UNIVERSITÀ
DEGLI STUDI
DI PADOVA

DIPARTIMENTO
DI INGEGNERIA

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

CORSO DI LAUREA MAGISTRALE IN BIOINGEGNERIA

MODELING THE EFFECT OF SGLT2 INHIBITORS IN PATIENT WITH TYPE 1 DIABETES

Relatore: Prof. Chiara Dalla Man

Laureanda: Giulia Nava

Correlatore: Prof. Roberto Visentin

ANNO ACCADEMICO 2024 – 2025

Data di laurea 27/02/2024

Index

ABSTRACT

In recent years, the prevalence of diabetes mellitus has been steadily increasing. Especially for type 1 diabetes, prevention and treatment are difficult to manage as the risk factors that interact with the genetic predisposition triggering the autoimmune reaction are still unclear.

This thesis investigates the potential impact of sodium-glucose cotransporter type 2 inhibitors (SGLT2i) in patients with type 1 diabetes (T1D). Despite being primarily indicated for type 2 diabetes, SGLT2 inhibitors have recently garnered attention for their potential benefits in adjunction to insulin therapy, also in individuals with type 1 diabetes.

This study employs mathematical modeling techniques to elucidate the mechanisms through which SGLT2 inhibitors modulate glucose metabolism and other related parameters in patients with type 1 diabetes. For the validation of the model, a database of 12 subjects involved in a double-blind, placebo-controlled crossover study with a 4-week washout period was used.

Model's performance was assessed in terms of data fit, precision of estimates and physiological plausibility. Satisfactory results have been obtained although additional studies and longer clinical trials are recommended.

SOMMARIO

Negli ultimi anni, la diffusione del diabete mellito è in costante aumento. Soprattutto per il diabete di tipo 1, la prevenzione e il trattamento sono difficili da gestire, poiché i fattori di rischio, che interagiscono con la predisposizione genetica che scatena la reazione autoimmune, non sono ancora chiari.

Questa tesi si propone di studiare il potenziale impatto degli inibitori del co-trasportatore sodio-glucosio di tipo 2 (SGLT2i) nei pazienti con diabete di tipo 1 (T1D). Nonostante siano stati inizialmente utilizzati nel trattamento di pazienti con diabete di tipo 2, gli inibitori SGLT2 hanno recentemente attirato l'attenzione per i loro potenziali benefici in aggiunta alla terapia insulinica anche nei pazienti diabetici di tipo 1.

Questo studio impiega tecniche di modellazione matematica per chiarire i meccanismi attraverso i quali gli inibitori SGLT2 modulano il metabolismo del glucosio e altri parametri correlati, in pazienti con diabete di tipo 1. Per la validazione del modello è stato utilizzato un database di 12 soggetti coinvolti in uno studio crossover in doppio cieco, controllato con placebo e con un periodo di wash out di 4 settimane.

Le prestazioni del modello sono state valutate in termini di adattamento ai dati, precisione delle stime e plausibilità fisiologica. Sono stati ottenuti risultati soddisfacenti, anche se si raccomandano ulteriori studi e trial clinici più lunghi.

Chapter 1
INTRODUCTION

1.1 Type 1 Diabetes.

Type 1 Diabetes Mellitus (T1D) is a chronic, autoimmune disease, characterized by immune-mediated destruction of pancreatic β-cells resulting in insulin deficiency. T1D represents almost 10-15% of all diabetes cases and it is also known as juvenile or insulin-dependent diabetes, as it historically was associate with the onset at a young age, even if this opinion has changed over the past decade, and the only possible treatment is with insulin [1].

The classic trio of symptoms associated with the disease onset are: polydipsia, polyphagia, and polyuria, which, together with hypoglycaemia constitute the main criteria for the diagnosis of the disease [1].

T1D seems to be more common in men than women and its incidence is strictly related with specific months of the year (more cases are diagnosed in autumn and winter [1]). In addition, the global incidence of T1D is strongly associated to the geographic area we are considering; this disorder seems to be uncommon in Cina, India and Venezuela whereas Finland, Sweden and Sardinia show a much higher incidence. In many countries the incidence of T1D has been increasing for several decades [2], even if this trend is not the same for all age groups. The reasons of this variability, in terms of distribution, incidence and age has been attributed to environmental influences, genetic predispositions and lifestyle [3].

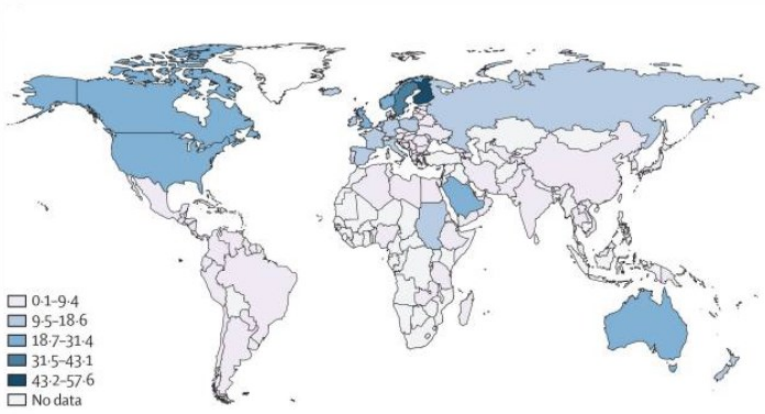

Figure 1.1.1: Distribution of Type 1 Diabetes worldwide. Adapted from [3]

In type 1 diabetes the body produces autoantibodies that attack the beta cells of the pancreas, which are responsible for insulin production, recognising them as foreigners. Therefore, the production of this hormone, whose task is to regulate the utilisation of glucose by the cells, is reduced to zero. This results in a situation of excess of glucose in the blood identified as hyperglycaemia. The lack of insulin, therefore, does not allow the body to utilise sugars introduced through the diet, which are thus eliminated with the urine. In this situation, the body is forced to produce energy in other ways, mainly through fat metabolism, which results in the production of so-called ketone bodies, which leads to diabetic ketoacidosis (DKA). The accumulation of ketone bodies in the body, if it is left untreated, can lead to very dangerous consequences, including coma.

Currently, there is no cure for type 1 diabetes mellitus and the only possible therapy is the exogenous administration of insulin, which involves basal insulin to stabilize blood glucose levels during fasting states and bolus insulin before carbohydrate (CHO) consumption, meals or beverages, to counteract respective blood glucose level rises.

Although it is normal for blood sugar levels to fluctuate throughout the day in nondiabetic individuals, but such fluctuations are modest. In diabetic patients these fluctuations are much more pronounced, but both hyper- and hypoglycaemia are conditions that should be avoided at all costs or at least reduced as much as possible. In fact, blood sugar levels that are too high or too low can cause health issues, in both short and long term, such as damage the body's organs, heart attack, stroke, diabetic coma, problems with kidneys, eyes, feet and nerves, which could then end up fatally.

Patients with type 1 diabetes should therefore try to keep their glucose levels in a range between 70-180 mg/dl of glucose, that is considered the "safety range".

<70 mg/dl	Hypoglycaemia
70-180 mg/dl	Target range
>180 mg/dl	Hyperglycaemia

Table 1.1: Blood sugar level ranges

Diagnosis of diabetes has historically included fasting blood glucose higher than 7 mmol/L (126 mg/dL), any blood glucose of 11,1 mmol/L (200 mg/dL) or higher with symptoms of hyperglycaemia, or an abnormal 2 horal glucose-tolerance test [4]. In 2009 the American Diabetes Association, decides to include the Glycated Haemoglobin

(HbA_{1C}) as a test for the diabetes diagnosis. People with HbA_{1C} ≥6,5% were treated as diabetic [5].

The diagnosis of T1D is a crucial issue since it can be easily mistaken with type 2 diabetes and the associated symptoms are not so evident in an early stage of the onset of the disease. The development of diabetes, in fact, can be divided into 3 stages. Stage 1 is characterized by the presence of autoantibodies and the absence of dysglycaemia (too low or too high blood sugar levels). In stage 2, both autoantibodies and dysglycaemia are present, whereas only stage 3 is characterized also by the presence of symptoms [6].

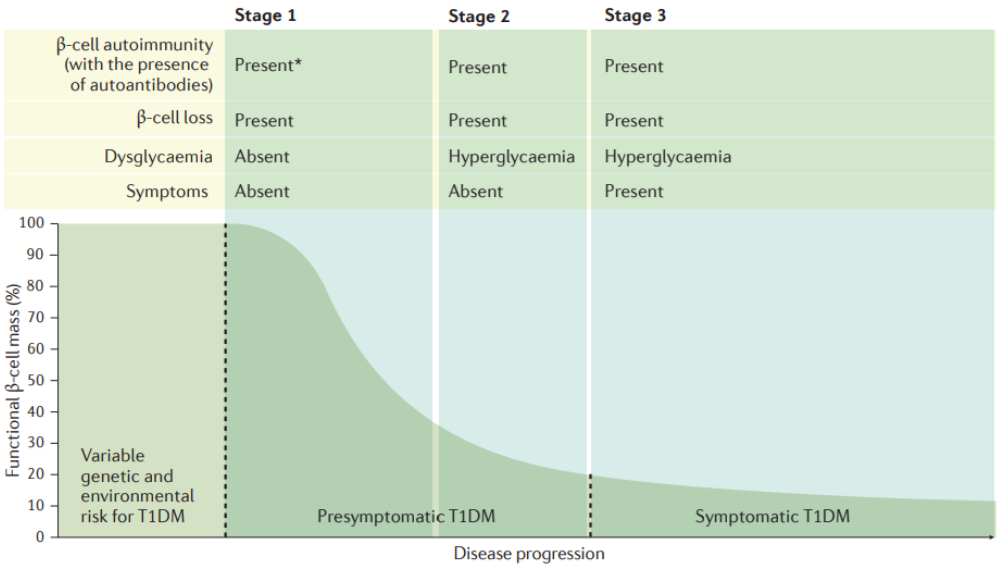

Figure 1.2: T1D stages. Adapted from [6]

1.2 Current therapies and future prospective.

Currently the first-line therapeutic option for treating T1D is the insulin replacement therapies, in which the exogenous insulin is used to achieve glucose-lowering action [7]. Depending on the type of treatment to be performed, different types of insulin can be used:

- Short-acting insulin: it starts its action around 30 minutes after injection. It reaches its peak effect at 90-120 minutes and lasts about 4 to 6 hours.

- Rapid-acting insulin: its action starts 15 minutes after the injection, it reaches its peak effect at 60 minutes and lasts about 4 hours. This type of insulin is generally used 15-20 minutes before meals.
- Intermediate-acting insulin: it starts working in about 1 to 3 hours, it reaches peak effect at 6-8 hours and lasts 12 to 24 hours.
- Long and ultra-long-acting insulin: which provides coverage for as long as 14 to 40 hours.

Insulin can be delivered by syringes or pens, inhaler or pumps.

- Syringes and pens are generally used for multiple daily injection (MDI) therapy and involve a combination of both short-acting (bolus) and long-acting (basal) insulin.
- Inhalers are usually used when insulin bolus is needed immediately, i.e. before a meal. Inhalation insulin powder is not a substitute for long-acting insulin and should therefore be used in combination with long-acting insulin. Unlike common inhaler sprays, the insulin dose is first released into a closed compartment that is an integral part of the inhaler and only then inhaled deeply by the patient. The diffusion of insulin into the chamber before inhalation has the advantage of avoiding the high-speed impact of insulin in the throat and first airways and facilitating a slow and deep inhalation.
- Insulin pumps are small devices able to deliver doses of insulin at specific times, thanks to a pre-programmed schedule. A traditional pump includes:
  - the pump, including controls, processing module and batteries.
  - a disposable reservoir for insulin, located inside the pump.
  - an infusion set, which includes a cannula for subcutaneous insertion and a tubing system to connect the insulin reservoir to the cannula. Insulin pumps can manage continuous subcutaneous insulin infusion (CSII) and can deliver both boluses and basal insulin.

All these treatments need actions by the user, in terms of blood sugar monitoring and carbohydrate counting, also known as self-monitoring of blood glucose (SMBG), which is a critical part of insulin therapy. Challenges that affect adherence to SMBG include pain, costs, behavioural and technical skills and motivation [8]. For those reasons,

attempts are being made to replace SMBG with real-time continuous glucose monitoring (rt-CGM), to improve patients' quality of life and limit pain as much as possible.

So far, the user's control is always required, both in terms of monitoring and the right dose of insulin to be delivered and these traditional delivery systems involves an invasive procedure and do not provide long-term insulin independence. The ultimate goal would be to make the whole procedure automatic, and this is exactly what is being attempted with the artificial pancreas. It combines the technology of a pump with that of continuous glucose monitoring (CGM) to adjust insulin delivery based on real-time glucose regarding. By automating insulin delivery and closely monitoring glucose levels, it can help regulate blood sugar more effectively than traditional methods that rely on manual injections or pump adjustments.

Others challenging future prospective are immunotherapy, islet cell transplantation or stem cell-based therapy, for which efforts are currently underway to enhance the outcomes achieved so far. Future endeavours will, therefore, require a novel focus, leveraging prior experience with regard to the immunopathophysiology of T1D, whilst also exploring the promise of combination therapies that integrate tried or new treatment modalities [9].

Firstly evaluated for type 2 diabetes, sodium–glucose co-transporter-type 2 (SGLT2) inhibitors treatment showed significant benefits also in T1D in reducing insulin dose requirements, improving glycaemic control and reducing the body weight, when added to insulin therapy. SGLT2 (sotagliflozing, dapagliflozin and empagliflozin) inhibitors are selective and reversible inhibitors of renal SGLT2, the major transporter responsible for renal glucose reabsorption which lower blood glucose levels by restraining the absorption of glucose in the small intestine and promoting the renal excretion of glucose [10].

So far, sotagliflozing and dapagliflozin are approved in Europe and Japan as adjuncts to insulin for the management of overweight people with T1D when insulin alone does not provide an adequate glycaemic control [9]. However, long-term clinical trials and additional studies are needed to evaluate all potential adverse effects and to further understand additional benefits of SGLT2 inhibitors [7].

1.3 Importance of models in physiology and medicine.

A model is a mathematical representation of a system, defining the relationship among all the possible inputs and related outputs of the system. A model is something which can simulate the reality, but it's poorer and more schematic than the reality or the actual objects. Models can be used for different purposes such as simulate, predict and gain insights, they allow us to avoid dangerous experiments, that cannot be done in vivo and to forecast what will happen in the future based on the past.

Mathematical modelling is now widely applied in physiology and medicine to support life scientists and clinical workers and finds applications in various fields, such as medical research, education and supporting clinical practice. They can help the users in yielding quantitative insights into the way physiological systems are controlled, in exploring the dynamic effects of pathophysiological processes or of drug therapy and enabling estimates to be made of physiological parameters that are not directly measurable. Mathematical models are also largely used in epidemiology, a field in which these models succeed in overcoming most of the ethical issues related to animal and human trial, by following the principles of the "three Rs alternatives": replacement, reduction and refinement.

Models can be divided into three categories:

- Black box: in which the internal structure is unknown or too complex to be represented by a set of equations. The parameters have no physical/physiological meaning and we can only know which are the inputs and the output. This kind of models are also known as Models of Data.
- White box: in which not only the input-output relationship is known, but also the internal behaviour of the system, which consist of a set of equations obtained by putting together all the subsystems known equations.
- Grey box: which is characterized by known or partially known structure but unknown parameters. Measured input and output and assumptions on the model structure are used to obtain values for the unknown parameters, this is a procedure also known as parameter estimation.

These last two models are also known as Models of System.

1.3.1 The UVA/Padova Simulator.

One of the most important and innovative models capable of simulating the glucose-insulin system is the UVA/Padova Simulator. This model is implemented in Matlab and allows the user to simulate, among many other things, metabolic dynamics, variation in glucose levels and impact on glycaemic control. The UVA/Padova Simulator emulated meal challenges and included a population of 300 in silico subjects (100 adults, 100 adolescents and 100 children), each of which is represented by a model parameter vector, randomly extracted from an appropriate joint parameter distribution.

The UVA/Padova Simulator puts in relation the measured plasma concentrations (Glucose G, and Insulin I) to glucose fluxes (rate of appearance Ra, production EGP, utilization U, renal excretion E) and insulin fluxes (secretion S, and degradation D).

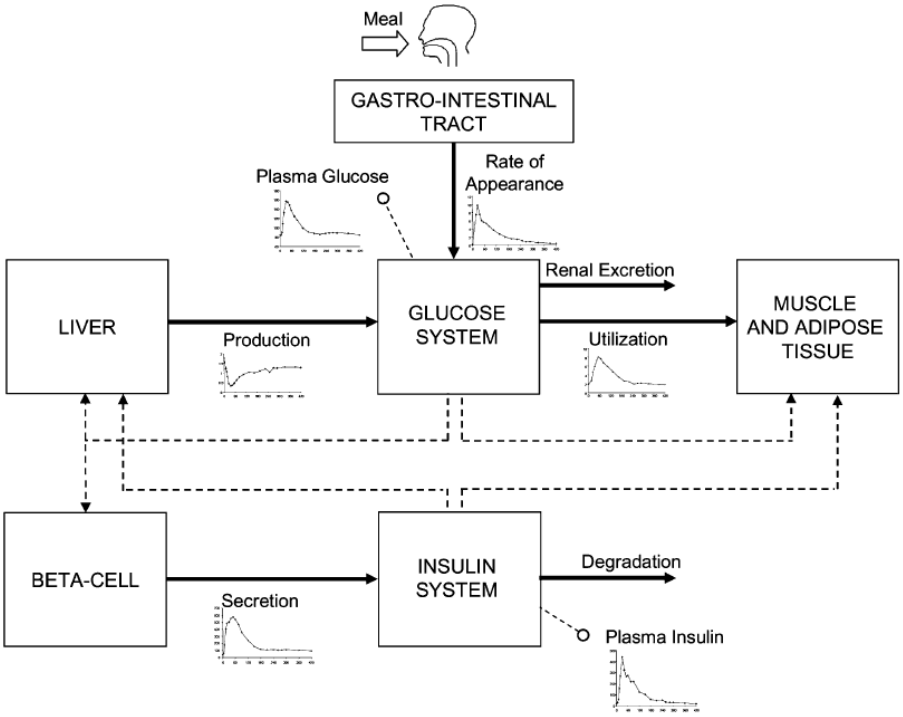

Figure 1.3: Scheme of the glucose-insulin control system. Adapted from [11]

The model can be divided into subsystems:

- Glucose subsystem:

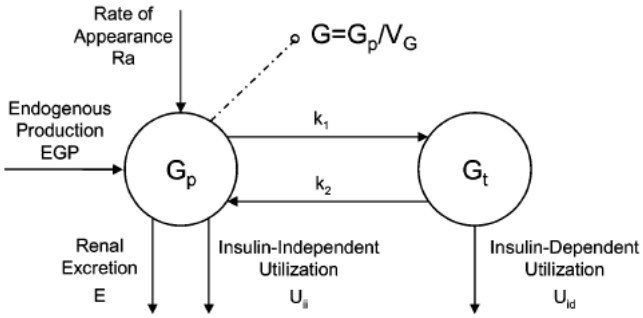

Figure 1.4: Glucose subsystem. Adapted from [11]

composed of two compartments G_p, glucose masses in plasma and rapidly equilibrating tissues and G_t, glucose masses in slowly equilibrating tissues. This subsystem is described by the following equations:

$$\begin{cases} \dot{G}_p(t) = EGP(t) + Ra(t) - U_{ii}(t) - E(t) - k_1 G_p(t) + k_2 G_t(t) & G_p(0) = G_{pb} \\ \dot{G}_t(t) = -U_{id}(t) + k_1 G_p(t) - k_2 G_t(t) & G_t(0) = G_{tb} \\ G(t) = \dfrac{G_p}{V_G} & G(0) = G_b \end{cases}$$

(1. 1)

Where EGP is the endogenous glucose production (mg/kg/min), Ra is the glucose rate of appearance in plasma (mg/kg/min), E is the renal excretion (mg/kg/min), U_{ii} and U_{id} are the insulin-independent and -dependent glucose utilizations (mg/kg/min), Vg is the distribution volume of glucose (dl/kg), k_1 and k_2 are the rate parameters and the suffix b denotes basal state.

