Diabetes Mellitus Glucose Prediction by Linear and Bayesian Ensemble Modeling

Ståhl, Fredrik

2012

Document Version:
Publisher's PDF, also known as Version of record

Diabetes Mellitus Glucose Prediction by Linear and Bayesian Ensemble Modeling

Fredrik Ståhl

Department of Automatic Control
Lund University
Lund, December 2012

Department of Automatic Control
Lund University
Box 118
SE-221 00 LUND
Sweden

ISSN 0280–5316
ISRN LUTFD2/TFRT--3255--SE

Printed in Sweden,
Lund University, Lund 2012

Preface

In 2002, I was diagnosed with diabetes type 1. Being an engineer with a control and systems oriented curriculum I realized that this was a control problem, however, set in an unfamiliar context, and I decided to put my newly earned skills to the test by trying to identify my own data in my Master Thesis. By then, frequent data sampling was less developed and not readily available, making identification an even harder task than it is today, which was a bit discouraging.

Six years later, Rolf invited me to join the department and the newly started EU FP7 IP DIAdvisor project. During the years since graduation, major leaps in sensor technology had been achieved, changing the playing field dramatically– well-sampled data could now be easily attained. Four years later, the project has ended, adding new knowledge to the rapidly growing body of knowledge of diabetes glucose metabolism, and bringing new hope of technical solutions to support the management of this often difficult medical condition. Lots of research and development is pursued both in academia and in the industry. Since 1999, two scientific journals on diabetes technology have been established, and both the U.S. and E.U. are putting major funding into diabetes technology oriented cross-disciplinary research projects. Being affected with type 1 diabetes I embrace this development, and I feel extremely fortunate to get a chance of being part of this endeavour, which in the end may make both mine and millions of other people's daily lives easier to manage.

Lund, December 3rd 2012

Fredrik Ståhl

Abstract

Diabetes Mellitus is a chronic disease of impaired blood glucose control due to degraded or absent bodily-specific insulin production, or utilization. To the affected, this in many cases implies relying on insulin injections and blood glucose measurements, in order to keep the blood glucose level within acceptable limits. Risks of developing short- and long-term complications, due to both too high and too low blood glucose concentrations are severalfold, and, generally, the glucose dynamics are not easy too fully comprehend for the affected individual—resulting in poor glucose control. To reduce the burden this implies to the patient and society, in terms of physiological and monetary costs, different technical solutions, based on closed or semi-closed loop blood glucose control, have been suggested.

To this end, this thesis investigates simplified linear and merged models of glucose dynamics for the purpose of short-term prediction, developed within the EU FP7 DIAdvisor project. These models could, e.g., be used, in a decision support system, to alert the user of future low and high glucose levels, and, when implemented in a control framework, to suggest proactive actions.

The simplified models were evaluated on 47 patient data records from the first DIAdvisor trial. Qualitatively physiological correct responses were imposed, and model-based prediction, up to two hours ahead, and specifically for low blood glucose detection, was evaluated. The glucose raising, and lowering effect of meals and insulin were estimated, together with the clinically relevant carbohydrate-to-insulin ratio. The model was further expanded to include the blood-to-interstitial lag, and tested for one patient data set. Finally, a novel algorithm for merging of multiple prediction models was developed and validated on both artificial data and 12 datasets from the second DIAdvisor trial.

Acknowledgements

First of all, I would like to express my gratitude towards my supervisor, professor Rolf Johansson, for introducing me to graduate studies and the DIAdvisor project, which he initiated, and for his encouragement over the last four years. Financial support was provided by the E.U. through the DIAdvisor project and the Swedish Research Council through the LCCC and ELLIT research excellence centers.

I am also fortunate to work at such a thriving and interesting workplace, where the theoretical research is matched by a wide range of applied science. The width and intensity of the scientific activities, in the form of workshops, focus periods and weekly seminars, are overwhelming for a part-time graduate student as myself, but create a very inspiring working atmosphere. If you want to, and have the time, something new can be learnt everyday! I would like to thank the senior research personal and the administrative personal for making this possible.

The DIAdvisor project has been a very exciting project, giving me the opportunity to work in a cross-disciplinary and cross-cultural context, with a subject matter of high personal interest. It has been a pleasure collaborating with our European colleagues and enjoying the hospitality of these partners. During this project I have also been fortunate to have been working with my highly praised colleagues Marzia Cescon and Meike Stemmann, with whom I have enjoyed many rewarding moments. Over these years, I have also had many fruitful discussions with the clinical researcher Mona Landin-Olsson, and I look forward to future research projects together.

I am grateful to all the people who have helped me with proof-reading of this manuscript; with special thanks to Rolf Johansson, Per Hagander, Mona

Landin-Olsson, Annika Ståhl-Delbanco, Marzia Cescon, Meike Stemmann and Jonas Dürago, and to Leif Andersson for providing his excellent LaTeX skills. The comments given by the reviewer have spurred interesting discussions and insights, which have improved the manuscript significantly.

Most of all, I thank my family for their support, and especially my wonderful wife Åsa for making it all possible.

Contents

1

Introduction

1.1 Motivation

Diabetes Mellitus is a chronic metabolic disease where the affected patients have disturbed glucose regulation, which, if left untreated results in elevated blood glucose levels. The disease is divided into two categories; type 1 diabetes and type 2 diabetes. In type 1 diabetes, the pancreas no longer produces insulin due to an auto-immune destruction of the pancreatic insulin producing β-cells. Type 2 diabetes is a common diagnosis for several different underlying causes to deteriorating glucose control, such as reduced insulin sensitivity and prolonged or deteriorated pancreatic insulin response. There is a strong genetic component to the risk of both type 1 and type 2. The etiology behind the sudden auto-immune attack leading to type 1 is still obscured, but some evidence point to that viral infections may play a key role in the triggering mechanism [Christen *et al.*, 2012]. Type 2 diabetes typically evolves over a number of years before diagnosis, and is strongly connected to sedentary life-style and overweight, but the incidence also increases with age.

The incidence of both types of diabetes, especially type 2, increases at an alarming rate on a global scale. In year 2000, WHO estimated 171 million to be affected [Wild *et al.*, 2004], and in 2011 the International Diabetes Federation (IDF) estimated the number to 366 million (of which 183 million are undiagnosed) [International Diabetes Federation, 2012], already exceeding the 2030 forecast from WHO in 2000. By 2030, the expected number exceeds 500 million in IDF's recent analysis [International Diabetes Federation, 2012]. In Sweden,

Table 1.1 Comparison of the cost structure for an average patient and a patient with both micro and macrovascular complications. Costs in SEK (1998) [Henriksson *et al.*, 2000].

	Average Patient	Patient with both micro- and macro-vascular complications
Hospitalization	10 599	29 555
Ambulatory Care	7 719	11 053
Drugs	6 665	9 520
Total	24 983	50 128

the total number is about 365.000, of which about 40.000 are type 1 patients [The Swedish National Board of Health and Welfare, 2012a]. In general, about 10% of the patients are of type 1. Along with the increasing numbers of affected, the total costs increase dramatically. In Sweden, the total direct cost of diabetes treatment was estimated to 7 billion SEK in 1998, considering only type 2 diabetes patients [Henriksson *et al.*, 2000], and globally, figures of 465 billion USD has been stated, amounting to 11% of the total healthcare expenditure for adults (20-79 years old) [International Diabetes Federation, 2012]. The main cost drivers are costs related to treatment of acute and late complications resulting from poor glycemic control [Henriksson *et al.*, 2000], see Table 1.1, and the indirect costs, related to loss of productivity resulting from mortality and disability from these complications—in Sweden estimated to 5.4 billion SEK [Bolin *et al.*, 2009].

These complications spring from either too low glucose values, hypoglycemia, which may result in acute seizure, coma and death, and too high blood glucose concentrations, hyperglycemia, which, over a longer time period may lead to impairment of the inner organs due to micro and macro cardiovascular implications, and result in, e.g., renal failure, amputation and blindness. Studies show that, in general, almost half of the diabetic population in Sweden have a mean glucose value, measured as HbA1c, above the guideline value, implying a significantly increased risk of the afore-mentioned long-term complications [The Swedish National Board of Health and Welfare, 2012b], and there is good reason to believe that these numbers translate globally. Thus, means to improve the metabolic control for these patients are seminal to cut back the dismaying rate of growth of monetary and physiological costs of this disease, and to lighten the

heavy burden this implies on the healthcare programmes and institutions. In order to contribute to this end, this thesis investigates individualized mathematical models of glucose dynamics—describing the relationship between carbohydrate and insulin ingestion and the blood glucose evolution—aiming for use in an artificial pancreas or a decision support system. These models have been developed within the European FP7 IP research project DIAdvisor [DIAdvisor, 2012].

Even though these problems affect both type 1 and type 2 patients, focus of this thesis will be on type 1 diabetes and insulin-treated type 2 patients, hereafter lumped together as a common cohort, insulin-dependent *diabetes mellitus* (IDDM). The reason for this is the following. In order to treat type 1 patients, external insulin must be administrated. For non-insulin dependant (NIDDM) type 2 diabetes, insulin sensitivity promoting oral agent, together with changes in lifestyle, may suffice to improve the metabolic control. However, for many type 2 patients, insulin is required after a few years. Thus, there is a fundamental and significant difference in treatment between the IDDM and NIDDM groups, which also implies differences in time perspectives of the glucose dynamics. For IDDM, the appropriate amount of insulin to administer is often hard to estimate and steep changes of the glucose level may suddenly arise. An undesirable, or even dangerous, situation may thereby quickly arise, calling for new treatment decisions. For NIDDM, the variations are slower, the number of decision points over the day are less and different, and the acute risks less pronounced. The need, prerequisites and type of decision support or automatic control is therefore very different between these two groups. For the IDDM, continuous support to optimize the insulin regime may have a profound affect on the possibility to maintain normal glucose levels, whereas management of diet, exercise and other lifestyle-related changes, and long-term follow-up thereof, is the core to improved NIDDM type 2 metabolic control.

1.2 Publications

This thesis is based on the following publications:

Ståhl F. and Johansson R., ”Diabetes Mellitus Modelling and Short-Term Prediction Based on Blood Glucose Measurements”, In *Mathematical Biosciences*, 217, pp. 101-117, January 2009.

Ståhl F., Johansson R. and Renard E.,”Models of Diabetic Glucose Dynamics:

Challenges of Identifiability and Physiological Correctness", Submitted to *IEEE Transactions on Biomedical Engineering*.

Ståhl F., Johansson R. and Renard E.,"Investigation of the relationship between elevated levels of insulin antibodies and prolonged insulin action", Accepted for presentation at *6th International Conference on Advanced Technologies and Treatments for Diabetes*, Paris, France, Feb 27-March 2, 2013.

Ståhl F. and Johansson R., "Observer Based Plasma Glucose Prediction in Type 1 Diabetes", In *Proc. 3rd IEEE Conf. on Systems and Control*, pp. 1620-1625, Yokohama, Japan, 8-10 Sept, 2010.

Ståhl F, Johansson R. and Renard E., "Bayesian Combination of Multiple Plasma Glucose Predictors", In *Proc. 34th Annual International IEEE EMBS Conference (EMBC 2012)*, pp. 2839-2844, San Diego, CA, U.S, Aug 28-Sept 1 2012.

Ståhl F. and Johansson R.,"Receding Horizon Prediction by Bayesian Combination of Multiple Predictors", Accepted for presentation at *51st Annual IEEE Conf. on Decision and Control (CDC2012)*, Maui, Hawaii, U.S, Dec. 10-13, 2012.

Other publications:

Ståhl F. and Johansson R., "Short-Term Diabetes Blood Glucose Prediction Based On Blood Glucose Measurements", In *Proc. 30th Annual International IEEE EMBS Conference (EMBC2008)*, pp. 291-294, Vancouver, British Columbia, Canada, August 20-24, 2008.

Ståhl, F., Cescon M., Johansson R., and Renard E., "Infinite horizon prediction of postprandial breakfast plasma glucose excursion." In *Proc. of the 9th Annual Diabetes Technology Meeting (DTM2009)*, p. A163, San Francisco, CA, U.S, Nov. 5-7, 2009.

Cescon, M., Ståhl F., Landin-Olsson M., and Johansson R., Subspace-based model identification of diabetic blood glucose dynamics. In *Proc. of the 15th Symposium on System Identification (SYSID2009)*, pp. 233-238. Saint-Malo, France, July 6-8, 2009.

Ståhl F., Johansson R., Renard E., "Post-Prandial Plasma Glucose Prediction in Type I Diabetes Based on Impulse Response Models", In *Proc. 32nd Annual*

International IEEE EMBS Conference (EMBC 2010), pp. 1324-1327, Buenos Aires, Argentine, Aug 31-Sep. 4, 2010.

Ståhl, F., Johansson R., and Renard E., "Can Blood Glucose Drops During Exercise Be Predicted From Heart Rate Data?", In *Proc. of the 11th Annual Diabetes Technology Meeting (DTM2011)*, p. A175, San Francisco, CA, U.S, Oct. 27-29, 2011.

1.3 Outline and Contributions

This thesis is organized as follows with the following contributions:

Chapter 2 provides an introduction to diabetes, a brief description of the metabolic physiology and a presentation of the DIAdvisor project.

Chapter 3 presents the DIAdvisor data and an analysis of these in terms of some specific data characteristics, typical of diabetic glucose data.

Chapter 4 describes simplified linear models with qualitatively correct responses to insulin and carbohydrate digestion. The models are validated for short-term glucose prediction, including hypoglycemic detection, on 47 patient data sets from the first trial of the DIAdvisor project, and estimates of the carbohydrate-to-insulin ratio are given.

Chapter 5 presents an augmented model including the dynamics related to the glucose measurements. The concept is evaluated for reduced lagging of the short-term prediction on one patient data set from the first trial of the DIAdvisor project.

Chapter 6 introduces a novel algorithm for ensemble prediction, using several models derived for short-term glucose prediction. The suggested method is validated on simulated data, as well as 12 patient data sets from the second DIAdvisor trial.

Chapter 7, finally, concludes the thesis and directions for future research are outlined.

2

Background

2.1 Diabetes Type 1 and the Glucoregulatory System

Diabetes type 1 is, as previously mentioned, a chronic disease where the β-cells of the pancreas have stopped to produce insulin. This is in most cases due to an auto-immune attack, but may in rare cases also be caused by sustained injuries from accidents or pancreatic cancer. In order to understand the disease, a brief overview of the glucoregulatory system is presented, see, e.g., [Nussey and Whitehead, 2001] for a more extensive review.

The Glucoregulatory System

The glucoregulatory system is concerned with glucose metabolism and the insulin/glucose mechanisms needed to maintain normoglycemia. Fig. 2.1 presents a simplified overview of the flow of glucose and insulin between the most important organs relevant for this system. Below, a short description of these organs and their role in the so-called absorptive state and the post-absorptive state, the two parts that make up the metabolic cycle, is given. A brief description of insulin absorption from insulin injections will also be presented. Emphasis will be put on the digestive system and insulin absorption from injections.

The absorptive state is the time following a meal during which the ingested carbohydrates are digested and absorbed. During this period, excess glucose is absorbed and stored for later use. The post-absorptive state is the time after a meal when the gastro-intestinal tract is empty and energy has to be provided by the body's own storages.

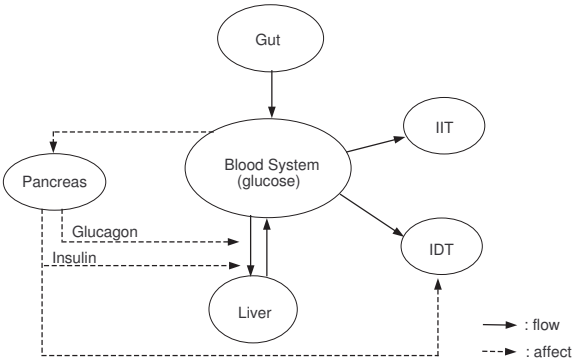

Figure 2.1 Overview of the glucoregulatory system describing the relationship between the flux from the gut into blood system and the interaction with the insulin-dependent tissue (IDT), the insulin-independent tissue (IIT), the pancreas and the liver.

During the absorptive stage, glucose is converted and stored as the polysaccharide glycogen, mainly in the liver, but also to some extent directly in the muscle cells. This process is stimulated by insulin. During the post-absorptive stage, the liver glycogen storage is broken down to glucose and released into the blood stream, providing energy for the body cells. This process is stimulated by glucagon and inhibited by insulin. Apart from converting glycogen to glucose, new glucose can be formed from protein and fat by gluconeogenesis. The metabolism of consumed alcohol inhibits this process [Siler *et al.*, 1998], which may result in severe hypoglycemia in IDDM patients [Turner *et al.*, 2001].

In the pancreas, two important hormones relevant to the glucoregulatory system are synthesized, namely insulin and glucagon. Insulin release is mainly stimulated by elevated blood glucose concentration. Therefore, substantial amounts are released in the absorptive stage, when the glucose level is raised due to the absorption from the gut. Glugacon, which has the opposite effect on the hepatic balance, is accordingly released when blood glucose concentration falls. These two hormones are thus in a feedback arrangement with the blood glucose concentration—controlling the glucose metabolism. In type 1 patients the insulin feedback is not functional. Another hormone group of importance during the absorbtive stage is the incretine gut hormone group. Incretine is secreted during meal uptake and stimulates pancreatic insulin release and inhibits the glucose flux from the gut into the blood stream. Impaired incretine function is believed to play an important role to the reduced and impaired insulin response

of type 2 patients [Nauck *et al.*, 2004].

Insulin-dependent tissue (IDT) is dependent on insulin to utilize glucose. This mechanism is discussed in the insulin section below. A significant portion of the insulin-dependent tissue is made up of skeletal muscles. In the absorptive state, skeletal muscle cells not only consume the glucose directly, but also convert some to glycogen, providing an energy storage for later use in a local depot.

Insulin independent tissue (IIT), such as the brain and the central nervous system, do not need insulin to utilize glucose.

Insulin

Insulin is the main hormone controlling the glucose metabolism. It is a protein consisting of three peptide parts; an A-, B- and C-chain. In healthy subjects it is produced in the β-cells in the pancreas, whereas IDDM patients depend mostly on injections of artificially produced insulin analogs. Three categories of different types of therapeutic insulins exist; rapid-, intermediate and long-acting insulins. The long-acting insulins are used to cover the basal metabolism, i.e., mainly to support the insulin-dependent tissue in the post-absorptive state. The most recent insulin types of this category, detemir [Levemir^{TM}, 2012] and glargine [Lantus^{TM}, 2012] type have almost flat pharmacokinetic profiles. Rapid-acting insulins, such as lispro [Humalog^{TM}, 2012], aspart [Novolog^{TM}, 2012] and glulisine [Apidra^{TM}, 2012] are designed to handle the glucose flux following a meal in the absoptive state. Therefore, these insulins have a short pharmacokinetic profile with a distinct peak after about 60 minutes. Intermediate-acting insulin are a mix of both, and are often used to support in cases when some insulin production is still left, i.e., insulin-dependent type 2 patients or the so-called latent auto-immune diabetes (LADA) patients [Landin-Olsson, 2002].

Insulin is normally injected in the subcutaneous tissue of the torso or legs. Rapid-acting insulin is injected in the abdominal fat layer, whereas long-lasting insulin is usually taken in the upper side of the thigh. From these depots the insulin is transferred to the blood system via the capillaries. The absorption rate depends on a series of factors. One contributing factor is the capillary density. A higher density results in a greater diffusion area between the depots and the capillaries. The abdominal region has the highest capillary density and the thigh the lowest [Home, 1997]. This explains why rapid-acting insulin is preferably infused in the abdominal fat layer and long-lasting in the thigh.

The size of the insulin molecules is a dominant rate limiter. Large molecules will have difficulties passing through the capillary pores. The structure of the

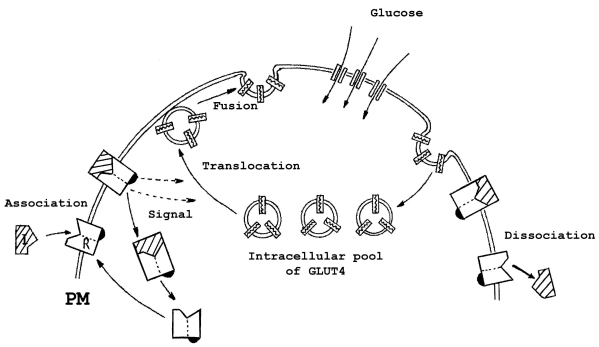

Figure 2.2 Insulin receptor and glucose transporter cycle. Reproduced from [Sasaki, 2002].

insulin molecules are either monomer, dimer or hexamer. Insulin will spontaneously form hexamers if the concentration is sufficiently high. This so-called self-association can be catalyzed by zinc ions. Therefore, zinc is added to the insulin solution in slow-acting insulins, thereby considerably reducing the absorption rate [Home, 1997]. In the rapid-acting insulins, the insulin molecules are mainly monomeric or dimeric. They have been modified so that hexamer formation is completely avoided [Shoelson and Halban, 1994], and are also called monomeric insulins. Another major factor affecting the absorption rate is the size of the injection dose. A large dose reduces the ratio between the absorption area and the depot volume, thus reducing the absorption. A number of studies have been undertaken, all indicating a linear relationship between insulin dose and absorption half-time [Hildebrandt *et al.,* 1984] and [Plank *et al.,* 2005]. These studies have been performed using slow-acting or intermediate-acting insulins. However, studies indicate that the linear relationship is not valid for monomeric insulin [Brange and Vølund, 1999]. Finally, blood flow and temperature of the injected site have a significant contribution to absorption rate. Raised temperature enhances the disassociation of hexameric insulin and accelerates insulin diffusion, and increased blood flow raises absorption rate. Thus, exercise plays a key role for absorption, since it raises both body temperature and blood flow. After the absorption from the depots, the insulin is circulated in the blood system and finally interacts with a insulin receptor at the cell surface.

The insulin receptors are so-called tetramers, consisting of two α- and two β-subunits. The α-subunits are entirely extracellular and serve as a binding site

for the insulin molecule. When the insulin has attached to the α-subunits, a signal process is initiated via the β-subunits, resulting in increased glucose transporter activity. The glucose transporters facilitate glucose cell membrane crossing, thereby reducing blood glucose concentration. The receptor/transporter cycle can be seen in Fig. 2.2. There are different types of glucose transporters and, so far, five different types have been found [White and Kahn, 1994]. Not all of these types require insulin to become active. Therefore, the glucose utilization is divided into insulin-dependent and insulin-independent utilization. It is a well-known fact that exercise enhances insulin sensitivity and is therefore one part of common type 2 therapy. However, what actually causes the increased insulin sensitivity is still not well understood. Studies indicate that the GLUT4 transporter activity is stimulated, resulting in increased insulin-dependent glucose utilization [Kahn, 1997].

Treatment

The most common therapy for IDDM patients is the multi-dose injection (MDI) basal-bolus regime. The patients use insulin pens, or perhaps the new Swedish mini pen—the DailyDose [Daily Dose, 2012]—to administer basal insulin, once or sometimes twice a day, and rapid-acting insulin for each meals as well as for additional corrections. An alternative therapy is to use an insulin pump, which, loaded with rapid-acting insulin, provides a continuous infusion, corresponding to the basal need and bolus doses accordingly.

Doses are based on heuristic rules derived from the patient's understanding of his/her metabolism, assessment of current glucose level from glucose meters and expected future evolution and estimates of carbohydrate content in digested meals. One common measure used in this regard is the carbohydrate-to-insulin ratio, which is an estimate of how many insulin units to administer to match the amount of digested carbohydrates. To avoid acute and long-term complications, the goal is to maintain normoglycemia (blood glucose G between 70-180 mg/dl) as far as possible, and especially to avoid insulin-induced hypoglycemia ($G < 70$ mg/dl) altogether, and to minimize time spent in hyperglycemia ($G > 180$ mg/dl). An extensively used evaluation criterion of the outcome is the glycosylated hemoglobin (HbA1c) blood measure, which provides an assessment of average blood glucose level over a 8-12 week period [Hanas and John, 2010].

Current research is focused on improving the therapy along two main directions; closed-loop control and semi-closed loop control by means of decision support systems.

