable, and α and β correlated much more. k_{abs} and k_{HL} was estimated perfectly with DS, but not so good with NR. k_{abs} and k_{HL} together with α could be estimated relatively good with DS ($ee_\alpha = 9.7\%$ the worst estimate), while those two together with β did not get estimated as good ($ee_\beta = 12.3\%$ the worst estimate).

When the exercise was 150 minutes after the meal β, k_{abs} and k_{HL} looked like an identifiable set, leaving α with little sensitivity. f would always correlate somewhat with k_{abs}, so it was also taken out. When estimating, β, k_{abs} and k_{HL} did not get a very precise estimation results, where $ee_\beta = 15\%$ was the worst.

5.1.4 One exercise session and one meal afterwards

Again, starting out with $P_{h,init}$, the parameters to investigate further based on the highest RSV and V_{sum} was α, k_{IL}, β, k_{abs}, k_{HL} and G_H. With only these parameters analyzed, only two parameters looked surely identifiable, β and k_{HL}. That these two looked most identifiable is not surprising, β because it is an exercise parameter and k_{HL} because it has high sensitivity for all inputs.

With a normal meal after the exercise, β and k_{HL} looked identifiable together with α or k_{abs}, which looked like they had some correlation with each other. When trying to estimate, it showed that for only β and k_{HL} the estimate was almost perfect with DS. The estimates were not so good, when they were estimated together with α or k_{abs}. For these, $ee_\alpha = 65.5\%$ and $ee_\beta = 64\%$ were the worst estimates for the two sets, respectively.

With a smaller meal during the exercise and no meal after, β and k_{HL} looked identifiable together with either α or G_H. Again, α got lower sensitivity with more time between the two inputs, but so did many of the other parameters. That G_H is there is a little surprising since it is a parameter determining the glucagon production rate. But since the small meal is not "eaten" before it is halfway through the exercise session, it might not be so surprising. This is because the glucose concentration can drop during the first half of the exercise session, before the meal is "eaten", and then glucagon is produced. Again, the estimates for β and k_{HL} alone were almost perfect, but not very good together with α or G_H, where $ee_\alpha = 88.9\%$ and $ee_{G_H} = 43\%$ for NR.

5.2 Diabetes Mellitus 1 subject

After removing the not interesting parameters, the resulting parameter set to investigate was:

$$P_{dm1,init} = [K_{mg}, \alpha, k_{Dia}, k_{IL}, \beta, G_{neo}, T_d, I_{Half}, n_h, T_Y, M_{lMx}, k_{abs}, S_i, k_{gb}, k_{HL},$$
$$G_H, p_2, f, U_{ii}, V_i, R_{Hbas}, R_{Hmax}]$$

When trying to estimate the parameters using DS and NR for the DM1 model, it became apparent that the NR algorithm could not get any good results for any input. Therefore NR is not mentioned in this section, and all estimation results mentioned are with use of the DS algorithm.

5.2.1 Insulin and one meal

First, the sensitivity for one meal with a rapid insulin dose was assessed. The insulin dose was taken 20 minutes before, at the same time as, or 20 minutes after the meal. The time of the insulin relative to the meal was one of the

things investigated when observing the parameter sensitivities. A slow insulin dose was also added as input at the start, just to keep the blood glucose in the simulation at a reasonable level. After removing the parameters only related to exercise (α, β and T_Y) from the initial parameter set $P_{dm1,init}$, the parameter sensitivities was assessed for the one meal input with rapid and slow insulin. For all three insulin injection times, the parameters G_{neo} and M_{lMg} correlated, making one singular value much smaller than the others. Therefore G_{neo} was taken out of the set.

For rapid insulin 20 minutes before the meal, T_d and k_{HL} contributed more to the highest singular values than the rest of the parameters. Then came f, S_i and k_{abs}. These five were looked at more for themselves. The highest singular value was much higher than the rest, and T_d and k_{HL} were the two main factors in the corresponding RSV. They also contributed most to the second highest singular value. Of the three other parameters, k_{abs} contributed most to the third highest singular value, so this could might be identifiable together with the two first parameters, as it did not correlate with them. That T_d was one of the parameters that looked identifiable was expected, as it determines how fast the rapid insulin is taken up by the body, and insulin of course affects the plasma glucose concentration a lot during a meal. The estimates were not good for T_d, k_{HL} and k_{abs} together, with $ee_{k_{abs}} = 24\%$ as the worst estimate. When removing k_{abs}, the estimates of T_d and k_{HL} were almost perfect.

With the rapid insulin dose at the same time as the meal, the same parameters were chosen to look at more (T_d, k_{HL}, f, S_i and k_{abs}), in addition to V_i. T_d and k_{HL} contributed most to the two highest singular values, which indicated identifiability for those two. Out of the other four, f and V_i contributed most to the third highest singular value. These looked correlated, so only one of them looked identifiable, together with T_d and k_{HL}. The estimates for T_d and k_{HL} were very good. The estimates of T_d and k_{HL} together with either f or V_i were not very good, with $ee_f = 33\%$ and $ee_{V_i} = 48\%$ as the worst estimates, respectively.