- Insulin subsystem:

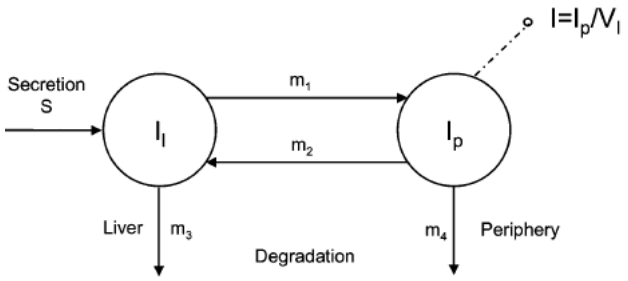

Figure 1.5: Insulin subsystem. Adapted from [11].

Composed by two compartments: I_l and I_p which are the insulin masses in liver and plasma respectively (pmol/l). This subsystem is described by the following equations:

$$\begin{cases} \dot{I}_l(t) = -(m_1 + m_3(t)) \cdot I_l(t) + m_2 I_p(t) + S(t) & I_l(0) = I_{lb} \\ \dot{I}_p(t) = -(m_2 + m_4) \cdot I_p(t) + m_1 \cdot I_l(t) & I_p(0) = I_{pb} \\ I(t) = \dfrac{I_p}{V_I} & I(0) = I_b \end{cases}$$

(1. 2)

Where S is the insulin secretion (pmol/kg/min), V_I is the distribution volume of insulin (l/kg) and $m_1, m_2\ and\ m_4$ are the rate parameters (min^{-1}).

- Unit process models and identification:

  The unit process of glucose and insulin subsystem was identified from average data with a forcing function strategy and are shown in figure 1.6.

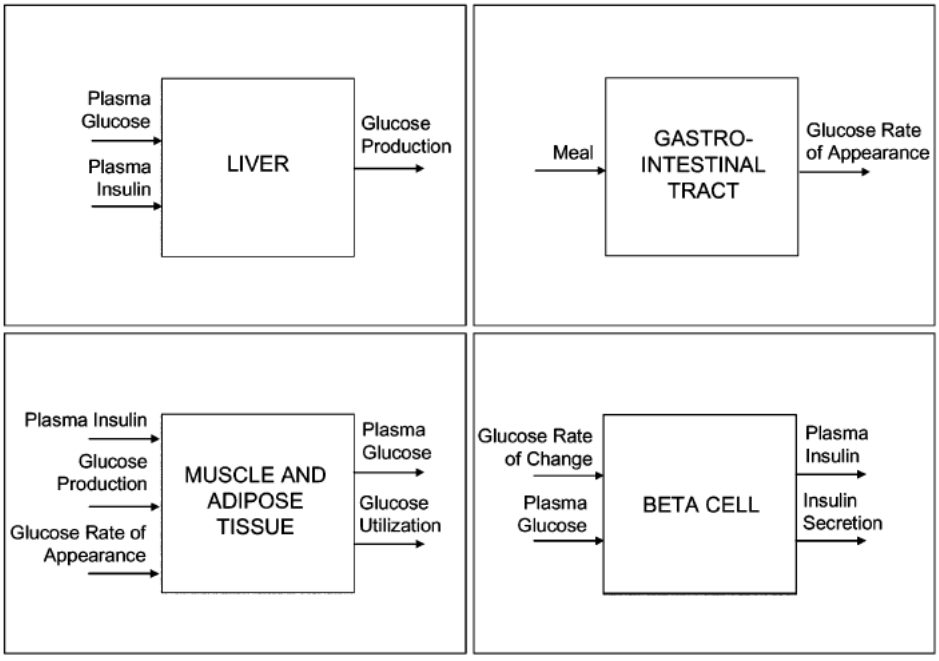

Figure 1.6: Unit process models and forcing function strategy: endogenous glucose production (top left panel), glucose rate of appearance (top right panel), glucose utilization (bottom left panel), insulin secretion (bottom right panel). Adapted from [11].

- Endogenous glucose production (constrained to be non-negative):

$$EGP(t) = k_{p1} - k_{p2} \cdot G_p(t) - k_{p3} \cdot I_d(t) - k_{p4} \cdot I_{po}(t) \qquad EGP(0) = EGP_b$$

(1. 3)

Where k_{p1} (mg/kg/min) is the extrapolated EGP at zero glucose and insulin, k_{p2} (min^{-1}) is the liver glucose effectiveness, k_{p3} (mg/kg/min per pmol/l) is the parameter governing amplitude of insulin action on the liver and k_{p4} (mg/kg/min/(pmol/kg)) is the parameter governing amplitude of portal insulin action on the liver.

I_{po} is the amount of insulin in the portal vein (pmol/kg), I_d (pmol/l) is a delayed insulin signal realized with a chain of two compartments:

$$\begin{cases} \dot{I}_1(t) = -k_i \cdot [I_1(t) - I(t)] & I_1(0) = I_b \\ \dot{I}_d(t) = -k_i \cdot [I_d(t) - I_1(t)] & I_d(0) = I_b \end{cases}$$

(1. 4)

where k_i (min^{-1}) is the rate parameter accounting for delay between insulin signal and insulin action.

- Glucose rate of appearance: describes the glucose transit through the stomach and intestine, by assuming the stomach to be represented by two compartments and a single compartment for the gut.

$$\begin{cases} Q_{sto}(t) = Q_{sto1}(t) + Q_{sto2}(t) & Q_{sto}(0) = 0 \\ \dot{Q}_{sto1}(t) = -k_{gri} \cdot Q_{sto1}(t) + D \cdot d(t) & Q_{sto1}(0) = 0 \\ \dot{Q}_{sto2}(t) = -k_{empt}(Q_{sto}) \cdot Q_{sto2}(t) + k_{gri} \cdot Q_{sto1}(t) & Q_{sto2}(0) = 0 \\ \dot{Q}_{gut} = -k_{abs} \cdot Q_{gut}(t) + k_{empt}(Q_{sto}) \cdot Q_{sto2}(t) & Q_{gut}(0) = 0 \\ Ra(t) = \dfrac{f \cdot k_{abs} \cdot Q_{gut}(t)}{BW} & Ra(0) = 0 \end{cases}$$

(1. 5)

where k_{gri} (min^{-1}) is the rate of grinding, $k_{empt}(Q_{sto})$ (min^{-1}) is the rate constant of gastric emptying, which is a non-linear function of Q_{sto}, $k_{abs}(min^{-1})$ is the rate constant of intestinal absorption, f is the fraction of intestinal absorption, D (mg) is the amount of ingested glucose, BW (kg) is the body weight and Ra (mg/kg/min) is the appearance rate of glucose in plasma.

- Glucose utilization: assumed to be made up of two components, insulin-independent and -dependent utilization. The first one is constant and represents glucose uptake by the brain and erythrocytes:

$$U_{ii}(t) = F_{cns}$$

(1. 6)

Insulin- dependent utilization depends nonlinearly from glucose in the tissues:

$$U_{id}(t) = \frac{V_m(X(t)) \cdot G_t(t)}{K_m(X(t)) + G_t(t)}$$

(1. 7)

Where $V_m(X(t))$ and $K_m(X(t))$ are assumed to be linearly dependent from a remote insulin, X(t):

$$V_m\big(X(t)\big) = V_{m0} + V_{mx} \cdot X(t)$$
$$K_m\big(X(t)\big) = K_{m0+}K_{mx} \cdot X(t)$$

(1. 8)

X (pmol/L) is insulin in the interstitial fluid and it is described by:

$$\dot{X}(t) = -p_{2u} \cdot X(t) + p_{2u}[I(t) - I_b] \qquad X(0) = 0$$

(1. 9)

Where I is plasma insulin and p_{2u} (min^{-1}) is the rate constant of insulin action on the peripheral glucose utilization.

- Insulin secretion: described by the following equations

$$S(t) = \gamma \cdot I_{po}(t)$$
$$\dot{I}_{po}(t) = -\gamma \cdot I_{po}(t) + S_{po}(t) \qquad I_{po}(0) = I_{pob}$$
$$S_{po}(t) = \begin{cases} Y(t) + K \cdot \dot{G}(t) + S_b & for\ \dot{G} > 0 \\ Y(t) + S_b & for\ \dot{G} \leq 0 \end{cases}$$

(1. 10)

Where γ (min^{-1}) is the transfer rate constant between portal vein and liver, K (pmol/kg per mg/dl) is the pancreatic responsivity to the glucose rate of change.

- Glucose renal excretion: which occurs if plasma glucose exceeds a certain threshold and can be modelled by a linear relationship with plasma glucose:

$$E(t) = \begin{cases} k_{e1} \cdot [G_p(t) - k_{e2} & if\ G_p(t) > k_{e2} \\ 0 & if\ G_p(t) \leq k_{e2} \end{cases}$$

(1. 11)

The complete model is given by the equations so far seen and represents one of the most comprehensive models of the glucose-insulin system [11].

In 2014, this model was extended (S2013) in order to overcome some previous limitations such as the absence of counteregulatory hormones (glucagon, epinephrine and growth hormone), the glucocentric nature of the model and the absence of intersubject variability. The model was modified as shown by the figure 1.7.

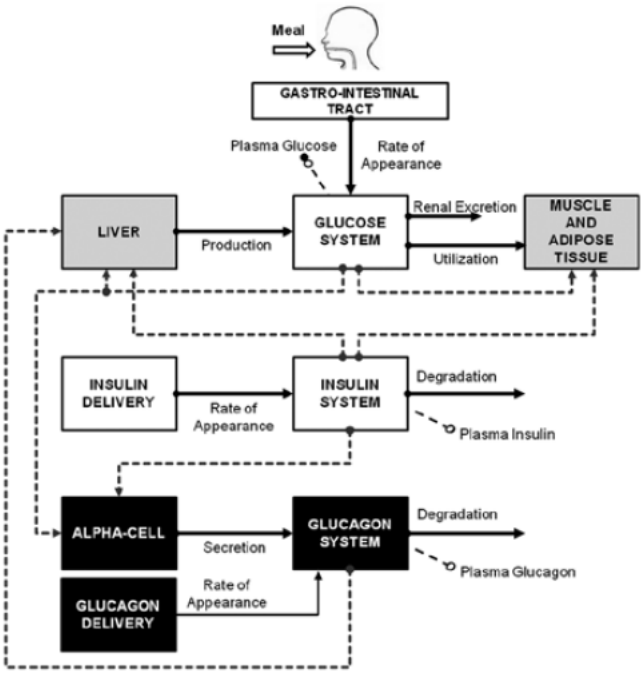

Figure 1.7: Scheme of the new version of the model. Gray blocks are those that were been updated and black block were new [12] with respect to the previous version [11]. Adapted from [12].

As reported by Dalla Man et al. [12] this version provides a more reliable framework for in silico trials for regulatory purposes, for testing glucose sensors and insulin augmented pump prediction methods and for closed-loop single/dual hormone controller design, testing and validation. Both S2008 and S2013 simulators have been validated and accepted by FDA for a single-meal scenario [12].

Mathematical models play a pivotal role in biomedical science because they enable the simulation and analysis of complex biological processes, offering a virtual overview, crucial in medical research and the development of new treatments. In fact, mathematical models allow the prediction of a biological system's behaviour under various conditions. In type 1 diabetes context, they can simulate glucose dynamics in response to different

therapies, aiding researches in assessing the potential effectiveness of new drugs or treatments.

By comparing predicted results with known experimental data, research can enhance and optimize models, making them more precise tool for prediction complex biological responses and long-term effects on patients' health conditions.

In conclusion, employing mathematical models is becoming more and more useful in medical research, allowing efficient and ethical prediction, validation and experimentation. This contributes to the development of new therapies and enhances the management of medical conditions.

1.4 Object of the thesis.

The purpose of this thesis is to build up a comprehensive and exhaustive model that examines and elucidates the effects of Sodium-Glucose Co-Transporter type 2 (SGLT2) inhibitors in patients with type 1 diabetes.

The aim of the study is to conduct an in-depth analysis of physiological and metabolic responses, with a specific emphasis on assessing the impact of these pharmaceutical agents on crucial aspects of glycaemic control and glucose-insulin variability. By analysing these effects in depth, the research aims not only to quantify the observed improvements, in terms of lowering plasma glucose and glycated haemoglobin (HbA_{1C}), reducing body weight and insulin requirements, but also to unravel the underlying mechanisms that contribute to the efficacy of SGLT2 inhibitors in patients with type 1 diabetes. This detailed exploration seeks to offer an understanding of the intricate interplay between the drug and the physiological process involved in glucose regulation.

The ultimate goal is to contribute valuable insights that can provide information and refine treatments strategies, potentially paving the way for optimised and personalised treatment approaches for patients with type 1 diabetes.

1.5 Content of the thesis.

The content of this thesis will encompass a multifaceted exploration of the effects of SGLT2 inhibitors in the context of type 1 diabetes.

The study will open with thorough review of the existing literature, providing a comprehensive understanding of current knowledge and gaps in the field. Subsequently, research will delve into the physiological and metabolic responses elicited by SGLT2 inhibitors, clarifying their impact on glycaemic control, insulin sensitivity and other relevant parameters. The investigation will involve the analysis of clinical data arising from a single-centre, investigator-led, double-blind, placebo-controlled crossover study conducted by Herring et al. in their paper "Metabolic effects of an SGLT2 inhibitor (Dapagliflozin) during a period of acute insulin withdrawal and development of ketoacidosis in people with type 1 diabetes".

Furthermore, the thesis will take a mechanistic approach, attempting to unravel the mechanisms through which SGLT2 inhibitors exert their influence. Special attention will be given to potential synergies with existing treatment modalities and the implications for personalized therapeutic strategies.

By presenting a comprehensive and detailed analysis, this thesis aims to contribute to the understanding of the intricate interplay between SGLT2 inhibitors and type 1 diabetes, fostering progress in the field of diabetes management.

Chapter 2
DATABASE AND PROTOCOL

The database used in this thesis is the one collected by Herring et al. and reported in their paper "Metabolic effects of an SGLT2 inhibitor (Dapagliflozin) during a period of acute insulin withdrawal and development of ketoacidosis in people with type 1 diabetes". This study was a double-blind, placebo-controlled crossover study with a four-week washout period. It was performed in 12 people (four male and eight female) with type 1 diabetes using insulin pump therapy. They were recruited between February and October 2018 via the diabetes insulin pump database at the "Royal Surrey Country National Health Service Trust". They were made aware of potential changes in glycaemic control and were asked to record any trial administration, concomitant medication (including insulin), hypoglycaemia frequency, fasting ketone levels and any adverse event. Ethics approval was granted from the National Research Ethics Service Committee.

Exclusion criteria included proliferative retinopathy, moderate-to-severe renal impairment, severe hepatic impairment, cardiac failure, uncontrolled cardiac arrhythmia and hypertension, mental incapacity, pregnancy or breastfeeding and suspected allergy to trial products.

Participants received dapagliflozin (10mg daily) or placebo in random order for seven days, and, on the last day, they came to the hospital for a metabolic study. They were asked not to consume food and to drink only water from 22:00 of the day before and not to undertake any strenuous exercise or consume alcohol from the day before the study day. In addition, they disconnected their subcutaneous insulin pump at 6:00 am on the day of the metabolic study. In order to maintain a blood glucose concentration of 90 mg/dl (5mmol/L) they received a soluble variable insulin infusion. Stable isotopes were also infused to measure glucose rate of appearance (Ra), glucose rate of disappearance (Rd) and lipolysis.

When the isotopic steady state was reached, insulin infusion was stopped, blood glucose was allowed to increase and the study terminated after 600 min or earlier if blood glucose reached 324mg/dl (18mmol/L), or bicarbonate was less than 15 mmol/L, or venous pH was less than 7.35 or if capillary ketones were higher than 5 mmol/L.

Blood samples were taken every 20 minutes until 180 min and then every 30 minutes. Even urine samples were collected every 2 hours to spot urinary ketones.

Information about the participants is summarized in the table below, as mean ± SEM (standard error of mean).

Participants	Duration of diabetes	Age	BMI	HbA_{1C}	C-peptide
12	23.3±4.1 years	40.7±3.9 years	26.8±1.4 kg/m^2	59.9±2.3 mmol/l	<0.2 mmol/L

Table 2.1: Participants' information

All subjects completed 180 min of each metabolic study. Results are shown in the figures below.

- Glucose concentration:

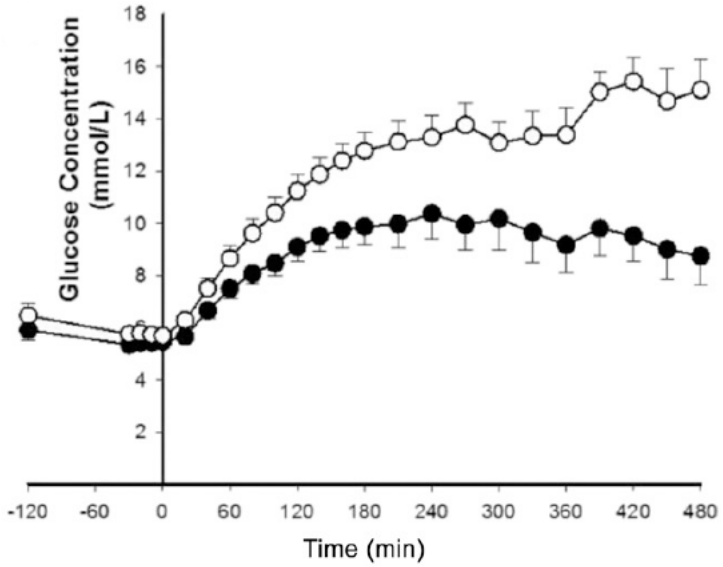

Figure 2.4: Plasma glucose concentration. White dots refer to the placebo group, instead black dots refer to dapagliflozin group. Adapted from [13]

At 0 min, glucose concentration was not different between treatments, then it increased in both groups reaching a final value of 8.5±0.7 mmol/L (mean ± SEM) in dapagliflozin group and 14.3±1.1 mmol/L with placebo. Urinary glucose excretion was 5.1±0.8 μmol/kg/min with dapagliflozin and 0.029±0.01 μmol/kg/min with placebo, proving the SGLT2 inhibitors effect of promoting the renal excretion of glucose.