Closed-loop control, or as it is often referred to in the diabetes context—the artificial pancreas—is most often suggested in the form of a regulator controlling an insulin pump by glucose sensor feedback. The earliest closed-loop system in this sense dates back to the ’60s and ’70s and the first commercial closed-loop system, the bed-side Biostator system, was introduced in 1977, relying on venous insulin infusion and glucose measurement. Today, the prerequisites have changed dramatically with major improvements in pump and sensor technology, and both academic researchers and biotechnology companies pursue the closed-loop control using primarily the subcutaneous route. The first step to implement an autonomous function in a commercial outpatient system has been made in the MedTronic Veo pump, which automatically suspends for two hours when a predefined hypoglycemic threshold is passed (this feature is so far only available in the E.U.) [MedTronic, 2012]. Reviews of current and historical development and of the challenges ahead can be found in, e.g., [Cobelli *et al.*, 2011] and [Bequette, 2012].

Closed-loop control is in many aspects a promising technology, but especially two major concerns need to be considered when evaluating the prospects of this technology to resolve glucose control for a larger part of the IDDM population. Firstly, using a closed autonomous system calls for a robust safe design. This aspect needs to be considered all throughout the system design, and identified hurdles, concerning, e.g., sensor accuracy and reliability, modeling and parameter estimation errors, disturbance detection and rejection and programming and software errors, still remain to be resolved. The second aspect is the cost aspect, as such a system relies on many expensive components. Unless prices are forcefully reduced, it is unlikely that an artificial pancreas system will be the default therapy for a majority of the IDDM population in the near future. An alternative technology is to provide the patient with decision support. The advantages of such an approach in comparison to the artificial pancreas, is that it is more flexible in underlying therapy format, as it is not locked to the pump technology, no injections are made automatically—providing an opportunity to detect wrongful and potentially dangerous actions—and that the total cost is lower, implying possible better cost effectiveness. On the other hand, the dependency on user interaction makes it more vulnerable in many aspects. Unless user confidence to the system is achieved, poor compliance to the suggested decisions may prove the system useless. Furthermore, in situations where the user is unable to respond, no action can be taken. Also, the potential risk reduction, associated with capturing dangerous actions, relies on an independent basic insight to the glucose dynamics of the user, and a sound non-authoritarian attitude

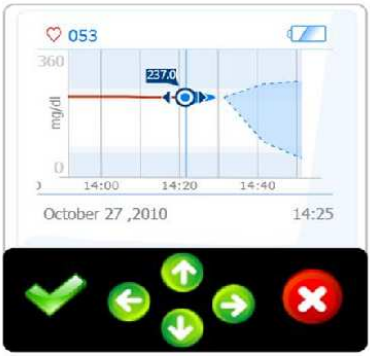

Figure 2.3 User Interface of the DIAdvisor system patient application implemented on the UMPC. On the screen, the user can follow the present and recent glucose values together with a projected future trajectory within specified uncertainty limits. Other vital signs, such as heart rate (see upper left screen corner), may also be possible to follow. User inputs, regarding, e.g., insulin and meal intake, are provided by a menu system controlled by the buttons at the bottom of the screen. Reproduced from [The DIAdvisor Consortium, 2012].

to the system. Of course, to the extent possible, self-monitoring and evaluation need to be implemented at a system level, to catch such errors before actions are suggested to the user.

2.2 The DIAdvisor Project

The DIAdvisor project [DIAdvisor, 2012] was an EU FP7 Integrated Project (IP) running between 2008 and 2012. The aim of the project was to develop a personal decision support system for IDDM patients using user-provided input, minimally invasive sensors and individualized models of glucose dynamics, in order to provide the user with short-term predictions of glucose evolution, together with insulin therapy decision support.

A mobile research system, incorporating these aspects, was developed and successfully evaluated under clinical conditions at three clinical sites covering 50 patients, with a significant reduction of time spent in hypoglycemia, and increase in time in normoglycemia [The DIAdvisor Consortium, 2012]. Using an Ultra Mobile PC (UMPC), the user could follow his/her glucose curve together

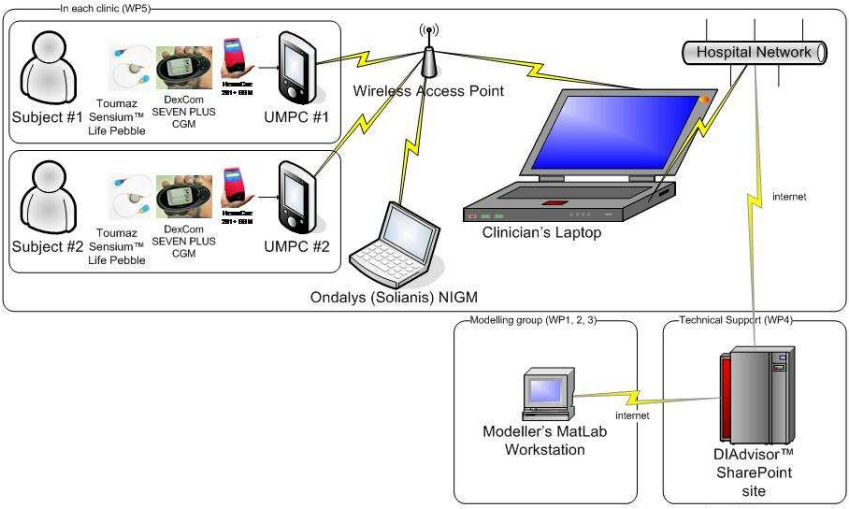

Figure 2.4 DIAdvisor system network. The UMPC communicated with the sensors attached to the patient, see Chapter 3, and transmitted the information to the clinician's laptop through a wireless network. After each visit the clinical team uploaded the data to a common FTP-server. Reproduced from [The DIAdvisor Consortium, 2012].

with an estimate of the near-time (2 hours) ahead projection, see Fig. 2.3. The same information was concurrently provided to the clinician's laptop application by a wireless network according to the network layout in Fig. 2.4.

The project consortium consisted in total of 14 partners, both academic institutions and commercial companies—each providing expertise in areas relevant for the development of the system. Especially noteworthy for the coming chapters are the three clinical partners where the data were collected; Montpellier University Hospital, Department of Clinical and Experimental Medicine (Montpellier), University of Padova, UNIPD (Padova) and the Institute for Clinical and Experimental Medicine, IKEM (Prague).

The models and algorithms presented in this thesis have been developed and used within this project, and were implemented in the system described above.

3

Protocol and Data characteristics

3.1 The DIAdvisor Data

The clinical part of the DIAdvisor project consisted of three clinical studies; the data acquisition (DAQ) trial (2009), the DIAdvisor I (2010) and DIAdvisor II (2011-2012) trials. The purpose of the first trial was to collect data in order to facilitate model and algorithmic development of the individual modules of the DIAdvisor system. The two following trials were set up for testing and validating the entire system in clinical settings. The results presented in this thesis are based on retrospective analysis of the data collected in the DAQ and DIAdvisor I trials.

A total of 90 patients participated (29 Montpellier, 31 Padova, 30 Prague) in the DAQ trial, including users of both MDI and subcutaneous pump therapy. For this thesis, the data were assessed for data completeness and data consistency. Exclusion criteria were missing bolus doses and missing meal data in the diary, missing continuous glucose measurement (CGM) data and large discrepancies between the CGM and the reference glucose meter data. Data segments not fulfilling the criteria were rejected, and only data records containing at least 48 hours of consecutive qualitative data were included in the study. In all, 47 out of the 90 patient data records reached the quality standards of inclusion (17 Montpellier, 19 Padova, 11 Prague). A summary of collected population statistics can be found in Table 3.1.

The DAQ trial was divided into two main parts; a three day hospitalized

study and an ambulatory second part, where the patients were allowed to bring the system home under normal living conditions. In this thesis, data from the hospitalized part of the trial were used in Chapter 4, 5 and 6.

In the second trial, the first configuration of the DIAdvisor system was tested for some of the patients that participated in the DAQ trial, as well as for some new patients. The trial was divided into six different sub-trials, DIAdvisor I A-F—each with a specific evaluation purpose. Trial A was a data collection study in order to validate that the system could retrieve data from the external sensors as expected, the B and C trials had identical protocols but with different purposes. The intention of trial B was to test the predictive performance, whereas trial C aimed at an assessment of the therapeutic advices provided by the system. In trial D, the patients underwent two different exercise tests, and in trial E, free meals, not regulated by the standardized procedure, were allowed. In the final F trial, periods of hypo- and hyperglycemia were induced. Trials A, B, D were conducted at the Montpellier hospital, trial E at the Padova site and trial F in Prague. Trial C was evaluated at all three sites. Data from the B and C trials were used in this thesis in Chapter 6. Six patients data records, namely patients 3, 7, 8, 18, 25 and 30 from Montpellier, who also participated in the DAQ trial, fulfilled the necessary quality standards outlined above.

The third trial, DIAdvisor II, was set up to validate the final performance against the project endpoints using on updated version of the DIAdvisor system. Data from this trial has not been analysed in this thesis.

Table 3.1 Population Statistics of the DAQ trial. Mean values and [min-max].

Parameter	**Montpellier**	**Padova**	**Prague**
Male/Female	13/4	10/9	6/5
Pump/MDI	9/8	10/9	8/3
Rapid Insulin	11 Aspart / 1 Glulisine / 5 Lispro	15 Aspart / 4 Lispro	4 Aspart / 7 Lispro
Age	44 [22-68]	42 [25-67]	33 [19-65]
BMI [kg/m^2]	24.2 [19.7-30.1]	24.5 [18.7-33.2]	25.0 [16.8-35.9]
HbA1c [mmol/mol]	7.7 [5.6-9.1]	8.0 [6.0-9.3]	7.8 [6.4-9.7]
Daily Total Insulin [IU]	47 [18-82]	44 [22-74]	22 [6-54]
Antibodies [% binding]	15.6 [0-62.1]	20.4 [0-75]	12.9 [0-53]

3.2 Equipment

During the trials, the patients were equipped with sensor devices in order to collect vital signs of potential interest in metabolic modeling.

Glucose Sensors

Blood glucose is generally measured manually by the individual patient using a personal glucose meter. A small blood sample is analysed in a test strip by the meter using enzymatically catalyzed-based electro-chemical or photometric methods [Hönes *et al.*, 2008]. Today, there exist more than 27 different personal glucose meters from 18 different manufacturers [Freckmann *et al.*, 2010]. The accuracy requirements of these is generally quite demanding, e.g., meters marked with the European CE mark should comply with the DIN EN ISO 15197 standard, specifying that the measurements may not differ more than 15 mg/dl for glucose concentration below 75 mg/dl and less than 20 % for glucose concentration above 75 mg/dl [Freckmann *et al.*, 2010], when evaluated against a laboratory equipment such as a Yellow Springs Instrument Analyzer [Yellow Springs Instrument, 2012]. Other norms and regulations have similar requirements [Tonyushkina and Nichols, 2009].

Self-monitored blood glucose (SMBG, BG or G) thus provides very accurate readings, but reveals little about the dynamics, unless sampled frequently enough. Generally, the diabetic population seem to measure their glucose level far too seldom, considering, e.g., the average HbA1c level [The Swedish National Board of Health and Welfare, 2012b], and numerous different studies show a definite positive correlation between increased testing frequency and lowered HbA1c [St John *et al.*, 2010].

The HemoCue Glucose 201+ Analyzer (Fig. 3.1, [HemoCue Glucose 201+ Analyzer, 2012]) is a high-quality glucose meter of laboratory precision [Stork *et al.*, 2005]. This meter was used as blood glucose reference in both trials.

Frequent automatic glucose measurements have become commercially available over the last ten years. Today, there are three companies with commercial systems, and this number will increase in the coming years, e.g., Roche and BD are researching and developing similar systems. These sensor systems are called Continuous Glucose Measurement (CGM) Systems and consist of a disposable sensor including a subcutaneous probe, a radio transmitter connected to the external part of the sensor and a receiver device to administrate and display the results. The sensor lasts for 3-7 days, depending on system, after which it is replaced. The measurements are made in the interstitial fluid and do not

Figure 3.1 The HemoCue 201+ Analyzer, [HemoCue Glucose 201+ Analyzer, 2012]

.

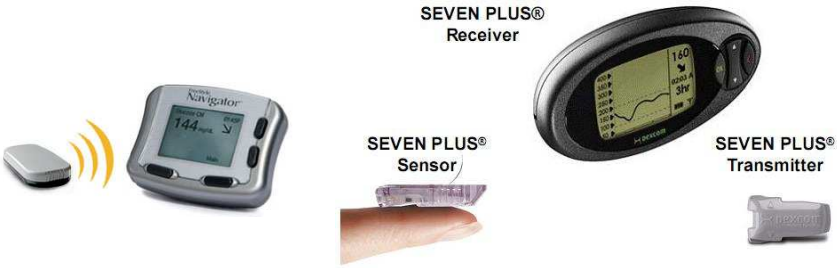

Figure 3.2 CGM systems used in the DIAdvisor project; the Abbott Freestyle CGM system, [Abbott Freestyle Navigator, 2012] (left), and the Dexcom Seven Plus CGM system [Dexcom Seven Plus, 2012] (right).

directly correspond to the blood glucose level, due to the first-order diffusion-like relationship between the blood stream and the interstitial compartment, see e.g., [Rebrin and Steil, 2000b]. The use of CGM has been shown to promote improved glycemic control with decreased level of HbA1c [Chetty *et al.*, 2008].

In the DAQ trial, the patients were equipped with the Freestyle CGM system (Fig. 3.2) from Abbott [Abbott Freestyle Navigator, 2012]. The system provided a CGM reading every 10 minutes, but the raw current signal from the sensor was also collected on a one-minute basis at the Montpellier and Padova sites. The sensors require initialisation during 10 hours and have a life time of 5 days, where after they need to be replaced. In the DIAdvisor I trial, the CGM system Seven Plus (Fig. 3.2) from Dexcom was used [Dexcom Seven Plus, 2012]. This sensor has an initial calibration time of 2 hours and is replaced after 7 days. Both systems need to be recalibrated every 12 hours.

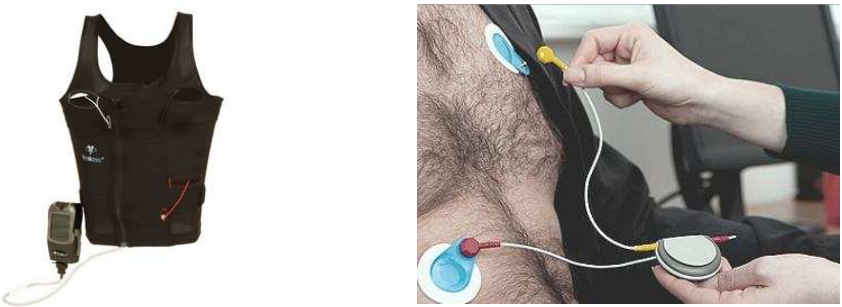

Figure 3.3 Vital signs sensor systems used in the DIAdvisor project; the VivoMetrics' LifeShirt system [VivoMetrics, 2012] (left), and the Sensium Life Pebble sensors by Toumaz [Toumaz, 2012] (right)

3.3 Vital Signs Sensors

During the DAQ trial the patients wore the Clinical LifeShirt (Fig. 3.3) from VivoMetrics [VivoMetrics, 2012], which is specially designed for clinical trials. This non-invasive monitoring system continuously collects, records and analyses several vital signs. To measure respiratory function, sensors are woven into the shirt around the wearer's chest and abdomen. A single-channel ECG measures heart rate, a three-axis accelerometer records posture and activity level, and a thermometer measures the skin temperature.

In the DIAdvisor trial, the LifeShirt was replaced by the Sensium Life Pebble sensors (Fig. 3.3) developed by Toumaz [Toumaz, 2012]. These continuously monitor ECG, heart rate, physical activity (3-axis accelerometer) and skin temperature, and stream the data using a wireless datalink over a short range (5 m).

3.4 Experimental Protocols and Conditions

The DAQ and the DIAdvisor I trials followed the same basic protocol. Standardized meals were served for breakfast (08:00), lunch (13:00) and dinner (19:00), according to the protocol. The amount of carbohydrates included in each meal was about 40 (45 in DAQ), 70 and 70 grams, respectively. Additional snacks, in some cases related to counter-act hypoglycemia, were also digested. No specific intervention on the usual diabetes treatment was undertaken during the studies, since a truthful picture of normal blood glucose fluctuation and insulin-glucose

interaction was pursued. Meal and insulin administration were noted in a logbook, glucose was monitored by the Continuous Glucose Measurement system and by frequent blood glucose measurements in the DAQ trial (37 measurements daily according to the protocol). The outcome, however, was that 39, 37 and 7 measurements (Montpellier, Padova and Prague) were made on average every day. In the DIAdvisor B and C trials, even more reference measurements were collected, making the average 43 measurements a day.

3.5 Graphical Data Evaluation Tool

The trial data was continuously uploaded into an Oracle database on a common FTP-server, from which the model developers could download data as they became available. In order to facilitate data overview and management, a standalone Graphical User Interface (GUI), see Fig. 3.4, was developed in Matlab code [MathWorks, 2012]. Using this GUI, different data channels and time periods could be selected for any individual patient in order to evaluate the data for completeness and correctness, before extracting and exporting them into a single Matlab data file. The evaluation described in section 3.1 was performed using this tool.

3.6 Glucose Data Characteristics

Before digging into modeling and prediction of glucose dynamics, some interesting features of the glucose data are worthwhile to explore a little more in-depth.

Optimal sampling frequency

An interesting question is how often sampling is needed in order to reconstruct the most important features of the glucose signal, and thus how important CGM measurements may be, and whether interpolation of frequent BG measurements can be used to reconstruct the glucose curve. According to [Worthington, 1990], at least 8 samples per day are needed to get the lowest essential dynamics of the system, namely the rise and fall of the blood glucose level due to the carbohydrate intake. This is a rigid assumption, relying on that the meal-related period is about 6 hours, and that the subject follows a strict schedule. In reality, people tend to have more irregular routines. This is generally overcome by

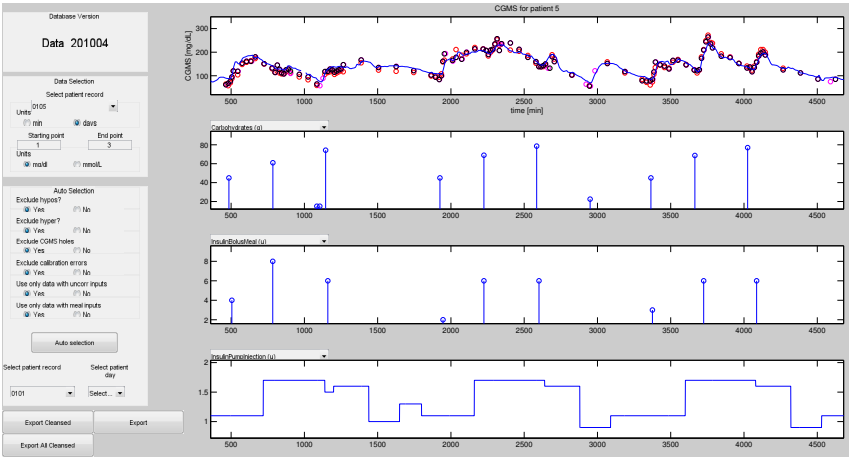

Figure 3.4 The Graphical User Interface (GUI) to manage the DAQ data. Different data sections from both the clinical and the home-monitored part of the DAQ trial can be analysed. The upper plot always shows the linearly interpolated CGM (blue curve) and the HemoCue reference measurements (circles). The lower windows can be used to display any of the recorded signals. In this example, the three days data from the clinical part of the trial has been selected. The second plot from the top shows timing and amount of ingested carbohydrates, the third plot depicts bolus and correction insulin doses, and in the bottom window the pump basal curve has been chosen for investigation.

non-equidistant sampling, collecting data on an event-driven basis, rather than a time-scheduled ditto.

Method In order to evaluate how much information is lost as the sampling rate decreases, the CGM data collected at the Montpellier hospital 3-day visit of the DAQ trial were used. The data was down-sampled to a sampling period of 20, 40, 60 min and then interpolated by piecewise splining. Likewise, the frequent BG measurements were also interpolated by the same method. Error analysis of the down-sampled signals in comparison to the original signal was done by frequency analysis, see [Johansson, 2009], and statistical analysis of the time-domain data.

Results Obviously, the frequency content diminished with increased sampling period, as seen in Fig. 3.5, where the periodogram of the original signal and the

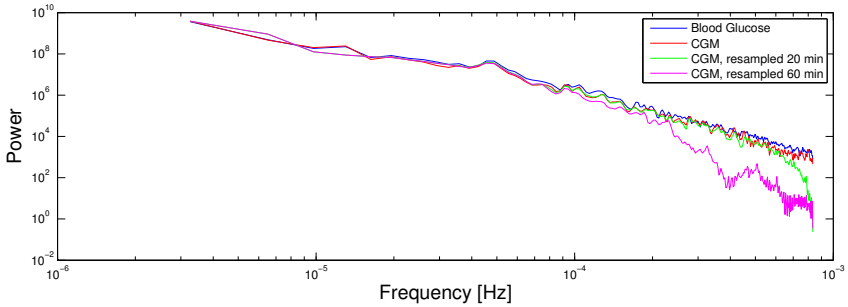

Figure 3.5 Periodogram of the SMBG, the original and the down-sampled CGM signals. Average for the Montpellier patients.

Table 3.2 Comparison between the original and the resampled CGM signals in terms of Root Mean Square Error (RMSE) and maximum error. Average over the DAQ population.

Criteria	**Sampling period [min]**				
	20	40	60	80	100
RMSE [mg/dl]	1.4	4.0	7.0	10.4	12.7
maximum error [mg/dl]	8	18	30	45	51

interpolated signals can be seen.

The spectrum of the blood glucose signal is very similar to that of the CGM signal. For the down-sampled CGM signals, the energy decreases for the higher frequencies as expected. However, frequency assessment does not easily translate to clinically relevant information. Turning to the time-domain, the difference between the signals deteriorates as depicted in Table 3.2. Already at a reduced sampling period of 60 min, the maximum average error amounts to more than 30 mg/dl. This is not surprising, as the glucose rise/drop over an hour can be in the magnitude of -35 mg/dl to +60 mg/dl (95 % conf. bound) at glucose levels of 100 mg/dl, and with an even wider spread for higher glucose levels, see Fig. 3.6.

Discussion It should be borne in mind that these values are under-estimated considering the low-pass character of the relationship between interstitial and

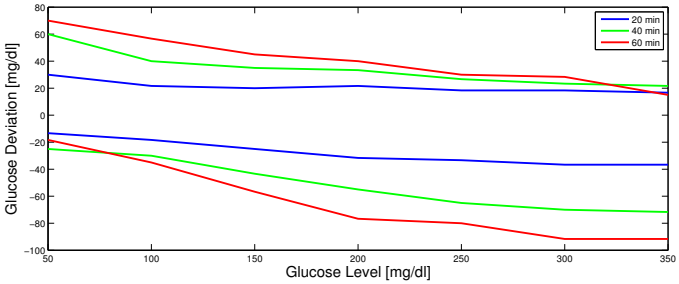

Figure 3.6 95 % confidence bounds of the deviation distribution over different time horizons. Average for all three clinical sites.

blood glucose value. This aspect also inhibits the possibility for direct comparison between these signals. However, the frequency response shows that the interpolated BG curve incorporates the same frequency content as the original CGM signal and should thus be a reasonable approximation of the true blood glucose evolution. Thus, even though only 37 samples were collected a day, making the average sample period about 40 min, the applied sampling schedule made it possible to capture the dynamical changes. In general, glucose self-monitoring does not follow a strict sampling schedule. Rapid changes in the blood glucose can be recognized by persons with normal hypoglycemic sensitivity as hypoglycemia, changes into hypoglycemia or hyperglycemia are often detected, and these circumstances call for unscheduled measurement to establish glycemic status. Therefore, the high and low peaks are, for many instances, represented in home-monitored data, but as the hypoglycemic sensitivity decreases over the years since diagnosis, the risk of undetected hypoglycemia increases [Mokan *et al.*, 1994].

Distribution

In order to investigate the range of excitation in the data in terms of glucose level, and to determine if there are any systematic differences in this aspect between the different sites, the distribution of the CGM data was analysed.

Method The distribution of the CGM data from the DAQ trial was assessed by standard statistical methods for all three clinical sites.