With insulin 20 minutes after the meal, V_i also had higher sensitivity than with insulin before the meal. The parameters T_d, k_{abs}, S_i, k_{HL}, f and V_i were again chosen to look more at. T_d, k_{abs} and k_{HL} were highest on sensitivity. f and V_i were the most sensitive of the three others, but also correlated with each other. One of these two could maybe be estimated together with the three first ones. When estimating T_d, k_{abs} and k_{HL}, the estimations were bad, with $ee_{k_{abs}} = 19.2\%$ as the worst estimate. If k_{abs} was removed however, the estimation of T_d and k_{HL} was perfect.

5.2.2 Exercise

When the input was a slow insulin dose and an exercise session, the parameters related to meal intake and rapid insulin was removed (k_{abs}, f and T_d).

For this input the three highest singular values each had one major contributor to each of their RSV. Those three parameters were β, k_{HL} and G_H. When analyzed alone, it could be observed that β, k_{HL} and G_H corresponded to the three highest singular values, as shown in figure 5.3, and therefore looked like they could be estimated correctly together. That G_H looked identifiable now was not a surprise, as glucagon secretion might happen during exercise. The estimates of β, k_{HL} and G_H together using DS however, were not very good. In this estimation, the estimate of both β and G_H were not so good, with $ee_\beta = 19\%$ and $ee_{G_H} = 39\%$.
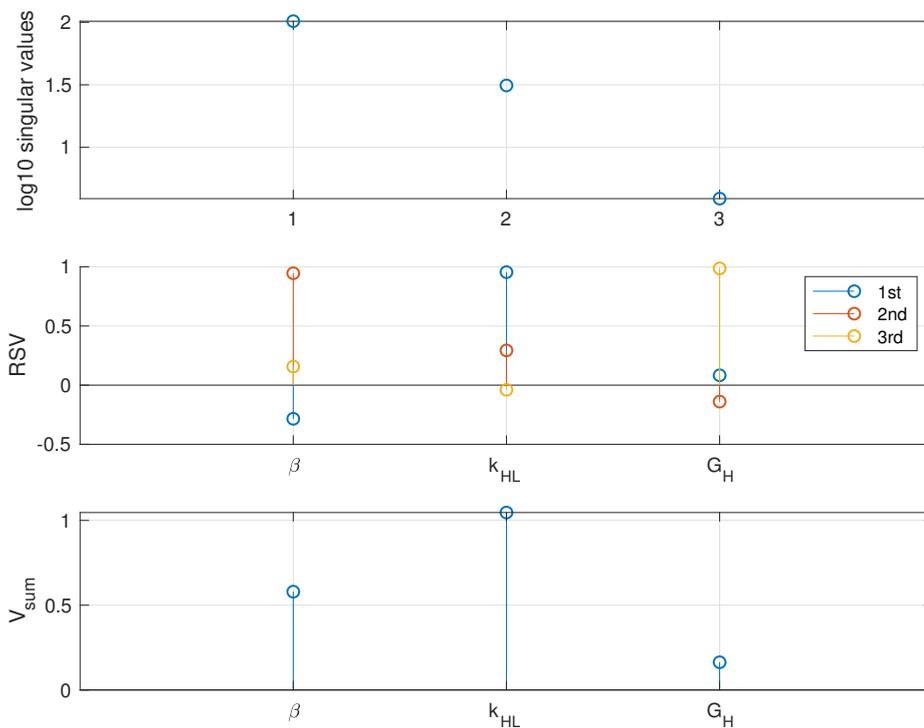
Figure 5.3: 1 exercise input sensitivity DM1

Singular values, RSV corresponding to the three highest singular values and Vsum, for 1 exercise input and DM1 subject.

5.2.3 One meal and rapid insulin, and one exercise session afterwards

With the exercise placed 60 minutes after the meal, the parameters selected for further investigation here were T_d, k_{abs}, k_{HL}, f and V_i. This led to an indication of T_d and k_{HL} being most identifiable together with one of f or V_i. When trying to estimate T_d and k_{HL} the estimation was perfect, but together with either f

or V_i the results were much worse, where $ee_{T_d} = 79\%$ and $ee_{V_i} = 49\%$ were the worst estimates, respectively.

Moving the exercise session closer to the meal gave k_{abs} higher identifiability, but when T_d and k_{HL} were estimated together with k_{abs} it was not accurate at all, and the worst estimation error was $ee_{T_d} = 81\%$.

5.2.4 One exercise session, and one meal and rapid insulin afterwards

The parameters to assess further for this input was α, β, T_d, k_{abs}, k_{HL}, f and V_i. When doing this it could be observed that β, T_d and k_{HL} were the most contributing parameters to the RSV of each of the three highest singular values. This indicated that these three parameters could be estimated precisely. Moving the insulin dose to before or after the meal did not change this very much. Delaying the meal to 150 minutes after the end of the exercise session or moving it closer to it made f and V_i more contributive to the third highest singular value, and therefore also more likely to be identifiable together with T_d and k_{HL}.