- Glucose rate of appearance:

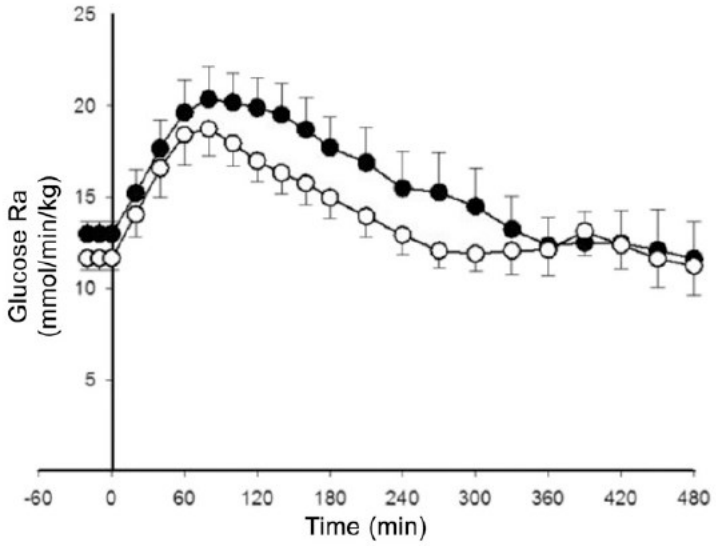

Figure 5.2: Glucose rate of appearance. White dots refer to the placebo group, instead black dots refer to dapagliflozin group. Adapted from [13]

At baseline glucose Ra was higher in dapagliflozin compared with the placebo, 13.0±0.77 μmol/kg/min and 11.7±0.7 μmol/kg/min respectively. During insulin withdrawal, glucose Ra increased, peaking at 90 min, and then declined, with no significant difference in the two groups.

- Glucose rate of disappearance:

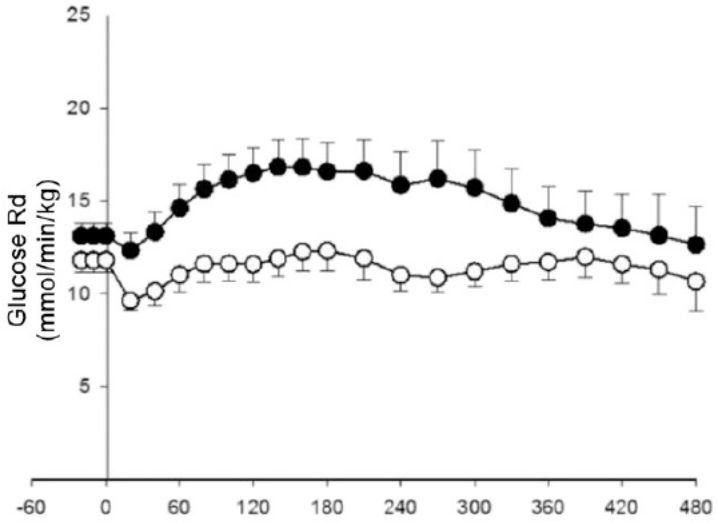

Figure 2.3: Glucose rate of disappearance. White dots refer to the placebo group, instead black dots refer to dapagliflozin group. Adapted from [13]

Glucose rate of disappearance showed to be higher in the dapagliflozin group than the placebo one for the duration of the study [13].

The data, reported as mean ±(SEM), were collected through WebPlotDigitalizer and arranged in a .mat file so that they could be used as database for our model implemented in Matlab.

Chapter 3
MODEL DEVELOPMENTS

3.1 Mathematical models

A very important class of Models of System are the Compartmental Models. They are based on mass-balance considerations and are widely used for quantitative studies of the kinetics of materials in physiological systems. Since the state variables represent masses, compartmental models are Positive Systems.

Materials can be either exogenous, such as drugs or tracers, or endogenous, such as a substrates or a hormones. Kinetics include processes such as production, distribution, transport, utilization and substrate-hormone interaction.

Compartmental Models consist of a finite number of compartments with specified interconnections among them. Each compartment represents a quantity of material that acts as though it is well mixed and kinetically homogeneous. Well-mixed means that any two samples taken at the same compartment at the same time would have the same concentration of the compound being studied, and therefore are equally representative. On the other hand, kinetically homogeneous means that every particle in a compartment has the same probability of taking a given pathways leaving the compartment.

Compartments can be accessible, if a measurement can be drawn from it, or non-accessible, if no measurement can be made.

Interconnections represent fluxes of material which physiologically constitute transport from one location to another or a chemical transformation or both.

Figure 3.1 represents a one-compartment model which can be mathematically formalized as follows.

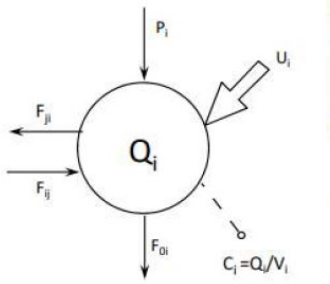

Figure 3.1: Mono-compartmental model

Each i-th compartment of a n-compartment model has a mass-balance equation:

$$\dot{Q}_i(t) = \sum_{\substack{j=1 \\ j\neq i}}^{n} F_{ij}(t) - \sum_{\substack{j=1 \\ j\neq i}}^{n} F_{ji}(t) - F_{0i}(t) + P_i(t) + U_i(t)$$

(3. 1)

With

- $Q_i(0) = Q_{i0}$
- $Q_i, F_{ij}, P_i, U_i \geq 0$
- $F_{ij}(t) = F_{ij}[Q_1(t), Q_2(t), \dots, Q_n(t)] = k_{ij}[Q_1(t), Q_2(t), \dots, Q_n(t)]Q_j(t)$
  $with\ k_{ij}[Q_1(t), Q_2(t), \dots, Q_n(t)] \geq 0$

Where U_i represent some external input and P_i represent materials produced by the body (endogenous production). F_{ij} and F_{ji} describe the fluxes entering and exiting the compartment. The subscript zero represents the external environment, thus F_{0i} depicts something that is being eliminated, and finally k_{ij} is the fractional transfer coefficient from compartment j to compartment i.

Summarizing, each compartmental model is described by n state equations, in the form of the mass-balance equations, and m measurement equations:

$$C_i(t) = Q_i(t)/V_i$$

(3. 2)

Defining the vectors:

$$P(t) = [P_1(t), P_2(t), \dots, P_n(t)]^T$$

$$U(t) = [U_1(t), U_2(t), \dots, U_n(t)]^T$$

$$C(t) = [C_1(t), C_2(t), \dots, C_m(t)]^T$$

(3. 3)

The compartmental model equations can be written as:

$$\begin{cases} \dot{Q}(t) = K[Q(t)]Q(t) + P[Q(t)] + U(t) & Q(0) = Q_0 \\ C(t) = HQ(t) \end{cases}$$

(3. 4)

In linear Compartmental Models, matrix K is constant, independent from Q, and it contains the rate transfer coefficients and their combination:

$$K = \begin{bmatrix} -(k_{01} + k_{21} + \cdots k_{n1}) & k_{12} & \dots k_{1n} \\ k_{21} & -(k_{02} + \cdots + k_{n2}) & \dots k_{2n} \\ k_{n1} & k_{n2} & -(k_{0n} + \cdots + k_{n-1\,n} \end{bmatrix}$$

(3. 5)

Non-diagonal elements correspond to the rate constants between compartments and are non-negative, whereas diagonal elements are linear combinations of rate constants and are non-positive.

K is a diagonally-dominant matrix, which means that the diagonal element of each column is not lower, in absolute value, than the sum of the non-diagonal elements of the column.

In presence of closed subsystems, part of the system in which material enter but is not able to exit, the model cannot be asymptotically stable, because the mass of compartments that belong to the subsystem, only increase or remain constant (not decrease), so that the state variables cannot converge to zero. On the other hand, in absence of closed subsystem, the model is asymptotically stable, because there is at least a flux towards the external environment.

Compartmental Models have been so widely used because they provide a simple and effective method for schematizing, simulating and predicting the properties of the spread of a disease, such as prevalence (total number of infected) or the duration of an epidemic. Compartmental models are largely used in epidemiology, a field in which these models succeed in overcoming most of the ethical issues related to animal and human trial, by

following the principles of the “three Rs alternatives”: replacement, reduction and refinement.

3.2 System Identifiability

Once the model is built, before trying to estimate the model parameters, we need to make sure that all the parameters can be estimated from the data, because only if the model is identifiable than it makes sense trying to numerically estimate the value of its parameters.

The a Priori Identifiability has the aim to theoretically establish, in the ideal case of error-free model and exact knowledge of all the continuous time model output and their derivatives, if it is possible to determine the unknown parameters.

If a model is not a priori identifiable we can try to make it identifiable by enriching the experiment or reducing the model complexity, or both.

For linear compartmental models, the identifiability can be verified by calculating the Laplace Transform of the input and output:

$$Y(s) = L[y(t)]$$

$$U(s) = L[u(t)] \tag{3.6}$$

Then calculating the transfer function of the system as the ratio between them:

$$H(s) = \frac{Y(s)}{U(s)} \tag{3.76}$$

Thus, to assess the a priori identifiability, one can check if the model parameters can be univocally determined from the coefficients of the transfer function.

For linear compartmental models, the transfer function H(s) is always a ratio of polynomials:

$$H(s) = \frac{\beta_n s^{n-1} + \beta_{n-1} s^{n-2} + \cdots + \beta_2 s + \beta_1}{s^n + \alpha_n s^{n-1} + \cdots + \alpha_2 s + \alpha_1}$$

(3.8)

Coefficients $\alpha_k \; and \; \beta_k$ function of model parameters, k_{ij} and V_i and may be thought as observable parameters. One can write down the algebraic relationship linking $\alpha_k \; and \; \beta_k$ with k_{ij} and V_i. If these set of equations, called exhaustive summary, can be solved in the unknowns k_{ij} and V_i, then the model is a priori identifiable. According to Cawley-Hamilton theorem the exhaustive summary will contain a number of equations equal to k=2n-1, in which n is the order of the system. This theorem is only valid for linear models and it is not for the non-linear ones.

A model can be:

- Uniquely identifiable, if all its parameters p_i are uniquely identifiable.
- Locally identifiable, if all its parameters p_i are uniquely identifiable apart from one or more who are not, but have a finite number of solutions.
- Not identifiable, if at least one parameter is not identifiable and has infinite solutions.

Dealing with non-linear models is more difficult because, even if we linearize the model, this would not give us information about the identifiability of the original non-linear model. There are several methods to verify the identifiability of a non-linear model such as the Taylor series expansion of y(t), Similarity Transformation or Differential Algebra.

The Taylor series expansion method requires the expression of the output as a function of the parameters, which are observable, thus obtaining an exhaustive summary as:

$$y^k(p, t_0) = \hat{\phi} \qquad k = 0, 1, 2, \ldots$$

(3.9)

in which y^k represent the output expressed as function of the parameters p.

In case of non-linear models, the number of needed equations k is not known (apart from the fact that k≥P, with P dimension of the vector p). The possibility to solve the exhaustive summary is a necessary and sufficient condition to claim the model identifiability.

However, due to the complexity of the exhaustive summary, if it is not possible to prove the uniquely identifiability, nothing can be claimed about the nonidentifiability.

A priori identifiability analysis allows to avoid doing experiments if the parameters of interest are not identifiable, which is particularly important in physiological and clinical studies where ethical and practical issues come into play and it is also an efficient instrument for defining the minimal input/output configuration.

In this project we didn't check the system a priory identifiability. We assumed instead that the model was identifiable, because the model was too complex to assess the a priory identifiability.

What it was possible to do instead, was to verify the a posteriori identifiability. Indeed, the a posteriori identifiability is a check on the feasibility of the estimate resulting in the so-called parametric identification of the model, which consists of determining whether it is possible, with measurements affected by error, to devise one or more experiments that allow the parameters to be estimated accurately.

3.3 Parameter estimation

The procedure known as parameter estimation aims to assign numerical values to the model's parameters, and it can be done with several optimization methods. The two most commonly used approaches are the Fisher and the Bayesian approach.

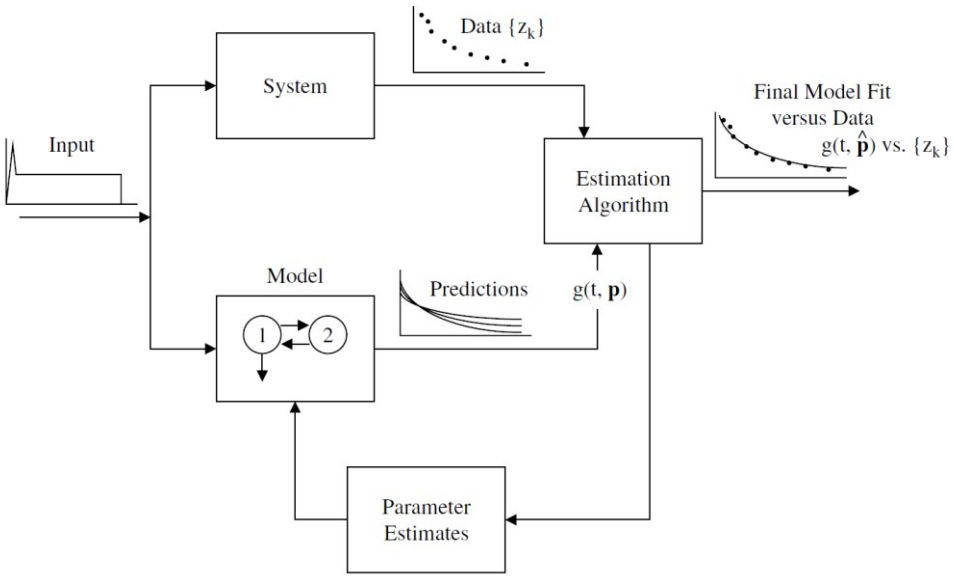

Figure 3.2: Parameter estimation scheme. Adapted from [14]

3.3.1 Fisher Approach

Let's consider the observable model output as:

$$y(t) = g(t, \boldsymbol{p}) \tag{3.10}$$

Where **p** are the parameters, we want to estimate.

We further assume that the N output measurements, z_i are affected by measurement error, which can be considered additive:

$$z_i = y_i + v_i = g(t_i, \boldsymbol{p}) + v_i \qquad i = 1, 2, \dots N \tag{3.11}$$

Where v_i is the measurement error corrupting the i-th measurement z_i and, since it is unknown, it can be considered as a random variable.

The problem is to assign a numerical value to **p** from the data z_i collected through experiments.

Weighted Least Squares (WLS) belongs to the so-called Fisherian Approach, which is a regression method based on adjusting the value of the parameters until we obtain the set of values which provide the best fit to the data. There are fundamentally two kind of regression: linear and non linear. The first one concerns the so-called models linear in the parameters and provide an exact solution. The second one regards models non-linear in the parameters, in which the optimization algorithm is computationally more complex and it provides only an approximated estimation of the parameters.

Let's start with the simplest case, in which we want to obtain the best fit to the data to the straight line. In this case the model output would be:

$$y(t) = \boldsymbol{p}t$$

(3.12)

For different values of **p**, different straight lines will be generated, but we want to find the one which provides the best fit to the data. In order to do that we make use of the so-called residual, which is the difference between the observed datum $z(t_i)$(▲) and the expected output $y(t_i)$:

$$r_k = r(t_i) = z(t_i) - y(t_i)$$

(3.13)

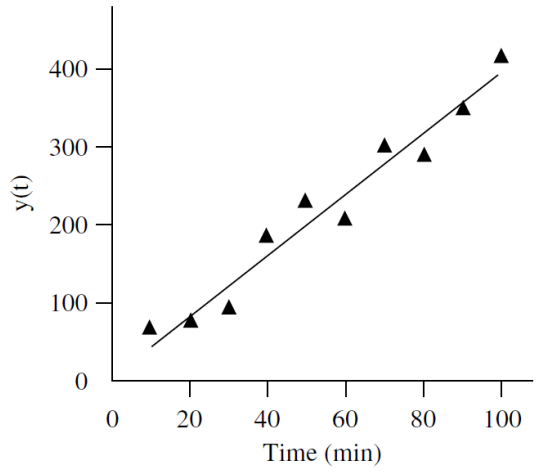

Figure 3.3: Model output y(t) and observable datum (▲). Adapted from [14]

Now we can calculate the Residual Sum of Squares (RSS), which is a measure of how good the fit is to the given set of data and it is also called objective function or cost function(J(p)):

$$RSS = \sum_{i=1}^{N} r_i^2 = \sum_{i=1}^{N} (z(t_i) - g(t_i, p))^2$$

(3.14)

The aim is to find the best **p** which minimize the RSS: the smaller the RSS the better the fit. This regression method is also known as Least Squares.

It is also possible to weight the data, so that some data are assigned more confidence than others, by changing the objective function.

The cost function (J(p)) to be minimized is now the Weighted Residual Sum of Squares (WRSS):

$$WRSS = \sum_{i=1}^{N} w_i r_i^2$$

(3.15)

Generally, if the measurement error variance is known, the optimal choice is to set $w_i = 1/{\sigma_i}^2$, with ${\sigma_i}^2$ the variance of the noise associated to the i-th measurement.

The most common assumptions on the noise are:

- Errors have zero mean: E[v_i]=0
- Errors are independent: Cov[v_i, v_j]=0 for i≠j
- The variance is known: Var[v_i]=σ_i^2

A standardized measure of the error is provided by the fractional standard deviation (FSD) or coefficient of variation (CV):

$$FSD[v_i] = CV[v_i] = \frac{\sqrt{\sigma_i^2}}{z_i}$$

(3.16)

Where $\sqrt{\sigma_i^2}$ is the standard deviation (SD) of the error.