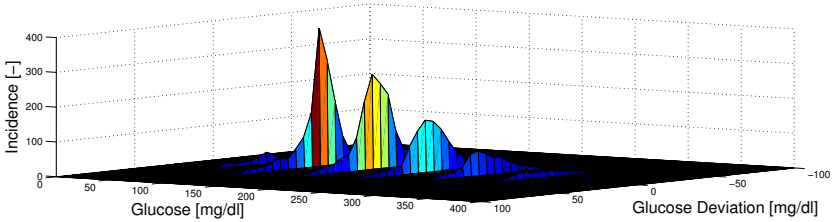

Figure 3.7 Total distribution of glucose level $G(t)$ and the 20 minute glucose deviation, $G(t+20)-G(t)$. Montpellier patients.

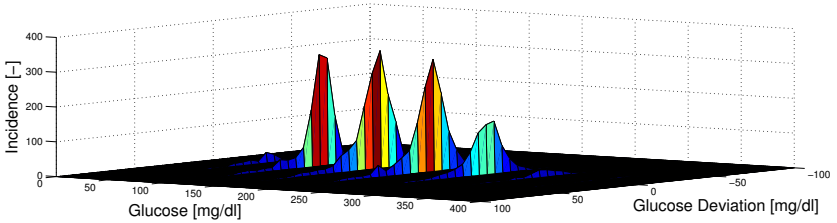

Figure 3.8 Total distribution of glucose level $G(t)$ and the 20 minute glucose deviation, $G(t+20)-G(t)$. Padova patients.

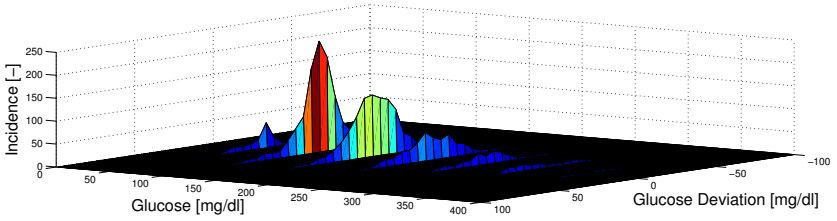

Figure 3.9 Total distribution of glucose level $G(t)$ and the 20 minute glucose deviation, $G(t+20)-G(t)$. Prague patients.

Table 3.3 Likelihood of each glycemic zone [%], and average mean glucose [mg/dl] for the patient data from each clinical site.

Glycemic Zone	Zone Limits [mg/dl]	Montpellier	Padova	Prague
Severe Hypoglycemia	$G \leq 50$	0	0	0
Hypoglycemia	$50 < G \leq 75$	3	2	4
Lower Euglycemia	$75 < G \leq 125$	32	23	37
Upper Euglycemia	$125 < G \leq 175$	31	30	34
Lower Hyperglycemia	$175 < G \leq 225$	20	27	15
Hyperglycemia	$225 < G \leq 275$	14	18	10
Upper Hyperglycemia	$G > 275$	0	0	0
Mean Glucose [mg/dl]	-	153	169	142

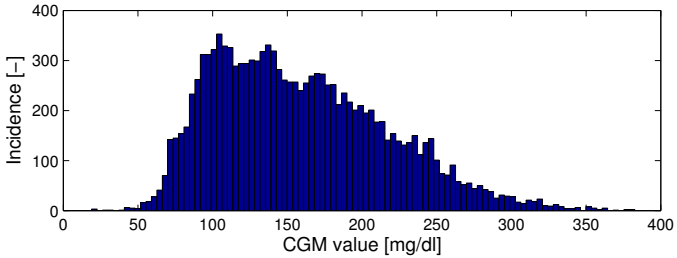

Figure 3.10 Total distribution of CGM glucose level. All DAQ patients.

Results The dynamical total distribution of glucose level $G(t)$ and glucose deviations over 20 minutes, $G(t+20)-G(t)$ can be seen in Fig. 3.7, 3.8 and 3.9. There is a clear difference in distribution between the clinical sites. The glucose range can be divided into 7 different zones of different clinical importance, and the likelihoods of each zone are found in Table 3.3.

The glucose data clearly do not follow a Gaussian distribution, as seen from Fig. 3.10, depicting the total distribution of the accumulated CGM readings collected at all three site. The samples fluctuate around an average of about 160 mg/dl, but the deviations are not normally scattered around this mean. This phe-

nomenon has been noted in [Kovatchev *et al.*, 1997] as well, where the following data transformation was suggested to transform the data into a Gaussian distributed variable with zero mean.

$$f(BG,\alpha,\beta) = (\log BG)^{\alpha} - \beta \tag{3.1}$$

The parameters α and β should be 1.084 and 5.381 when using the milligram-per-dl scale. The accumulated data from each site was transformed in this manner and the distributions can be seen in Fig. 3.12, 3.13 and 3.14. The data from Padova do not fit the normal distribution very well, but the data from the other sites show better resemblance. However, the normal hypothesis was rejected in every case (Kolmogorov-Smirnov test, [Johansson, 2009]), contrary to the results in [Kovatchev *et al.*, 1997]. From Fig. 3.12, 3.13 and 3.14 it can be seen that the upper tail of the normal distribution is missing or deformed, which is due to the low incidence of hyperglycemia, see Table 3.3.

Discussion The Prague patients have the most aggressive glucose control, with fewer high values and more time spent in hypoglycemia. The Padova patients have more hyperglycemic events, but also half as much time spent in hypoglycemia compared to the Prague patients. This is also reflected in the average mean glucose values, which are statistically significantly different from each other ($p<0.01$ for all possible comparisons).

The total distribution was found to be non-Gaussian, but the log-normal like distribution suggested by [Kovatchev *et al.*, 1997] could not be confirmed. Under free-living conditions, the hyperglycemia tendency is generally higher than for the DIAdvisor DAQ data evaluated here, which may explain why [Kovatchev *et al.*, 1997] found that 203 out of 205 transformed home-monitored SMBG data sets confirmed the normal hypothesis.

The data transformation stems from an intention to create a risk value describing the increased clinical risk associated with hypoglycemia and hyperglycemia. By taking the square of the transformated glucose level and multiplying by 10, the risk function of [Kovatchev *et al.*, 2000] is retrieved, see Fig. 3.11. This function forms the basis for the cost function used in Chapter 6.

Time-variability

Another important aspect of diabetic glucose data is the question of time-variability. The circadian rhythm may have a significant impact on insulin sensitivity over the course of the day [Van Cauter *et al.*, 1997a], especially in the early

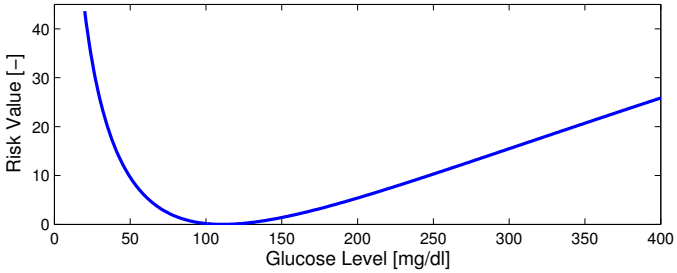

Figure 3.11 Risk function using Kovatchev's transformation.

morning, when counter-regulatory hormones (primarily growth hormone, cortisol and adrenalin) are released—triggering increased hepatic production [Perriello *et al.*, 1991].

Variability over longer time horizons has not been thoroughly investigated in the literature, which may be explained by the scarcity and difficulty of obtaining qualitative longer data records. Very few longitudinal data sets longer than a few days, or weeks at best, seem to be available for type 1 diabetics in the research community. The data set used in [Ståhl and Johansson, 2009] is thus quite unique in this aspect. This data set was collected during the first months of a newly diagnosed type I patient (the author). This period of time is generally referred to as the 'honey-moon period', during which the pancreatic β-cells recover somewhat, resulting in temporary remission with considerably varying insulin doses and glycemic response [Abdul-Rasoul *et al.*, 2006]. Mathematically, this translates into time-varying model parameters.

Method The honey-moon data were analysed. In order to estimate and validate different models, data segments with constant parameter values are needed. To find such segments, the data were investigated using the Adaptive Forgetting Multiple Model change detection algorithm (AFMM), implemented in the Matlab command "SEGMENT" [MathWorks, 2012].

Results In Fig. 3.15, the variations of the estimated ARMAX parameter over the time period can be seen.

Discussion The model parameters shifted a number of times during the honey-moon period, giving an indication of both more stable and unstable data sections,

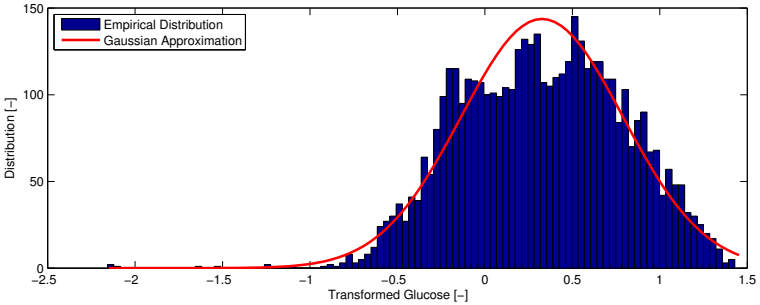

Figure 3.12 Empirical and Approximated Distribution of transformed CGM data. Montpellier patients.

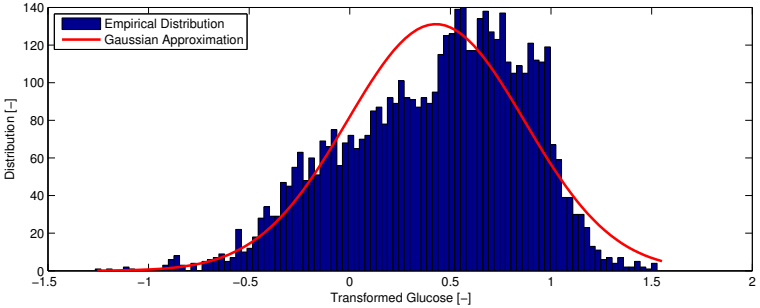

Figure 3.13 Empirical and Approximated Distribution of transformed CGM data. Padova patients.

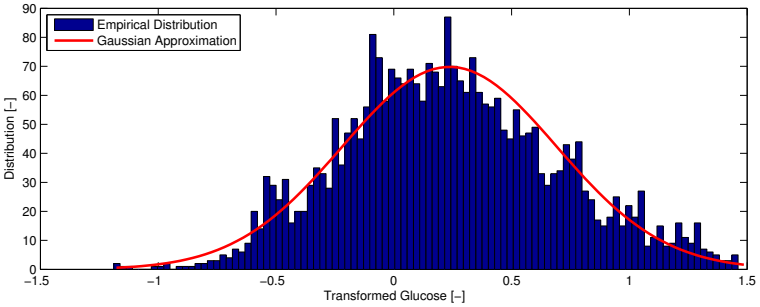

Figure 3.14 Empirical and Approximated Distribution of transformed CGM data. Prague patients.

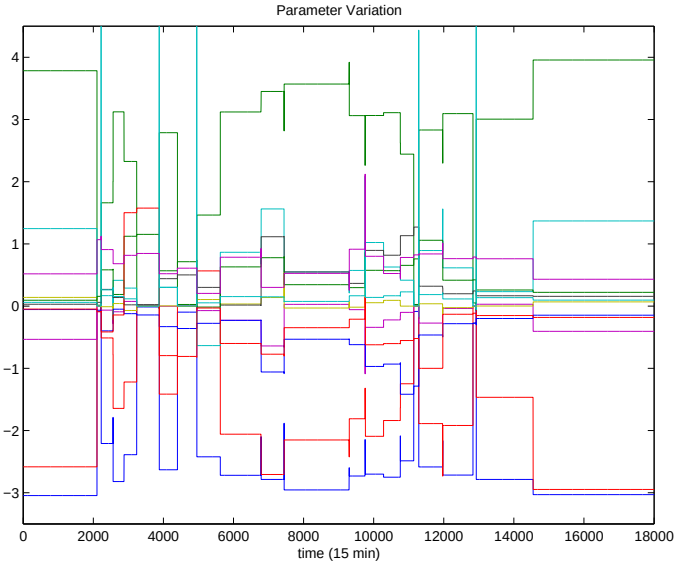

Figure 3.15 Data segmentation using the Matlab command SEGMENT. Variability of the parameters of the recursive ARMAX model over approximately 200 days.

and this behavoir is expected during this remission phase. The last stable parameter section is more than a month in length, signalling the end of the honey-moon period. It may also be noted that the parameter values end up close to the original values, which may be another indication that the temporary β-cell recovery has ended. Longer time-variability in non-newly-diagnosed patients is generally less dramatic, but should not be overlooked, especially for the so-called 'brittle' patients [Voulgari *et al.*, 2012].

Blood-to-Interstitial Glucose Delay

The diffusion-like relationship between the blood and interstitial compartments implies a low-pass character in the response to glucose changes, which means lagging glucose levels in the CGM sensor in comparison to the reference SMBG.

Methods The CGM signal and the blood glucose reference measurements from the DAQ trial were analysed as follows. To retrieve an initial non-parametric

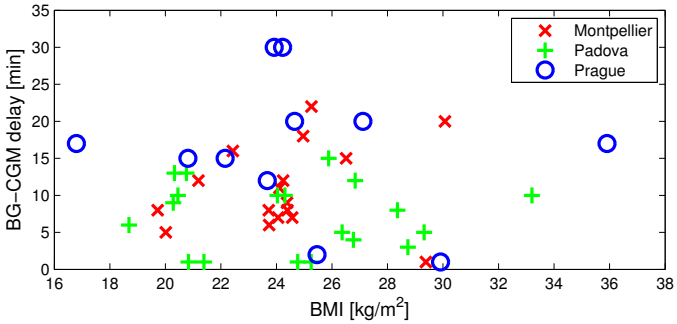

Figure 3.16 Delay between blood glucose reference measurement and the corresponding CGM measurement vs. BMI for the DAQ data from the three clinical sites.

estimate of the magnitude of the lagging between the blood glucose reference $BG(t)$ and the CGM signal, the lag was approximated to a delay, and was found by finding the delay Δ which minimized the Root Mean Square Error (RMSE) between the blood glucose measurements $BG(t_{BG})$ and the corresponding backward-translated CGM measurements $CGM(t_{BG}+\Delta)$ for the time point t_{BG}, corresponding to time points when the blood glucose reference measurements were sampled. The measurement error was also assessed by RMSE between the untranslated CGM signal and the blood glucose reference, and a possible correlation between sensor delay and Body Mass Index (BMI) was investigated.

Results In Tables 3.4, 3.5 and 3.6 the estimated delay and RMSE between the CGM signal $CGM(t)$ and the HemoCue reference $G(t)$ is given for every included patient. The BG-CGM delay was statistically larger for the Prague patients than for the Montpellier patients ($p<0.05$) and for the Padova patients ($p<0.001$), and the BG-CGM delay of the Montpellier patients was larger than that of the Padova patients ($p<0.02$). In terms of BG-CGM measurement error, the Prague patients had a significantly larger BG-CGM RMSE than the Padova patients ($p<0.002$) and the Montpellier patients ($p<0.003$). No correlation between BMI and BG-CGM delay was found, see Fig. 3.16.

Discussion The differences between the Prague data and the other sites could be explained by the significantly lower number of reference measurements at the Prague site (7 measurements) compared to the Padova (37 measurements)

and the Montpellier (39 measurements) patients, and the simplified assumption of estimating the delay and not the lag. The low number of samples, mainly collected during periods of substantial glucose changes, where the low-pass filter relationship causes long delays and mismatches, result in that the delay estimate is biased. Thus, the estimates for the Prague patients should be disregarded.

Intuitively, a correlation was expected between high BMI, often indicating a thicker abdominal layer, and longer BG-CGM time delay, as a possible explaination for the large interpersonal differences. However, the diffusion may be more dependent on other factors than mere amount of abdominal fat, such as capillary density and blood turn-over rate.

Table 3.4 Glucose Data Statistics Montpellier

Patient ID	BG-CGM Delay [min]	BG-CGM RMSE [mg/dl]
102	15	19.8
103	6	11.8
104	5	22.1
105	7	14.7
106	12	27.7
107	22	28.1
108	8	15.3
111	9	19.6
112	18	23.9
115	8	15.4
117	20	27.6
118	11	23.7
120	16	24.9
122	7	17.9
126	12	21.5
127	8	14.4
130	1	34.3
Mean (std)	**10.9**(**5.7**)	**21.3**(**6.0**)

Table 3.5 Glucose Data Statistics Padova

Patient ID	BG-CGM Delay [min]	BG-CGM RMSE [mg/dl]
201	1	22.8
202	10	14.5
203	5	27.2
205	10	30.1
209	1	21.5
211	12	24.6
212	6	21.9
213	10	20.6
214	1	16.2
215	5	15.6
216	3	24.0
217	8	14.8
219	13	25.6
220	13	19.8
221	15	23.3
222	4	30.0
226	10	28.4
227	9	24.2
231	1	18.3
Mean (std)	**7.2**(**4.6**)	**22.3**(**4.9**)

Table 3.6 Glucose Data Statistics Prague

Patient ID	BG-CGM Delay [min]	BG-CGM RMSE [mg/dl]
301	20	30.7
310	20	22.0
313	15	41.0
316	17	36.2
317	30	50.3
318	17	16.8
322	12	34.5
324	15	28.9
325	1	17.7
326	30	44.5
328	2	23.3
Mean (std)	**16**.**3**(**9**.**2**)	**31**.**4**(**11**.**6**)

4

Modeling and Prediction

Prediction of glucose changes in type 1 *Diabetes Mellitus* has received a considerable amount of scientific and commercial interest over the last decade. The driving force behind this surge in research can in large be explained by the recent advances in sensor technology [Vaddiraju *et al.,* 2010], and the thereto attached promises and hopes of closed, or semi-closed, loop control of diabetic glucose dynamics. Predicting models play a key role in many of these concepts—providing the essential simulation tool in MPC-oriented closed loop arrangements of an artificial pancreas [Cobelli *et al.*, 2011], or as a component in a decision support system—providing predictions directly to the user [Poulsen *et al.*, 2010].

However, even though many of these models indicate good predictive performances, less attention has been put into establishing whether the physiological responses are qualitatively correct and safe. Evaluation under strict clinical protocols, or even during normal life style, may not reveal short-comings due to a flawed identified model that may produce dangerous or suboptimal prediction under less strict conditions. Furthermore, identifying models with these features may be non-trivial, with regards to the opposite influence on the glucose level from the main driving inputs– carbohydrates and insulin.

In this chapter, low-level models of glucose dynamics are investigated, considering the aspects of identifiability and physiological correctness outlined above, and concerning the ability to detect hypoglycemia in advance.

4.1 Previous Work

Models of glucose dynamics for predictive purposes can mainly be divided into two categories; physiologically-oriented models and data-driven black-box approaches. The latter sometimes incorporate physiological sub models of insulin and glucose infusion following insulin administration and meal intake, but the main part of the dynamics stem from the statistically derived relationships.

The development of physiological diabetic glucose modeling started with the simple models of [Bolie, 1961] and [Ackerman *et al.*, 1965], aiming at describing the relationship between glucose and insulin utilization. External meal and insulin administrations were not considered, and the models found little use beyond basic insight to the dynamics of this interaction. Following these efforts, the slightly more complex and well-established minimal model [Bergman and Cobelli, 1980] was suggested as a means to estimate insulin sensitivity from an intravenous glucose tolerance test (IVGTT). Detailed models of the glucose metabolism; separating insulin and non-insulin dependent glucose utilization, incorporating models of hepatic balance, renal clearance, and in some cases pancreatic insulin synthesis and release, have surged since then. In [Lehmann and Deutsch, 1992], a simulation model was presented, and later the decision support system AIDA [Lehmann, 1994] was developed using this model. The system was validated on a set of 24 subjects with parameter convergence achieved in 80% of the cases [Lehmann *et al.*, 1994].

A large model with 19 tunable parameters was proposed in the Sorensen thesis [Sorensen, 1985], a model often used as a verification tool to assess different control approaches, e.g., [Eren-Oruklu *et al.*, 2009a]. The web-based educational simulation model GlucoSim [Agar *et al.*, 2005] has been developed based on another thesis [Puckett, 1992]. Generally, these models are difficult to fit to an individual person, and may lack structural identifiability. This makes them unsuitable for predictive purposes, but synthetic subjects may be created for simulation studies. Currently, the most influential simulation model is the University of Virginia and Padova University (UVa/Padova) model described in [Dalla Man *et al.*, 2007a] and [Dalla Man *et al.*, 2007b], which has been accepted by the Federal Drug Administration of the U.S. (FDA) to be used as a substitute for animal trials in preclinical trials of closed-loop development [Kovatchev *et al.*, 2008]. To this purpose, 300 artificial subjects have been derived from estimated parameters from population studies, and used in, e.g., [Lee *et al.*, 2009].

A simpler model, with only five tunable parameters, is the Hovorka model

[Hovorka *et al.*, 2004], later extended and altered for the critically ill in [Hovorka *et al.*, 2008]. The former model has been validated for predictive capacity on 15 subjects with a RMSE of 3.6 mg/dl for a prediction horizon of 15 minutes. Parameter estimates were retrieved recursively from a sliding data window using a Bayesian approach. This model is also used extensively for MPC-oriented closed-loop validation, e.g., in the evaluation of PID control in [Farmer *et al.*, 2009], which also make use of the Sorensen [Sorensen, 1985] and the minimal model [Bergman and Cobelli, 1980].

Data-driven models have been investigated on CGM time-series alone, or by considering inputs as well. The meal sub models of [Dalla Man *et al.*, 2007b] and [Lehmann and Deutsch, 1992] are furthermore often used as input generating components in data-driven models to approximate the glucose flux input from the gut following a meal intake. Here, the focus has been prediction for the purpose of early hypoglycemic detection, e.g., to be used for alarm triggering in CGM devices, or temporary insulin pump shut-off, as well as establishing models suitable for model-based control.

Time-series analysis by Auto-regressive (AR) models started with [Bremer and Gough, 1999], who evaluated the basic underlying assumptions concerning stationarity and auto-covariance that AR modeling is based upon, concluding that diabetic data generally is non-stationary, but highly auto-correlated, thus recommending the models to be recurrently re-estimated. Following this, AR and ARMA models were developed in [Ståhl, 2003] and [Ståhl and Johansson, 2009] using glucose data from a recently diagnosed type 1 diabetic. In [Sparacino *et al.*, 2007], first-order recursive AR models were investigated for 28 subjects using a low-pass filtered CGM signal from the GlucoDay CGM system. The results indicate that hypoglycemia can be detected by the model 25 min before the CGM signal passes the same threshold. Another example of recursive AR and ARMA models of third order, incorporating a change detection feature for more rapid parameter re-estimation when large changes in the dynamics are detected, is found in [Eren-Oruklu *et al.*, 2009b]. The models were evaluated for 30 healthy, 7 glucose-intolerant and 25 type II diabetic subjects, with less than 4% mean Relative Average Deviation (RAD) and almost no values in D or E zones of the Clarke Error Grid (p-CGA, see section 4.5 below for definition) for the 30-minute predictions in comparison to the CGM Medtronic Gold reference [MedTronic, 2012]. Contrary to the above, the authors of [Gani *et al.*, 2009] claim that a generic patient- and time-invariant AR model of order 30 can be identified from any patient and used for glucose prediction for any other patient. Very promising results were achieved in [Gani *et al.*, 2010], where

the model was evaluated for three different datasets, each utilizing a different CGM device, and the patient cohorts included both type I and type II diabetes. The prediction error was on average, in terms of RMSE, less than 3.6 mg/dl for a 30-minute prediction, with negligible delay, and with 99% of the paired prediction-reference points in the A and B zones of the p-CGA. However, these results were achieved by filtering the CGM signal in both training and test data using a non-causal filter, removing the high frequency components. In [Lu *et al.*, 2011] the causality aspect of the input filtering was addressed. The AR model, here reduced to order 8 after model complexity considerations, was reformulated as a linear model with a Kalman filter, and the filter parameters were adjusted to account for the filtering of the CGM signal. For evaluation purposes, the reference was however still filtered in the same non-causal way as before. Using this approach on the same data set as in [Gani *et al.*, 2010], yielded more moderate results with an average prediction error of 16 mg/dl, and a 9 minute lag for the 20-minute prediction.