When trying to estimate these parameters, the only combination that got good results were T_d and k_{HL}. When trying to estimate these two together with either β, f or V_i, the estimations were not accurate. For example, the estimation error for β was $ee_\beta = 90\%$ when estimated together with T_d and k_{HL}, no matter how long or short after the exercise the meal was placed.

5.3 Diabetes Mellitus 2 subject

After removing the not interesting parameters, the resulting parameter set to investigate was:

$$P_{dm2,init} = [K_{mg}, \alpha, k_{Dia}, k_{IL}, \beta, G_{neo}, R_m, T_d, I_{Half}, R_i, n_h, T_Y, M_{lMx}, k_{abs}, S_i,$$
$$k_{gb}, k_{HL}, G_I, G_H, p_2, f, U_{ii}, V_i, R_{Hbas}, R_{Hmax}]$$

Also for this model the NR algorithm struggled to get any good results, so unless otherwise stated, the estimation results are from the DS method.

5.3.1 One meal

First for the preadapted model for diabetes 2, only one meal was examined, and the parameters related only to exercise were taken out (α, β and T_Y). For only one meal, without rapid insulin, the parameters T_d and V_i were also taken out. With this input, six parameters were chosen based on V_{sum} and RSV for the highest singular values. Those were k_{Dia}, k_{IL}, k_{abs}, S_i, k_{HL} and f. In the sensitivity analysis of these six parameters, k_{Dia} contributed the most to the

lowest singular value. Removing this it showed that S_i, k_{HL} and f contributed most to the highest singular value, which was much higher than the rest. That S_i showed to be one of the most identifiable parameters was not surprising, as the insulin sensitivity is a very low, but very important parameter for people with DM2. The parameter f was not surprising either, as it describes how much of the carbohydrates in the gut that appears in plasma as glucose.

All three combinations of pairs between the parameters S_i, k_{HL} and f got almost perfect estimation results. All three of them together could not be estimated perfectly however, with $ee_f = 12.9\%$ as the worst estimate.

5.3.2 Insulin and one meal

For this input, the parameters α, β and T_Y were still taken out, but T_d and V_i were taken in, as rapid insulin were now part of the input. This gave the parameters k_{IL}, T_d, k_{abs}, S_i, k_{HL} and f to examine further. This set was the same no matter if the rapid insulin was placed before, at the same time as, or after the meal. Just like with only one meal without insulin, the three parameters S_i, k_{HL} and f stood out in V_{sum}, in addition to T_d which contributed most to the second highest singular value. Looking at only these four parameters, one could observe that S_i and T_d contributed most to RSVs of the first and second highest singular values. These two are both related to insulin, and it is not surprising that they looked possible to estimate with this input.

When estimating, T_d and S_i did not get good estimation results together, with $ee_{T_d} = 79.5\%$. On the other hand, all the following parameter pairs did get accurate estimations: (T_D, k_{HL}), (T_D, f), (S_i, k_{HL}) and (S_i, f). In addition, the three parameters S_i, k_{HL} and f did get relatively good estimation accuracies together, with $ee_f = 2.2\%$ as the worst estimate.

5.3.3 Exercise

First, the meal and insulin related parameters were removed, namely k_{abs}, f, T_d and V_i. When analyzing this set, β and k_{HL} stood out as the biggest factors in the first and second singular values. k_{IL} contributed most to the third highest parameter and G_H had a relatively high V_{sum} value, so these were also chosen. These four parameters were looked at further. This showed that k_{HL} contributed to the singular value much higher than the others. k_{IL} and β contributed most to singular value two and three, but they were much lower than the first singular value, showed in figure 5.4. Therefore it looked like maybe only k_{HL} could be estimated alone. It was a little surprising that no exercise related variable was higher in identifiability.

The estimate of k_{HL} got a good result with the NR algorithm, but with the DS algorithm, the error was $ee_{k_{HL}} = 15\%$. Together with β the estimation were
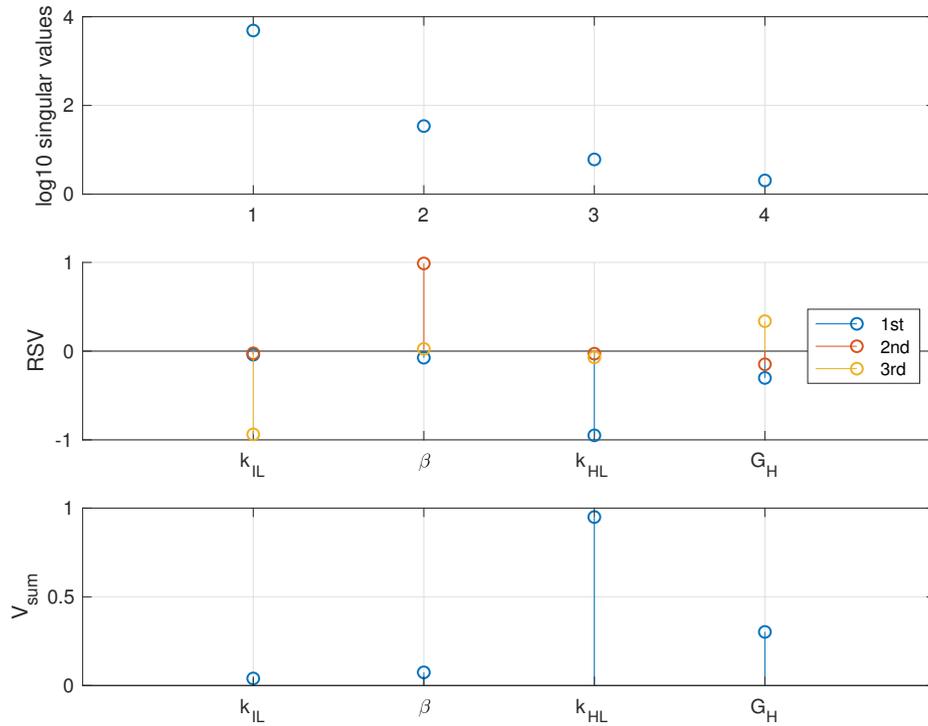
Figure 5.4: 1 exercise input sensitivity DM2