It is possible now to extend the simple scalar case, previously seen, to the vector one, considering the model z=Gp+v (where G is a matrix n x M, with n numbers of data and M numbers of parameters), where v has the covariance matrix $\Sigma_v = diag(\sigma_1^2, \sigma_2^2, \dots, \sigma_N^2)$. The cost function will be:

$$J(p) = \sum_{i=1}^{N} \frac{1}{\sigma_i^2} r_i^2 = ||r^2||_{\Sigma_v^{-1}} = r^T \Sigma_v^{-1} r = (z - Gp)^T \Sigma_v^{-1} (z - Gp)$$

(3.17)

WLS estimate is the value of p which minimizes J(p):

$$\hat{p}_{WLS} = arg \min_p (z - Gp)^T \Sigma_v^{-1} (z - Gp)$$

(3.18)

Since it is a quadratic problem there is a closed form analytical solution for $\hat{p}_{WLS}$:

$$\begin{aligned} WRSS(p) &= (z - Gp)^T \Sigma_v^{-1} (z - Gp) \\ &= z^T \Sigma_v^{-1} z - 2p^T G^T \Sigma_v^{-1} z + p^T G^T \Sigma_v^{-1} Gp \\ \frac{dWRSS(p)}{dp} &= -2G^T \Sigma_v^{-1} z + 2G^T \Sigma_v^{-1} Gp = 0 \\ \hat{p}_{WLS} &= (G^T \Sigma_v^{-1} G)^{-1} G^T \Sigma_v^{-1} z \end{aligned}$$

(3. 197)

At this point we have to figure out how much reliable is $\hat{p}$. Since the estimation of p depends on measurements z which are effected by errors also the estimated $\hat{p}$ would be affected by error ($\tilde{p}$), defined as the distance between the true value (p) and the estimated one ($\hat{p}$):

$$\tilde{p} = p - \hat{p}$$

(3.20)

We can compute the variability of the estimation error by calculating its covariance matrix:

$$\Sigma_{\tilde{p}} = cov(\tilde{p}) = E[\tilde{p}\tilde{p}^T] = \Sigma_{\hat{p}}$$

(3.21)

Since $\hat{p}_{WLS} = (G^T\Sigma_v^{-1}G)^{-1}G^T\Sigma_v^{-1}z$, we get:

$$\tilde{p} = p - \hat{p} = p - (G^T\Sigma_v^{-1}G)^{-1}G^T\Sigma_v^{-1}z = p - (G^T\Sigma_v^{-1}G)^{-1}G^T\Sigma_v^{-1}(Gp + v)$$
$$= [I_M - (G^T\Sigma_v^{-1}G)^{-1}G^T\Sigma_v^{-1}G]p - (G^T\Sigma_v^{-1}G)^{-1}G^T\Sigma_v^{-1}v$$

(3.22)

By using the definition (3.21), can be immediately obtained as the covariance of the random addendum:

$$\Sigma_{\tilde{p}} = (G^T\Sigma_v^{-1}G)^{-1}$$

(3.23)

Given the model G, the more inaccurate the data (larger Σ_v) more inaccurate the parameter estimates (larger Σ_p) will be. The covariance matrix provides a measure of the precision with which *p* is estimated, from it we can calculate the:

- Standard deviation of the estimate: $SD(\hat{p}_i) = \sqrt{var(\hat{p}_i)}$
- Confidence interval: $\hat{p}_i \pm SD(\hat{p}_i)$
- Coefficient of variation, which represents the precision of the estimated parameter: $\text{CV}(\hat{p}_i) = 100 \times \frac{SD(\hat{p}_i)}{\hat{p}_i}$

In case of models nonlinear in the parameters the problem of how to estimate the parameters becomes more complex. The solution is computed through a number of iterations that draws on the linear regression theory. The most commonly used approach is the Gauss-Newton Method, which can be summarized in the following steps:

1) Given the non-linear model $z_i = g(t_i, \boldsymbol{p}) + v_i$ with i=1, 2, …, N, we firstly have to choose an initial value $\hat{p}^{(0)}$

2) Linearize the model $g(t_i, \boldsymbol{p})$ around $\hat{p}^{(k)}$ at any time

$$t_i g(t_i, p) = g(t_i, p^{(k)}) + \left[\frac{\partial g(t_i, p^{(k)})}{\partial p_1} \; \frac{\partial g(t_i, p^{(k)})}{\partial p_2} \; \dots \frac{\partial g(t_i, p^{(k)})}{\partial p_M}\right] \begin{bmatrix} p_1 - p_1^{(k)} \\ p_2 - p_2^{(k)} \\ \dots \\ p_M - p_M^{(k)} \end{bmatrix}$$

(3.24)

The quantity $\begin{bmatrix} p_1 - p_1^{(k)} \\ p_2 - p_2^{(k)} \\ \dots \\ p_M - p_M^{(k)} \end{bmatrix}$ is called Δp and represents the deviation of p in respect of the initial assigned value $\hat{p}^{(0)}$

3) Solve the problem in Δp using the WLS

$$\begin{bmatrix} z_1 - g(t_1, p^{(k)}) \\ z_2 - g(t_2, p^{(k)}) \\ z_3 - g(t_3, p^{(k)}) \\ \dots \\ z_N - g(t_N, p^{(k)}) \end{bmatrix} = \begin{bmatrix} \left.\frac{\partial g(t_1,p)}{\partial p_1}\right|_{p=p^{(k)}} & \left.\frac{\partial g(t_1,p)}{\partial p_2}\right|_{p=p^{(k)}} & \dots & \left.\frac{\partial g(t_1,p)}{\partial p_M}\right|_{p=p^{(k)}} \\ \left.\frac{\partial g(t_2,p)}{\partial p_1}\right|_{p=p^{(k)}} & \left.\frac{\partial g(t_2,p)}{\partial p_2}\right|_{p=p^{(k)}} & \dots & \left.\frac{\partial g(t_2,p)}{\partial p_M}\right|_{p=p^{(k)}} \\ & & \dots & \\ & & \dots & \\ \dots & \dots & \dots & \left.\frac{\partial g(t_{N-1},p)}{\partial p_M}\right|_{p=p^{(k)}} \\ \left.\frac{\partial g(t_N,p)}{\partial p_1}\right|_{p=p^{(k)}} & \left.\frac{\partial g(t_N,p)}{\partial p_2}\right|_{p=p^{(k)}} & \dots & \left.\frac{\partial g(t_N,p)}{\partial p_M}\right|_{p=p^{(k)}} \end{bmatrix} \begin{bmatrix} p_1 - p_1^{(k)} \\ p_2 - p_2^{(k)} \\ \dots \\ p_M - p_M^{(k)} \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ \\ \\ \dots \\ v_N \end{bmatrix}$$

(3.25)

Which in compact form becomes:

$$\Delta z = S\Delta p + v$$

(3.26)

4) Update $\hat{p}^{(k+1)} = \hat{p}^{(k)} + \Delta\hat{p}$

5) If $||z\text{-}G(\hat{p})||^2_{\Sigma_v^{-1}}$ has changed significantly, k=k+1 and we have to start again from step 2.

Since we are dealing with a model nonlinear in the parameters, the objective function is no more a quadratic problem (parabola), but it contains several local minima. Among all the local minima we need to find the smallest one, called global minimum.

The difference between the linear and nonlinear case is that in linear regression there is a unique minimum for WRSS that can be calculated with a closed-form (eq. 3.24) while, in the nonlinear case, no closed-form is available, therefore, one has to use iterative method to find and approximate solution, and initial values are needed to start the algorithm. Moreover, since there may be several local minima for WRSS and the choice of initial values plays a key role on the final result.

In order to estimate the precision of our model we have to calculate the covariance matrix, but in this case we do not have a closed form solution like for linear WLS. An approximate solution may be:

$$\Sigma_{\tilde{p}} = \Sigma_{\hat{p}} \cong (S^T \Sigma_v^{-1} S)^{-1}$$

(3.27)

In which S is the sensitivity matrix calculated at the optimal value p and represents the gradient of the model in respect to the parameters. Given Σ_v, the larger is S, the lower is Σ_p, the more precise are the parameter estimates.

Once determined **p** the steps to follow to assess the quality of the results are:

- Analysis of residuals: which should reflect the assumption on the measurement error (zero mean, uncorrelated and with a known covariance matrix).
- Parameter precision: which depends on the coefficients of variation (CV) of the parameters.
- Parsimony criteria: based on principle of parsimony. It is used to balance between the need of good fit and good precision. In case of time invariant systems and Gaussian error it can be calculated as:
  Akaike criterion: $AIC = WRSS + 2M$
  Schwartz criterion: $SC = WRSS + M \log N$
  In which M is the number of parameters and N is the number of data.

3.3.2 Bayesian Approach

Whereas Fisher approach was based on knowledge of data (z) and the noise characteristics, the Bayesian approach is based not only on the experimental data (a posteriori information), but also on some a priori information (data independent) available on the unknown parameters.

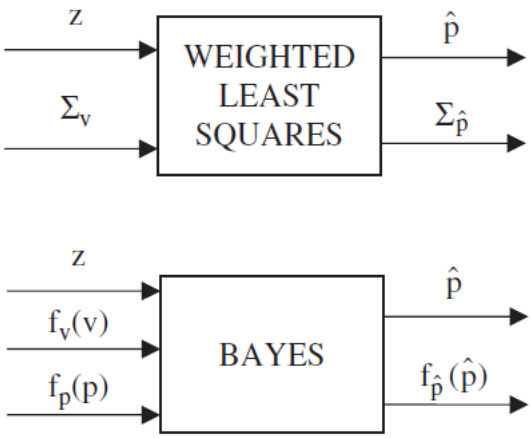

Figure 3.4: Comparison of the requirements for WLS and Bayesian estimations. Adapted from [14]

A Bayesian estimator assumes that the a priori probability distribution of **p**, $f_p(p)$, is available. This probability density changes after seeing the data z, and is therefore called a posteriori probability density $f_{p|z}(p|z)$ (conditioned by the data) and it is the key function in Bayes estimation.

From the Bayes' theorem we can obtain the a posteriori probability density as:

$$f_{p|z}(p|z) = \frac{f_{z|p}(z|p)f_p(p)}{f_z(z)} \tag{3.28}$$

In which $f_p(p)$ is the a priori probability density of p, $f_{z|p}(z|p)$ is the likelihood of z which depends on the model g(t, p), and $f_z(z)$ is the probability density of the measurement error.

The Maximum a Posteriori estimator (MAP) is given by:

$$\hat{p}_{MAP} = arg \max_{p} f_{p|z}(p|z)$$

(3.29)

Since $f_z(z)$ does not depend on p the MAP estimator can be written as:

$$\hat{p}_{MAP} = arg \max_{p} f_{z|p}(z|p) f_p(p) = arg \max_{p} J(p)$$

(3.30)

In general, an analytic expression for $f_{p|z}(p|z)$ is either not available or is simply intractable. Hence, the Markov Chain Monte Carlo method can be adopted.

The expression of $f_{p|z}(p|z)$ can be simplified under specific assumption regarding the distribution of $f_z(z)$ and $f_p(p)$. Assuming that both **v** and **p** are independent and normally distributed, we obtain:

$$f_p(p) = \frac{1}{[(2\pi)^M det(\Sigma_p)]^{1/2}} exp\left(-\frac{1}{2}(p-\mu_p)^T \Sigma_p^{-1}(p-\mu_p)\right)$$

$$f_{z|p}(z|p) = \frac{1}{[(2\pi)^N det(\Sigma_v)]^{1/2}} exp\left(-\frac{1}{2}[z-G(p)]^T \Sigma_v^{-1}[z-G(p)]\right)$$

$$\begin{aligned} f_{z|p}(z|p) f_p(p) = {} & \frac{1}{[(2\pi)^N det(\Sigma_v)]^{1/2}[(2\pi)^M det(\Sigma_p)]^{1/2}} \; exp\left(-\frac{1}{2}(p \right. \\ & \left. - \mu_p)^T \Sigma_p^{-1}(p-\mu_p)\right) \cdot exp\left(-\frac{1}{2}[z-G(p)]^T \Sigma_v^{-1}[z-G(p)]\right) \\ = {} & \frac{1}{[(2\pi)^N det(\Sigma_v)]^{1/2}[(2\pi)^M det(\Sigma_p)]^{1/2}} \\ & \cdot exp\left\{-\frac{1}{2}\left([z-G(p)]^T \Sigma_v^{-1}[z-G(p)] + (p-\mu_p)^T \Sigma_p^{-1}(p \right.\right. \\ & \left.\left. - \mu_p)\right)\right\} \end{aligned}$$

(3.31)

In order to maximize $f_{z|p}(z|p) f_p(p)$ we have to minimize the argument of the exponential, obtaining:

$$\hat{p}_{MAP} = arg\min_{p}\left([z - G(p)]^T \Sigma_v^{-1}[z - G(p)] + (p - \mu_p)^T \Sigma_p^{-1}(p - \mu_p)\right)$$

(3.32)

In which the first term represents the weighted distance between data and the prediction (a posteriori information) and the second one represents the distance between the estimate and the prior (a priori information).

Compared to WLS, MAP worsens the fit but betters the precision. It realizes a compromise between a priori and a posteriori information. For instance, if a priori information is 'poor', i.e. Σ_p is large, the second term becomes negligible.

As well as for the Fisher approach, for the Bayesian approach we can use the parsimony principle to establish the precision of the estimate, by using the Generalized Information Criteria (GEN-IC):

$$GEN - IC = \frac{2M}{N} + J_{MAP}(\hat{p})$$

(3. 83)

where

$$J_{MAP}(p) = f_{z|p}(z|p) f_p(p)$$

(3. 94)

As previously stated, the Bayesian Approach is based on the experimental data and some a priori information. This a priori information (the prior) is usually summarized in the distribution of **p**, $f_p(p)$, which in turn is completely defined by the average vector μ_p (of dimension M) and the covariance matrix Σ_p (of dimension M×M, with M number of parameters), if the distribution is Gausian. Such information may come from literature, previous studies or experiments. MAP approach can be also used in case of availability of a priori information for a subset of parameters only. Parameters for which no a priori information is available will be assigned a 'fake' prior with infinite variance, which is equivalent of having no prior.

3.4 Model selection

Once we have the model structure and we have assigned a numerical value to the parameters, selecting an optimal model involves analysing three fundamental aspects: the fit, the coefficient of variations and the parsimony criteria.

The first involves a visual analysis of the results obtained. The model must in fact be able to adapt to the available data without under- or over-estimating them. Finding the best fit therefore means finding the curve that best approximates a distribution of points or a function.

The coefficient of variation is a dispersion index, which represents the precision of the estimated parameter. It is a number expressed as a percentage and the larger it is, the lower the precision with which that parameters are estimated will be.

The last one is the so-called Parsimony Criterion, based on principle of parsimony. It is used to balance between the need of good fit and good precision. The parsimony principle states that a simpler model with fewer parameters is preferred over more complex models with more parameters, provided the models fit the data similarly well. One of the most frequently used methods is the Akaike information criterion, which is an estimator of prediction error and is calculated as: $AIC = WRSS + 2M$, where M is the number of parameters. *AIC* deals with the trade-off between the goodness of fit of the model and the simplicity of the model. In other words, AIC deals with both the risk of overfitting and the risk of underfitting. If a set of models performs similarly in terms of ability to fit the data and precision of the estimates, one should choose the one providing the lowest *AIC*.

3.5 Model Implementation

To develop our model, we started from the UVA/Padova Simulator, which unfortunately was not able to reproduce the results shown by Herring et al. [13] in the two experiments conducted. Thus, we decided to consider only the Glucose subsystem (Figure 3.5), described by the equations below, and try to estimate the values of the parameters in order to reproduce the pattern observed in the data.

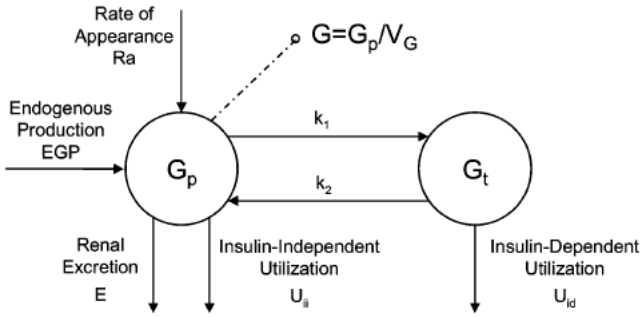

Figure 3.5: Glucose subsystem. Adapted from [11]

$$\begin{cases} \dot{G}_p(t) = EGP(t) + Ra(t) - U_{ii}(t) - E(t) - k_1 G_p(t) + k_2 G_t(t) & G_p(0) = G_{pb} \\ \dot{G}_t(t) = -U_{id}(t) + k_1 G_p(t) - k_2 G_t(t) & G_t(0) = G_{tb} \\ G(t) = \dfrac{G_p}{V_G} & G(0) = G_b \end{cases}$$

(3. 35)

Where G_p and G_t (mg/kg) are glucose masses in plasma and rapidly equilibrating tissues, and slowly equilibrating tissues, respectively, G (mg/dl) is plasma glucose concentration, and the suffix b denotes basal state. EGP (mg/kg/min) is the endogenous glucose production; Ra (mg/kg/min) is the glucose rate of appearance in plasma; E is the renal excretion; U_{ii} and U_{id} (mg/kg/min) are insulin-independent and -dependent glucose utilizations; V_G (dl/kg) is the distribution volume of Glucose and k_1 and k_2 (min^{-1}) are the rate parameters.

Since we had no information about insulin, except that it is withdrawn at 6am in the morning of the metabolic study and replaced with a soluble variable insulin infusion to maintain a glucose concentration at 5 mmol/L, we modelled the insulin trend as shown in the Figure 3.6.

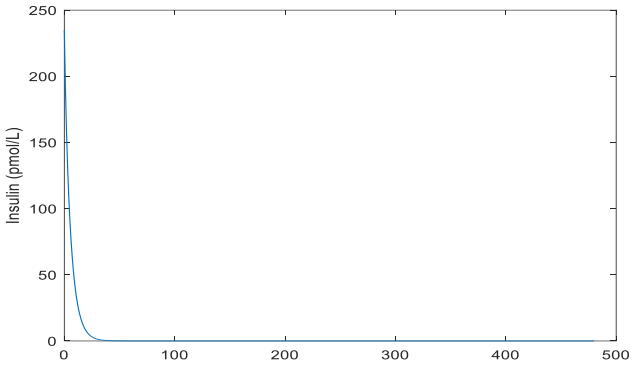

Figure 3.6: Modelled insulin trend

In order to build the model, other equations were taken into account, such as:

- Glucose Renal Excretion:

$$E(t) = \begin{cases} k_{e1} \cdot [G_p(t) - k_{e2}] & if\ G_p(t) > k_{e2} \\ 0 & if\ G_p(t) \leq k_{e2} \end{cases}$$

(3.36)

Where k_{e1} (min^{-1}) is the glomerular filtration rate and k_{e2} (mg/kg) is the renal threshold of glucose.

- Glucose Independent Utilization:

$$U_{ii}(\mathrm{t}) = F_{cns}$$

(3.37)

It takes place in the first compartment (G_p), is constant and represents glucose uptake by the brain and erythrocytes.

- Glucose Dependent Utilization:

$$U_{id}(t) = \frac{V_m(X(t)) \cdot G_t(t)}{K_m(X(t)) + G_t(t)}$$

(3.38)

Where

$$V_m(X(t)) = V_{m0} + V_{mx} \cdot X(t)$$
$$K_m\big(X(t)\big) = K_{m0} + K_{mx} \cdot X(t)$$
$$\dot{X}(t) = -p_{2u} \cdot X(t) + p_{2u}[I(t) - I_b] \qquad X(0) = 0$$

(3.39)

Where X (pmol/L) is insulin in the interstitial fluid, I is plasma insulin, and p_{2u} (min^{-1}) is the rate constant of insulin action on the peripheral glucose utilization [11].

The model was identified simultaneously on plasma glucose and rate of disappearance data, using simulated insulin (I), rate of appearance pattern (Ra), reported in the paper (Figure 2.2) as forcing function (known input).

We assumed that measurement error was independent, Gaussian, with zero mean and standard deviation equal to 2% for glucose and 4% for rate of disappearance. This last assumption was the result of some trial, since no information was available on the precision of Rd.

The Fisher approach was the tested first. The model has been built in Matlab and the optimization algorithm used was 'lsqnonlin', which is able to solve Non Linear Weighted Least Squares problems. The goal of 'lsqnonlin' is to minimize the sum of the weighted square residuals, which represent the distance between the data point and the values predicted by the model. The optimization is done by varying the model's parameters to find the combination that minimize the sum of weighted squared residuals.

One critical point was the choice of the initial values of the parameters. In fact, it turns out that the optimization algorithm is very sensitive that, due to the model complexity. Initially we decided to use as initial values those reported in the UVA/Padova Simulator: 'adult #9999', which represents the average data of 100 adult patients collected in the Simulator.

Parameter	Value	Unit of measurement
$\mathbf{V_g}$	1,84	dl/kg
$\mathbf{V_{mx}}$	0,07	mg/kg/min per pmol/l
$\mathbf{K_{m0}}$	0,10	mg/kg
$\mathbf{k_2}$	0,06	min^{-1}
$\mathbf{k_1}$	230,96	min^{-1}
$\mathbf{p_{2u}}$	0,04	min^{-1}
$\mathbf{k_{e1}}$	0,0005	min^{-1}
$\mathbf{k_{e2}}$	339	mg/kg
$\mathbf{F_{cns}}$	1	mg/kg/min

Table 3.1: Initial parameters values

Since, the choice of initial values is crucial, we decided to precede the parameters estimation with two different optimization functions: 'fmincon' and 'fminsearch'. The first one is used for constrained optimization problem and can handle problems with both equality and inequality constraints on the variables. It uses different algorithms (interior-point, active-set, sqp) depending on the problem structure and the user's preferences. These algorithms are iterative and use information about the objective function and constraints, including derivative when provided. The second instead is used for unconstrained optimization problem and it finds the minimum of a scalar function of several variables without any constraints on those variables. It uses the Nelder-Mead algorithm, which does not require the calculation of the derivative of the objective function, so it is less expensive to compute. It iteratively modifies the parameters until it converges to the minimum. Both optimization functions return us the minimum value of the objective function and the optimal values of the parameters, which will be then used by 'lsqnonlin' for parameters estimation. Tests have been done with both 'fminsearch' and 'fmincon' but the first one has shown better results in terms of precision. Since with 'fminsearch' we obtained negative values of some parameters, we decided to transform them into logarithmic scale, thus overcoming the problem.