Algorithms specifically developed for hypoglycemic detection has also been proposed. In [Palerm *et al.*, 2005], a Kalman filter approach was proposed, estimating the states corresponding to the interstitial glucose level, and the first and second derivative thereof, i.e., rate of glucose change and acceleration. In [Palerm and Bequette, 2007] this method was evaluated for 13 hypoglycemic clamp data sets. Using a hypoglycemic threshold of 70 mg/dl, the sensitivity and specificity were 90 and 79%, respectively, with unknown alarm time. Combining three different methods for hypoglycemic detection with the ARMA model of [Eren-Oruklu *et al.*, 2009b], data from insulin-induced hypoglycemic tests for 54 type 1 subjects were evaluated in [Eren-Oruklu *et al.*, 2010]. With a hypoglycemic threhold of 60 mg/dl, sensitivity of 89, 88, and 89% and specificity of 67, 74, and 78% were reported for each method, respectively. Mean values for time to detection were 30, 26, and 28 minutes. In [Dassau *et al.*, 2010], five different algorithms were used together in a voting based detection system called hypoglycemic prediction algorithm (HPA). The system was developed using 21 datasets from a 24-hour Abbott Navigator CGM trial for children with type 1 diabetes, and was validated on hypoglycemic induced studies on 22 type 1 patient records. With a voting scheme of 3-out-of-5, and a hypoglycemic defined as 60 mg/dl, a sensitivity of 91% was achieved, and when 4-out-of-5 positive alarms were required, the sensitivity dropped to 82%.

A short-coming of the AR models and the algorithms above is the lack of input-output relationship, excluding them from being used in a model-based control framework. A natural extension to the AR concept is to include exter-

nal inputs, transforming the model to an ARX model. This type of model has been considered in, e.g., [Finan *et al.*, 2009a], where both batch-wise and recursively identified patient-specific ARX-models have been analysed for 9 patients with a mean 30-minute prediction error RMSE of 26 mg/dl. In [Cescon, 2011] both ARX, ARMAX and state space models were investigated using different identification methods for 30-, 60-, 90- and 120-minute prediction for 9 Montpellier patients from the DAQ trial. The best performance was achieved with the ARX and the ARMAX models. The ARX model gave a standard deviation of the prediction error of 17, 34, 46 and 56 mg/dl on average for the 30-, 60-, 90- and 120-minute prediction, respectively. The corresponding results for the ARMAX model were 16, 30, 39 and 44 mg/dl.

Another type of transfer function model, cast in the continuous domain, was approached in [Percival *et al.*, 2010], where it was evaluated for 9 type I subjects on separated meal and insulin intakes. Model parameters were determined both heuristically and by least-squares estimation. The carbohydrate and insulin impacts of the model, i.e., the steady-state rise and drop of glucose following these intakes, were further compared to the corresponding practically used estimates of these factors. No independent prediction validation was given. This model was later evaluated in a control framework in [Percival *et al.*, 2011], where two data sets were created by the Hovorka (4 subjects) and Padova (10 subjects) simulation models. Here, the model could approximate the simulated data very well, with a 3-hour look-ahead prediction error of 26 mg/dl reported. A very similar model structure was used in [Kirchsteiger *et al.*, 2011], the difference being a time delay changed into a time lag. In this paper, breakfast glucose excursion prediction was addressed for 10 Montpellier patients from the DAQ trial. For each patient, model parameters were determined by constrained least squares for two breakfast meals and validated on a third breakfast, with an average fit value of 42%.

Neural network (NN) models have been shown to be a competitive approach in [Daskalaki *et al.*, 2012], where a feed-forward NN model was compared against an AR and an ARX model on a 30 patient dataset, retrieved from the Padova simulation model. Here, the NN clearly outperformed the competing models with an average RMSE of 4.9 mg/dl versus 29 mg/dl (AR) and 26 mg/dl (ARX) for the 45-minute prediction. Apart from meal and insulin information, emotional factors, hypoglycemic/hyperglycemic symptoms and lifestyle/ activities, were collected in an electric diary and used as inputs in the NN model of [Pappada *et al.*, 2011]. Training was performed on a dataset from 17 patients, and performance was evaluated on 10 patient data sets not included in the training

set, with a RMSE of 44 mg/dl for the 45-minute prediction.

A fully connected three-layer (5,10,1 neuron per layer) NN, with sigmoidal transfer functions in the first two layers and a linear for the output block was used in [Pérez-Gandía *et al.*, 2010]. No insulin nor meal information were used, but the concurrent and previous CGM values, up to 20 minutes back, acted as inputs. The model was evaluated on two datasets with different CGM devices (Abbott Freestyle and MedTronic Guardian). Three subject data sets were used for training for each patient group and were thereafter excluded from the validation data. For the 6 Guardian patients and the 3 Abbott Freestyle patients the performance was 10, 18 and 27 mg/dl for the 15, 30 and 45-minute prediction, with a delay of around 4, 9, and 14 min for upward trends, and 5, 15, and 26 min for downward trends. In [Zecchin *et al.*, 2011], the linear predictor from [Sparacino *et al.*, 2007] worked in a cascade-like configuration with a NN model, which also used both CGM and glucose flux from the meal model of [Dalla Man *et al.*, 2007b] into account as inputs. Training and validation was done using 15 patient records from the 7-day free-living conditions set of the DAQ trial, see Chapter 3. The NN was trained and validated on 25 time series, each one of three days, selected so as to ensure a wide variety of glycemic dynamics. Nine daily profiles, containing several hypo- and hyperglycemic events, were used to test the NN with an average of 14 mg/dl and a 14 min delay for the 30-minute prediction. For an assessment on 20 simulated subjects using the UVa/Padova model, the corresponding metrics were 9.4 mg/dl and 5 min. Both insulin and carbohydrate digestion were considered by incorporating input-generating sub models in the support vector machine of [Georga *et al.*, 2011]. Additionally, exercise-induced glucose and insulin absorption variations were also considered as inputs by processing a metabolic equivalent (MET) estimate, derived from a SenseWear body monitoring system (BodyMedia Inc.) used in the study, in a model by [Roy and Parker, 2007]. The NN was trained individually for 7 type 1 patients with RMSE of 9.5, 16, 25 and 36 mg/dl for the 15, 30, 60 and 120-minute prediction.

Deeper reviews can be found in [Makroglou *et al.*, 2006] and [Balakrishnan *et al.*, 2011] and [Georga *et al.*, 2011].

4.2 Identifiability

The challenges of identifiability in physiologically-based models (structural identifiability) have been widely recognized [Chis *et al.*, 2011], and specifically for the diabetic glucose dynamics [Docherty *et al.*, 2011], and optimal experimental

design to facilitate parameter estimation [Galvanin *et al.*, 2009; Galvanin *et al.*, 2011]. Empirical black-box identification problems have received less attention, but the problems associated to identification of ARX models of glucose dynamics have been considered in [Finan *et al.*, 2009b].

In diabetic real world data, the problem is especially important, since the two main inputs affecting the dynamics, meal and insulin intake, have opposing impact and similar dynamics, and generally act simultaneously. The aspect is further problematic since safety concerns impose constraints on the possibility to excite the system sufficiently (which of course does not apply to simulated data). Thus, from an identification viewpoint, the impact from inputs may be entangled with one another, and it may be impossible to separate the impact of each input without considering constraints to the identification routine, incorporating prior information of the expected qualitative response. In [Percival *et al.*, 2010], this was resolved by applying an experimental protocol, where a small meal and the corresponding bolus dose were separated by a few hours. However, such an approach yields only short data sets and may be infeasible, e.g., if re-estimation recurrently is required due to, e.g., shifting dynamics.

4.3 Data

Data from the three day hospitalized part of the DAQ trial was assessed for data completeness and data consistency. Exclusion criteria were missing bolus doses and missing meal data in the diary, missing CGM data and large discrepancies between the CGM and the reference glucose meter data. Data segments not fulfilling the criteria were rejected, and only data records containing at least 42 hours of consecutive qualitative data were included in the study. Thereafter, the data was divided in batches of 24 hours, two hours apart, thus providing 9 data sets for a 42-hour period and 24 data sets for a 72-hour period.

A total of 47 out of the 90 patient data records reached the quality standards of inclusion (17 Montpellier, 19 Padova, 11 Prague).

The CGM data were interpolated to a 5-minute sampling period by linear interpolation, giving the discrete-time CGM glucose signal $G_{CGM}(k)$ sampled at time instances $t_k \in (5, 10, \ldots)$ min. For the Montpellier and Pdavoa patients, the blood glucose reference was also interpolated to a five minute sampling rate by piece-wise splining, for use in the hypoglycemic detection assessment.

4.4 Modeling

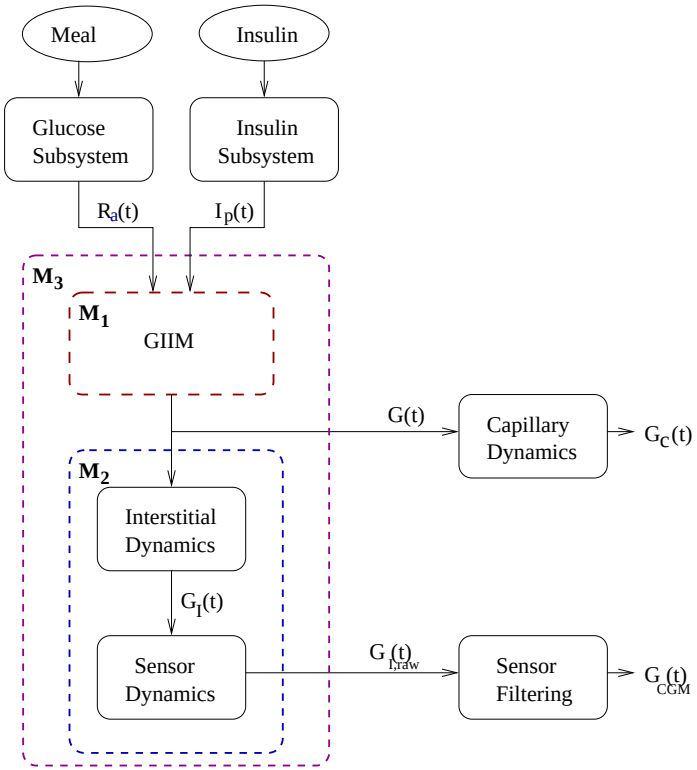

Figure 4.1 Overview of the modeling approach. Notation: Plasma Insulin $I_p(t)$, Rate of Glucose Appearance following a meal $R_a(t)$, Blood glucose $G(t)$, Capillary glucose $G_C(t)$, Interstitial Glucose $G_I(t)$, CGM raw current signal $G_{I,raw}(t)$ and CGM signal $G_{CGM}(t)$. M_1 represent the model describing the glucose-insulin interaction in the blood and inner organs (GIIM), the M_2 model represents the diffusion-like relationship between blood and interstitial glucose and the CGM sensor dynamics, and M_3 is the joint model of M_1 and M_2.

For modeling purposes, the system was considered to consist of three main parts: the Glucose Sub Model (GSM), the Insulin Sub Model (ISM) and the Glucose/Insulin interaction Model (GIIM), including blood-to-interstitial dynamics,

as outlined in Fig. 4.1. The GSM describes the absorption of glucose from meal, the ISM the absorption of insulin from insulin injections and the GIIM the interaction of glucose and insulin in the blood system and organs. For now, we ignore the division of M_3 into M_1 and M_2, and let M_3 represent the GIIM as a single model. In other words, we use the CGM signal $G_{CGM}(t)$ as a proxy for blood glucose $G(t)$, and ignore the lag, described in Chapter 3, between these signals.

Insulin Sub Model

The transport of rapid-acting insulin from the subcutaneous injection site to the blood stream has been described in quite a few models of insulin pharmacokinetics, see [Nucci and Cobelli, 2000] and [Wilinska *et al.*, 2005] for reviews. Among these, the Insulin Sub Model (ISM), was based on the compartment model in [Dalla Man *et al.*, 2007a] and [Dalla Man *et al.*, 2007b], as follows.

$$\dot{I}_{sc1}(t) = -(k_{a1} + k_d) \cdot I_{sc1}(t) + D(t) \tag{4.1}$$

$$\dot{I}_{sc2}(t) = k_d \cdot I_{sc1}(t) - k_{a2} \cdot I_{sc2}(t) \tag{4.2}$$

$$\dot{I}_p(t) = k_{a1} \cdot I_{sc1}(t) + k_{a2} \cdot I_{sc2}(t) - (m_2 + m_4) \cdot I_p(t) + m_1 \cdot I_l(t) \tag{4.3}$$

$$\dot{I}_l(t) = m_2 \cdot I_p(t) - (m_1 + m_3) \cdot I_l(t) \tag{4.4}$$

$$m_2 = 0.6 \frac{C_L}{H_{Eb} \cdot V_i \cdot M_{BW}} \tag{4.5}$$

$$m_3 = \frac{H_{Eb} \cdot m_1}{1 - H_{Eb}} \tag{4.6}$$

$$m_4 = 0.4 \frac{C_L}{V_i \cdot M_{BW}} \tag{4.7}$$

Following the notation in [Dalla Man *et al.*, 2007a] and [Dalla Man *et al.*, 2007b], I_{sc1} is the amount of non-monomeric insulin in the subcutaneous space, I_{sc2} is the amount of monomeric insulin in the subcutaneous space, k_d is the rate constant of insulin dissociation, k_{a1} and k_{a2} are the rate constants of non-monomeric and monomeric insulin absorption, respectively, $D(t)$ is the insulin infusion rate, I_p is the level of plasma insulin, I_l the level of insulin in the liver, m_3 is the rate of hepatic clearance, and m_1, m_2, m_4 are rate parameters. The parameters m_2, m_3, m_4 are determined based on steady-state assumptions—relating them to the constants in Table 4.1 and the body weight M_{BW}.

Only rapid-acting insulins were considered. This means that the dynamics of the basal doses of the MDI patients were not included in the insulin signal.

Glucose Sub Model

The initial stages of glucose metabolism, describing the digestive process and the flux of glucose from the intestines, have been modeled less extensively. However, two models have been widely used; the model by [Lehmann and Deutsch, 1992] and [Dalla Man *et al.*, 2006]. The latter, a nonlinear compartment model, was used in this study:

$$q_{sto}(t) = q_{sto1}(t) + q_{sto2}(t) \tag{4.8}$$

$$\dot{q}_{sto1}(t) = -k_{gri} \cdot q_{sto1}(t) + C(t) \tag{4.9}$$

$$\dot{q}_{sto2}(t) = k_{gri} \cdot q_{sto1}(t) - k_{empt} \cdot q_{sto}(t) \cdot q_{sto2}(t) \tag{4.10}$$

$$\dot{q}_{gut}(t) = -k_{abs} \cdot q_{gut}(t) + k_{empt} \cdot q_{sto}(t) \cdot q_{sto2}(t) \tag{4.11}$$

$$R_a(t) = \frac{f \cdot k_{abs} \cdot q_{gut}(t)}{M_{BW}} \tag{4.12}$$

where, again following the notation in [Dalla Man *et al.*, 2006], q_{sto} is the amount of glucose in the stomach (q_{sto1} solid, and q_{sto1} liquid phase), q_{gut} is the glucose mass in the intestine, k_{gri} the rate of grinding, k_{empt} is the rate constant of gastric emptying, k_{abs} is the rate constant of intestinal absorption, f is the fraction of intestinal absorption which actually appears in the blood stream, $C(t)$ is the amount of ingested carbohydrates and $R_a(t)$ is the appearance rate of glucose in the blood. k_{empt} is a non-linear function of q_{sto} and $C(t)$:

$$k_{empt}(q_{sto}) = k_{min} + k \cdot \{\tanh[\alpha(q_{sto} - b \cdot G(t))] + \tag{4.13}$$

$$- \tanh[\beta(q_{sto} - d \cdot G(t))] + 2\} \tag{4.14}$$

with $k = (k_{max} - k_{min})/2$, $\alpha = 5/2D(1b)$, $\beta = 5/2Dc$, with parameters k_{max}, k_{min}, b, and d.

Both models were evaluated using generic population parameter values according to Table 4.1.

GIIM

The outputs $I_p(t_k)$ and $R_a(t_k)$ from these models were fed, using the generic parameter values in Table 4.1, as inputs $u(k) = [I_p(t_k)\ R_a(t_k)]^T$ into a linear state space model of the Glucose-Insulin Interaction (GIIM), generating the final output - the blood glucose $G(k)$ at time $t_k \in (5, 10, \ldots)$ min. The model equations, with model order n, are:

$$x(k+1) = Ax(k) + Bu(k) + w(k) \tag{4.15}$$

$$G(k) = Cx(k) + v(k) \tag{4.16}$$

Table 4.1 Generic parameter values used for the GSM and ISM.

Parameter	Value	Unit	Parameter	Value	Unit
$\mathbf{k_{gri}}$	0.0558	[min^{-1}]	$\mathbf{k_{a1}}$	0.004	[min^{-1}]
$\mathbf{k_{max}}$	0.0558	[min^{-1}]	$\mathbf{k_{a2}}$	0.0182	[min^{-1}]
$\mathbf{k_{min}}$	0.008	[min^{-1}]	$\mathbf{k_{d}}$	0.0164	[min^{-1}]
$\mathbf{k_{abs}}$	0.0568	[min^{-1}]	$\mathbf{k_{d}}$	0.0164	[min^{-1}]
b	0.82	[-]	$\mathbf{m_1}$	0.1766	[min^{-1}]
d	0.01	[-]	$\mathbf{V_i}$	0.05	[L/kg]
f	0.9	[-]	$\mathbf{C_L}$	1.1069	[L/min]

with system matrices $A \in \mathbb{R}^{nxn}$, $B \in \mathbb{R}^{nx2}$, $C \in \mathbb{R}^{1xn}$, and process and measurements noises $w(k)$ and $v(k)$.

4.5 Identification

Subspace Identification

To identify the GIIM $\Sigma : \{A, B, C, D\}$, including a prediction feedback vector K, the subspace algorithm N4SID was used. The CGM glucose level G_{CGM} was normalized by subtracting the mean value over the data section, and the stationary basal insulin level, equal to the minimum insulin level over the data section, was removed from this input.

For each 24-hour data segment, models of model order n (2-4) were identified. In order to fulfill the a priori constraints of physiological correctness, models that exhibited initial improper input response (rising glucose due to insulin administration or lowered glucose after meal intake) were discarded. Thereafter, the best model was determined by the Minimum Description Length (MDL) criterion [Johansson, 2009], for the 2-hour ahead prediction error for all estimated models. These models will hereafter be referred to as the automatically identified models.

Models were also identified by manual selection of suitable data sections, no longer than 24 hours. These models were subject to the same screening criteria as the automatically identified models described above. To facilitate manual model

identification, a graphical user interface was developed in Matlab, see Fig. 4.2. These models will hereafter be referred to as the manually identified models.

Short-term predictions, p steps ahead, were evaluated using the Kalman filter:

$$\hat{x}(k+1) = A\hat{x}(k) + Bu(k) + K(y(k) - C\hat{x}(k)) \tag{4.17}$$

$$\hat{x}(k+p) = A\hat{x}(k+p-1) + Bu(k+p-1) \tag{4.18}$$

$$\hat{G}(k+p) = C\hat{x}(k+p) \tag{4.19}$$

where meal and insulin announcement was assumed at least T_{PH} minutes ahead, implying that $u(k+l)$ was known for all $0 \leq l \leq p$. For validation purposes, the predictions were evaluated on the entire feasible data set, by means of RMSE and by Clarke Error Grid Analysis (p-CGA, see below), for prediction horizons $T_{PH} = 20 - 120$ min ($p = 4 - 24$).

The Clarke Error Grid Analysis The Clarke Error Grid Analysis (p-CGA) [Clarke *et al.*, 1987] is a metric originally developed to evaluate blood glucose meters, relating the measurement error to clinical implications. This metric is also often used to rate CGM precision, and recently to assess prediction performance as well. It will be used in this aspect in this and the coming Chapters 5 and 6. Estimated glucose is plotted against the reference measurements and evaluated according to how the points fall into the different error zones, each with a different clinical interpretation, see Fig. 4.3. For comparative purposes, the trivial zero-order hold predictor (ZOH), $G(t+T_{PH}) = G(t)$, was used as a reference of a non-informative, minimum performance predictor. Predictions with similar or worse performance than the ZOH, measured as

$$\gamma = \frac{\text{RMSE}_{GIIM}}{\text{RMSE}_{ZOH}} \tag{4.20}$$

are not providing any predictive value and were thus considered as flawed.

The ability to detect hypoglycemic events beforehand was also assessed. Hypoglycemia was defined as glucose values below 72 mg/dl for at least 20 minutes. The hypoglycemic episode was considered to have ended when euglycemia was restored, here defined as glucose level above 100 mg/dl. The hypoglycemia was considered detected by the predictor, at time T_{th}, when the predicted value was below the hypoglycemic alarm threshold G_{th}, and the same alarm lasted until the predicted value was above 100 mg/dl for at least 20 minutes. Assessment was performed by calculating the sensitivity, the rate of false alarms and

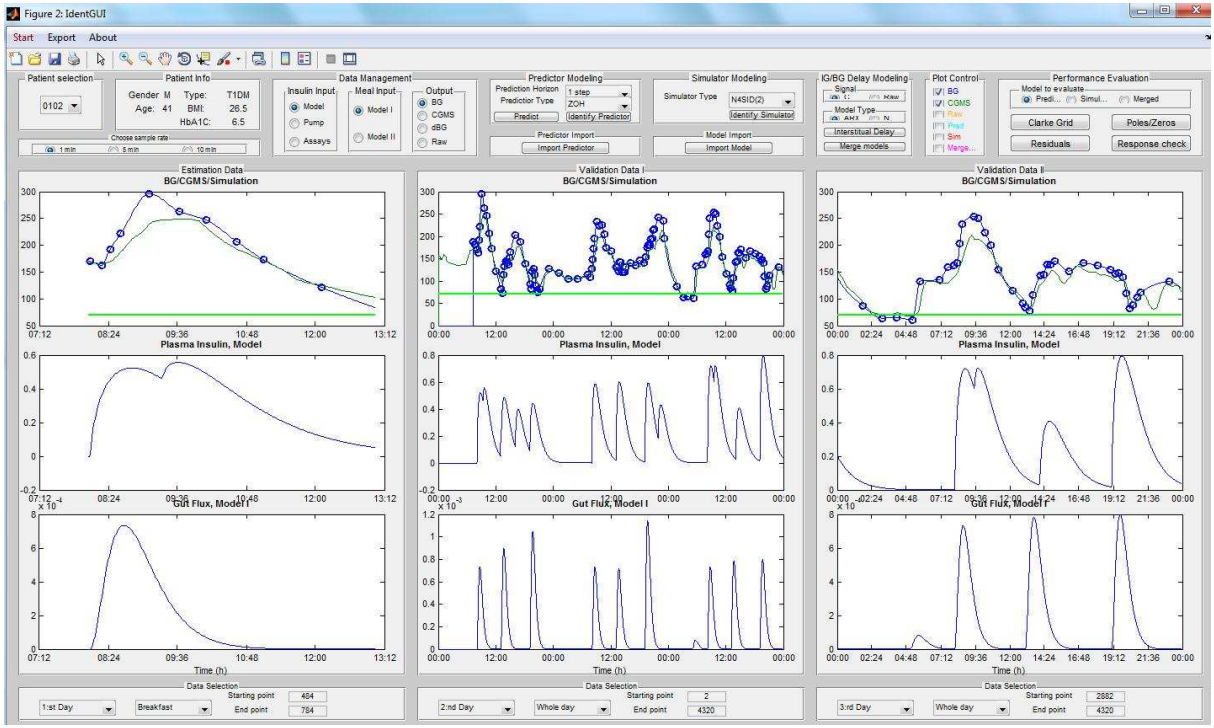

Figure 4.2 A Graphical User Interface (GUI) developed in order to facilitate manual model identification. Patient specific data can be loaded from each of the DIAdvisor trials and displayed in three different columns of data windows. The top windows depict glucose data, where the blue circles correspond to blood glucose reference values, the blue line represents the splined interpolation of these values G and the dark green line is the interpolated Continuous Glucose Measurements (G_{CGM}). The light green bar corresponds the hypoglycemic threshold, 72 mg/dl. The middle plots show the plasma insulin level I_p, given by the Insulin Sub Model (ISM), derived from basal and bolus doses. The lower plots describe the corresponding results from the Glucose Sub Model (GSM), yielding the rate of appearance of glucose R_a following meal intakes. Different types of models can be evaluated by changing in the scroll-down menu in the header of the GUI, and previous developed models can also be imported for comparative purposes. Model evaluation plots can be requested using the push buttons in the upper right corner of the GUI header.