Singular values, RSV corresponding to the three highest singular values and Vsum, for
1 exercise input and DM2 subject.

perfect for the DS algorithm. k_{HL} and k_{IL} together did not give any good esti-
mations, with $ee_{k_{IL}} = 48\%$.

5.3.4 One meal and exercise session afterwards

The insulin related parameters, T_d and V_i, were taken out from $P_{dm2,init}$. For
exercise 10 minutes, 60 minutes and 150 minutes after a meal, the parameters
chosen to look at more was α, k_{IL}, β, k_{abs}, S_i, k_{HL} and f.

For exercise 10 minutes after meal the parameters α, β and k_{HL} had highest
sensitivity. However, the estimates of these three together were not accurate,
with $ee_\alpha = 43\%$ and $ee_\beta = 42\%$. The three combinations of pairs between them
all got pretty perfect estimates.

For exercise 60 minutes after the meal, the parameters α and β had highest
sensitivity. Also k_{abs}, k_{HL} and f looked promising. α and β together gave almost
perfect estimations, but together with either of k_{abs}, k_{HL} and f, the estimates
were poor. Here the estimates of k_{abs}, k_{HL} and f were not the problem, as these

got estimated quite good. Is was α and β that got estimated with low precision, with $ee_\alpha = 24\%$ and $ee_\beta = 31\%$ as the worst ones.

For exercise 150 minutes after the meal, the parameters S_i, k_{HL} and f looked like they most possibly could be estimated. Since the first singular value was so much higher than the rest, they might have been a little correlated. In addition α and β looked promising, but they correlated completely, so only one of them at the time. When estimating, the parameters S_i, k_{HL} and f together got a relatively accurate result, with $ee_{S_i} = 2.2\%$ as the worst estimation error. Them, together with either α or β did not get any good estimates at all.

5.3.5   One exercise session and one meal afterwards

The insulin related parameters were taken out from $P_{dm2,init}$, just like in the previous section. The parameters chosen to investigate further when it was 30 minutes between the exercise end and the meal were α, k_{IL}, β, k_{abs}, S_i, k_{HL} and f. When doing this, the parameters α, β and k_{HL} looked most identifiable. Also the parameter f looked promising. The estimates of α, β and k_{HL} were not perfect, as $ee_\alpha = 9.8\%$ and $ee_\beta = 8\%$.

When it was 100 minutes between the end of the exercise and the meal, the parameters α, k_{IL}, β, k_{abs}, S_i, k_{HL} and f was looked at more. This time α, k_{HL} and f looked like they could be estimated together. Also β and S_i looked possibly identifiable together with them. However, α, k_{HL} and f together did not get any accurate estimation results, with $ee_\alpha = 11.2\%$. The three combinations of pairs between them did get good estimations, and for (α, f) and (k_{HL}, f), the NR algorithm got pretty accurate estimates.

With a smaller meal during the exercise α, k_{IL}, β, S_i, k_{HL}, G_H and f was chosen to look more at. Of those, α, β and k_{HL} looked identifiable together with either k_{IL} or G_H. When trying to estimate α, β and k_{HL}, the accuracy was not good, with $ee_\alpha = 73\%$ as the worst estimate. When removing α, the accuracy of the estimates for β and k_{HL} were very good. Neither of the sets (k_{IL}, β, k_{HL}) or (β, k_{HL}, G_H) gave any good estimates.