After testing the WLS approach and realising that it was not able to return adequate results we switched to the Bayes approach. This approach was used to estimate the parameters of both the placebo group and the SGLT2i group separately.

Then a simultaneous identification of the two visits was performed, which allows us to see which parameters changed significantly after the use of SGLT2 inhibitor. The aim was to use only one model to identify simultaneously both placebo and SGLT2 inhibitor group, maintaining most of the parameters the same for both, but using two different values for parameters k_{e1} and k_{e2}, which are the parameters involved in the glucose renal excretion. In fact, in the literature it is reported that the use of SGLT2 inhibitors has an impact on the renal excretion subsystem, thus assuming that these two parameters were different in the two occasions (placebo and SGLT2i) was reasonable.

As already stated, the model is very sensitive to the initial condition. Thus, after several trials using different set of initial conditions, we concluded that the best set of initial condition were the one reported in the table below.

Parameter	Value	Unit of measurement
$\mathbf{V_g}$	1,21	dl/kg
$\mathbf{V_{mx}}$	0,11	mg/kg/min per pmol/l
$\mathbf{K_{m0}}$	236,94	mg/kg
$\mathbf{k_2}$	0,07	min^{-1}
$\mathbf{k_1}$	0,05	min^{-1}
$\mathbf{p_{2u}}$	0,05	min^{-1}
$\mathbf{k_{e1}}$	0,008	min^{-1}
$\mathbf{k_{e2}}$	216	mg/kg
$\mathbf{F_{cns}}$	1,15	mg/kg/min
$\mathbf{k_{e1}}$**_SGLT2i**	0,015	min^{-1}
$\mathbf{k_{e2}}$**_SGLT2i**	129,55	mg/kg

Table 3.2: Best set of initial conditions

Chapter 4
RESULTS

4.1 Model assessment

To assess which model would be the best, several tests have been conducted. The tests performed can be divided into four phases. The first aimed at estimating the parameters of the placebo group. The second was similar to the first but aimed at estimating the parameters of the SGLT2i group. The third, named simultaneous identification, aimed to find a single parameter set (apart from a couple of parameters that are allowed to vary between visits) able to fit both placebo and SGLT2i group simultaneously. The last one aimed at finding the best set of initial conditions tested again in the simultaneous identification.

For the first three phases the set of initial conditions considered was the one reported in the Table 3.1, which was taken from the UVA/Padova Simulator, representing the average data of 100 adult patients.

The results of the first phase are reported below.

- Test n.1.
  Fisher approach:
  This approach provided a satisfactory fit of the data (Figure 4.1) but also a very bad precision of the estimated parameters (Tabel 4.1). The precision of a parameter is represented by the coefficient of variation (CV), which is a number expressed as a percentage and the larger it is, the lower the precision with which that parameters are estimated will be.

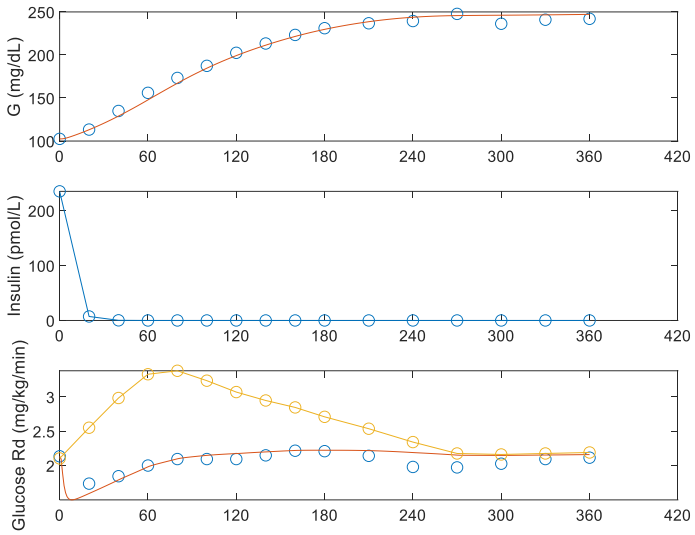

Figure 4.1: Result obtained with the Fisher Approach on the placebo group. Upper panel: plasma glucose concentration data (open blue circle) vs model prediction (continuous red line). Middle panel: simulated plasma insulin concentration (open blue circle and continuous blue line). Lower panel: glucose rate of disappearance data (open blue circle) vs model prediction (continuous red line) and endogenous glucose production data (open yellow circle and continuous yellow line)

Parameter	**Value**	**CV%**
$\mathbf{V_g}$ $[dl/kg]$	1,35	4,26
$\mathbf{V_{mx}}$ $[\ mg/kg/min\ per\ pmol/l]$	0,5	>1000
$\mathbf{K_{m0}}[\ mg/kg\ ]$	$8{,}967\cdot 10^{-7}$	$1{,}190\cdot 10^{14}$
$\mathbf{k_2}[\ min^{-1}\ ]$	0,38	>1000
$\mathbf{k_1}\ [\ min^{-1}\ ]$	0,14	>1000
$\mathbf{p_{2u}}\ [\ min^{-1}\ ]$	0,11	>1000
$\mathbf{k_{e1}}\ [\ min^{-1}\ ]$	0,005	>1000
$\mathbf{k_{e2}}$ $[mg/kg]$	3,27	>1000
$\mathbf{F_{cns}}$ $[mg/kg/min]$	0,5	>1000

Tabel 4.1: Values of parameters estimated with the Fisher approach and their precision in the placebo group.

Moreover, some of the estimated values turn out to be non-physiological (i.e. $k_{e2}, k_{m0}, p_{2u}, F_{cns}$). Obviously, this approach cannot be considered for the construction of an optimal model, so we decided to try to estimate the parameters through the Bayesian Maximum a Posteriori Approach.

- Test n.2.

  This test consisted in using Bayesian estimator preceded by 'fminsearch' optimization algorithm:

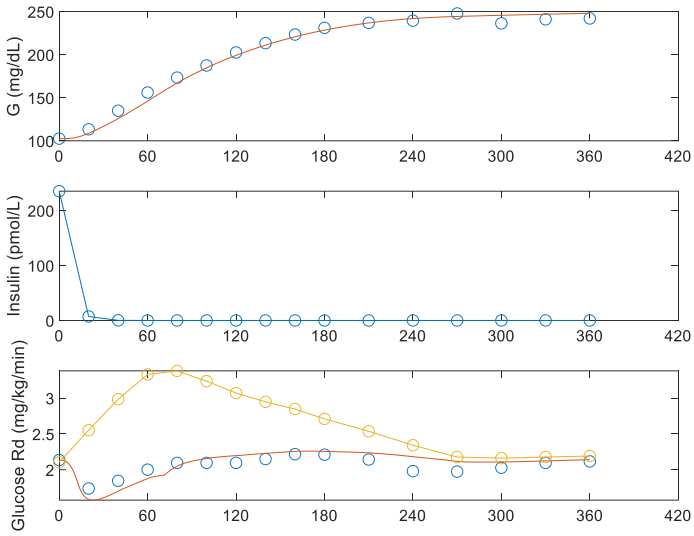

Figure 4.2: Result obtained with the Bayesian estimator preceded by 'fminsearch' optimization algorithm on the placebo group. Upper panel: plasma glucose concentration data (open blue circle) vs model prediction (continuous red line). Middle panel: simulated plasma insulin concentration (open blue circle and continuous blue line). Lower panel: glucose rate of disappearance data (open blue circle) vs model prediction (continuous red line) and endogenous glucose production data (open yellow circle and continuous yellow line)

Parameter	**Value**	**CV%**
$\mathbf{V_g}$ [*dl/kg*]	1,35	2,42
$\mathbf{V_{mx}}$ [*mg/kg/min per pmol/l*]	0,12	252,6
$\mathbf{K_{m0}}$[mg/kg]	230	0,04
$\mathbf{k_2}$[min^{-1}]	0,09	337
$\mathbf{k_1}$ [min^{-1}]	0,09	226
$\mathbf{p_{2u}}$ [min^{-1}]	0,05	625,5
$\mathbf{k_{e1}}$ [min^{-1}]	0,01	>1000
$\mathbf{k_{e2}}$ [*mg/kg*]	213	0,045
$\mathbf{F_{cns}}$ [*mg/kg/min*]	0,6	117,06

Table 4.2: Values of parameters estimated with the Batesian approach preceded by 'fminsearch' optimization algorithm and their precision in the placebo group.

This second test provided a good fit of the data but a bad precision of some estimates.

- Test n.3.

  This test consisted in using 'fminsearch' optimization algorithm, MAP estimator and a standard deviation (SD) for F_{cns}. Having noticed that, in the previous case, the Fcns value was far from the physiological one, it was decided to use a very tight variance associated to this parameter (0,001), so that the estimated value will not deviate much from the initial value.

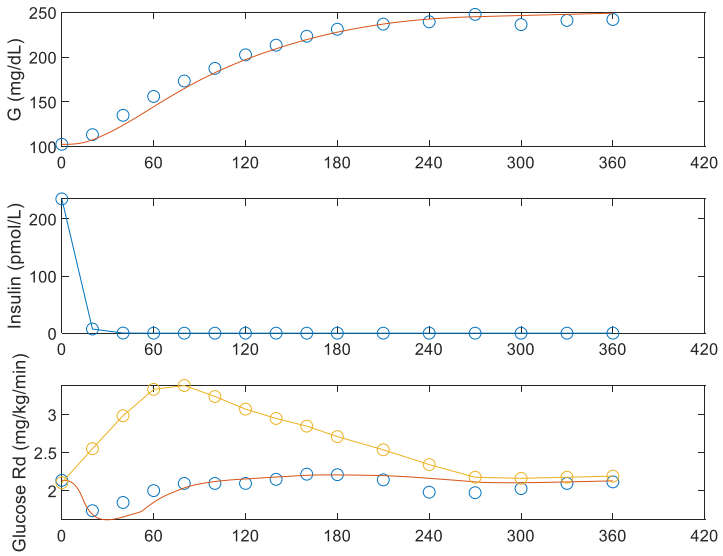

Figure 4.3: Result obtained with the Bayesian estimator preceded by 'fminsearch' optimization algorithm and SD for F_{cns} of 0,001 on the placebo group. Upper panel: plasma glucose concentration data (open blue circle) vs model prediction (continuous red line). Middle panel: simulated plasma insulin concentration (open blue circle and continuous blue line). Lower panel: glucose rate of disappearance data (open blue circle) vs model prediction (continuous red line) and endogenous glucose production data (open yellow circle and continuous yellow line)

Parameter	**Value**	**CV%**
$\mathbf{V_g}$ $[dl/kg]$	1,4	2,7
$\mathbf{V_{mx}}$ $[\ mg/kg/min\ per\ pmol/l]$	0,45	60,7
$\mathbf{K_{m0}}[\ mg/kg\ ]$	218	0,04
$\mathbf{k_2}[\ min^{-1}\ ]$	0,08	199
$\mathbf{k_1}$ $[\ min^{-1}\ ]$	0,05	407
$\mathbf{p_{2u}}$ $[\ min^{-1}\ ]$	0,01	>1000
$\mathbf{k_{e1}}$ $[\ min^{-1}\ ]$	0,007	>1000
$\mathbf{k_{e2}}$ $[mg/kg]$	188	0,07
$\mathbf{F_{cns}}$ $[mg/kg/min]$	1	0,1

Table 4.3: Values of parameters estimated with the Batesian approach preceded by 'fminsearch' optimization algorithm and a SD for F_{cns} of 0,001 and their precision in the placebo group.

Since every test performed involved the use of both the optimisation function ('fminsearch') and the MAP, from now on it will be implied that they were used, and any additions or changes made will be highlighted exclusively.

- Test n.4.

  To try to lower the high CV values a SD of 0,2 was assumed for k_{e1}.

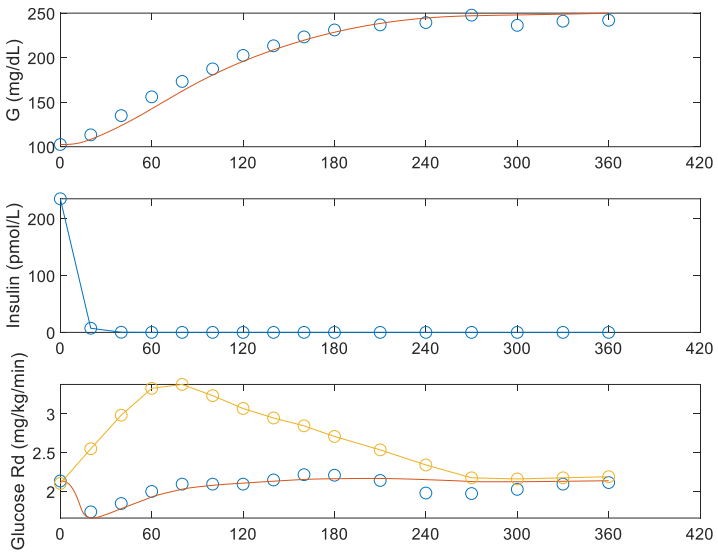

Figure 4.4: Result obtained with the Bayesian estimator preceded by 'fminsearch' optimization algorithm and SD for F_{cns} of 0,001 and for k_{e1} of 0,2 on the placebo group. Upper panel: plasma glucose concentration data (open blue circle) vs model prediction (continuous red line). Middle panel: simulated plasma insulin concentration (open blue circle and continuous blue line). Lower panel: glucose rate of disappearance data (open blue circle) vs model prediction (continuous red line) and endogenous glucose production data (open yellow circle and continuous yellow line)

Parameter	**Value**	**CV%**
$\mathbf{V_g}$ [*dl/kg*]	1,4	1,9
$\mathbf{V_{mx}}$ [*mg/kg/min per pmol/l*]	0,13	274,3
$\mathbf{K_{m0}}$[mg/kg]	227	0,04
$\mathbf{k_2}$[min^{-1}]	0,2	120,3
$\mathbf{k_1}$ [min^{-1}]	0,06	349
$\mathbf{p_{2u}}$ [min^{-1}]	0,06	625
$\mathbf{k_{e1}}$ [min^{-1}]	0,004	>1000
$\mathbf{k_{e2}}$ [*mg/kg*]	38	3
$\mathbf{F_{cns}}$ [*mg/kg/min*]	1	0,1

Table 4.4: Values of parameters estimated with the Batesian approach preceded by 'fminsearch' optimization algorithm and a SD for F_{cns} of 0,001 and for k_{e1} of 0,2 and their precision in the placebo group.

As shown in the Table 4.4, this test returned a non-physiological k_{e2} value (37,88 mg/kg), which leads us to say that this case cannot be considered.

- Test n.5.

  To avoid too low k_{e2} values its SD in the prior was reduced to 0,2.

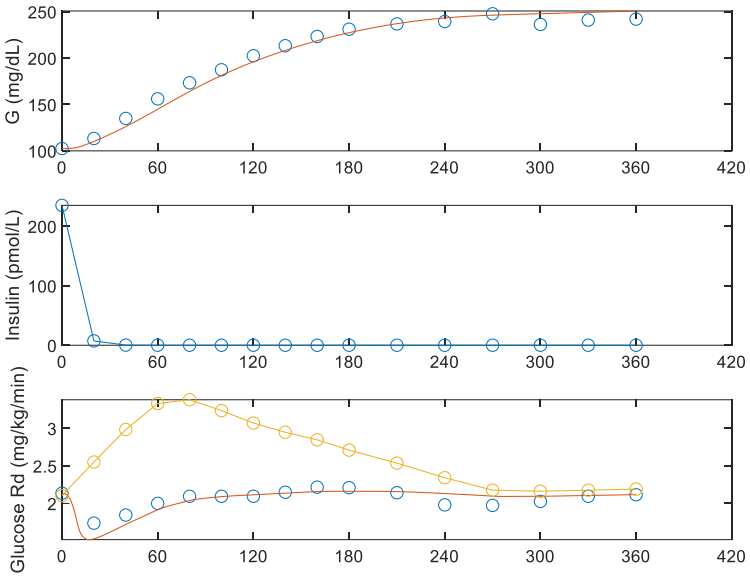

Figure 4.5: Result obtained with the Bayesian estimator preceded by 'fminsearch' optimization algorithm and SD for F_{cns} of 0,001, for k_{e1} of 0,2 and for k_{e2} of 0,2 on the placebo group. Upper panel: plasma glucose concentration data (open blue circle) vs model prediction (continuous red line). Middle panel: simulated plasma insulin concentration (open blue circle and continuous blue line). Lower panel: glucose rate of disappearance data (open blue circle) vs model prediction (continuous red line) and endogenous glucose production data (open yellow circle and continuous yellow line)

Parameter	**Value**	**CV%**
$\mathbf{V_g}$ $[dl/kg]$	1,4	1,8
$\mathbf{V_{mx}}$ $[\ mg/kg/min\ per\ pmol/l]$	0,2	186,3
$\mathbf{K_{m0}}[\ mg/kg\ ]$	224	0,04
$\mathbf{k_2}[\ min^{-1}\ ]$	0,22	108,3
$\mathbf{k_1}$ $[\ min^{-1}\ ]$	0,09	236,8
$\mathbf{p_{2u}}$ $[\ min^{-1}\ ]$	0,06	560,6
$\mathbf{k_{e1}}$ $[\ min^{-1}\ ]$	0,005	>1000
$\mathbf{k_{e2}}$ $[mg/kg]$	133	0,12
$\mathbf{F_{cns}}$ $[mg/kg/min]$	1	0,1

Table 4.5: Values of parameters estimated with the Batesian approach preceded by 'fminsearch' optimization algorithm and a SD for F_{cns} of 0,001, for k_{e1} of 0,2 and for k_{e2} of 0,2 and their precision in the placebo group.

Finally, the last three tests performed involved a reduction of the covariance matrix, as the smaller the covariance of the prior, the more precise the estimates will be.

- Test n.6.

  First, the covariance matrix was multiplied by 0,75.