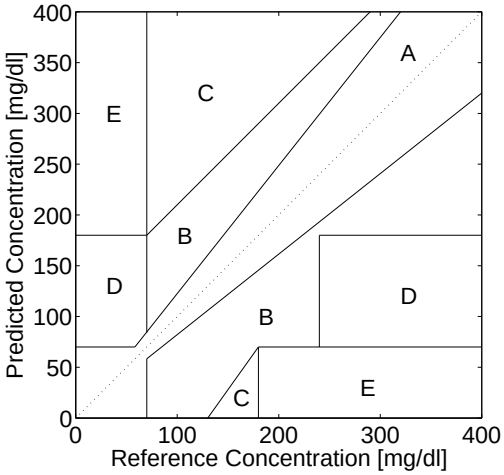

Figure 4.3 The Clarke Error Grid [Clarke *et al.*, 1987] is divided into different zones; the A zone corresponds to error of little clinical significance, the B zone represent values that deviate more than 20% from the reference, but would lead to benign or no treatment decisions if acted upon, the C zone error could result in overcorrection, the D zone represents failure to detect dangerously low or high glucose values and zone E corresponds to predictions that would lead to erroneous and dangerous treatment decisions (e.g. administrating insulin when already hypoglycemic).

the warning time for each prediction horizon. The warning time was calculated for the issued alarms,

$$T_{warn} = T_{72} - T_{th} + T_{PH} \tag{4.21}$$

where T_{72} is the time instant when the glucose level first drops below 72 mg/dl. The sensitivity is defined as

$$\text{Sensitivity} = \frac{TP}{TP+FN} \tag{4.22}$$

where TP corresponds to the true positive alarms, i.e., the number of times when an alarm was triggered before an event when the glucose level dropped below 72 mg/dl. FN, false negatives, correspond to that the predictor missed the hypoglycemic event or raised an alarm too late, i.e., if no positive warning time T_{th} is

provided. The false alarm ratio ρ was defined as:

$$\rho = \frac{FP}{TP+FP} \tag{4.23}$$

where, the false positives FP correspond to the number of times the predicted glucose level $\hat{G}(k)$ was below G_{th}, while the reference blood glucose $G(k)$ was above 72 mg/dl.

Input Impact Modification

All models identified in this manner will have a pole close to the integrating pole at $z = 1$ on the unit circle. Not surprisingly they are all slightly off, since the data series is finite, and it only takes a small perturbation, to shift the pole in either direction of '1'. This means that the total stationary impact from either input goes to zero (or explodes, if the model is unstable), which of course is non-physiological for a diabetes patient. In order to have models suitable not only for short-term prediction, but also for longer predictions, the total impact over longer time horizons should be correct. Thus, the integrating term has to be fixed. Simply moving the closest pole to '1' would alter the gain and dynamics in an unacceptable way.

Physiologically qualitatively correct models, incorporating the necessary integrating term, are retrieved based on the models identified by the subspace algorithm as follows for the second order system. New poles of the dynamics are determined from the characteristic polynomial as

$$(z - p_{new})(z-1) = (z-p_1)(z-p_2) \tag{4.24}$$

where p_1 and p_2 represent the existing poles, and p_{new} is the second pole in the modified model. The equation lacks solutions, and p_{new} is approximated as the average of the second and third coefficient of the characteristic polynomial:

$$p_{new} = \frac{p_1 + p_2 + p_1 p_2 - 1}{2} \tag{4.25}$$

The new poles are introduced by transforming the system into the companion canonical form and replacing the terms of the characteristic polynomial in the right column of the new A-matrix.

Now, in order to determine the new B matrix, the following constrained regularized Least Squares (LS) optimization is undertaken

$$\min_{b} \quad ||Y - \Lambda X||_2 + \sum_{k=1}^{3} ||H(k) - CA^{3k}B||_2 \tag{4.26}$$

$$X = \Phi X + \Gamma U \tag{4.27}$$

$$G_{i,j} = CA^{i}B(:,j) \tag{4.28}$$

$$G_{i,1} < G_{i-1,1} < 0, \quad i \in \{1,\ldots,20\} \tag{4.29}$$

$$G_{i,2} > G_{i-1,2} > 0, \quad i \in \{1,\ldots,20\} \tag{4.30}$$

where $Y = [y_1 \ldots y_N]^T$ is the 24-hour CGM estimation record used to identify the GIIM, $X = [x_1 \ldots x_N]^T$ is the corresponding stacked state matrix, $U = [u_1 \ldots u_N]^T$ is the vectorized input record, $H(1)$, $H(2)$, $H(3)$ is the 15-, 30- and 45-minute impulse response of the GIIM, $G_{i,j}$ is the i:th term of the impulse response from input j of the modified model, and b is the vectorized B matrix. Φ, Γ and Λ represent the dynamical relationship of the model as:

$$\Phi = \begin{bmatrix} 0 & \ldots & 0 \\ A & 0 & \ldots \\ 0 & A & \ddots \end{bmatrix}_{n\cdot N, n\cdot N} \quad \Gamma = \begin{bmatrix} B & \ldots & 0 \\ 0 & B & \ldots \\ 0 & \ldots & \ddots \end{bmatrix}_{n\cdot N, n\cdot N} = \Pi \cdot b \tag{4.31}$$

$$\Lambda = \begin{bmatrix} C & \ldots & 0 \\ 0 & C & \ldots \\ 0 & \ldots & \ddots \end{bmatrix}_{N, \cdot n\cdot N} \tag{4.32}$$

where Π is a suitable transformation matrix. The regularization allows for a sound balance between approximating the model to the GIIM initial response (the first 45 minutes) and fitting the long term impact to the data. The constraints guarantee qualitatively correct impulse responses. Here, constraints over 100 minutes, $i = 20$ in (4.29) and (4.30), was considered long enough to guarantee sign-correct impulse responses.

4.6 Results

The selected model order was $n = 2$ for all models. The summarizing results of the predictive performance for 20-, 40-, 60- and 120-minute ahead prediction

for the automatically identified GIIM in terms of RMSE are found in Tables 4.2. The corresponding results for the p-CGA is found in Tables 4.3, 4.4, 4.5. In Fig. 4.4, an example of a 40-minute ahead prediction can be seen, and in Fig. 4.5 the relative performance in comparison to the ZOH predictor is presented. The performance slowly deteriorates as the prediction horizon increases, and seems to converge to a value of 0.7 for long prediction horizons. The predictions were similar when the manually identified GIIM models or the modified models were used.

Table 4.2 Mean RMSE prediction results for the 20-, 40-, 60- and 120-minute ahead predictions [minimum and maximum values]. Automatically identified models.

	RMSE [mg/dl]		
T [min]	**Montpellier**	**Padova**	**Prague**
20	9 [6-13]	8 [5-11]	9 [4-11]
40	17 [13-23]	15 [10-21]	18 [9-23]
60	25 [18-33]	21 [14-31]	24 [14-33]
120	40 [20-57]	35 [26-49]	37 [25-58]

Table 4.3 Montpellier mean p-CGA results for the 20-, 40-, 60- and 120-minute ahead predictions [minimum and maximum values]. Zone A and B are presented separately, but the erroneous zone C and the dangerous zones D and E were lumped together. Automatically identified models.

	Montpellier		
T [min]	**A [%]**	**B [%]**	**CDE [%]**
20	98.6 [95.1-100]	1.0 [0-4.2]	0.4 [0-1.6]
40	94.6 [83.7-99.7]	4.4 [0.4-15.9]	0 [0-3.3]
60	91.6 [79.0-99.3]	7.1 [0.7-17.9]	0 [0-4.3]
120	87.8 [65.9-99.3]	9.6 [0.7-30.5]	0.1 [0-14.7]

Depending on the length of acceptable data, between 9 and 24 models were automatically identified for each patient. On average, about 80% of the models

Table 4.4 Padova mean p-CGA results for the 20-, 40-, 60- and 120-minute ahead predictions [minimum and maximum values]. Zone A and B are presented separately, but the erroneous zone C and the dangerous zones D and E were lumped together. Automatically identified models.

	Padova		
T [min]	**A [%]**	**B [%]**	**CDE [%]**
20	99.1 [96.4-100]	0.5 [0-2.8]	0.3 [0-2.6]
40	96.1 [92.3-99.7]	2.6 [0.3-7.1]	1.3 [0-3.8]
60	93.5 [86.5-99.3]	4.5 [0.9-11.1]	2.0 [0-5.3]
120	89.9 [80.0-99.3]	6.2 [1.6-11.4]	3.9 [0-11.7]

Table 4.5 Prague mean p-CGA results for the 20-, 40-, 60- and 120-minute ahead predictions [Minimum and maximum values]. Zone A and B are presented separately, but the erroneous zone C and the dangerous zones D and E were lumped together. Automatically identified models.

	Prague		
T [min]	**A [%]**	**B [%]**	**CDE [%]**
20	99.2 [97.6-100]	0.6 [0-1.4]	0.2 [0-1.2]
40	96.0 [90.6-98.7]	3.0 [0.8-5.5]	1.1 [0-3.8]
60	93.1 [87.3-98.4]	5.2 [1.2-10.1]	1.7 [0-4.2]
120	91.9 [83.7-97.4]	6.5 [2.6-13.6]	1.6 [0-6.2]

fulfilled the screening criteria, but with large interpersonal differences. For two patients only one, out of all evaluated models, was acceptable, whereas for 21 patients, all models were satisfactory in terms of qualitative response.

In Fig. 4.6, a typical example of the model response for meal and insulin can be seen for both the modified and the original model derived from the automatically identified model. The modified model responses follow the original models closely up until about 100 minutes after insulin or meal intake.

The values of the stationary estimated insulin and carbohydrate impact levels are found in Fig. 4.7. The distribution of the ratio between these was estimated by a log-normal distribution and by the Parzen estimate method using a Gaussian

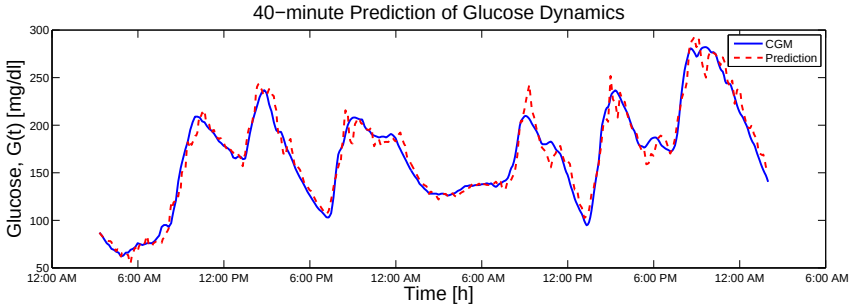

Figure 4.4 Example of a 40-minute ahead prediction $\hat{G}$ compared to the reference CGM signal G_{CGM}.

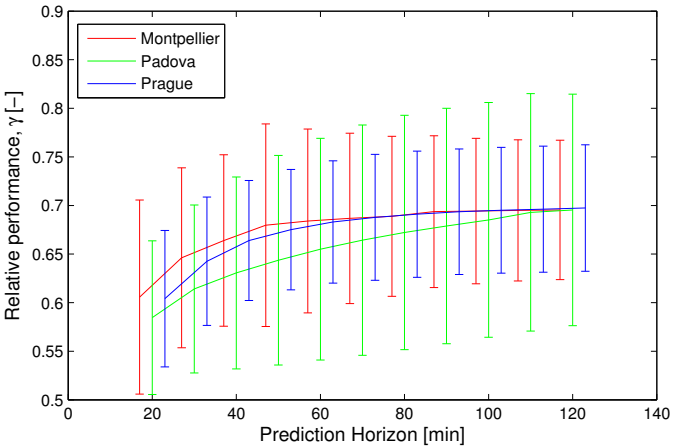

Figure 4.5 Mean relative performance in comparison to the ZOH predictor including standard deviations.

kernel ($\sigma = 2.0$) [Bishop, 2006]. The results are found in Fig. 4.8. Similar results were achieved when the modified models were derived from the automatically and the manually identified models.

The Carbohydrate-to-Insulin Ratio (CIR) was estimated from the modified models, and is a common metric in clinical diabetological practice, used to estimate the insulin need for different meals [Davidson *et al.*, 2008]. The actual CIR was calculated for each individual, by dividing the summarized amount of

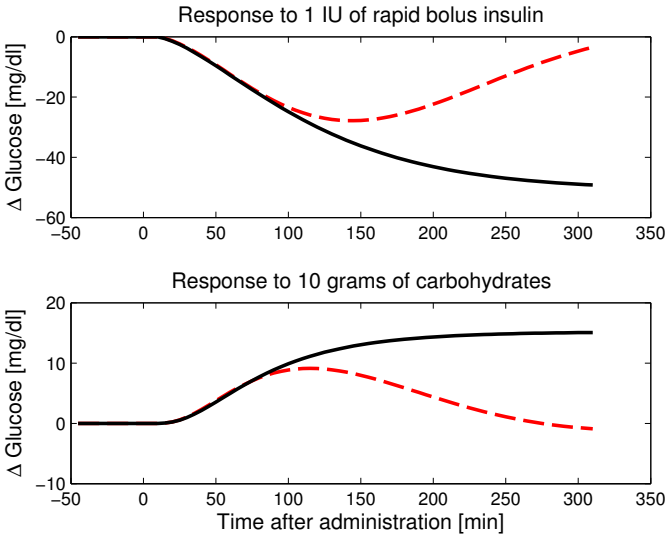

Figure 4.6 Example of response to 1 Insulin Unit (IU) of rapid insulin and 10 g of carbohydrates. Black curve: Modified model, Red dotted curve: Original prediction model (automatically identified).

digested carbohydrates and insulin over the study. In Fig. 4.9, the actual and the estimated CIR were compared. The similarity is quite good (regression: $y = a \cdot x, a = 1.05 \pm 0.05$, $R = 0.8, p < 0.05$), apart for some outliers and for some of the Prague patients, where there is a clear bias.

The dynamical aspect of the insulin and carbohydrate impact described by the time constants τ_{Ins} and τ_{Carb} (corresponding to the time it takes to reach 63% of the stationary level), as determined by the modified model, are illustrated in Fig. 4.10. The time constants were heavily distributed on $\tau_{Ins} = 110$ minutes and $\tau_{Carb} = 70$ minutes. There was no principal difference in distribution between the three sites.

The data were reviewed for hypoglycemic events. In total, 57 (35 Montpellier, 22 Padova, 0 Prague) hypoglycemic events were found for 21 patients (13 Montpellier, 10 Padova, 0 Prague), when reviewing the blood glucose reference records. Due to the low number of blood glucose samples collected at the Prague site, no interpolation could be undertaken, and no analysis in terms of hypoglycemic detection could therefore be performed from those data records.

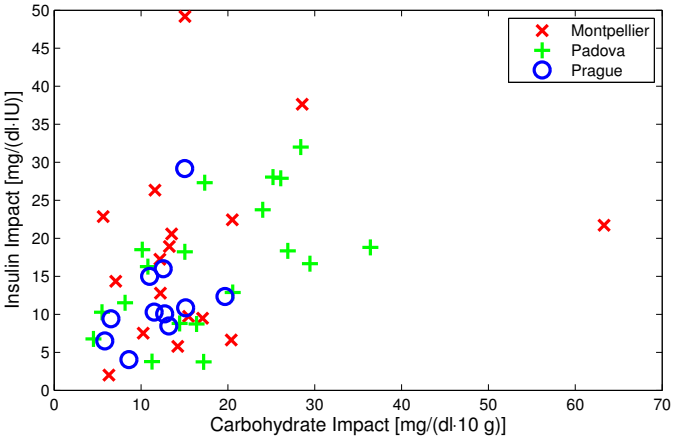

Figure 4.7 Levels of stationary carbohydrate and insulin impacts.

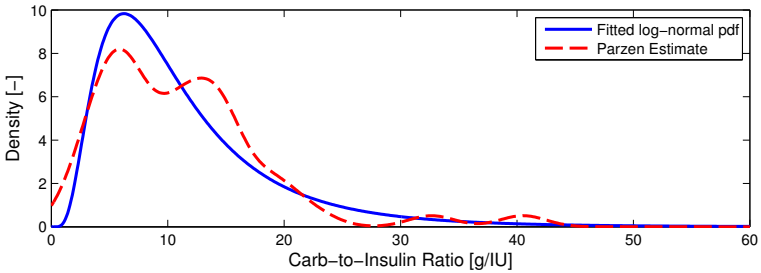

Figure 4.8 Distribution of Carbohydrate-to-Insulin Ratio (CIR) for the data from all three sites.

Four different settings of the alarm threshold were analysed; 72, 78, 84 and 90 mg/dl, for prediction horizons $10, 15, \ldots, 60$ min of the underlying predictors. In Fig. 4.11 and 4.12, the sensitivity, false alarm ratio and warning time are plotted for every combination of threshold level and prediction horizon for the Montpellier and the Padova sites. The Montpellier results shown are based on the automatically identified models, and the Padova results have been retrieved using the manually identified models.

For the Montpellier patients, the automatically identified models were slightly better than the manually identified models. A threshold of $G_{th} = 90$ mg/dl and

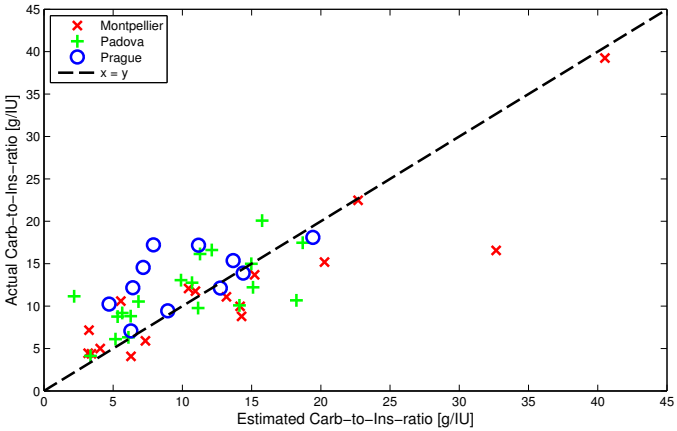

Figure 4.9 Actual versus estimated Carbohydrate-to-Insulin Ratio (CIR).

$T_{PH} = 40$ gave a sensitivity of 70%, a false alarm rate of 49% and a warning time of 24 minutes. On the contrary, the manually identified models were the better choice in this aspect for the Padova patients. A threshold of G_{th} =90 mg/dl and $T_{PH} = 20$ gave a sensitivity of 76%, a false alarm ratio of 47% and a warning time of 34 minutes. Using the ZOH, the sensitivity, false alarm ratio and warning time results can be found in Tables 4.6 and 4.7.

Table 4.6 Performance metrics for the hypoglycemic detection using the comparative ZOH model, triggered at the different alarm threshold levels G_{th}. Montpellier patients

	G_{th} [mg/dl]			
	72	78	84	90
Sensitivity [%]	3	15	17	37
False Alarm Ratio, ρ [%]	89	63	81	79
Warning time [min]	10	8	12	8

The relationship between level of insulin antibodies and time constant of the ISM was investigated, with no positive outcome, see Fig. 4.13.

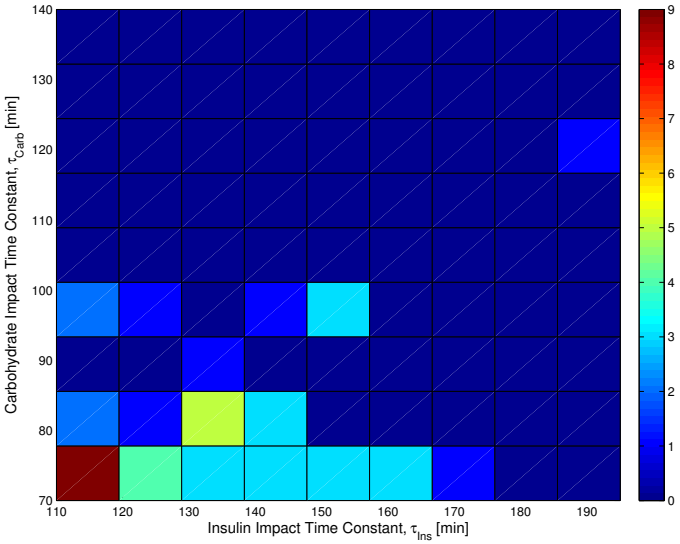

Figure 4.10 Distribution of Carbohydrate and Insulin Time constants. The color of the boxes, corresponding to different combinations of carbohydrate and insulin impact constants, represents is described by the

Table 4.7 Performance metrics for the hypoglycemic detection using the comparative ZOH model, triggered at the different alarm threshold levels G_{th}. Padova patients

	G_{th} [mg/dl]			
	72	78	84	90
Sensitivity [%]	0	15	31	34
False Alarm Rate, ρ [%]	100	78	80	82
Warning time [min]	N/A	5	12	12

4.7 Discussion

Prediction

The selected linear model structure of the GIIM is in many aspects a crude approximation to known non-linearities, e.g., in terms of the inverse relationship

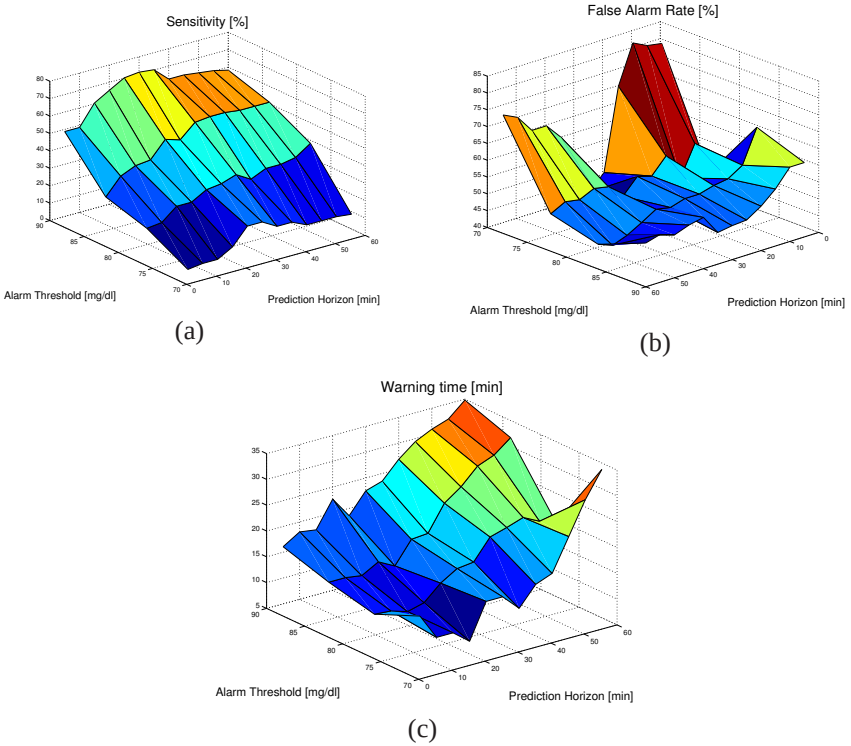

Figure 4.11 Average hypoglycemic sensitivity (a), specificity (b) and warning time (c), using different thresholds and prediction horizons. Montpellier patients.

between insulin sensitivity and glucose level [Chan *et al.*, 2010], and different aspects of time-variability, see Chapter 3. Furthermore, the generic parameters used in the sub models are also non-optimal, and can be expected to reduce the predictive performance (see discussion on the insulin sub model's influence on the input response further down). However, these choices are to some extent driven by the available data. On average, the investigated patient records contained 58 hours of data, or about 700 samples collected with a 5-minute sampling period. Considering splitting these data into estimation and validation data, further reduces the available data for identification—setting a limit for the parameter complexity of the model. The sub models were chosen in order to transform

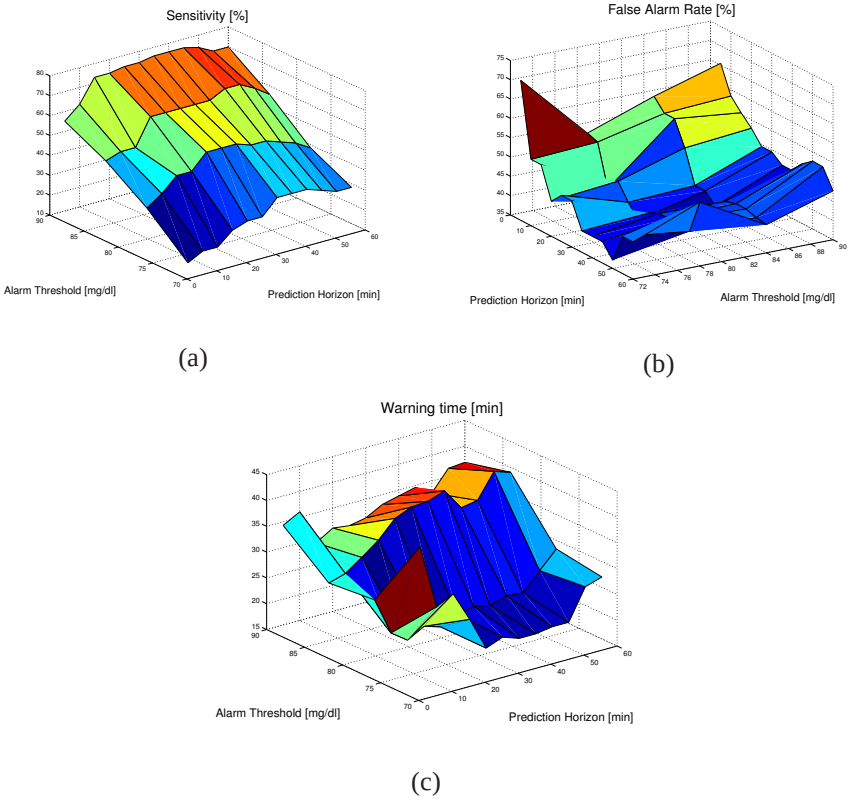

Figure 4.12 Average hypoglycemic sensitivity (a), specificity (b) and warning time (c), using different thresholds and prediction horizons. Padova patients.

the meal and insulin impulses into a continuous signal to be usable in the chosen state-space framework. Individual parametrization of the sub models, even on the individual meal composition, see e.g. [Cescon *et al.*, 2009], would have been preferred, but the modeling errors introduced by the generic parameters in these sub models can to some extent be handled by the GIIM model, following the sub models in a cascade-like manner. Furthermore, the protocol lacked identifiability concerns in terms of that it was not designed to, e.g., provide excitation to the system over the entire glucose range, to excite the system with a single input

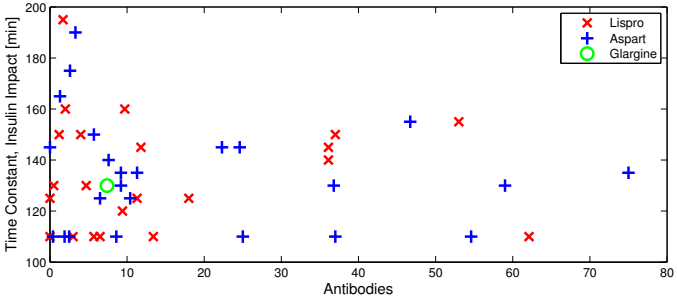

Figure 4.13 Time constant of the insulin impact in relation to the level of antibodies.

channel separately, or to test a wider range of input magnitudes. This, together with the short data set, makes assessment of time-variability or non-linear effects very difficult, and should therefore not be addressed in a first modeling attempt.