5.4   Sensitivity analysis result table

Subject	Input	Most identifiable	Together with
Healthy	Meal	k_{abs}, k_{HL}	S_i or p_2
Healthy	Exercise	β, k_{HL}	G_H
Healthy	Exercise 10 min after meal	α, β, k_{abs}	
Healthy	Exercise 60 min after meal	k_{abs}, k_{HL}	α or β
Healthy	Exercise 150 min after meal	β, k_{abs}, k_{HL}	f
Healthy	Meal after exercise	β, k_{HL}	α or k_{abs}
Healthy	Small meal during exercise	β, k_{HL}	α or G_H
DM1	Insulin 20 min before meal	T_d, k_{abs}, k_{HL}	
DM1	Insulin at same time as meal	T_d, k_{HL}	f or V_i
DM1	Insulin 20 min after meal	T_d, k_{abs}, k_{HL}	f or V_i
DM1	Exercise	β, k_{HL}, G_H	
DM1	Exercise right after insulin and meal	T_d, k_{abs}, k_{HL}	f or V_i
DM1	Exercise 60 min or more after insulin and meal	T_d, k_{HL}	f or V_i
DM1	Insulin and meal 10 min after exercise	T_d, k_{HL}	β or f or V_i
DM1	Insulin and meal 60 min after exercise	T_d, k_{HL}	β
DM1	Insulin and meal 150 min after exercise	T_d, k_{HL}	β or f or V_i
DM2	Meal	S_i or k_{HL} or f	k_{abs}
DM2	Insulin and meal	T_d, S_i	k_{HL} or f
DM2	Exercise	k_{HL}	k_{IL}, β
DM2	Exercise 10 min after meal	α, β, k_{HL}	
DM2	Exercise 60 min after meal	α, β	k_{abs}, k_{HL}, f
DM2	Exercise 150 min after meal	S_i, k_{HL}, f	α or β
DM2	Meal 30 min after exercise	α, β, k_{HL}	f
DM2	Meal 100 min after exercise	α, k_{HL}, f	β, S_i
DM2	Meal during exercise	α, β, k_{HL}	k_{IL} or G_H

Table 5.1: Sensitivity analysis result.

Table showing which parameters looked like they could form an identifiable set, in the "Most identifiable" column, given preadapted model type "Subject" and input "Input". The "Together with" column shows the parameters which maybe could be estimated together with the parameters in "Most identifiable".

5.5   Sensitivity and estimation observations

When looking only briefly at the estimation results, one can observe a trend that of the parameter sets with only two parameters, most of them got estimated

accurately. One the other hand, most of the parameter sets consisting of three or more parameters did not get any good estimation results.

For the healthy model, for the parameter sets that were tried, k_{abs} got estimated pretty accurately four out eight times when it was in a parameter set with three parameters. That was much better than any of the other parameters, which failed much more times than they succeeded when part of a parameter set containing three parameters. The only exception from this was α, which got estimated correctly two out of five times when in a parameter set of three parameters.

For the DM1 model, only the parameter sets with two parameters got accurate estimates, while the parameter sets with three parameters never got any good estimation results at all.

k_{abs} was an important parameter for the healthy model, but barely among the most sensitive for the DM2 model. This can be because of the difference in the model structures. One can for example see that for the one meal input in the healthy model, k_{abs} was one of the two parameters with highest sensitivity. For the DM2 model, k_{abs} was not in the possibly identifiable set, but S_i and f was there instead.

Another interesting observation about the DM2 model was that when S_i, k_{HL} and f got estimated together with only one meal input, the estimates were not good. When the input was both insulin and a rapid insulin dose on the other hand, the estimates of those three were much better. Especially S_i improved a lot, from an estimate of 0.088 to 0.0011, when the real value of S_i was 0.001. Also for exercise 150 minutes after a meal, the estimates of these three parameters together were better. This gives an indication that an input of meal and rapid insulin is more useful than only a meal. It also shows that meal and exercise input gives more information than only a meal.

6 Discussion

In the initial parameter sensitivity analysis described in section 4.3, an assumption was made that parameters with very low sensitivity for the inputs in table 4.1, also would have low sensitivity for other combinations of the same input. Would different different input combinations in this initial parameter sensitivity analysis have given different initial parameter sets ($P_{h,init}$, $P_{dm1,init}$ and $P_{dm2,init}$) to work with? When looking at the parameters that were chosen to investigate further for the different inputs, all of them had pretty high sensitivity values for either meal, exercise or both in the initial analysis. This may lead to the conclusion that the assumption was a correct one to make. Both based on the time saved by not trying many different input combinations for every parameter in the initial analysis and that it does not seem like any important parameters were left out in this initial sensitivity analysis. Another approach could have been to try two different input combinations (For example meal before exercise, and exercise before meal) to be sure that no important parameters were discarded.

One letdown in the results was how poor the NR algorithm worked for estimation. One of the reasons for that may be the approximations of the gradient and Jacobian of the gradient used in the algorithm, explained in section 4.5.2. If these approximations were not accurate enough, it might sometimes have lead the algorithm to make wrong conclusions and search in the wrong directions in the parameter space. To make this better one could have tried to actually derive the gradient and Jacobian analytically, or put more emphasis on making sure the derivative approximations were accurate.

Another reason for poor NR estimation may have been the Jacobian matrix update made for every iteration. This was done instead of starting from scratch when deriving the Jacobian for each iteration, as explained in section 4.5.2. An improvement could have been to do an approximation for every iteration, just like in the first iteration, to get a more accurate Jacobian matrix.

Also, in the NR algorithm, a line search optimization could have been performed for every iteration when the search direction was found. This way it would make sure that the optimal point along the search trajectory was used when starting on the next iteration. The way the algorithm was implemented in this study, the step along the search direction was not optimized for every iteration.