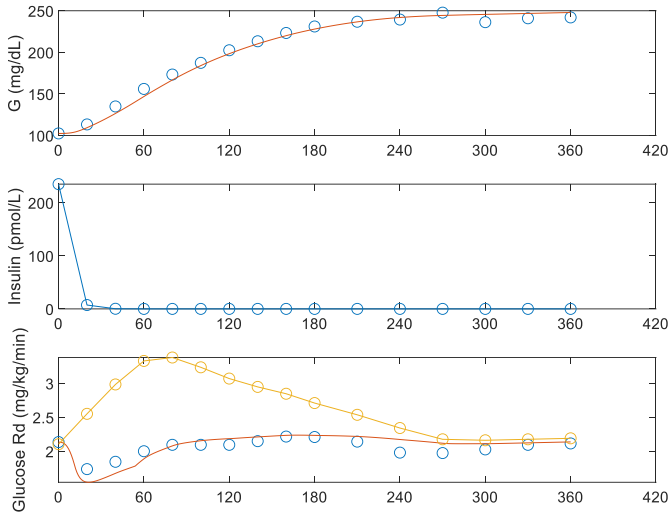

Figure 4.6: Result obtained with the Bayesian estimator preceded by 'fminsearch' optimization algorithm and the covariance matrix multiplied by 0,75 on the placebo group. Upper panel: plasma glucose concentration data (open blue circle) vs model prediction (continuous red line). Middle panel: simulated plasma insulin concentration (open blue circle and continuous blue line). Lower panel: glucose rate of disappearance data (open blue circle) vs model prediction (continuous red line) and endogenous glucose production data (open yellow circle and continuous yellow line)

Parameter	**Value**	**CV%**
$\mathbf{V_g}$ [*dl/kg*]	1,4	1,9
$\mathbf{V_{mx}}$ [*mg/kg/min per pmol/l*]	0,14	234,3
$\mathbf{K_{m0}}$[mg/kg]	231	0,03
$\mathbf{k_2}$[min^{-1}]	0,13	203,4
$\mathbf{k_1}$ [min^{-1}]	0,09	203,6
$\mathbf{p_{2u}}$ [min^{-1}]	0,06	535,8
$\mathbf{k_{e1}}$ [min^{-1}]	0,009	>1000
$\mathbf{k_{e2}}$ [*mg/kg*]	192,3	0,06
$\mathbf{F_{cns}}$ [*mg/kg/min*]	0,8	38,8

Table 4.6: Values of parameters estimated with the Batesian approach preceded by 'fminsearch' optimization algorithm and the covariance matrix multiplied by 0,75 and their precision in the placebo group.

- Test n.7.

  Then, the covariance matrix was multiplied by 0,5

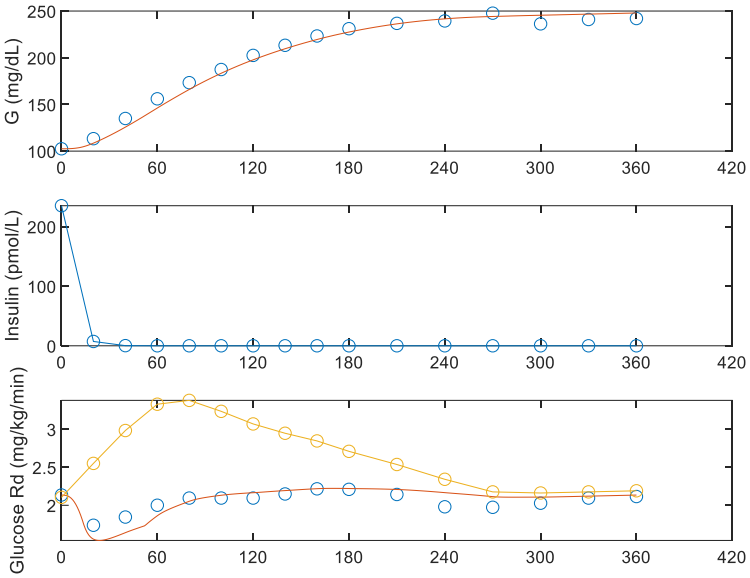

Figure 4.7: Result obtained with the Bayesian estimator preceded by 'fminsearch' optimization algorithm and the covariance matrix multiplied by 0,5 on the placebo group. Upper panel: plasma glucose concentration data (open blue circle) vs model prediction (continuous red line). Middle panel: simulated plasma insulin concentration (open blue circle and continuous blue line). Lower panel: glucose rate of disappearance data (open blue circle) vs model prediction (continuous red line) and endogenous glucose production data (open yellow circle and continuous yellow line)

Parameter	**Value**	**CV%**
$\mathbf{V_g}$ $[dl/kg]$	1,4	1,8
$\mathbf{V_{mx}}$ $[\ mg/kg/min\ per\ pmol/l]$	0,11	228
$\mathbf{K_{m0}}[\ mg/kg\ ]$	233,5	0,03
$\mathbf{k_2}[\ min^{-1}\ ]$	0,13	191,1
$\mathbf{k_1}$ $[\ min^{-1}\ ]$	0,09	186,6
$\mathbf{p_{2u}}$ $[\ min^{-1}\ ]$	0,05	495,1
$\mathbf{k_{e1}}$ $[\ min^{-1}\ ]$	0,008	>1000
$\mathbf{k_{e2}}$ $[mg/kg]$	192,5	0,06
$\mathbf{F_{cns}}$ $[mg/kg/min]$	0,8	38

Table 4.7: Values of parameters estimated with the Batesian approach preceded by 'fminsearch' optimization algorithm and the covariance matrix multiplied by 0,5 and their precision in the placebo group.

- Test n.8.

  The last test is performed by multiplying the covariance matrix by 0,25.

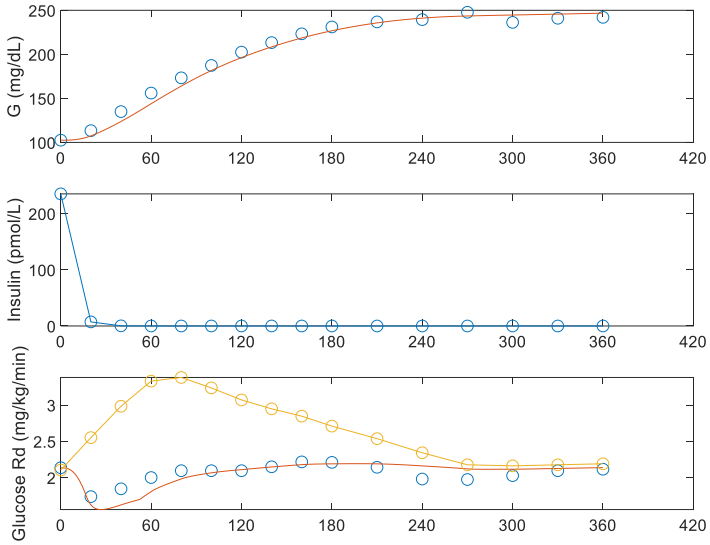

Figure 4.8: Result obtained with the Bayesian estimator preceded by 'fminsearch' optimization algorithm and the covariance matrix multiplied by 0,25 on the placebo group. Upper panel: plasma glucose concentration data (open blue circle) vs model prediction (continuous red line). Middle panel: simulated plasma insulin concentration (open blue circle and continuous blue line). Lower panel: glucose rate of disappearance data (open blue circle) vs model prediction (continuous red line) and endogenous glucose production data (open yellow circle and continuous yellow line)

Parameter	**Value**	**CV%**
$\mathbf{V_g}$ [*dl/kg*]	1,45	1,6
$\mathbf{V_{mx}}$ [*mg/kg/min per pmol/l*]	0,09	158,8
$\mathbf{K_{m0}}$[mg/kg]	237,6	0,02
$\mathbf{k_2}$[min^{-1}]	0,15	134,2
$\mathbf{k_1}$ [min^{-1}]	0,07	153,54
$\mathbf{p_{2u}}$ [min^{-1}]	0,04	374,37
$\mathbf{k_{e1}}$ [min^{-1}]	0,007	>1000
$\mathbf{k_{e2}}$ [*mg/kg*]	197,35	0,07
$\mathbf{F_{cns}}$ [*mg/kg/min*]	1	16,34

Table 4.8: Values of parameters estimated with the Batesian approach preceded by 'fminsearch' optimization algorithm and the covariance matrix multiplied by 0,25 and their precision in the placebo group.

As shown in figures 4.2 to 4.8, all tests performed provide a good fit of the data but this is not the only aspect to consider when choosing an optimal model. Another important issue to take into account is the precision of the estimates, expressed as coefficient of variation (CV).

The next step was to repeat all the tests done for the placebo group on the SGLT2 inhibitor group.

- Test n.9.

  We started with the simplest case: ‘fminsearch’ optimization algorithm, followed by MAP estimator and a SD for Fcns of (0,001). It should also be mentioned that the case of 'fminsearch' and MAP was tested but unfortunately returned infinite values for CVs. This means that the algorithm was not able to find a good set of parameters that minimize the objective function that resulted in very imprecise parameter estimates.

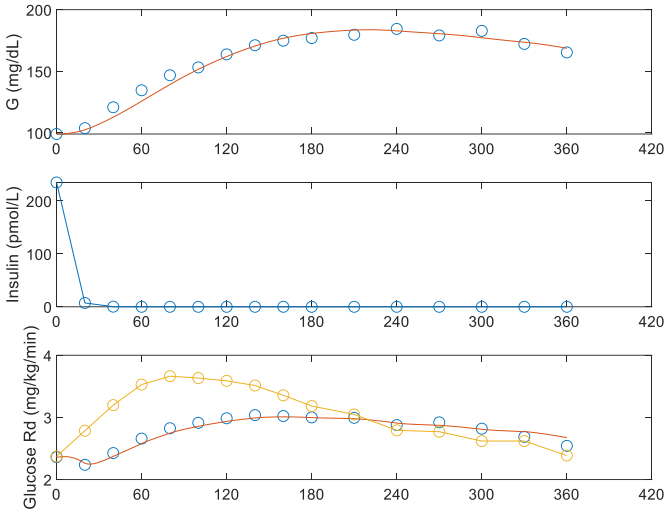

Figure 4.9: Result obtained with the Bayesian estimator preceded by ‘fminsearch’ optimization algorithm and SD for F_{cns} of 0,001 on the SGLT2i group. Upper panel: plasma glucose concentration data (open blue circle) vs model prediction (continuous red line). Middle panel: simulated plasma insulin concentration (open blue circle and continuous blue line). Lower panel: glucose rate of disappearance data (open blue circle) vs model prediction (continuous red line) and endogenous glucose production data (open yellow circle and continuous yellow line)

Parameter	**Value**	**CV%**
$\mathbf{V_g}$ $[dl/kg]$	1,37	3,2
$\mathbf{V_{mx}}$ $[\ mg/kg/min\ per\ pmol/l]$	0,04	892,2
$\mathbf{K_{m0}}[\ mg/kg\ ]$	244,7	0,04
$\mathbf{k_2}[\ min^{-1}\ ]$	0,14	189,7
$\mathbf{k_1}$ $[\ min^{-1}\ ]$	0,05	412,6
$\mathbf{p_{2u}}$ $[\ min^{-1}\ ]$	0,05	705,2
$\mathbf{k_{e1}}$ $[\ min^{-1}\ ]$	0,009	>1000
$\mathbf{k_{e2}}$ $[mg/kg]$	28	4,4
$\mathbf{F_{cns}}$ $[mg/kg/min]$	1	0,1

Table 4.9: Values of parameters estimated with the Batesian approach preceded by ‘fminsearch’ optimization algorithm and a SD for F_{cns} of 0,001 and their precision in the SGLT2i group.

As shown in the table 4.9, the value of k_{e2} is by no means physiological, so in the next test a variance of 0.2 was assumed for k_{e1} to see if changing a parameter closely related to k_{e2} also changes the value of the latter.

- Test n.10.

  'fiminsearch' optimization algorithm, MAP and SD for F_{cns} and k_{e1} of 0,001 and 0,2 respectively:

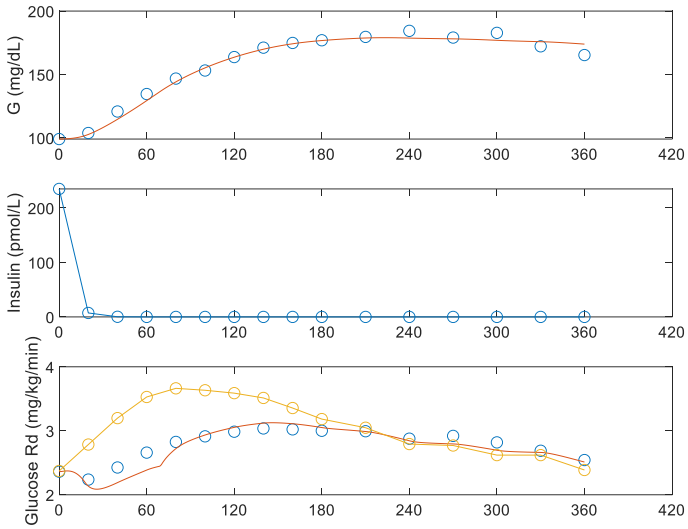

Figure 4.10: Result obtained with the Bayesian estimator preceded by 'fminsearch' optimization algorithm and SD for F_{cns} of 0,001 and for k_{e1} of 0,2 on the SGLT2i group. Upper panel: plasma glucose concentration data (open blue circle) vs model prediction (continuous red line). Middle panel: simulated plasma insulin concentration (open blue circle and continuous blue line). Lower panel: glucose rate of disappearance data (open blue circle) vs model prediction (continuous red line) and endogenous glucose production data (open yellow circle and continuous yellow line)

Parameter	**Value**	**CV%**
$\mathbf{V_g}$ $[dl/kg]$	1,46	2,8
$\mathbf{V_{mx}}$ $[mg/kg/min\ per\ pmol/l]$	0,12	259
$\mathbf{K_{m0}}[mg/kg]$	236,4	0,04
$\mathbf{k_2}[min^{-1}]$	0,07	278
$\mathbf{k_1}$ $[min^{-1}]$	0,09	198,4
$\mathbf{p_{2u}}$ $[min^{-1}]$	0,02	>1000
$\mathbf{k_{e1}}$ $[min^{-1}]$	0,03	608
$\mathbf{k_{e2}}$ $[mg/kg]$	199,6	0,03
$\mathbf{F_{cns}}$ $[mg/kg/min]$	1	0,1

Table 4.10: Values of parameters estimated with the Batesian approach preceded by 'fminsearch' optimization algorithm and a SD for F_{cns} of 0,001 and for k_{e1} of 0,2 and their precision in the SGLT2i group.

As can be seen, this strategy succeeded in bringing the k_{e2} value in a physiological range. Unfortunately, however, the problem of high CVs still persisted.

As in the case of the placebo group, also for the SGLT2i all the tests conducted involved the use of both 'fminsearch' and MAP. Therefore, in the following, only additional assumptions or changes made are higilighted.

- Test n.11.

  To try to lower the CV values, a SD of 0,2 for ke2 was assumed.

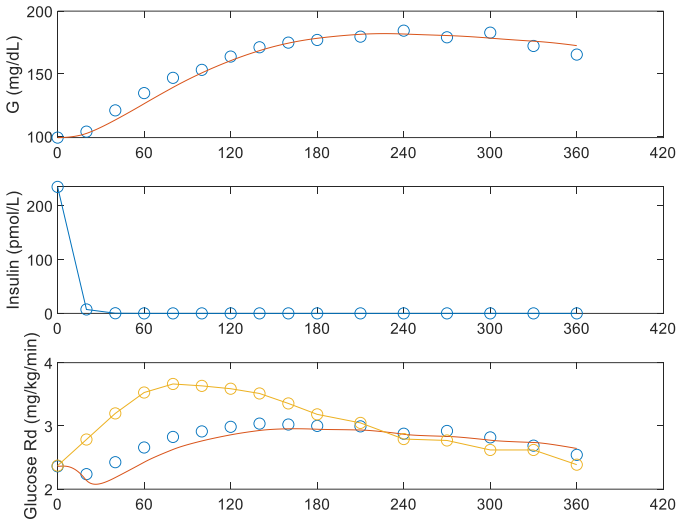

Figure 4.11: Result obtained with the Bayesian estimator preceded by 'fminsearch' optimization algorithm and SD for F_{cns} of 0,001, for k_{e1} of 0,2 and for k_{e2} of 0,2 on the SGLT2i group. Upper panel: plasma glucose concentration data (open blue circle) vs model prediction (continuous red line). Middle panel: simulated plasma insulin concentration (open blue circle and continuous blue line). Lower panel: glucose rate of disappearance data (open blue circle) vs model prediction (continuous red line) and endogenous glucose production data (open yellow circle and continuous yellow line)

Parameter	Value	CV%
$\mathbf{V_g}$ [*dl/kg*]	1,62	2,4
$\mathbf{V_{mx}}$ [*mg/kg/min per pmol/l*]	0,09	332,1
$\mathbf{K_{m0}}$[mg/kg]	240	0,04
$\mathbf{k_2}$[min^{-1}]	0,14	180,3
$\mathbf{k_1}$ [min^{-1}]	0,05	391,3
$\mathbf{p_{2u}}$ [min^{-1}]	0,03	>1000
$\mathbf{k_{e1}}$ [min^{-1}]	0,01	898,10
$\mathbf{k_{e2}}$ [*mg/kg*]	102	0,2
$\mathbf{F_{cns}}$ [*mg/kg/min*]	1	0,1

Table 4.11: Values of parameters estimated with the Batesian approach preceded by 'fminsearch' optimization algorithm and a SD for F_{cns} of 0,001, for k_{e1} of 0,2 and for k_{e2} of 0,2 and their precision in the SGLT2i group.

The last three tests performed involved the reduction of the covariance matrix, as the smaller the covariance of the prior, the more precise the estimates will be.

- Test n.12.

  The covariance matrix was multiplied by a factor 0,75

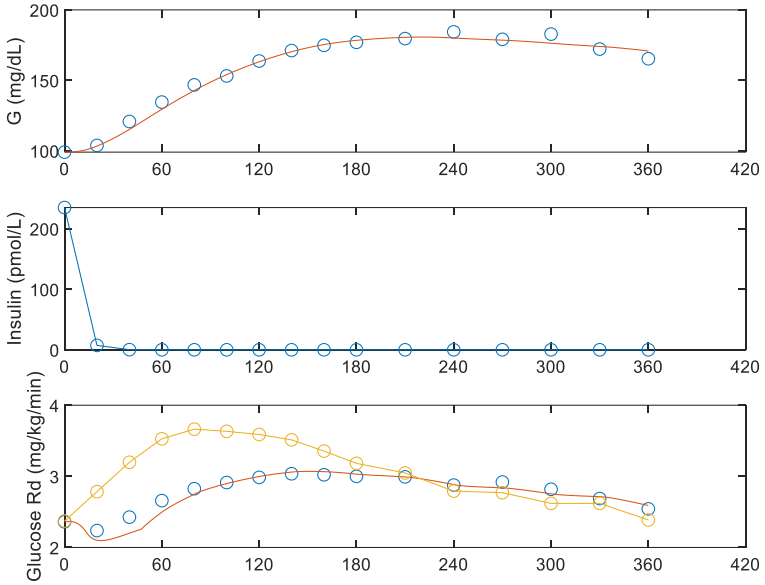

Figure 4.12: Result obtained with the Bayesian estimator preceded by 'fminsearch' optimization algorithm and the covariance matrix multiplied by 0,75 on the SGLT2i group. Upper panel: plasma glucose concentration data (open blue circle) vs model prediction (continuous red line). Middle panel: simulated plasma insulin concentration (open blue circle and continuous blue line). Lower panel: glucose rate of disappearance data (open blue circle) vs model prediction (continuous red line) and endogenous glucose production data (open yellow circle and continuous yellow line)

Parameter	**Value**	**CV%**
$\mathbf{V_g}$ $[dl/kg]$	1,4	2,8
$\mathbf{V_{mx}}$ $[mg/kg/min\ per\ pmol/l]$	0,07	456,5
$\mathbf{K_{m0}}[mg/kg]$	239,7	0,03
$\mathbf{k_2}[min^{-1}]$	0,11	220
$\mathbf{k_1}[min^{-1}]$	0,07	260,4
$\mathbf{p_{2u}}[min^{-1}]$	0,04	772,9
$\mathbf{k_{e1}}[min^{-1}]$	0,015	>1000
$\mathbf{k_{e2}}$ $[mg/kg]$	174,7	0,06
$\mathbf{F_{cns}}$ $[mg/kg/min]$	1,6	67

Table 4.12: Values of parameters estimated with the Batesian approach preceded by 'fminsearch' optimization algorithm and the covariance matrix multiplied by 0,75 and their precision in the SGLT2i group.