The aspects outlined above may also be responsible for the variable degree of identifiability, experienced for some data segments of the patient data. Non-linearity, time-variability and large inter-variability between meal responses may be perceived as data inconsistencies to the model. However, on the positive side, it should be noted that the patient data contained extra snacks and insulin corrections during the three day course (on average 5 meals and 4 bolus doses in total over a day). These unscheduled inputs, often separated in time, provided some extra excitation to the system identification.

Despite these short-comings in terms of modeling, the predictive quality of the simplified model is quite good for short prediction horizons. The Clarke Grid evaluation (Tables 4.3, 4.4, 4.5) indicates that clinically acceptable results (here defined as $A+B > 98\%$) in general could be achieved for prediction horizons up to 60 minutes.

Comparing the three clinical sites, no significant difference in predictive quality was found. However, it was found that the predictive performance was significantly better for the MDI patients than for the pump patients (0.05 significance level, $\forall T \leq 90$ min, when Montpellier and Padova where considered, $\forall T \leq 40$ min, when the Prague patients were included. Data not shown).

The main difference between these two groups, with regards to modeling, is in terms of the insulin signal. Here, the basal dose is included for the pump patients, whereas the insulin signal used for the MDI patients only take bolus doses from the rapid-acting insulin into account. The basis for this choice was that the

action profiles of the long-acting insulins used by the MDI patients (glargine and detemir) are very flat, with low dynamic impact, and is not expected to directly influence the short-term dynamics investigated. The basal dose for pump patients on the other hand, utilizes the same rapid insulin as used for bolus purposes. The short action profile of the rapid insulins allows for sudden changes in basal dose, a feature that often is exploited to optimize the 24-hour basal regime for these patients. The fast response to these changes is a strong argument for including the basal level in the insulin signal for the pump patients. However, a difficulty in modeling and identifying pump patients with shifting basal level is to establish the true reference level, i.e., the basal level corresponding to a stationary glucose level when no other external inputs affect the system. An argument for not including the basal level for pump patients is that the intention of the shifting basal therapy regime is to match the circadian rhythm of the metabolism [Van Cauter *et al.*, 1997b], and that the selected basal dose thus matches the true reference. In reality, mismatches are expected. Considering these aspects, the models were identified both with, and without, the basal level, but without principal difference in result. The underlying reason for better predictive performance of the MDI patients thus remains unknown.

Comparing the results to previously published predictors is not always easy due to the use of different evaluation metrics. However, below some comparable results have been found. In [Finan *et al.*, 2009a] both batch-wise and recursively identified patient-specific ARX models have been analysed for 9 patients and compared to a ZOH predictor for 30-, 45- and 60-minute ahead prediction. The corresponding mean relative performance was 0.91 for all three prediction horizons, and the absolute RMSE values are summarized in Table 4.8 together with results for the following studies. A neural network approach was utilized in [Zecchin *et al.*, 2012] for 30-minute predictions on 9 subjects, and support vector regression was applied to the problem in [Georga *et al.*, 2011] for 7 patients. In this latter study, the results were also evaluated with the p-CGA, see Table 4.9. Comparing these results with the results from this study in Tables 4.2, 4.3, 4.4 and 4.5, show that the results achieved are competitive.

For hypoglycemic detection, a sound balance between sensitivity, false alarm rate and maximum warning time is crucial. The results show a very high false alarm rate. However, this metric depends on the binary counting of whether the blood glucose level passed the hypoglycemic threshold, which makes it sensitive to small changes. As it turns out, many of the false alarms were close misses. The mean value of the blood glucose reference at these instances, was 88 and 93 mg/dl for the Montpellier and the Padova site, respectively, with many values

Table 4.8 Comparative values of previous publications.

RMSE [mg/dl]						
Publication		**Prediction Horizon [min]**				
	Nr of subjects	15	30	45	60	120
[Finan *et al.*, 2009a]	9	N/A	26	34	40	N/A
[Zecchin *et al.*, 2012]	9	N/A	14	N/A	N/A	N/A
[Georga *et al.*, 2011]	7	9.6	16.2	N/A	24.9	35.8

Table 4.9 Comparative Clarke values [%] from [Georga *et al.*, 2011].

Zone	**Prediction Horizon [min]**			
	15	**30**	**60**	**120**
A	98.8	92.5	80.0	62.9
B	1.1	7.0	18.5	33.7
C	0	0	0.1	0.4
D	0.1	0.5	1.4	3
E	0	0	0	0

close to the hypoglycemic threshold. Considering these results, the high false alarm rate should not be overstated, and comparing the predictor alarms' overall results to the corresponding results for ZOH predictor, significant improvements are apparent. Furthermore, it should be borne in mind that the low incidence of hypoglycemic events has a high impact on the estimate of these performance metrics. Considering the strong amplification of the insulin action in the hypoglycemic zone, reported in [Chan *et al.*, 2010], a specialized low blood glucose model may well improve these results. However, the hypoglycemic episodes of the data sets used here are short, and are shadowed by the data in euglycemia and hyperglycemia in the identification process.

Long-term Impact Response

It is interesting to note that the impact responses of the GIIM models are similar to the modified models up until about 100 min (Fig. 4.6). This is sufficient to ensure short-term predictions within this range, as the prediction results indicate. However, the entanglement of the inputs after this point makes the models unsuitable for control strategies dependent on long-term impact, such as manual control, and the modified models should be employed instead.

The estimated time constants of the model from each input were heavily distributed in the lower end at $\tau_{ins} = 110$ and $\tau_{Carb} = 70$ min, see Fig. 4.10. Looking at the dynamics of the sub models in section 4.4, it can be concluded that the ISM has a time constant close to 110 minutes. Thus, it seems likely that the generic model was too slow for some of the patients. The GSM, evaluated in the same manner, has a time constant of 58 min, which is lower than 70 min, and this sub model is therefore not subject to the same problem.

Case studies suggest that high levels of insulin antibodies could have a negative effect on glucose regulation in insulin-dependent diabetic patients, such as post-prandial hyperglycemia, followed by, in some cases, hypoglycemia, long time after the expected duration of the insulin action [Van Haeften, 1989]. The causal pathways to such an influence would be the binding of insulin, resulting in prolonged, and initially dampened, insulin impact. These adverse effects have not been found in large cross trial analysis, [Lindholm *et al.*, 2002], but the numerous case studies suggest that the problem can be significant for some individuals, see e.g. [Hirata and Uchigata, 1994]. The individual specific conditions explaining the case reports remain to the revealed. If such binding effect was present in the data at hand, the estimated insulin impact would possibly have a large time constant for patients with high antibody levels. However, the results, see Fig. 4.13, suggest the opposite—the large estimated time constants belong to the patients with low antibody levels, possibly suggesting that most of antibodies identified in our patients are of high affinity for insulin, thus forming stable immune-complexes, which are not prone to on/off binding to insulin, and consequently have minimal effect on insulin action.

4.8 Conclusion

In this study, patient-specific models were identified for 47 different patients. The predictive quality was competitive to previously published results, and it

was found that there is a significant difference in prediction quality between the MDI and pump patient groups for short prediction horizons. The reason for this is unknown. No difference was found between the different clinical sites, but large interpersonal differences. Further in-depth analysis should be undertaken to investigate whether any stratifications are possible based on basic patient characteristics, as those collected in the study, see Chapter 3. Hypoglycemic alarm triggering was evaluated with an average sensitivity of 73%, 28 minutes in advance with a 48% false-alarm rate. However, the average glucose value when the false alarms were raised was 90 mg/dl, with many values close to the hypoglycemic threshold. Thus, most of the false alarms were near misses, reducing the significance of the high false alarm rate. The models provide a significant improvement in the possibility to detect hypoglycemia in advance, in comparison to relying on CGM data alone. Furthermore, the estimated impacts of the modified models seem plausible and the estimated Carbohydrate-to-Insulin Ratio (CIR) matched the true ratio well. Overall, the results indicate that the used modeling and identification approach may prove useful for short-term prediction, utilized in a decision support system or an artificial pancreas.

5

Augmented Model Incorporating Sensor Dynamics

Predicting the glucose level in a real-time setting means relying on CGM data. In Chapter 4 the delay between the blood glucose and the measured interstitial glucose level was ignored, and the CGM signal was used as a proxy for blood glucose. Actually, this is the most common way of glucose modeling and prediction, and applies to all the models listed in Chapter 4. However, in many cases there is a significant lag between the interstitial glucose and the blood glucose due to physiological and sensor dynamics [Keenan *et al.*, 2009]. Ignoring this delay in the modeling implies corresponding delays in the prediction, an aspect of special importance during falling glucose levels and impending hypoglycemia, when an hypoglycemic alarm, based on the prediction, could warn the patient and instigate corrective actions. For an assessment of the delay between these signals in the data at hand, see Chapter 3. The capillary and sensor characteristics of the finger-stick measurement sensors are, in this context negligible, and are generally disregarded (and the delay is indicated to be small [Dye *et al.*, 2010]). In this chapter, the GIIM, M_1 in Fig. 5.1, and the interstitial and sensor dynamics (here treated as one model M_2, see [Boyne *et al.*, 2003] for a brief discussion on the contribution of each term to the delay) are identified separately, and thereafter merged together into one single grey-box model. Using an observer, the blood glucose evolution is predicted ahead, based on the raw sensor output.

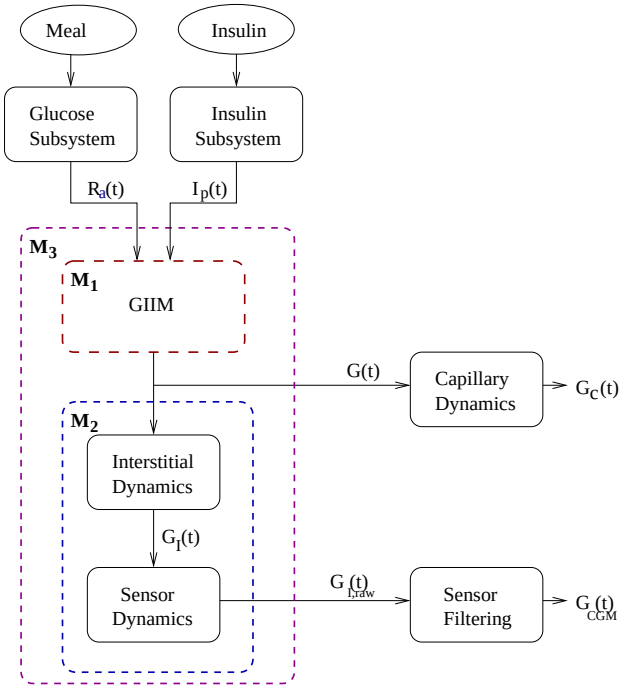

Figure 5.1 Overview of the modeling approach. Notation: Plasma Insulin $I_p(t)$, Rate of Glucose Appearance following a meal $R_a(t)$, Blood glucose $G(t)$, Capillary glucose $G_C(t)$, Interstitial Glucose $G_I(t)$, CGM raw current signal $G_{I,raw}(t)$ and CGM signal $G_{CGM}(t)$. M_1 represent the model describing the glucose-insulin interaction in the blood and inner organs (GIIM), the M_2 model represents the diffusion-like relationship between blood and interstitial glucose and the CGM sensor dynamics, and M_3 is the joint model of M_1 and M_2. Same figure as Fig. 4.1.

The interstitial and CGM sensor dynamics have been investigated assuming a first-order diffusion model in [Kovatchev *et al.*, 2006] and [Facchinetti *et al.*, 2007]. In [Facchinetti *et al.*, 2007], the blood glucose level was recovered from the CGM signal using deconvolution, and in [Bequette, 2004] an early attempt at observer-based estimation was presented. In [Leal *et al.*, 2010] a third order Box-Jenkins model was used to estimate the glucose level from the raw sensor signal. However, so far (to the best of the author's knowledge) no attempts have been made on merging all the modules together for the purpose of blood glucose prediction.

5.1 Data

Based on the assessment of blood-to-interstitial delay in Chapter 3 and data completeness, one patient was chosen from the total DAQ data set. In order to show significant results, a patient with large lag was chosen (patient 107 from Montpellier).

Signals

The HemoCue measurements were interpolated using a shape preserving spline interpolation method (pchip in Matlab [MathWorks, 2012]) to retrieve an equidistant sampled signal $G(t)$ with sampling period 5 minutes.

Apart from the $G_{CGM}(t)$ signal (10 min sampling rate), an intermediate signal $G_{I,raw}(t)$ from the glucose sensor was collected (1 min sample rate). The signal, corresponding to the electrical current measured by the sensor, was normalized to the same amplitude as the blood glucose data and resampled to a 5 minute basis, and was used in the identification instead of the CGM signal $G_{CGM}(t)$.

5.2 GIIM Modeling - M1

Denoting the blood glucose $G(t_k)$, at sampling time t_k, with $y(k)$, and the raw CGM signal $G_{I,raw}(t_k)$ with $z(k)$:

$$\zeta = \begin{bmatrix} y(k) \\ z(k) \end{bmatrix} = \begin{bmatrix} G(t_k) \\ G_{I,raw}(t_k) \end{bmatrix} \tag{5.1}$$

and the filtered inputs $u_k = [I_p(k) \quad R_a(k)]^T$, the GIIM is modeled with a discrete-time state space model $\mathbf{M_1}$.

$$x(k+1) = A_1 x(k) + B_1 u(k) + \omega(k) \tag{5.2}$$

$$y(k) = C_1 x(k) + \upsilon(k) \tag{5.3}$$

where $x(k) \in \mathbb{R}^n$ is the state vector and ω is process noise and υ is the fingerstick measurements noise with covariances:

$$E\left\{ \begin{pmatrix} \omega \\ \upsilon \end{pmatrix} \begin{pmatrix} \omega \\ \upsilon \end{pmatrix}^T \right\} = \begin{bmatrix} Q_1 & 0 \\ 0 & R_1 \end{bmatrix} \tag{5.4}$$

The model order was determined using the Akaike criterion [Johansson, 2009].

5.3 Interstitial and Sensor Model - M2

The dynamics between blood glucose z and interstitial glucose y, as measured by the sensor, was modeled as an ARX process.

$$A(z) \cdot z(k) = B(z) \cdot y(k-d) + e(k) \tag{5.5}$$

where A, B are polynomials of the zero-order-hold operator z, d is a delay, and $e(k)$ is the CGM measurement noise. The model orders n_A, n_B and d evaluated for values according to Table 5.1 are determined using the MDL criterion. The choice of evaluated model orders covers the compartment model suggested in [Rebrin and Steil, 2000a].

Table 5.1 Evaluated model orders

Parameter	Value
n_A	1-2
n_B	1-2
d	1-4

5.4 Model Merging - M3

Converting the sensor ARX model into a state-space model $\mathbf{M_2} : \{A_2, B_2, C_2\}$ with process and measurement noises Q_2 and R_2, the GIIM and sensor models are merged into one model $\mathbf{M_3} : \{A_3, B_3, C_3\}$, with the augmented state vector ξ and the output ζ.

$$A_3 = \begin{bmatrix} A_1 & 0_{[n_{A_1} \times n_{A_2}]} \\ B_2 \cdot C_1 & A_2 \end{bmatrix}, \quad B_3 = \begin{bmatrix} B_1 \\ 0_{[n_{A_2} \times 2]} \end{bmatrix} \tag{5.6}$$

$$C_3 = \begin{bmatrix} C_{31} \\ C_{32} \end{bmatrix} = \begin{bmatrix} C_1 & 0_{[1 \times n_{C_2}]} \\ 0_{[1 \times n_{C_1}]} & C_2 \end{bmatrix} \tag{5.7}$$

$$Q_3 = \begin{bmatrix} Q_1 & 0 \\ 0 & Q_2 \end{bmatrix}, \quad R_3 = \begin{bmatrix} R_1 & 0 \\ 0 & R_2 \end{bmatrix} \tag{5.8}$$

5.5 State Estimation and Sensor Fusion

Data is available at different rates from the two measurement devices, and at least from the finger-stick measurements, in a non-equidistant manner. Thus, combinatorially there are 3 (4) possibilities; (1) data from both, (2) Data from HemoCue and (3) Data from the CGM sensor, ((4) No data). This calls for time-varying system of switched dynamics. The boolean variables δ_1 and δ_2 are used to keep track of which signal that is present in the feedback, and the new system becomes:

$$\hat{\xi}(k+1) = A_3\hat{\xi}(k) + K(\zeta(k) - C_3\hat{\xi}(k)) \tag{5.9}$$

$$\hat{\zeta}(k) = \begin{bmatrix} \delta_1 & 0 \\ 0 & \delta_2 \end{bmatrix} C_3\hat{\xi}(k) \tag{5.10}$$

where the time-varying Kalman gain K depends on the unknown covariance of the process noise Q and measurement noises R_1 and R_2. The accuracy of the finger-stick HemoCue glucose monitor [HemoCue Glucose 201+ Analyzer, 2012] has been studied in [Stork *et al.*, 2005], which indicate a standard deviation in the area of 10-15 mg/dl when compared to a state of the art laboratory device (Yellow Spring Instrument [Yellow Springs Instrument, 2012]). The study indicates a linear relationship between noise and glucose level, which is common for glucose meters. No information on the measurement noise of the relatively new Abbott CGM system [Abbott Freestyle Navigator, 2012] has been found, but a standard deviation of 20 mg/dl is not an unrealistic assumption (compare to the BG-CGM deviation in Chapter 3). Also for CGM systems a proportional increase in noise level to the glucose level is found. Current evaluation methods to assess the performance of CGM systems is based on comparing the CGM signal to a blood glucose reference. As the previous discussion shows, the signal to reference deviation incorporates deviation due to the time lag between the signals and does not accurately capture the stochastic variation in the CGM signal. Recent developments in CGM error assessment aim to quantify these error dynamics, but do not address the estimate of CGM variation per se [Clarke and Kovatchev, 2009]. In this thesis, the initial guess for noise level standard deviation was chosen to correspond to 15 mg/dl for the HemoCue device and 20 mg/dl for Abbott CGM. The measurement errors were considered to be uncorrelated.

Given the initial guesses $\hat{Q}_0$ and $\hat{R}_0$, Q and R can be iteratively estimated by first calculating the state estimation sequence $\hat{\Xi}_N = [\hat{\xi}_1 \dots \hat{\xi}_N]$ and the estimation error sequence $\hat{W}_N = [\hat{w}_1 \dots \hat{w}_N]$ from the estimation data $\{Y_N, U_N, \hat{\Xi}_0\}$

[Johansson, 2009].

$$\hat{\Xi}_{N+1} = A\hat{X}_N + BU_N + K(Y_N - C\hat{\Xi}_N) \tag{5.11}$$

$$\hat{W}_N = C\hat{\Xi}_N - Y_N \tag{5.12}$$

Thereafter the covariance estimates

$$S = E\{(\hat{\xi} - \xi)(\hat{\xi} - \xi)^T\}, \quad R = E\{\hat{w}\hat{w}^T\} \tag{5.13}$$

are determined. Given that the sequence is stationary

$$\lim_{N\to\infty} S_N = S \tag{5.14}$$

and

$$\lim_{N\to\infty} R_N = R \tag{5.15}$$

Now $\{A, B, C\}$ may be re-estimated again by recognizing that:

$$\hat{\xi}_{k+1} = (A - KC)\hat{\xi}_k + [B \quad K][u_k^T \quad \gamma_k^T] \tag{5.16}$$

$$\hat{w}_k - \gamma_k = -C\hat{\xi}_k \tag{5.17}$$

Finally,

$$\hat{Q} = S_N - AS_NA^T - KR_NK^T \tag{5.18}$$

$$\hat{R} = CS_NC^T - R_N \tag{5.19}$$

Note that this computation may result in sign-indefinite solutions [Johansson, 2009].

Estimation and Validation

The overnight data between the first and the second day were used together with breakfast meal data from the second day for estimation. It was decided to use overnight data together with meal data, in order to have a data set with sufficient amount of excitation. Using meal data alone is problematic, since both inputs act simultaneously during these circumstances. An assessment of the importance of input excitation to identification using simulated diabetic data sets is made in [Finan *et al.*, 2009b]. The first and third days' breakfasts were used for cross validation. Additionally, to challenge the predictor, all HemoCue measurements were removed from the validation data sets.

5.6 Evaluation Criteria

To evaluate the predictive performance of the model, 20, 40 and 60 minute predictions were considered. The correspondence to the reference HemoCue measurements were assessed using the Clarke Pointwise Error Grid Analysis (p-CGA), see Chapter 3, RMSE and maximum absolute error. The performance was compared to the CGM signal's ability to measure the blood glucose.

5.7 Results

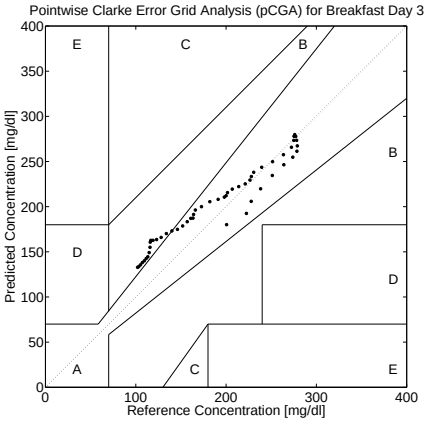

Figure 5.2 Example of Clarke Error Grid Diagram, 40 min prediction Day 3. Prediction versus the interpolated HemoCue blood glucose reference.