Obviously the objective function $d_{tot}(\mathbf{x}, \mathbf{y})$ was not a convex objective function, because if it were, the algorithms would find the global minima pretty consistently. This means that it must have been several local minima around the global minimum that prevented the estimations from getting there. This prevention might be worse for a derivative based approach like the NR, because the derivative information can trick it into getting stuck in local minima that is not as good as the global minima. Also when taking into consideration

that the NR just makes approximation for the gradients and Jacobians matrices, one can derive that this might be the reason for the poor NR results. While DS of course can be stuck in local minima, the search technique of the simplex can also help it to avoid them.

The objective function gets more complicated the more parameters added. That may be the reason that two parameters together very often gave accurate estimates, but three or four parameters together very often gave much worse estimates.

Even when an estimation algorithm finds the global minimum of the objective function, the parameter estimations can still be wrong. This indicates that these parameters are correlated or indistinguishable from each other in estimation. This did not seem to be the case during estimation here, because every time the desirability was high and the curve fitting was good (this is the same thing), the parameter estimations were pretty accurate. This means that the global minimum of the objective function was the only point which would give a perfect curve fitting.

When trying to estimate against real measurement data, the objective function space might be more complicated, because of the noisy nature of measurement data. For the estimations done in this study, one knows that it exists one global minimum where the curve fitting is perfect and the parameter estimation is accurate. This might not be the case for estimation with real data, so curve fitting can be a harder problem.

k_{HL} was a parameter which proved to have high sensitivity for almost all kinds of input combinations. It featured in almost all of the parameter sets that seemed identifiable for the different inputs. That is because it is a parameter not dependent on input, because glycogenolysis happens more or less all the time no matter what the person eats or how much he exercises.

One limitation of this study is that only simulated measurement data from the model itself are used, not real measurement data. This mostly limits the estimation part. Since the data is from the same model, we know that all the dynamics in it is modelled. More about this in chapter 8: Suggestions for future work.

7 Conclusion

Since the desired use of the GlucoPred metabolism model is for simulation and prediction of future plasma glucose concentration values, one wants the model parameters to be calibrated individually. The main goal of this project was to analyze the identifiability of a reduced complexity version of the model. That is, the identifiability of a small subset of the model parameters. This was to be done with respect to different model excitations and the different model types, healthy, DM1 and DM2.

For the sensitivity analysis, some of the parameters had high sensitivity for many different excitations and model types, like k_{HL}, α, β, T_d and f. Other parameters, like S_i, K_{abs}, p_2, G_H, V_i and k_{IL} had high sensitivity only for one model type or one specific input. The rest of the parameters had lower sensitivity for most of the input and model types. The sensitivity analysis confirmed that input is an important factor to consider with respect to identifiability. This because the different input combinations often gave completely different sensitivity plots, showed by singular values, RSV and V_{sum}. Also the different preadapted model types gave different results for the same inputs. This shows that the same excitation scheme cannot be used when trying to estimate parameters for a DM2 subject and a DM1 subject, for example.

k_{HL} was a model parameter which had very high sensitivity for almost all kinds of input, and it was not dependent on the input or the model type to be estimated with accuracy.

S_i was a parameter with high sensitivity during meals for the DM2 model. This is good because this is a parameter that is very important for people with DM2. It is also an important parameter to try to estimate for people who is in the risk zone of getting DM2 and/or have prediabetes, as it can indicate if one is getting better or worse in terms of getting DM2.

The estimation techniques did not work very well for more than two parameters at a time. For three parameters together, very few of the sets were estimated correctly. For two parameters together, very few of the parameter sets were estimated inaccurately.

For all three models, an input of only exercise looked like it gave the worst sensitivity and identifiability. The only time a model parameter set of two parameter did not get a good estimation result was for k_{IL} and k_{HL} in the DM2 model, and then the input was only exercise. Also the parameter set of only k_{HL} alone got a bad estimation result with the DS method, only exercise input and DM2 model.

For the rest of the model input combinations (only meal, meal with insulin, meal after exercise, exercise after meal), it was not easy to see a pattern for which inputs were most useful. For the healthy model, both the only meal input and the exercise after meal input, each gave a good estimation of a parameter set containing three parameters. For the DM1 model, none of the parameter

sets with three parameters included could be identified for any input. The DM2 model gave good estimation results for a parameter set of three parameters for both the input of rapid insulin + meal, and the input of exercise 150 minutes after a meal.

The two different estimation methods were also compared. The DS algorithm worked much better than NR as a parameter estimation algorithm for this model. This was especially for the DM1 model, where NR could not get any good result for any parameter set tested, even though DS could for some.

This study does not provide any results in the form of recommended excitation schemes for parameter estimation or parameter sets that should definitely be estimated together. These results can be used as a starting point for an analysis of this model more in-depth than what is done here, with respect to identifiability.

8 Suggestions for future work

FIM One could also use FIM and a covariance matrix in the identifiability analysis to gain a better understanding of which parameters have information that the others do not and which parameters that correlate.

Real glucose measurement data Try to do parameter estimation against real measurement data. This might reveal effects in the real human not modelled in the GlucoPred model.

More simulated output measurements The identifiability relies on the different outputs of the model, simulating sensor measurements of the real system. In this study, only the subcutaneous glucose measurements were assessed when analyzing the identifiability. More measurements together will give other results.