- Test n.13.

  The covariance matrix was multiplied by a factor 0,5

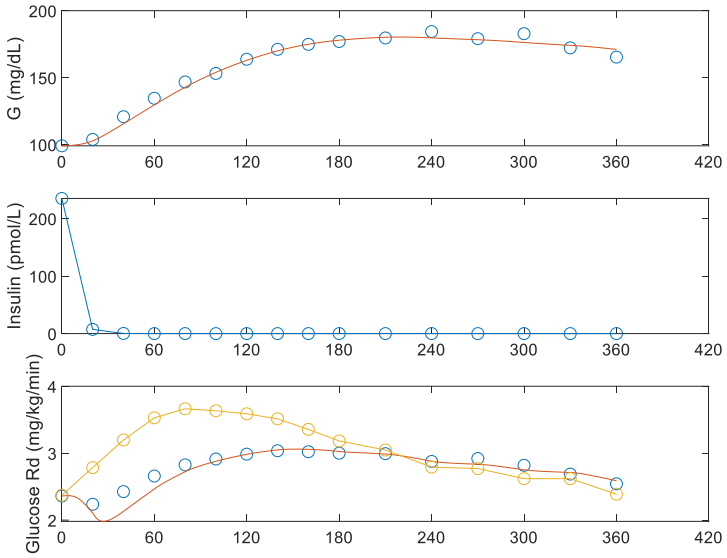

Figure 4.13: Result obtained with the Bayesian estimator preceded by 'fminsearch' optimization algorithm and the covariance matrix multiplied by 0,5 on the SGLT2i group. Upper panel: plasma glucose concentration data (open blue circle) vs model prediction (continuous red line). Middle panel: simulated plasma insulin concentration (open blue circle and continuous blue line). Lower panel: glucose rate of disappearance data (open blue circle) vs model prediction (continuous red line) and endogenous glucose production data (open yellow circle and continuous yellow line)

Parameter	**Value**	**CV%**
$\mathbf{V_g}$ $[dl/kg]$	1,5	2,4
$\mathbf{V_{mx}}$ $[\ mg/kg/min\ per\ pmol/l]$	0,06	480,3
$\mathbf{K_{m0}}[\ mg/kg\ ]$	241,7	0,03
$\mathbf{k_2}[\ min^{-1}\ ]$	0,11	196,3
$\mathbf{k_1}$ $[\ min^{-1}\ ]$	0,07	236,4
$\mathbf{p_{2u}}$ $[\ min^{-1}\ ]$	0,04	726,4
$\mathbf{k_{e1}}$ $[\ min^{-1}\ ]$	0,02	>1000
$\mathbf{k_{e2}}$ $[mg/kg]$	152,3	0,08
$\mathbf{F_{cns}}$ $[mg/kg/min]$	1,05	41

Table 4.13: Values of parameters estimated with the Batesian approach preceded by 'fminsearch' optimization algorithm and the covariance matrix multiplied by 0,5 and their precision in the SGLT2i group.

- Test n.14.

  The covariance matrix was multiplied by a factor 0,25

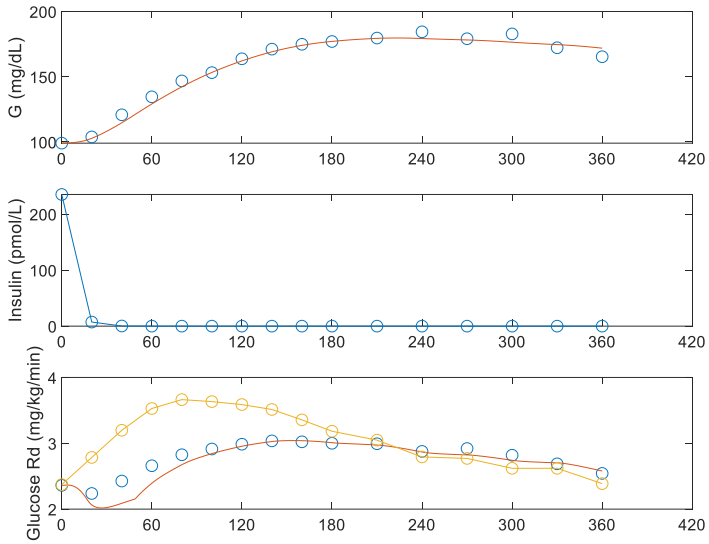

Figure 4.14: Result obtained with the Bayesian estimator preceded by 'fminsearch' optimization algorithm and the covariance matrix multiplied by 0,25 on the SGLT2i group. Upper panel: plasma glucose concentration data (open blue circle) vs model prediction (continuous red line). Middle panel: simulated plasma insulin concentration (open blue circle and continuous blue line). Lower panel: glucose rate of disappearance data (open blue circle) vs model prediction (continuous red line) and endogenous glucose production data (open yellow circle and continuous yellow line)

Parameter	**Value**	**CV%**
$\mathbf{V_g}$ [*dl/kg*]	1,6	1,9
$\mathbf{V_{mx}}$ [*mg/kg/min per pmol/l*]	0,06	276,7
$\mathbf{K_{m0}}$[mg/kg]	237	0,02
$\mathbf{k_2}$[min^{-1}]	0,09	177,5
$\mathbf{k_1}$ [min^{-1}]	0,07	167,45
$\mathbf{p_{2u}}$ [min^{-1}]	0,04	450
$\mathbf{k_{e1}}$ [min^{-1}]	0,02	>1000
$\mathbf{k_{e2}}$ [*mg/kg*]	193,9	0,04
$\mathbf{F_{cns}}$ [*mg/kg/min*]	1,4	10,28

Table 4.14: Values of parameters estimated with the Batesian approach preceded by 'fminsearch' optimization algorithm and the covariance matrix multiplied by 0,25 and their precision in the SGLT2i group.

So far, the best compromise in terms of data fitting and precision of the parameters, for both placebo and SGLT2i group, even if some of them were still estimated with poor precision, was obtained with the MAP estimator preceded by 'fminsearch' optimization function and the covariance matrix multiplied by 0.75 (figure and table 4.6 and 4.12). Unfortunately, despite we aimed to find a balance between good fit and good precision of the estimate, we had to deal with high CV, especially for some parameters. This was

in part due to the fact that few data points were available to estimate well the slope of glucose and glucose rate of disappearance's curves.

To improve the precision of the estimates we resorted to the so-called simultaneous identification. It is a way which allow us to identify simultaneously both the placebo and the SGLT2i group. In this way, it was easier to understand which parameters changed significantly before and after the use of SGLT2i. Since what is reported in the literature is that the use of SGLT2 inhibitors impacts the renal excretion, it was reasonable to consider that k_{e1} and k_{e2}, which are the parameters involved in the glucose renal excretion, were different in the two occasions (placebo and SGLT2i), while the remaining parameters were assumed to be identical.

- Test n.15.

  'fminsearch' optimization algorithm, MAP estimator and a SD for V_g and F_{cns} of 0,5 and 0,001 respectively were used in this case.

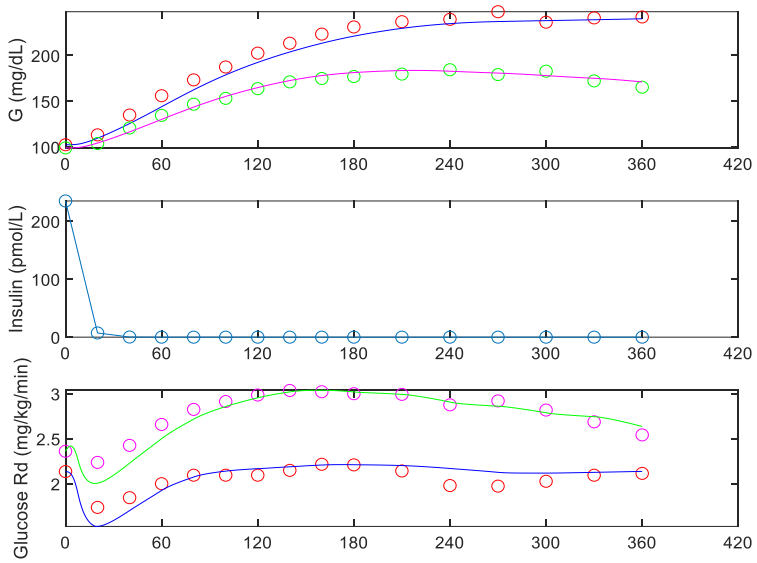

Figure 4.15: Result obtained with the simultaneous identification using Bayesian estimator preceded by 'fminsearch' optimization algorithm and a SD of 0,5 for V_g and of 0,001 for F_{cns} . Upper panel: plasma glucose concentration data of placebo group (open red circle) vs model prediction of placebo group (continuous blue line), plasma glucose concentration data of SGLT2i group (open pink circle) vs model prediction of SGLT2i group (continuous green line). Middle panel: simulated plasma insulin concentration (open blue circle and continuous blue line). Lower panel: glucose rate of disappearance data of placebo group (open red circle) vs model prediction of placebo group (continuous blue line) and glucose rate of disappearance data of SGLT2i group (open pink circle) vs model prediction of SGLT2i group (continuous green line).

Parameter	**Value**	**CV%**
$\mathbf{V_g}$ $[dl/kg]$	1,46	1,8

$\mathbf{V_{mx}}$ [*mg/kg/min per pmol/l*]	0,2	105,15
$\mathbf{K_{m0}}$[mg/kg]	213,4	0,04
$\mathbf{k_2}$[min^{-1}]	0,11	196,5
$\mathbf{k_1}$ [min^{-1}]	0,05	336,75
$\mathbf{p_{2u}}$ [min^{-1}]	0,09	363,8
$\mathbf{k_{e1}}$ [min^{-1}]	0,006	>1000
$\mathbf{k_{e2}}$ [*mg/kg*]	159	0,06
$\mathbf{F_{cns}}$ [*mg/kg/min*]	1	0,09
$\mathbf{k_{e1}}$**_SGLTi**[min^{-1}]	0,01	834,8
$\mathbf{k_{e2}}$**_SGLTi** [*mg/kg*]	102,2	0,12

Table 4.15: Values of parameters estimated in the simultaneous identification with the Batesian approach preceded by 'fminsearch' optimization algorithm and and a SD of 0,5 for V_g and of 0,001 for $F_{cns}$and their precision.

- Test n.16.

  To try to lower some CVs, SD of p_{2u} and to $k_{e1}$was assumed to be 0,75 and 0,2 respectively.

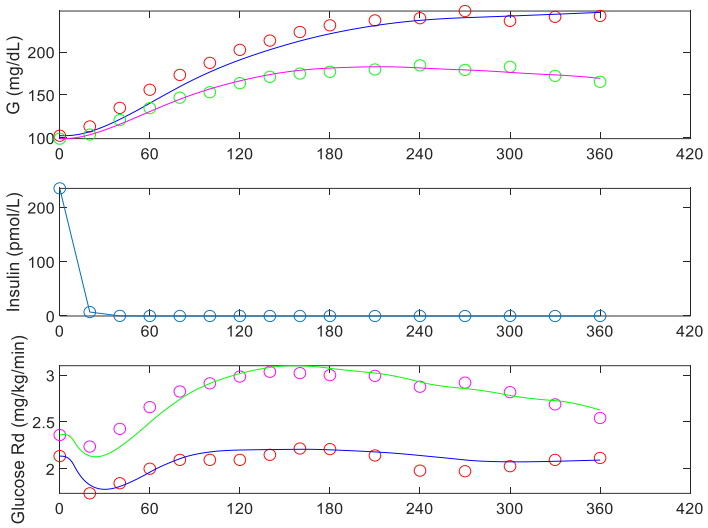

Figure 4.16: Result obtained with the simultaneous identification using Bayesian estimator preceded by 'fminsearch' optimization algorithm and a SD of 0,5 for V_g, of 0,001 for F_{cns}, of 0,75 for p_{2u} and of 0,2 for k_{e1} . Upper panel: plasma glucose concentration data of placebo group (open red circle) vs model prediction of placebo group (continuous blue line), plasma glucose concentration data of SGLT2i group (open pink circle) vs model prediction of SGLT2i group (continuous green line). Middle panel: simulated plasma insulin concentration (open blue circle and continuous blue line). Lower panel: glucose rate of disappearance data of placebo group (open red circle) vs model prediction of placebo group (continuous blue line) and glucose rate of disappearance data of SGLT2i group (open pink circle) vs model prediction of SGLT2i group (continuous green line).

Parameter	**Value**	**CV%**
$\mathbf{V_g}$ [*dl/kg*]	1,36	2
$\mathbf{V_{mx}}$ [*mg/kg/min per pmol/l*]	0,5	56,8
$\mathbf{K_{m0}}$[mg/kg]	222,6	0,04
$\mathbf{k_2}$[min^{-1}]	0,04	448,8
$\mathbf{k_1}$ [min^{-1}]	0,02	800

$\mathbf{p_{2u}}$ [min^{-1}]	0,03	>1000
$\mathbf{k_{e1}}$ [min^{-1}]	0,005	>1000
$\mathbf{k_{e2}}$ [*mg/kg*]	114,2	0,25
$\mathbf{F_{cns}}$ [*mg/kg/min*]	1	0,09
$\mathbf{k_{e1}}$**_SGLTi**[min^{-1}]	0,01	856,8
$\mathbf{k_{e2}}$**_SGLTi** [*mg/kg*]	104,5	0,16

Table 4.16: Values of parameters estimated in the simultaneous identification with the Batesian approach preceded by 'fminsearch' optimization algorithm and and a SD of 0,5 for V_g, of 0,001 for F_{cns}, of 0,75 for p_{2u} and of 0,2 for k_{e1} and their precision.

As already stated, the model is very sensitive to the initial conditions, thus several simulations were carried out with different sets of initial conditions to find the best one. The new set of initial conditions to be used was chosen by assessing the goodness of the fit it generated, and it is reported in the table below.

Parameter	**Value**
$\mathbf{V_g}$ [*dl/kg*]	1,2
$\mathbf{V_{mx}}$ [*mg/kg/min per pmol/l*]	0,11
$\mathbf{K_{m0}}$[mg/kg]	236,9
$\mathbf{k_2}$[min^{-1}]	0,08
$\mathbf{k_1}$ [min^{-1}]	0,05
$\mathbf{p_{2u}}$ [min^{-1}]	0,05
$\mathbf{k_{e1}}$ [min^{-1}]	0,008
$\mathbf{k_{e2}}$ [*mg/kg*]	216
$\mathbf{F_{cns}}$ [*mg/kg/min*]	1
$\mathbf{k_{e1}}$**_SGLTi**[min^{-1}]	0,015
$\mathbf{k_{e2}}$**_SGLTi** [*mg/kg*]	129,55

Table 4.17: New set of initial conditions

In addition, the value of Fcns was varied from 1 to 1.5 with an increment of 0.05, and the best result was achieved with an Fcns value of 1.15.

Taking this change into account, then, the final set of initial conditions is:

Parameter	**Value**
$\mathbf{V_g}$ [*dl/kg*]	1,20
$\mathbf{V_{mx}}$ [*mg/kg/min per pmol/l*]	0,11
$\mathbf{K_{m0}}$[mg/kg]	236,9
$\mathbf{k_2}$[min^{-1}]	0,08
$\mathbf{k_1}$ [min^{-1}]	0,05
$\mathbf{p_{2u}}$ [min^{-1}]	0,05
$\mathbf{k_{e1}}$ [min^{-1}]	0,008
$\mathbf{k_{e2}}$ [*mg/kg*]	216
$\mathbf{F_{cns}}$ [*mg/kg/min*]	1,15
$\mathbf{k_{e1}}$**_SGLTi**[min^{-1}]	0,015

k_{e2}_SGLTi [*mg/kg*]	129,55

Table 4.18: Final set of initial conditions

With these new considerations other tests were performed.

- Test n.17.

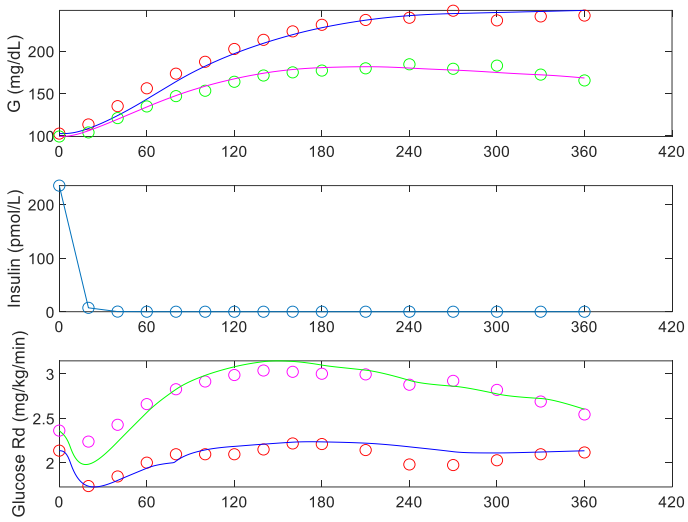

Figure 4.17: Result obtained with the simultaneous identification using Bayesian estimator preceded by 'fminsearch' optimization algorithm and a SD of 0,5 for V_g, of 0,009 for F_{cns}, of 0,001 for $k_{e2}$and k_{e2}_SGLT2i and of 0,1 for k_{e1} and k_{e1}_SGLT2i. Upper panel: plasma glucose concentration data of placebo group (open red circle) vs model prediction of placebo group (continuous blue line), plasma glucose concentration data of SGLT2i group (open pink circle) vs model prediction of SGLT2i group (continuous green line). Middle panel: simulated plasma insulin concentration (open blue circle and continuous blue line). Lower panel: glucose rate of disappearance data of placebo group (open red circle) vs model prediction of placebo group (continuous blue line) and glucose rate of disappearance data of SGLT2i group (open pink circle) vs model prediction of SGLT2i group (continuous green line).

Parameter	**Value**	**CV%**
$\mathbf{V_g}$ [*dl/kg*]	1,33	1,3
$\mathbf{V_{mx}}$ [*mg/kg/min per pmol/l*]	0,15	117
$\mathbf{K_{m0}}$[mg/kg]	207,3	0,04
$\mathbf{k_2}$[min^{-1}]	0,07	153,2
$\mathbf{k_1}$ [min^{-1}]	0,05	214,5
$\mathbf{p_{2u}}$ [min^{-1}]	0,09	344,35
$\mathbf{k_{e1}}$ [min^{-1}]	0,008	675,7
$\mathbf{k_{e2}}$ [*mg/kg*]	216	0,0005
$\mathbf{F_{cns}}$ [*mg/kg/min*]	1,151	0,78
$\mathbf{k_{e1}}$**_SGLTi**[min^{-1}]	0,02	256,7
$\mathbf{k_{e2}}$**_SGLTi** [*mg/kg*]	129,54	0,0007

Table 4.19: Values of parameters estimated in the simultaneous identification with the Batesian approach preceded by 'fminsearch' optimization algorithm and and a SD of 0,5 for V_g, of 0,001 for F_{cns}, of 0,001 for $k_{e2}$and k_{e2}_SGLT2i and of 0,1 for k_{e1} and k_{e1}_SGLT2i and their precision.

- Test n.18.