First, the GIIM M_1 was identified. Using the interpolated HemoCue data and the meal and insulin sub models to retrieve the filtered inputs, a second-order state-space model was identified using the N4SID command of the System Identification Toolbox in Matlab [MathWorks, 2012]. The model was stable and responded qualitatively correctly to input (not shown). The interstitial mode M_2 was thereafter identified from the interpolated blood glucose data G and the raw CGM signal $G_{I,raw}$. The model order chosen according to the MDL criteria was $n_A = 2$, $n_B = 1$ and $d = 1$. Converting the M_2 model to state space format, the merged model M_3 was retrieved.

Since only CGM data were available in the validation data ($\delta_1 = 0$), the system became time-invariant and a stationary Kalman filter was designed.

Using the initial guess for Q and R produced noisy predictions. The attempt to estimate the noise characteristics from the estimation data broke down into non-positive definite covariance matrices. Instead, Q and R were heuristically chosen to strike a sound balance between signal smoothness and responsiveness to model-to-feedback mismatch.

In Fig. 5.3, the 20, 40 and 60 minutes predictions together with normalized $G_{I,raw}$ signal and the G_{CGM} signal can be seen, and in Fig. 5.2 an example of p-CGA can be seen. All performance metrics have been summarized in Table 5.2.

Table 5.2 Performance evaluation for the M_3 predictor and the G_{CGM} in comparison to the blood glucose reference G on validation data.

Prediction	**p-CGA[%]**			**RMSE**	**max** $\lvert$ **e** $\rvert$
Horizon	A	B	CDE	[mg/dl]	[mg/dl]
20	84.2	15.8	0	19	42
40	84.9	15.1	0	20	46
60	83.7	16.3	0	21	45
G_{CGM}	45.9	51.6	2.5	47	90

5.8 Discussion

Error Analysis

To determine the source of the prediction error, the simulation errors of the sensor model,

$$\varepsilon_z = z - \hat{z} \tag{5.20}$$

and of the GIIM model

$$\varepsilon_y = y - \hat{y} \tag{5.21}$$

were investigated separately. In Fig. 5.4, the simulation error between the simulated raw CGM signal $\hat{G}_{I,raw}$ and the true signal can be seen. The error distribu-

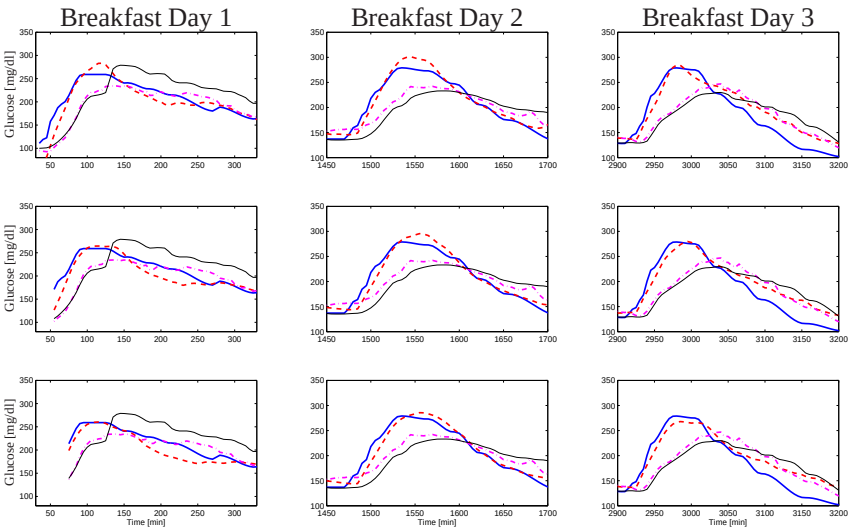

Figure 5.3 Plasma glucose predictions. Interpolated HemoCue measurements (thick solid blue), G_{CGM} (solid black), $G_{I,raw}$ (dash dotted magenta) and $M3$ predictions (dashed red) for estimation data (middle plot) and validation data (left and right plots).

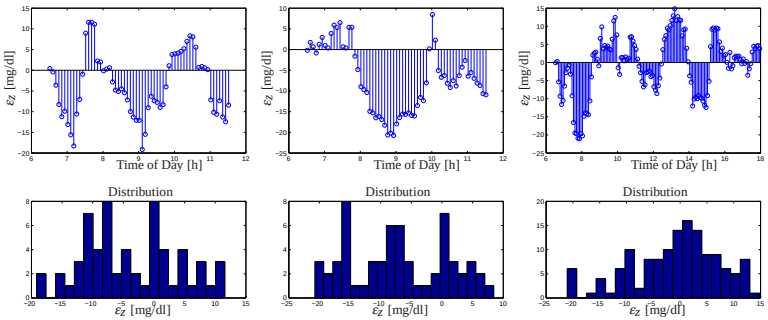

Figure 5.4 Simulation error ε_z of the simulated raw CGM signal $\hat{z}(k)$ given blood glucose $y(k)$ using the sensor model M_2. Estimation data (middle plot) and validation data (left and right plots).

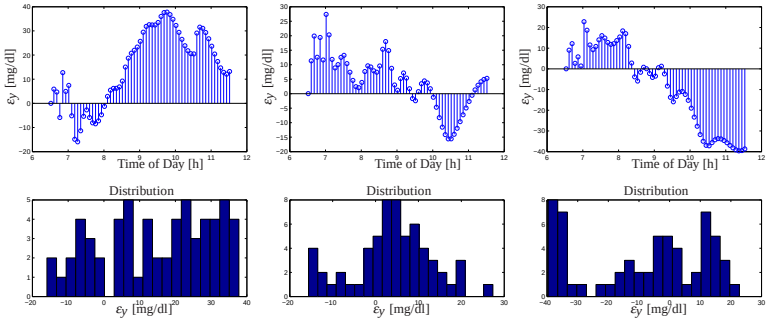

Figure 5.5 Simulation error ε_y of blood glucose $y(k)$ given inputs $u(k)$ using M_1. Estimation data (middle plot) and validation data (left and right plots).

tion is clearly non-Gaussian. This could be explained by time-varying dynamics, and in [Sparacino *et al.*, 2007] a recursive sensor model is used to handle such occurences. However, the evaluated time periods are short, and applying the model over the entire data record gives a more even distribution (Fig. 5.6). Given a tolerance interval of ± 20 mg/dl, corresponding to the p-CGA A zone for a 100 mg/dl blood glucose value, the model error can be considered acceptable.

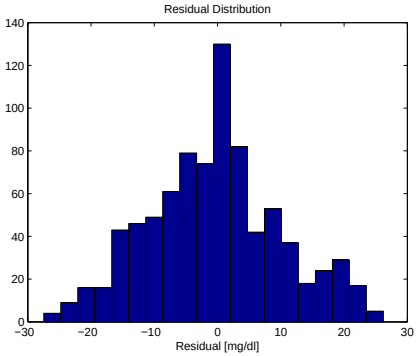

Figure 5.6 Distribution of the simulation error of the sensor model over the entire data record.

The simulation error of the GIIM can be seen in Fig. 5.5. The contribution is significantly larger. For breakfast day one, the model underestimates the glucose

drop after the peak. On day three on the other hand, the model overestimates the same drop. Given that these are infinite-horizon prediction without any measurement feedback, a maximum error in the magnitude of 40 mg/dl should be considered to be a very good result. In fact, the error is almost within the p-CGA A zone at all times.

Looking at the predictions in Fig. 5.3, the behavior of the prediction error can be understood from the error contribution from the sensor model, sensor errors and the GIIM. As the prediction horizon increases, the GIIM error becomes more and more dominant.

The corresponding prediction error for the model from Chapter 4, without the sensor model and with CGM as feedback signal, in terms of RMSE, can be found in Table 5.3. Compared to the results in Table 5.2, an improvement can be seen for every evaluated prediction horizon, with a relatively larger improvement as the prediction horizon increases.

Table 5.3 Prediction error assessment for the model without incorporated sensor model in terms of RMSE.

Prediction	**RMSE [mg/dl]**	
Horizon [min]	vs G_{CGM}	vs G
20	8.0	25.0
40	16.3	31.4
60	24.4	37.6

Glucose and Insulin Sub Models

Major sources of uncertainty are the intermediate inputs R_a and I_p and the assumptions made to retrieve them. Unfortunately, these obstacles are hard to overcome in the applied modeling framework. Neither the rate of glucose appearance following a meal nor the plasma insulin level are normally available for measurement. Estimates of R_a have been made in [Dalla Man *et al.*, 2006] and require a tracer based experiment. I_p can be obtained from lab assays of blood samples. Obviously, such arrangements cannot be expected in a normal day setting. Further work to assess the intra- and inter-individual variations of these processes, and on mitigations to handle these principle obstacles, is needed.

IG Dynamics

These dynamics were assumed to be time invariant, and homogeneous in direction and magnitude of glucose change and glucose level. The assumption of independence of the sign of the glucose change has been shown to be questionable, see [Kovatchev *et al.*, 2009], where statistically significant differences in response time, depending on the direction of glucose change, are presented. However, in this study no such differences could be observed.

5.9 Conclusions

The comparison of the merged prediction to the CGM signal clearly shows that the augmented model manages to significantly reduce the delay, that otherwise is present when only relying on the CGM signal to estimate the blood glucose. Furthermore, the results indicate that the underlying GIIM model seems to, with acceptable accuracy, describe the combined impact of a breakfast and the subsequent insulin injection. Further research is needed to evaluate the concept on more patient data, to investigate whether generic sensor models can be utilized—thereby reducing the necessity of well-sampled reference blood glucose data—and to investigate potential non-linear sensor characteristics and time-variability associated with sensor drift.

6

Ensemble Prediction

Diabetic glucose dynamics is known to be subject to time-shifting dynamics, as indicated in Chapter 3. Considering this, and the vast number of models developed in the literature, as described in Chapter 4, it is unclear if a single model can be determined to be optimal under every possible situation. This raises the question whether it is more useful to use one of the models solely, or if it is possible to gain additional prediction accuracy by combining their outcomes. Accuracy may be gained from merging, due to mismodeling or to changing dynamics in the underlying data creating process, where a single model capturing the system behavior may be infeasible, e.g., for practical identification concerns. Thus, by an ensemble approach, robustness and performance may be improved. In this chapter, a novel merging approach—combining elements from both switching and averaging techniques, forming a 'soft' switcher in a Bayesian framework—is presented for the glucose prediction application.

6.1 Related Research

Merging models for the purpose of prediction has been developed in different research communities. In the meteorological and econometric communities regression-oriented ensemble prediction has been a vivid research area since the late '60s, see e.g. [Raftery *et al.*, 2005] and [Elliott *et al.*, 2006].

Also in the machine learning community, the question of how different predictors or classifiers can be used together for increased performance has been investigated, and different algorithms have been developed, such as the bagging, boosting [Breiman, 1996] and weighted majority [Littlestone and Warmuth, 1994] algorithms, and online versions of these [Oza, 2005; Kolter and Maloof, 2003].

In most approaches the merged prediction $\hat{y}_k^e$ at time k is formed by a linear weighted average of the individual predictors $\hat{\mathbf{y}}_k$.

$$\hat{y}_k^e = \mathbf{w}_k^{\mathrm{T}} \hat{\mathbf{y}}_k \tag{6.1}$$

It is also common to restrict the weights $\mathbf{w}_k$ to $[0,1]$. The possible reasons for this are several, where the interpretation of the weights as probabilities, or rather Bayesian beliefs, is the dominating. Such restrictions are however not always applicable, e.g. in the related optimal portfolio selection problem, where negative weight (short selling) can reduce the portfolio risk [Elton *et al.*, 1976].

A special case, considering distinct switches between different linear system dynamics, has been studied mainly in the control community. The data stream and the underlying dynamic system are modelled by pure switching between different filters derived from these models, i.e., the weights w_k can only take value 1 or 0. A lot of attention has been given to reconstructing the switching sequence, see e.g. [Gustafsson, 2000; Ohlsson *et al.*, 2010]. From a prediction viewpoint, the current dynamic mode is of primary interest, and it may suffice to reconstruct the dynamic mode for a limited section of the most recent time points in a receding horizon fashion [Alessandri *et al.*, 2005].

Combinations of specifically adaptive filters has also stirred some interest in the signal processing community. Typically, filters with different update pace are merged, to benefit from each filter's specific change responsiveness, respectively steady state behaviour [Arenas-Garcia *et al.*, 2006].

Finally, in fuzzy modeling, soft switching between multiple models is offered using fuzzy membership rules in the Takagi-Sugeno systems [Takagi and Sugeno, 1985].

6.2 Problem Formulation

A non-stationary data stream $z_k : \{y_k, u_k\}$ arrives with a fixed sample rate, set to 1 for notational convenience, at time $t_k \in \{1,2,\ldots\}$. The data stream contains a variable of primary interest called $y_k \in \mathbb{R}$ and additional variables u_k. The data stream can be divided into different periods T_{S_i} of similar dynamics $S_i \in S = [1 \ldots n]$, and where $s_k \in S$ indicates the current dynamic mode at time t_k. The system changes between these different modes according to some unknown dynamics.

Given m number of expert q-steps-ahead predictions, $\hat{y}^j_{k+q|k}, j \in \{1,..m\}$ of the variable of interest at time t_k, each utilizing different methods, and/or different training sets; how is an optimal q-steps-ahead prediction $\hat{y}^e_{k+q|k}$ of the primary variable, using a predefined norm and under time-varying conditions, determined?

6.3 Sliding Window Bayesian Model Averaging

Apart from conceptual differences between the different approaches to ensemble prediction, the most important difference is how the weights are determined. Numerous different methods exist, ranging from heuristic algorithms [Takagi and Sugeno, 1985; Arenas-Garcia *et al.*, 2006] to theory based approaches, e.g., [Hoeting *et al.*, 1999]. Specifically, in a Bayesian Model Averaging framework [Hoeting *et al.*, 1999], which will be adopted in this chapter, the weights are interpreted as partial beliefs in each predictor M_i, and the merging is formulated as:

$$p(y_{k+q}|D_k) = \sum_i p(y_{k+q}|M_i, D_k)p(M_i|D_k) \tag{6.2}$$

where $p(y_{k+q}|D_k)$ is the conditional probability of y at time t_{k+q} given the data, $D_k : \{z_{1:k}\}$ received up until time k, and if only point-estimates are available, one can, e.g., use:

$$\hat{y}^e_{k+q|k} = \mathbb{E}(y_{k+q}|D_k) \tag{6.3}$$
$$= \sum_i \mathbb{E}(M_i|D_k)\mathbb{E}(y_{k+q}|M_i, D_k) \tag{6.4}$$
$$= \mathbf{w}_k^T \hat{\mathbf{y}}_k \tag{6.5}$$
$$\mathbf{w}_k^{(i)} = \mathbb{E}(M_i|D_k) \tag{6.6}$$
$$p(\mathbf{w}_k^{(i)}) = p(M_i|D_k) \tag{6.7}$$

where $\hat{y}^e_{k+q}$ is the combined prediction of y_{k+q} using information available at time k, and $\mathbf{w}_k^{(i)}$ indicates position i in the weight vector. The conditional probability of predictor M_i can be further expanded by introducing the latent variable Θ_j.

$$p(M_i|D_k) = \sum_j p(M_i|\Theta_j, D_k)p(\Theta_j|D_k) \tag{6.8}$$

or in matrix notation

$$\mathbf{p}(\mathbf{w}_k) = \left[\mathbf{p}(\mathbf{w}_{k|\theta_k=1}) \dots \mathbf{p}(\mathbf{w}_{k|\theta_k=n})\right] \mathbf{p}(\Theta|D_k) \tag{6.9}$$

Here, Θ_j represents a *predictor mode* in a similar sense to the dynamic mode S_j, and likewise θ_k represents the prediction mode at time k. $\mathbf{p}(\Theta|D_k)$ is a row vector of $p(\Theta_j|D_k), j = \{1 \dots m\}$ and $\mathbf{p}(\mathbf{w}_{k|\Theta_i})$ is a row vector of the joint prior distribution of the conditional weights of each predictor model given the predictor mode Θ_i.

Data for estimating the distribution for $\mathbf{p}(\mathbf{w}_{k|\Theta_i})$ is given based upon using a constrained optimization on the training data. In cases of labelled training data sets, the following applies:

$$\begin{aligned} \{\mathbf{w}_{k|\Theta_i}\}_{T_{S_i}} &= \arg\min \sum_{m=k-N/2}^{k+N/2} \mathscr{L}(y(t_m), \mathbf{w}_{k|\Theta_i}^T \hat{\mathbf{y}}_\mathbf{i}), \quad k \in T_{S_i} \\ &\text{s.t.} \sum_j \mathbf{w}_{k|\Theta_i}^{(j)} = 1 \end{aligned} \tag{6.10}$$

where T_{S_i} represents the time points corresponding to dynamic mode S_i, N is the size of the evaluation window and $\mathscr{L}(y,\hat{y})$ is a cost function. From these data sets, the prior distributions can be estimated by the Parzen window method [Bishop, 2006], giving mean $\mathbf{w}_{0|P_i} = \mathbb{E}(\mathbf{w}_{k|\Theta_i})$ and covariance matrix $\mathbf{R}_{\Theta_i}$. An alternative to the Parzen approximation is of course to estimate a more parsimoniously parametrized probability density function (pdf) (e.g., Gaussian) for the extracted data points. For unlabelled training data, with time points T, the corresponding datasets $\{\mathbf{w}_{k|\Theta_i}\}_T$ are found by cluster analysis, e.g., using a Gaussian Mixture Model (GMM) [Bishop, 2006].

Now, in each time step k, the $\mathbf{w}_{k|\theta_{k-1}}$ is determined from the sliding window optimization below, using the current active mode θ_{k-1}. For reasons soon explained, only $\mathbf{w}_{k|\theta_{k-1}}$ is thus calculated:

$$\begin{aligned} \mathbf{w}_{k|\theta_{k-1}} = \arg\min & \sum_{j=k-N}^{k-1} \mu^{k-j} \mathscr{L}(y_j, \mathbf{w}_{k|\theta_{k-1}}^T \hat{\mathbf{y}}_j) \\ & + (\mathbf{w}_{k|\theta_{k-1}} - \mathbf{w}_{0|\theta_{k-1}}) \Lambda_{\theta_{k-1}} (\mathbf{w}_{k|\theta_{k-1}} - \mathbf{w}_{0|\theta_{k-1}})^T \\ \text{s.t.} & \sum_j \mathbf{w}_{k|\theta_{k-1}}^{(j)} = 1 \end{aligned} \tag{6.11}$$

Here, μ_j is a forgetting factor, and Λ_{Θ_i} is a regularization matrix. To infer the posterior $\mathbf{p}(\Theta|D_k)$ in (6.9), it would normally be natural to set this probability function equal to the corresponding posterior pdf for the dynamic mode $\mathbf{p}(S|D_k)$. However, problems arise if $\mathbf{p}(S|D_k)$ is not directly possible to estimate from the dataset D_k. This is circumvented by using the information provided by the $\mathbf{p}(\mathbf{w}_{k|\theta_k})$ estimated from the data retrieved from equation (6.10) above. The $\mathbf{p}(\mathbf{w}_{k|\theta_k})$ prior density functions can be seen as defining the region of validity for each predictor mode. If the $\mathbf{w}_{k|\theta_{k-1}}$ estimate leaves the current active mode region θ_{k-1} (in a sense that $\mathbf{p}(\mathbf{w}_{k|\theta_{k-1}})$ is very low), it can thus be seen as an indication of that a mode switch has taken place. A logical test is used to determine if a mode switch has occurred. The predictor mode is switched to mode Θ_i, if:

$$\begin{cases} p(\Theta_i|\mathbf{w}_k, D_k) > \lambda, \text{ and} \\ p(\mathbf{w}_k|\Theta_i, D_k) > \delta \end{cases} \tag{6.12}$$

where

$$p(\Theta_i|\mathbf{w}_k, D_k) = \frac{p(\mathbf{w}_k|\Theta_i, D_k)p(\Theta_i|D_k)}{\sum_j p(\mathbf{w}_k|\Theta_j, D_k)p(\Theta_j|D_k)} \tag{6.13}$$

A λ somewhat larger than 0.5 gives a hysteresis effect to avoid chattering between modes, and δ assures that non-conclusive situations, evaluated on the outskirts of the probability functions, don't result in switching. Unless otherwise estimated from data, the conditional probability of each prediction mode $p(\Theta_i|D_k)$ is set equal for all possible modes, and thus cancels in (6.13). The logical test is evaluated using the priors received from the pdf estimate and the $\mathbf{w}_{\mathbf{k}|\theta_\mathbf{k}}$ received from (6.11). If a mode switch is considered to have occurred, (6.11) is rerun using the new predictor mode.

Now, since only one prediction mode θ_k is active; (6.9) reduces to $\mathbf{p}(w_k) = \mathbf{p}(w_{k|\theta_k})$.

Parameter choice

The length N of the evaluation period is, together with the forgetting factor μ, a crucial parameter determining how fast the ensemble prediction reacts to sudden changes in dynamics. A small forgetting factor will put much emphasis on recent data, making it more agile to sudden changes. However, the drawback is of course that noise sensitivity increases.

Λ_{Θ_i} should also be chosen, such that a sound balance between flexibility and robustness is found, i.e., a too small $||\Lambda_{\Theta_i}||_2$ may result in over-switching,

whereas a too large $||\Lambda_{\Theta_i}||_2$ will give a stiff and inflexible predictor. Furthermore, Λ_{Θ_i} should force the weights to move within the perimeter defined by $p(\mathbf{w}|\Theta_i)$. This is approximately accomplished by setting Λ_{Θ_i} equal to the inverse of the covariance matrix $\mathbf{R}_{\Theta_i}$, thus representing the pdf as a Gaussian distribution in the regularization.

Nominal mode

Apart from the estimated prediction mode centres, an additional predictor mode can be added, corresponding to a heuristic fall-back mode. In the case of sensor failure, or other situations where loss of confidence in the estimated predictor modes arises, each predictor may seem equally valid. In this case, a fall-back mode to resort to may be the equal weighting. This is also a natural start for the algorithm. For these reasons, a nominal mode $p(\overline{\mathbf{w}}^{\mathbf{0}}) \in N(\mathbf{1}/\mathbf{n}, \mathbf{I})$ is added to the set of predictor modes.

Summary of algorithm

1. Estimate n numbers of predictors according to best practice.
2. Run the constrained estimation (6.10) on labelled training data and retrieve the sequence of $\{\mathbf{w}_{k|\Theta_i}\}_{T_{\theta_i}}, \forall i \in \{1,..,n\}$.
3. Classify different predictor modes, and determine density functions $\mathbf{p}(\mathbf{w}_{k|\Theta_i})$ for each mode Θ_i from the training results by supervised learning. If possible; estimate $p(\Theta_i|D)$.
4. Initialize mode to the nominal mode.
5. For each time step; calculate $\mathbf{w_k}$ according to (6.11).
6. Test if switching should take place by evaluating (6.12) and (6.13), and switch predictor mode if necessary and recalculate new $\mathbf{w}_k$ according to (6.11).
7. Go to 5.

The ensemble engine outlined above will hereafter be referred to as Sliding Window Bayesian Model Averaging (SW-BMA) Predictor.

6.4 Choice of Cost Function $\mathscr{L}$

Cost function should be chosen by the specific application in mind. A natural choice for interpolation is the 2-norm, but in certain situations asymmetric cost functions are more appropriate. For the glucose prediction application, a suitable cost function for determining appropriate weights should take into account that the consequences of acting on too high glucose predictions in the lower blood glucose (G) region ($<$90 mg/dl) could possibly be life threatening. The margins to low blood glucose levels, that may result in coma and death, are small, and blood glucose levels may fall rapidly, as seen in Chapter 3. Hence, much emphasis should be put on securing small positive predictive errors and sufficient time margins for alarms to be raised in due time in this region. In the normoglycemic region (here defined as 90-200 mg/dl), the predictive quality is of less importance. This is the glucose range that healthy subjects normally experience, and thus can be considered, from a clinical viewpoint in regards to possible complications, a safe region. However, due to the possibility of rapid fluctuation of the glucose into unsafe regions, some considerations of predictive quality should be maintained.