The insulin sensitivity S_i as a parameterized function of time (and possibly other states) has not been implemented and could have made an impact one the result for input over more than one day. This parameter is such an important one that this is a task that one could do in the future.

The meal parameters Parameters describing a meal (meal size in carbohydrates $[g]$, glycemic index $SlowFact$ and time of the meal) have not been part of the analysis, but could have had an impact on it.

9 Bibliography

[1] I. D. Federation, *IDF Diabetes Atlas, Eighth edition*, 2017. [Online]. Available: http://www.diabetesatlas.org/

[2] J. Bass and J. Takahashi, "Circadian integration of metabolism and energetics," *Science (New York, N.Y.)*, vol. 330, pp. 1349–54, 12 2010.

[3] R. Visentin, C. Dalla Man, Y. C Kudva, A. Basu, and C. Cobelli, "Circadian variability of insulin sensitivity: Physiological input for in silico artificial pancreas," *Diabetes technology & therapeutics*, vol. 17, 12 2014.

[4] J. Qian and F. Scheer, "Circadian system and glucose metabolism: Implications for physiology and disease," *Trends in Endocrinology & Metabolism*, vol. 27, 04 2016.

[5] A. Saad, C. Dalla Man, D. Nandy, J. Levine, A. E Bharucha, R. Rizza, R. Basu, R. Carter, C. Cobelli, Y. C Kudva, and A. Basu, "Diurnal pattern to insulin secretion and insulin action in healthy individuals," *Diabetes*, vol. 61, pp. 2691–700, 06 2012.

[6] J. Yoshino, P. Almeda-Valdes, B. W Patterson, A. Okunade, S.-I. Imai, B. Mittendorfer, and S. Klein, "Diurnal variation in insulin sensitivity of glucose metabolism is associated with diurnal variations in whole-body and cellular fatty acid metabolism in metabolically normal women," *The Journal of clinical endocrinology and metabolism*, vol. 99, p. jc20141579, 05 2014.

[7] C. D. Man, R. A. Rizza, and C. Cobelli, "Meal simulation model of the glucose-insulin system," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 10, pp. 1740–1749, Oct 2007.

[8] R. Bergman, Y. Ider, C. Bowden, and C. Cobelli, "Quantitative estimation of insulin sensitivity," *The American journal of physiology*, vol. 236, pp. E667–77, 07 1979.

[9] R. N Bergman, "Minimal model: Perspective from 2005," *Hormone research*, vol. 64 Suppl 3, pp. 8–15, 02 2005.

[10] J. Sorensen, "A physiological 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.

[11] J. Chase, G. Shaw, J. Lin, C. V Doran, C. Hann, T. Lotz, G. Wake, and B. Broughton, "Targeted glycemic reduction in critical care using closed-loop control," *Diabetes technology & therapeutics*, vol. 7, pp. 274–82, 05 2005.

[12] R. Hovorka, V. Canonico, L. Chassin, U. Haueter, M. Massi-Benedetti, M. Orsini Federici, T. Pieber, H. C Schaller, L. Schaupp, T. Vering, and M. Wilinska, "Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes," *Physiological measurement*, vol. 25, pp. 905–20, 09 2004.

[13] O. M. Staal, Private Communication, 2018.

[14] H. Miao, X. Xia, A. Perelson, and H. Wu, "On identifiability of nonlinear ode models and applications in viral dynamics," *SIAM review. Society for Industrial and Applied Mathematics*, vol. 53, pp. 3–39, 01 2011.

[15] J. A. Jacquez and T. Perry, "Parameter estimation: local identifiability of parameters," *American Journal of Physiology-Endocrinology and Metabolism*, vol. 258, no. 4, pp. E727–E736, 1990, pMID: 2333964. [Online]. Available: https://doi.org/10.1152/ajpendo.1990.258.4.E727

[16] J. Garcia Tirado, C. Zuluaga-Bedoya, and M. Breton, "Identifiability analysis of three control-oriented models for use in artificial pancreas systems," *Journal of diabetes science and technology*, vol. 12, pp. 937–952, 08 2018.

[17] L. Ljung, "Convergence analysis of parametric identification methods," *IEEE Transactions on Automatic Control*, vol. 23, no. 5, pp. 770–783, October 1978.

[18] H. Stigter, D. Joubert, and J. Molenaar, "Observability of complex systems: Finding the gap," *Scientific Reports*, vol. 7, 11 2017.

[19] E. C. Harrington, "The desirability function," *Industrial Quality Control*, vol. 21, no. 10, pp. 494–498, 1965.

[20] S. N. Deming, "Multiple-criteria optimization," *Journal of Chromatography A*, vol. 550, pp. 15–25, 12 1991.

[21] D. Fuller and W. Scherer, ""the desirability function: underlying assumptions and application implications"," vol. 4, 11 1998, pp. 4016 – 4021 vol.4.