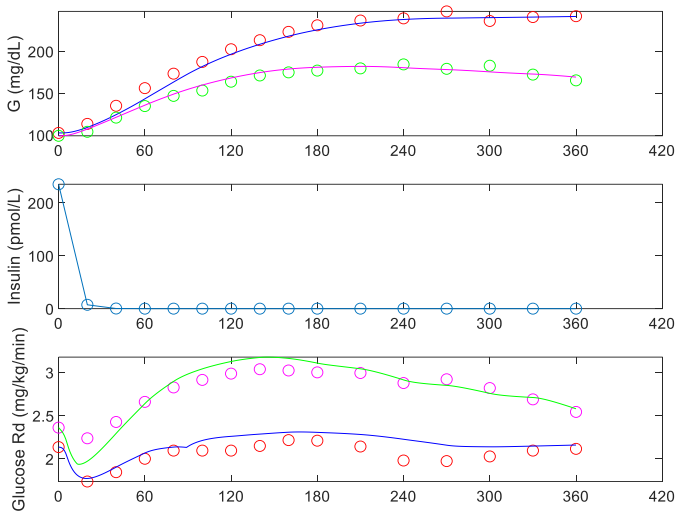

Figure 4.18: Result obtained with the simultaneous identification using Bayesian estimator preceded by 'fminsearch' optimization algorithm and a SD of 0,5 for V_g, of 0,009 for F_{cns}, of 0,01 for $k_{e2}$and k_{e2}_SGLT2i and of 0,1 for k_{e1} and k_{e1}_SGLT2i. Upper panel: plasma glucose concentration data of placebo group (open red circle) vs model prediction of placebo group (continuous blue line), plasma glucose concentration data of SGLT2i group (open pink circle) vs model prediction of SGLT2i group (continuous green line). Middle panel: simulated plasma insulin concentration (open blue circle and continuous blue line). Lower panel: glucose rate of disappearance data of placebo group (open red circle) vs model prediction of placebo group (continuous blue line) and glucose rate of disappearance data of SGLT2i group (open pink circle) vs model prediction of SGLT2i group (continuous green line).

Parameter	**Value**	**CV%**
$\mathbf{V_g}$ [*dl/kg*]	1,24	1,34
$\mathbf{V_{mx}}$ [*mg/kg/min per pmol/l*]	0,23	91,2
$\mathbf{K_{m0}}$[mg/kg]	209,4	0,04
$\mathbf{k_2}$[min^{-1}]	0,07	236,7
$\mathbf{k_1}$ [min^{-1}]	0,06	284,6
$\mathbf{p_{2u}}$ [min^{-1}]	0,08	370,7
$\mathbf{k_{e1}}$ [min^{-1}]	0,01	566,1
$\mathbf{k_{e2}}$ [*mg/kg*]	213,2	0,005
$\mathbf{F_{cns}}$ [*mg/kg/min*]	1,15	0,77
$\mathbf{k_{e1}}$**_SGLTi**[min^{-1}]	0,02	239,24
$\mathbf{k_{e2}}$**_SGLTi** [*mg/kg*]	127,5	0,008

Table 4.20: Values of parameters estimated in the simultaneous identification with the Batesian approach preceded by 'fminsearch' optimization algorithm and and a SD of 0,5 for V_g, of 0,001 for F_{cns}, of 0,01 for $k_{e2}$and k_{e2}_SGLT2i and of 0,1 for k_{e1} and k_{e1}_SGLT2i and their precision.

- Test n.19.

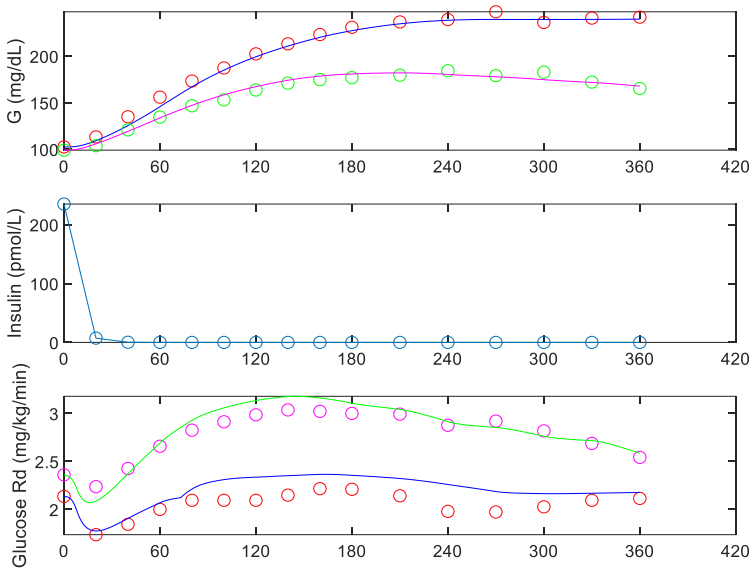

Figure 4.19: Result obtained with the simultaneous identification using Bayesian estimator preceded by 'fminsearch' optimization algorithm and a SD of 0,5 for V_g, of 0,009 for F_{cns}, of 0,05 for $k_{e2}$and k_{e2}_SGLT2i and of 0,1 for k_{e1} and k_{e1}_SGLT2i. Upper panel: plasma glucose concentration data of placebo group (open red circle) vs model prediction of placebo group (continuous blue line), plasma glucose concentration data of SGLT2i group (open pink circle) vs model prediction of SGLT2i group (continuous green line). Middle panel: simulated plasma insulin concentration (open blue circle and continuous blue line). Lower panel: glucose rate of disappearance data of placebo group (open red circle) vs model prediction of placebo group (continuous blue line) and glucose rate of disappearance data of SGLT2i group (open pink circle) vs model prediction of SGLT2i group (continuous green line).

Parameter	**Value**	**CV%**
$\mathbf{V_g}$ [*dl/kg*]	1,15	1,88
$\mathbf{V_{mx}}$ [*mg/kg/min per pmol/l*]	0,17	146,4
$\mathbf{K_{m0}}$[mg/kg]	226,89	0,04
$\mathbf{k_2}$[min^{-1}]	0,08	217,88
$\mathbf{k_1}$ [min^{-1}]	0,07	271,3
$\mathbf{p_{2u}}$ [min^{-1}]	0,07	418,15
$\mathbf{k_{e1}}$ [min^{-1}]	0,01	654,1
$\mathbf{k_{e2}}$ [*mg/kg*]	184,1	0,02
$\mathbf{F_{cns}}$ [*mg/kg/min*]	1,15	0,8
$\mathbf{k_{e1}}$**_SGLTi**[min^{-1}]	0,02	298,8
$\mathbf{k_{e2}}$**_SGLTi** [*mg/kg*]	104,05	0,04

Table 4.21: Values of parameters estimated in the simultaneous identification with the Batesian approach preceded by 'fminsearch' optimization algorithm and and a SD of 0,5 for V_g, of 0,001 for F_{cns}, of 0,05 for $k_{e2}$and k_{e2}_SGLT2i and of 0,1 for k_{e1} and k_{e1}_SGLT2i and their precision.

- Test n.20.

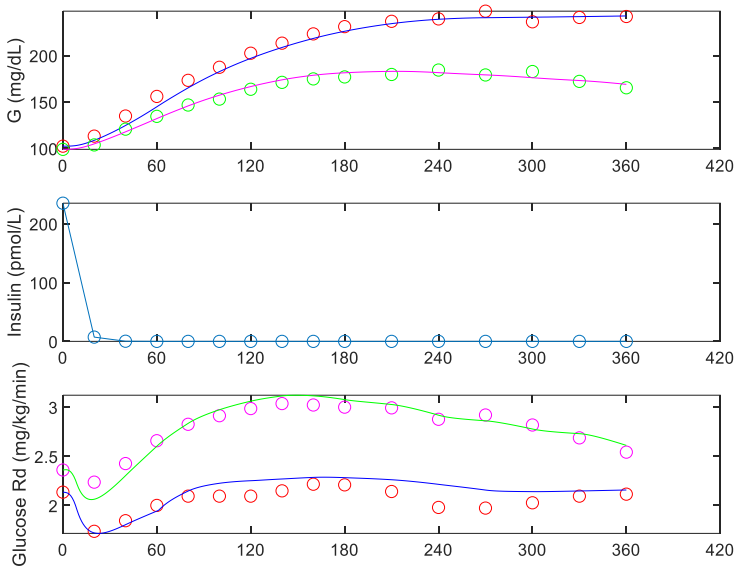

Figure 4.20: Result obtained with the simultaneous identification using Bayesian estimator preceded by 'fminsearch' optimization algorithm and a SD of 0,5 for V_g, of 0,009 for F_{cns}, of 0,1 for $k_{e2}$and k_{e2}_SGLT2i and of 0,001 for k_{e1} and k_{e1}_SGLT2i. Upper panel: plasma glucose concentration data of placebo group (open red circle) vs model prediction of placebo group (continuous blue line), plasma glucose concentration data of SGLT2i group (open pink circle) vs model prediction of SGLT2i group (continuous green line). Middle panel: simulated plasma insulin concentration (open blue circle and continuous blue line). Lower panel: glucose rate of disappearance data of placebo group (open red circle) vs model prediction of placebo group (continuous blue line) and glucose rate of disappearance data of SGLT2i group (open pink circle) vs model prediction of SGLT2i group (continuous green line).

Parameter	**Value**	**CV%**
$\mathbf{V_g}$ [*dl/kg*]	1,3	1,76
$\mathbf{V_{mx}}$ [*mg/kg/min per pmol/l*]	0,17	148,7
$\mathbf{K_{m0}}$[mg/kg]	216,2	0,04
$\mathbf{k_2}$[min^{-1}]	0,09	200,4
$\mathbf{k_1}$ [min^{-1}]	0,05	331,9
$\mathbf{p_{2u}}$ [min^{-1}]	0,06	451,9
$\mathbf{k_{e1}}$ [min^{-1}]	0,008	12,65
$\mathbf{k_{e2}}$ [*mg/kg*]	186,4	0,02
$\mathbf{F_{cns}}$ [*mg/kg/min*]	1,16	0,76
$\mathbf{k_{e1}}$**_SGLTi**[min^{-1}]	0,015	6,83
$\mathbf{k_{e2}}$**_SGLTi** [*mg/kg*]	108,34	0,04

Table 4.22: Values of parameters estimated in the simultaneous identification with the Batesian approach preceded by 'fminsearch' optimization algorithm and and a SD of 0,5 for V_g, of 0,001 for F_{cns}, of 0,1 for $k_{e2}$and k_{e2}_SGLT2i and of 0,001 for k_{e1} and k_{e1}_SGLT2i and their precision.

4.2 Model selection

As already stated in chapter 3, the selection of an optimal model involves the evaluation of several aspects. The first step involves the analysis of the fit, the evaluation of which involves a visual inspection of the plots comparing data and model prediction (figure from 4.1 to 4.20). The model should be able to well fit the data without under- or over-estimating them.

The second aspect to be taken into account is the value of CVs. In fact, they represent the precision of an estimate and are expressed as a percentage. The higher the CV value, the more inaccurate the estimate will be. Each test performed includes both the value of the estimated parameters and their CV (table from 4.1 to 4.22). Unfortunately, in our model we had to deal with high CV values for some parameters, but this may be due to the accuracy of the measurements taken and used in the model as input.

The last parameter to consider is the so-called Akaike information criterion, which is based on the principle of parsimony.

The table below provides a summary of the above-mentioned aspects, with their respective values, involved in the choice of an optimal model.

Test N°	**Test n.6**	**Test n.8**	**Test n.12**	**Test n.14**	**Test n.17**	**Test n.18**	**Test n.19**	**Test n.20**
N°. of param	9	9	9	9	8	10	10	8
Median CV%	485,29%	433,85%	346,03%	246,37%	160,33%	162,79%	182,68%	105,01%
RMSE	5,37	6,56	60,13	58,73	5,31	5,35	4,49	4,81
AIC	23,37	24,56	78,13	76,73	21,31	25,35	24,49	24,91

Table 4.23: Model comparison

Having compared the different models and having reported the relative AIC, we can conclude that the best model can be considered the number 17, which is the one with the lowest AIC coefficient.

4.3 Role of SGLT2i in glucose control

SGLT2 are responsible for reabsorption of more than 90% of glucose in the proximal tubule of the kidney. In diabetic patients the expression of SGLT2 paradoxically increases, so that the renal threshold for glycosuria increases and therefore excess glucose is retained.

Sodium-glucose cotransporter type 2 inhibitors are selective and reversible inhibitors of renal SGLT2. Their action is totally independent from insulin. They affect the kidneys directly by lowering the threshold for renal excretion of glucose and increasing its urinary excretion, they are able to bring about a significant reduction in glycaemia (both fasting and post-prandial) and, consequently, a lower glucotoxicity.

The model constructed contains a parameter strictly responsible for the renal absorption threshold. In fact, k_{e2} divided by V_g gives us exactly the value of the renal threshold in *mg/dl*.

As reported in the literature, under physiological conditions the renal threshold limit value should be around 180mg/dl. When blood glucose exceeds 180/200 mg/dl the patient starts to have glycosuria, which indicates the presence of glucose in the urine. Glycosuria is therefore an expression of hyperglycaemia related to an excessive sugar intake or to a disease that alters its regulation, as type 1 diabetes.

Since SGLT2i reduce the reabsorption of glucose and increase its excretion through the urine, the renal threshold will therefore be lower than 180 mg/dl, but unfortunately by how much is not yet known.

Model number 17 returns a lower k_{e2}/V_g value in the SGLT2i group confirming that the drug acts on the renal threshold.

Chapter 5
CONCLUSION

The purpose of this thesis is to build up a comprehensive and exhaustive model that describes the effects of Sodium-Glucose Co-Transporter2 (SGLT2) inhibitors in patients with type 1 diabetes.

The database collected in the paper by Herring et al. [13] was used to test the model. This study was a double-blind, placebo-controlled crossover study with a 4-week washout period, and it was performed in 12 people with type 1 diabetes using insulin pump therapy.

Initially, an attempt was made to see if the UVA/Padova simulator was able to reproduce data observed by Herring et al. [13] but, unfortunately, that was not the case. We thus switched to another approach. Only the glucose subsystem was considered and the model was identified using the simulated insulin (I) and rate of appearance (Ra) profiles as forcing function.

Several tests have been conducted, using different approaches and trying to adopt strategies aimed at improving results. Some of them reported satisfactory results, while some others did not, due to poor ability to fit the data or high CVs or non-physiological values of the estimated parameters.

As already mentioned in paragraph 4.2, the best results were achieved with Test n. 17, since it provided the best compromise in terms of data fitting, CVs and AIC. Nevertheless, the experiments carried out are not to be considered definitive, nor are they the only ones possible; indeed, the possibility of carrying out further experiments would help to improve the results.

From a physiological perspective, what we expected was that after using SGLT2i the renal threshold would decrease, as these inhibitors have the primary effect of decreasing this threshold. This is exactly what happened in most of the tests carried out. However, the values obtained for the renal threshold with or without SGLT2 inhibitors were lower than expected. This can be due to some limitations of our approach.

A fundamental aspect that must be taken into account in the development of this thesis is the fact that we did not have individual data available but only averages of data obtained

from the clinical trial carried out by Herring et al. [13]. Furthermore, we had no information on the insulin time course. In fact, we had to model insulin concentration knowing only that subcutaneous infusion was withdrawn at 6am in the morning of the metabolic study and replaced with a soluble variable insulin infusion to maintain a glucose concentration at 5 mmol/L; then, such infusion was stopped to allow glucose rising. Finally, both Rd and Ra time courses were not directly measured, but estimated from a tracer experiment, which provided virtually model independent profiles. However, the goodness of such estimates is difficult to assess, since tracer-to-tracee ratio are not reported in the original paper [13].

In conclusion, through our modeling approach, we were able to assess the impact of these drugs on glycemic dynamics and their effects on renal excretion. As already reported in the literature, SGLT2i may serve as a valuable addition to the standard treatment of type 1 diabetes, aiding in reducing blood glucose levels and thus enhancing disease management. However, it is important to recognize that the efficacy and safety of these drugs may vary among patients, and further research is needed to fully understand their long-term impact on the health and well-being of diabetic individuals. SGLT2i have been shown to stimulate the release of glucagon and the reabsorption of ketones in the renal tubules, thus increasing the concentrations of ketone bodies, which leads patients with type 1 diabetes into a life-threatening situation of ketoacidosis, which can lead to diabetic coma.

Ultimately, SGLT2 inhibitors represent a promising perspective in type 1 diabetes management and warrant further investigation, such as long-term clinical trial, additional studies, or new customised models, to maximize benefits and mitigate potential risks.

BIBLIOGRAPHY

[1] EV Moltchanova, N Schreier, N Lammi, and M Karvonen. “Seasonal variation of diagnosis of Type 1 diabetes mellitus in children world-wide”. In: *Diabetic Medicine* 26.7 (2009), pp. 673–678.

[2] Dana Dabelea. “The accelerating epidemic of childhood diabetes”. In: *The Lancet* 373.9680 (2009), pp. 1999–2000.

[3] Mark A Atkinson, George S Eisenbarth, and Aaron W Michels. “Type 1 diabetes”. In: *The Lancet* 383.9911 (2014), pp. 69–82.

[4] American Diabetes Association. “Diagnosis and classification of diabetes mellitus”. In: *Diabetes care* 33.Supplement 1 (2010), S62–S69.

[5] The International Expert Committee. “International Expert Committee report on the role of the A1C assay in the diagnosis of diabetes”.In: *Diabetes care* 32.7 (2009), p. 1327.

[6] Anastasia Katsarou et al. “Type 1 diabetes mellitus”. In: *Nature reviews Disease primers* 3.1 (2017), pp. 1–17.

[7] Varun Pathak, Nupur Madhur Pathak, Christina L O’Neill, Jasenka Guduric-Fuchs, and Reinhold J Medina. “Therapies for type 1 diabetes: current scenario and future perspectives”. In: *Clinical Medicine Insights: Endocrinology and Diabetes* 12 (2019), p. 1179551419844521.

[8] Hsin-Chieh Yeh et al. “Comparative effectiveness and safety of methods of insulin delivery and glucose monitoring for diabetes mellitus: a systematic review and meta-analysis”. In: *Annals of internal medicine* 157.5 (2012), pp. 336–347.

[9] Bernt Johan von Scholten, Frederik F Kreiner, Stephen CL Gough, and Matthias von Herrath. “Current and future therapies for type 1 diabetes”. In: *Diabetologia* 64 (2021), pp. 1037–1048.

[10] Ele Ferrannini. “Sodium-glucose co-transporters and their inhibition: clinical physiology”. In: *Cell metabolism* 26.1 (2017), pp. 27–38.

[11] Chiara Dalla Man, Robert A Rizza, and Claudio Cobelli. “Meal simulation model of the glucose-insulin system”. In: *IEEE Transactions on biomedical engineering* 54.10 (2007), pp. 1740–1749.

[12] Chiara Dalla Man, Francesco Micheletto, Dayu Lv, Marc Breton, Boris Kovatchev, and Claudio Cobelli. “The UVA/PADOVA type 1 diabetes simulator: new features”. In: *Journal of diabetes science and technology* 8.1 (2014), pp. 26–34.

[13] Roselle A Herring et al. “Metabolic effects of an SGLT2 inhibitor (dapagliflozin) during a period of acute insulin withdrawal and development of ketoacidosis in people with type 1 diabetes”. In: *Diabetes Care* 43.9 (2020), pp. 2128–2136.

[14] Claudio Cobelli and Ewart Carson. “8 - Parametric models—The estimation problem”. In: *Introduction to Modeling in Physiology and Medicine*. Ed. by Claudio Cobelli and Ewart Carson. Biomedical Engineering. Burlington: Academic Press, 2008, pp. 195–234. isbn: 978-0-12-160240-6. doi: https : / / doi . org / 10 . 1016 / B978 - 012160240 -6. 50009 - 0. url: https: / / www . sciencedirect . com / science /article/pii/B9780121602406500090.