[22] R. L. Plackett and J. P. Burman, "The design of optimum multifactorial experiments," *Biometrika*, vol. 33, no. 4, pp. 305–325, 1946. [Online]. Available: http://www.jstor.org/stable/2332195

[23] H. Stigter and J. Molenaar, "A fast algorithm to assess local structural identifiability," *Automatica,* vol. 58, 08 2015.

Appendix A

Appendix A.1   Values for preadapted parameter sets

Parameter	Unit	Healthy	DM1	DM2
K_{mg}	mg/dL	120	120	120
α	(mg L)/(mU dL min)	0.02	0.02	0.02
k_{Dia}	pg/mL	0.0	10.0	0.0
k_{IL}	mg/min	750.0	750.0	750.0
β	mg/(dL min)	1.0	1.0	1.0
G_{neo}		0.3	0.3	0.3
R_m	(mU dL min)/(mg L)	21.0	0.0	21.0
T_d	min	1000.0	15.0 (30.0)*	10.0
I_{Half}	mU/L	40.0	40.0	40.0
R_i	mU/L	118.0	0.0	118.0
U_b	mU/min	0.0	0.0	0.0
n_h		6.4	6.4	6.4
n_i		4.2	4.2	4.2
T_Y	min	6.0	6.0	6.0
I_0	mU/L	15.0	15.0	15.0
a_P	L/mU	0.1	0.1	0.1
M_{lMx}	g	100.0	150.0	100.0
HR_B	beats/min	71.0	71.0	71.0
k_{abs}	1/min	0.05	0.04	0.062
T_{max}	min	600.0	600.0	600.0
T_{ds}	min	1000.0	1000.0	1000.0
S_i	L/(mU min)	0.2	0.2	0.001
k_{gb}	1/min	0.0012	0.0012	0.0012
k_{HL}	(mg/min) / (pg/mL)	3.6	2.0	2.0
k_{Del}	1/min	0.1	0.1	0.1
E_{neo}	mg/min	50.0	50.0	50.0
k_{gm}	1/min	0.4	0.4	0.4
G_I	mg/dL	150.0	150.0	150.0
G_H	mg/dL	55.0	55.0	55.0
p_2	1/min	0.02	0.02	0.01
f		0.8	0.8	0.8
U_{ii}	mg/(dL min)	0.78	0.78	0.78
V_i	L	15.7	15.7	15.7
R_{Hbas}	pg/mL	130.0	130.0	130.0
s_{Comp}		2.0	2.0	2.0
n	1/min	0.142	0.142	0.142

R_{Hmax}	pg/mL	665.0	665.0	665.0
r		1.1	1.1	1.1
k_{glg}	1/min	0.6	0.6	0.6
V_g**	dL	128.0	128.0	128.0
HR_M***	beats/min	184.0	184.0	184.0

Preadapted parameter values.

Empty unit means unitless parameter.

* T_d had value 15 in the original preadapted set, but was changed to 30 in this analysis.

** V_g approximated from weight with $V_g = 1.6 * w$, where weight $w = 80kg$ was used in this study.

*** HR_M approximated from age with $HR_M = 208 - 0.8 * age$, where $age = 30$ was used in this study.

Appendix A.2   Parameter ranges

Parameter	Unit	Minimum	Maximum
K_{mg}	mg/dL	100	120
α	(mg L)/(mU dL min)	0.01	0.1
k_{Dia}	pg/mL	0.0	27.1
k_{IL}	mg/min	300.0	1715.0
β	mg/(dL min)	0.2	10.0
G_{neo}		0.3	0.45
R_m	(mU dL min)/(mg L)	0.0	21.0
T_d*	min	5.0	100.0
I_{Half}	mU/L	25.0	60.0
R_i	mU/L	0.0	118.0
U_b	mU/min	0.0	0.016424
n_h		6.4	8.0
n_i		4.19	4.21
T_Y	min	6.0	10.0
I_0	mU/L	14.9	15.1
a_P	L/mU	0.09	0.11
M_{lMx}	g	100.0	175.0
HR_B	beats/min	70.9	71.1
k_{abs}	1/min	0.005	0.15
T_{max}	min	600.0	1000.0
T_{ds}	min	999.9	1000.1
S_i	L/(mU min)	0.001	0.4
k_{gb}	1/min	0.0	0.0012
k_{HL}	(mg/min) / (pg/mL)	1.1	14.2
k_{Del}	1/min	0.09	0.11
E_{neo}	mg/min	49.9	50.1
k_{gm}	1/min	0.39	0.41
G_I	mg/dL	100.0	154.0
G_H	mg/dL	10.0	88.0
p_2	1/min	0.007	0.08
f		0.11	1.7
U_{ii}	mg/(dL min)	0.0	0.78
V_i	L	7.92	35
R_{Hbas}	pg/mL	60.0	130.0
s_{Comp}		1.9	2.1
n	1/min	0.1419	0.1421
R_{Hmax}	pg/mL	200.0	665.0
r		1.09	1.11

k_{glg}	1/min	0.59	0.61
V_g	dL	105.6	467.0
HR_M	beats/min	140.0	205.0

Parameter ranges.

Empty unit means unitless parameter.
*T_d is not relevant for healthy subjects and that is why its value, 1000 min, are outside the range for healthy subjects.