Based on the cost function in [Kovatchev *et al.*, 2000], the selected cost function incorporates these features; asymmetrically increasing cost of the prediction error depending on the absolute glucose value and the sign of the prediction error.

In Fig. 6.1 the cost function can be seen, plotted against relative prediction error and absolute blood glucose value.

Correspondence to the Clarke Grid Error Plot

A de facto accepted standardized metric of measuring the performance of CGM signals in relation to reference measurements, and often used to evaluate glucose predictors, is the Clarke Grid Plot [Clarke *et al.*, 1987], described in Chapter 3. This metric meets the minimum criteria raised earlier. However, other aspects makes it less suitable; no distinction between prediction errors within error zones, instantaneous switches in evaluation score, etc.

In Fig. 6.2, the isometric cost of the chosen cost function for different prediction errors at different G values has been plotted together with the Clarke Grid Plot. The boundaries of the A/B/C/D/E areas of the Clarke Grid can be regarded as lines of isometric cost according to the Clarke metric. In the figure, the isometric cost of the cost function has been chosen to correspond to the lower edge, defined by the intersection of the A and B Clarke areas at 70 mg/dl. Thus, the

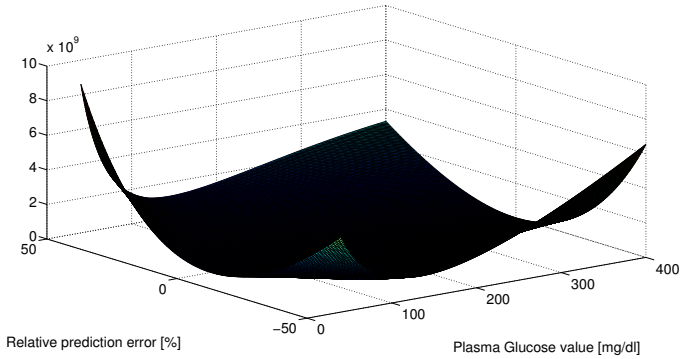

Figure 6.1 Cost function of relative prediction error.

area enveloped by the isometric cost can be regarded as the corresponding A area of this cost function. Apparently, much tougher demands are imposed both in the lower and upper *G* regions in comparison to the Clarke Plot.

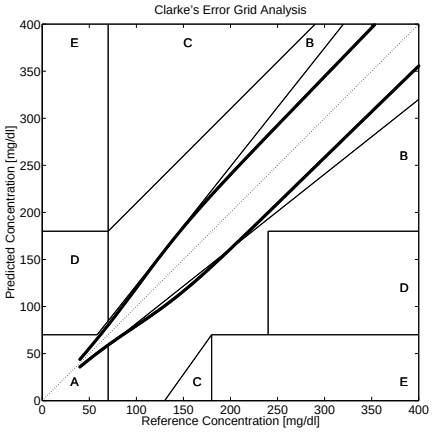

Figure 6.2 Isometric cost in comparison to the Clarke Grid.

6.5 Example I: Approximate Lower-Order Models

Data

Data were generated using a switched fourth-order ARX system, where the A-polynomial switches between three different models M_A, M_B, M_C, with poles according to Table 6.1. The B-polynomial was simply a one step delay, and white noise $N(0, 0.25)$ was added to the output channel. A PRBS signal was used for input.

Table 6.1 Poles of the data generating processes.

Model	Poles
M_A	$0.8, 0.1, -0.3 + i\sqrt{0.41}, -0.3 - i\sqrt{0.41}$
M_B	$0.9, 0.2, -0.2, -0.5$
M_C	$0.8, -0.2, -0.4, -0.4$

The active dynamic mode $s_k \in S$ switches between dynamic mode A, B and C according to a Markov Chain with transition matrix M.

$$M = \begin{bmatrix} 0.99 & 0.005 & 0.005 \\ 0.005 & 0.99 & 0.005 \\ 0.005 & 0.005 & 0.99 \end{bmatrix} \tag{6.14}$$

A labelled training set of 2000 samples and a 2000 sample validation set were simulated in 40 different batches. An example of a training data set can be seen in Fig. 6.3.

Predictors

To simulate modeling errors, three prediction models $M_I - M_{III}$ were set up as reduced order approximations of the corresponding state-space models of the data generating processes. Model reduction was undertaken by singular value evaluation to the second order [Johansson, 2009]. Using these models and their associated Kalman filters, 50 step prediction length was evaluated.

Cost function

For this example the 2-norm was used.

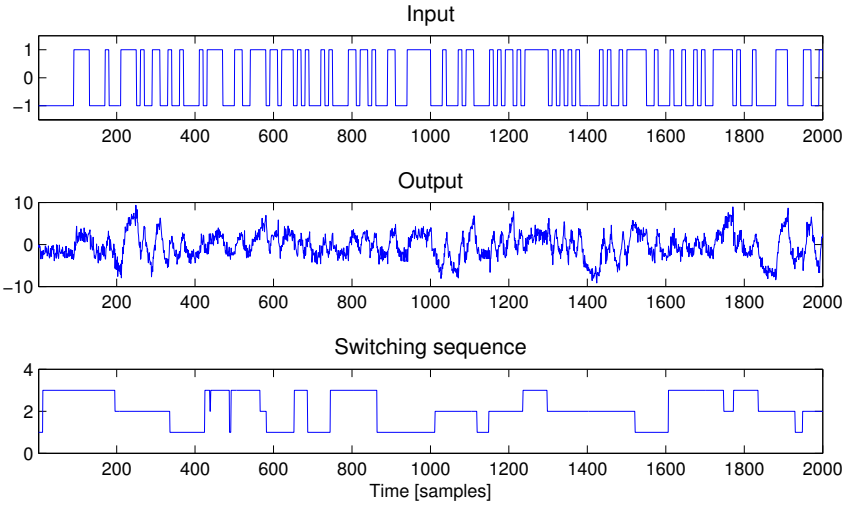

Figure 6.3 Training Data. Upper plot: input, middle plot: output and lower plot: switching sequence of dynamic mode. Example I.

Parameter Choices

Different values for the tunable parameters N and μ were evaluated; 20 batches for the combinations of $\{10, 20, 30\}$ and $\{0.8, 0.9, 1\}$, and 20 batches for the combination of $\{25, 50, 75\}$ and $\{0.7, 0.8, 0.9\}$. The parameters λ and δ were fixed to 0.6 and $3 \cdot 10^{-3}$.

Evaluation Metric

To evaluate the predictive performance, the squared sum of prediction errors was compared to the squared sum of prediction errors using a pure switching strategy, where it (optimally) has been assumed that the dynamic mode at the time of prediction was known.

Results

Training the Mode Switcher Using the labelled training set, the pdf:s $\mathbf{p}(\mathbf{w}_{k|\Theta_i})$ were estimated for each batch using the different N values. For this example the best evaluation record length for the estimation task was 10. In Fig. 6.4, an

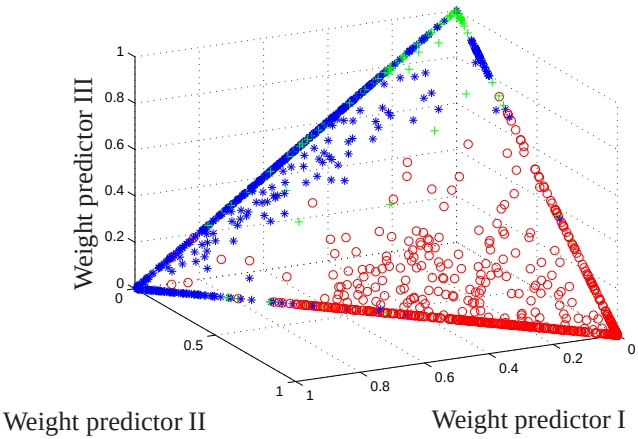

Figure 6.4 Distribution of weights in the training data retrieved by (6.10). Blue stars represents $t_k \in T_A$ for Mode A, Red circles: T_B for Mode B and Green crosses: T_C for Mode C. Example I.

example of the distribution of $\{\mathbf{w}_{k|\Theta_i}\}_{S_i}$ along the $\{w_1+w_2+w_3=1, 0 \leq w_i \leq 1\}$ plane can be seen for one representative training batch.

The corresponding probability distribution for each mode, projected onto the (w_1, w_2)-plane, estimated by Parzen window technique, can be seen in Fig. 6.5 together with the pdf of the nominal mode. The densities have higher values in the corners $[1,0,0]$, $[0,1,0]$ and $[0,0,1]$, and the means $\mathbf{w_{0|1}} = [0.57, 0.03, 0.4]$, $\mathbf{w_{0|2}} = [0.18, 0.76, 0.06]$ and $\mathbf{w_{0|3}} = [0.25, 0.03, 0.72]$, define the expected weights for each predictor mode.

Evaluation of Parameter Choices Comparing the predictive performance for the different value combinations of N and μ, the slightly better choice over the others was $[25, 0.8]$. Table 6.2 summarizes the predictive performance for each combination of N and μ.

Predictive Performance The merged prediction was compared on the validation data, using the best choices of $N = 25$ and $\mu = 0.8$, to; 1) each individual predictor, 2) an unregularized version of (6.11) without switching functionality, and 3) to the optimal pure switching strategy. The results are summarized in Table 6.3.

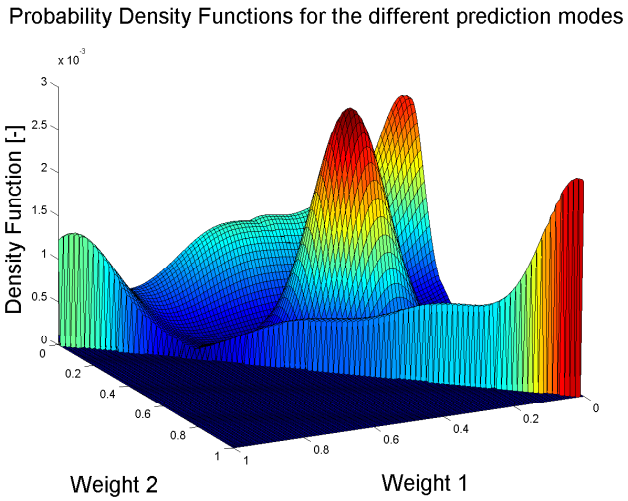

Figure 6.5 Estimated probability density functions for the weights in the training data, including the nominal mode. Batch 4. Example I.

Table 6.2 Summary of predictive performance using different N and μ on validation data over all simulated batches, evaluated as mean $\sum e^2_{N,\mu} / \sum e^2_{opt}$, where $e_{N,\mu}$ is corresponding prediction error(pe) and e_{opt} is the pe when using the optimal switching strategy. Example I.

N/μ	1	0.9	0.8	0.7
10	0.92	0.91	0.90	-
20	0.95	0.94	0.92	-
25	-	0.91	**0.87**	0.87
30	0.94	0.92	0.88	-
50	-	1.05	1.04	1.01
75	-	0.95	0.90	0.88

Compared to the other approaches a 7% improvement can be seen to the unregularized version, and a 13% improvement to the optimal switching scheme.

Looking at the distribution of the weights for the validation data in Fig. 6.6, it is apparent that the merging mechanism has concentrated these around the pre-

Table 6.3 Summary of predictive performance on validation data over all simulated batches. Example I.

Predictor	$\sum e^2 / \sum e^2_{opt}$
Predictor I	1.07
Predictor II	2.76
Predictor III	1.39
Merged Predictor	0.87
Unregularized Merged Predictor	0.93
Optimally Switched Predictor	1.0

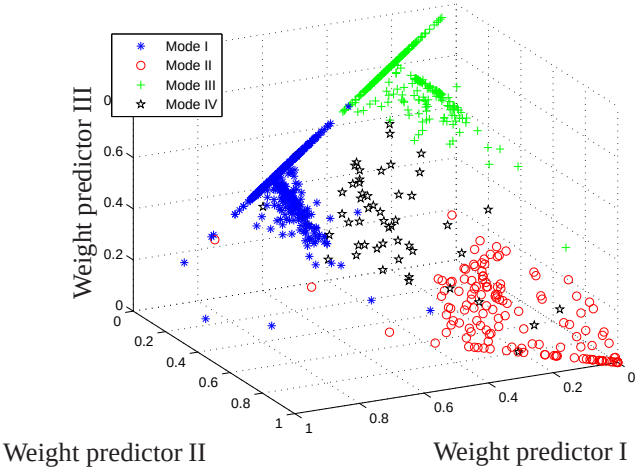

Figure 6.6 Distribution of weights in the test data using the estimated pdf:s and expected weights. Batch 4. Example I.

diction mode centres, especially if comparing to the corresponding distribution for the unregularized version, see Fig. 6.7.

Switching between the different prediction modes, in comparison to the dynamic mode for the validation data, can be seen in Fig 6.8 for a representative batch.

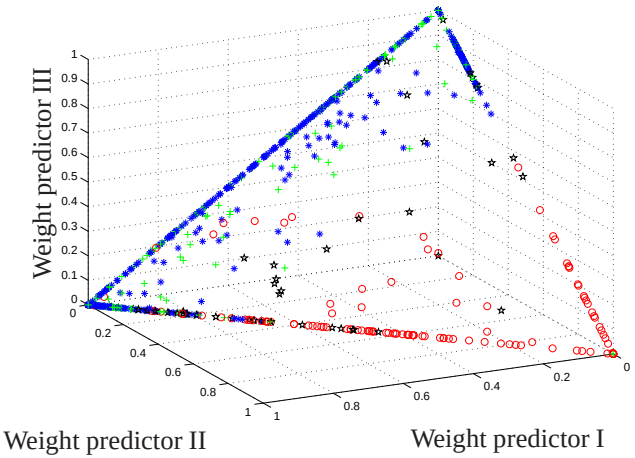

Figure 6.7 Distribution of weights in the test data using the unregularized merging predictor. Batch 4. Example I.

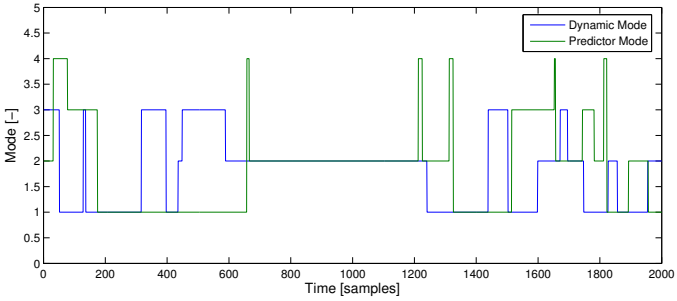

Figure 6.8 Example of switching between different predictor modes in the validation data. Predictor mode 4 represents the nominal mode. Example I.

Discussion

Parameter Choice The optimal choices of N and μ, are unsurprisingly, closely connected. These parameters must be set with the specific dynamics in mind, and are probably difficult to determine beforehand. λ and δ should probably not be

set too low in order to avoid uncalled for switching, and the values used are deemed correct from this aspect.

Predictive Performance The merged predictor clearly outperformed each of the individual predictors, and also the unregularized version as well as the optimal pure switching predictor. The latter can be explained by that the merged predictor offers some extra robustness to sudden dynamical changes, as all predictors to some extent are used in all situations. The unregularized version has quite good performance, but the regularization in the proposed merging mechanism reduces the impact of noise, making it slightly better.

6.6 Example II: The UVa/Padova Simulation Model

Data

Data was generated using the non-linear metabolic simulation model, jointly developed by the University of Padova, Italy and University of Virginia, U.S. (UVa) and described in [Dalla Man *et al.*, 2007b], with parameter values obtained from the authors. The model consists of three parts that can be separated from each other. Two sub models are related the influx of insulin following an insulin injection and the rate of appearance of glucose from the gastro-intestinal tract following meal intake, respectively, and have been described in Chapter 4. The final part of the total model is concerned with the interaction of glucose and insulin in the blood stream, organs and tissue, including renal extraction, endogenous glucose production and insulin and non-insulin dependant glucose utilization. The model equations are found in [Dalla Man *et al.*, 2007b].

Twenty datasets, each corresponding to 8 days, were generated. Two dynamic modes A and B were simulated by, after 4 days, changing four model parameters (following the notation in [Dalla Man *et al.*, 2007b]) k_1, k_i, k_{p3} and p_{2u}, related to endogenous glucose production and insulin and glucose utilization. One data set was used for training and the others were considered test data.

A section of four days, including the period when the dynamic change takes place, of an example training and test data set can be seen in Fig. 6.9.

Timing and size of meals were generated with some normal randomization for each data set, according to Table 6.4. The amount of insulin administered for each meal was based on a fixed carbohydrate-to-insulin ratio, perturbed by normally distributed noise, with a 20% standard deviation.

Table 6.4 Meal amount and timing randomization. Standard deviation in parenthesis.

Meal	Time	Amount carbohydrates (g)
Breakfast	08:00 (30 min)	45 (5)
Lunch	12:30 (30 min)	70 (10)
Dinner	19:00 (30 min)	80 (10)

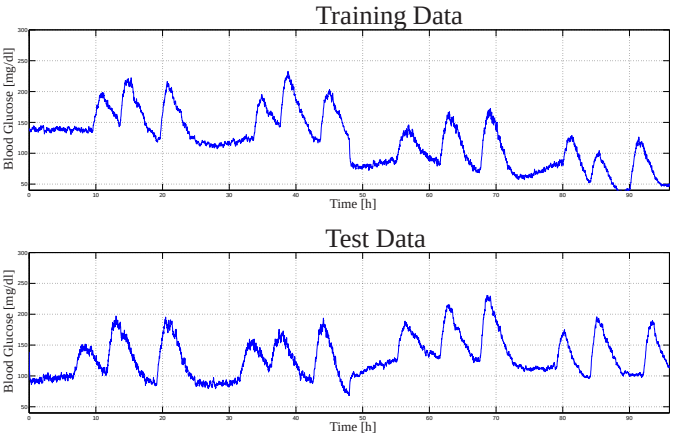

Figure 6.9 Four days of the blood glucose data of the training and test data for one of the 20 generated datasets. The upper plot correspond to the training data, and the lower plot represents the test data. For both plots, a mode switch takes place after 2 days ($t = 48$h). Example II: UVa/Padova Model.

Process noise was added by perturbing some crucial model parameters p_i in each simulation step; $p_i(t) = (1+\delta(t))p_i^0$, where p_i^0 represent nominal value and $\delta(t) \in N(0, 0.2)$. The affected parameters are (again following the notation in [Dalla Man *et al.*, 2007b])) $k_1, k_2, p_{2u}, k_i, m_1, m_{30}, m_2, k_{sc}$, and represents natural variability in the underlying physiological processes.

Predictors

Three models, based on subspace based technique, were identified using the N4SID algorithm of the Matlab System Identification Toolbox. Model order $[2-4]$ was determined by the Akaike criterion [Johansson, 2009]. The first model I

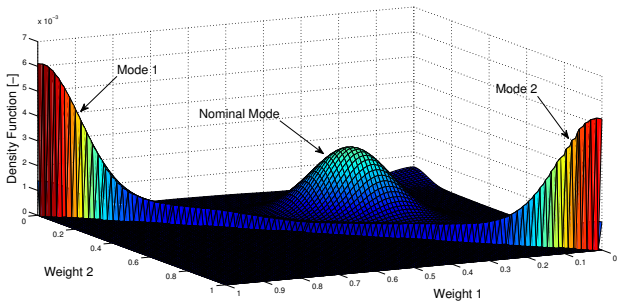

Figure 6.10 Estimated probability density functions for the weights in the training data, including nominal mode. Example II.

was estimated using data from dynamic mode A in the training data, and the third *III* from the mode *B* data, and the final model *II* from the entire training data set. Thus, model *I* and *III* are each specialized, whereas *II* is an average of the two dynamic modes. The models were evaluated for a prediction horizon of 60 min.

Results

Training the Mode Switcher The three predictors were used to create three sets of 60 minute ahead predictions for the training data. Using (6.10) with $N = 10$, the weights $\mathbf{w}_k$ were determined. The corresponding probability distribution for each mode, projected onto the (w_1, w_2)-plane, was estimated by Parzen window technique. The densities are well concentrated to the corners $[1,0,0]$ and $[0,0,1]$, with means $\mathbf{w_{0|1}} = [0.83, 0.11, 0.06]$ and $\mathbf{w_{0|2}} = [0.25, 0.1, 0.65]$ defining the expected weights for each predictor mode. The nominal mode probability density function was set to $N(\frac{\mathbf{1}}{3}\frac{\mathbf{1}}{3}\frac{\mathbf{1}}{3}, \mathbf{0.1I})$. In Fig. 6.10 all density functions, including the nominal mode, projected onto the (w_1, w_2)-plane, can be seen together.

Ensemble Prediction vs Individual Predictions Using the estimated pdf:s and expected weights $\mathbf{w}$ of the identified predictor modes, the ensemble machine was run on the test data. An example of the distribution of the weights for the two dynamic modes *A* and *B* can be seen in Fig. 6.11.

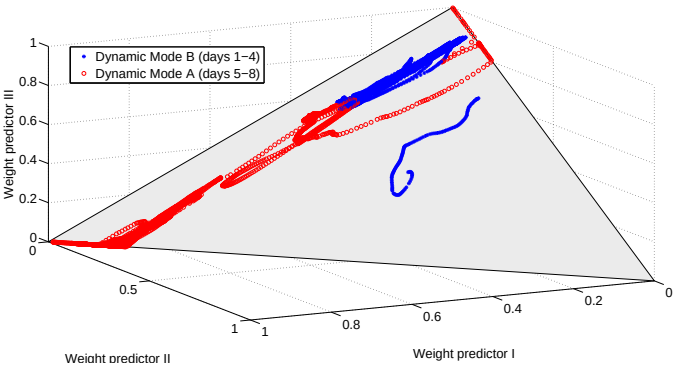

Figure 6.11 Example of the distribution of weights in the test data using the estimated pdf:s and expected weights. Example II.

An example of how switching between the different modes occurs over the test period can be found in Fig 6.12.

For evaluation purposes, all predictors were run individually. In Table 6.5, a comparative summary of the predictive performance between the different approaches, in terms of mean Root Mean Square Error (RMSE) over the test batches, is given.

Table 6.5 Performance evaluation by RMSE for the 60 minute predictors using different approaches.

	RMSE [mg/dl]		
Predictor Type	Section A	Section B	A+B
Predictor I	8.3	16.3	13.0
Predictor II	9.1	11.2	10.9
Predictor III	14.3	7.9	12.6
Merged prediction	8.7	10.5	9.6

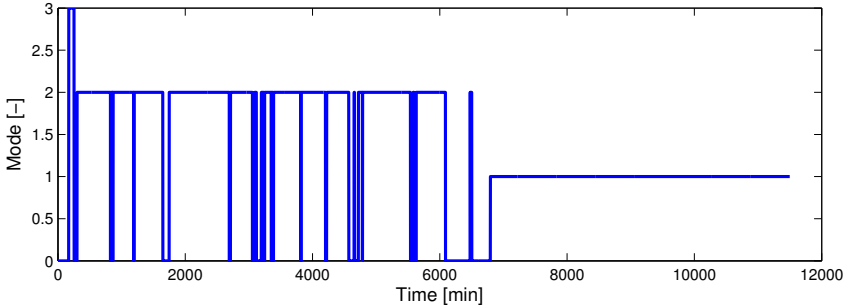

Figure 6.12 Example of switching between different predictor modes in the test data. The transition from dynamic mode *B* to mode *A* takes place at 5760 min (4 days). Mode 3 represents the nominal mode, and the gaps (mode=0) correspond to time instances when no $p_i(w|D)$ fulfilled the criteria (in which case the mode stays at the the current predictor mode). Example II.

6.7 Example III: The DIAdvisor Data

Data and evaluation criteria

Predictors Three different predictors of different structure were used in this study; the state-space-based model (SS) of Chapter 4, a recursive ARX model [Estrada *et al.*, 2010] and a kernel-based predictor [Naumova *et al.*, 2011], all developed within the DIAdvisor project. The SS and ARX models utilized inputs, generated by the population parametrized sub models describing the flux and digestion of insulin and glucose following an insulin injection or meal intake [Dalla Man *et al.*, 2007b], described in Chapter 4. For further information regarding the ARX and the kernel-based predictors, see [Estrada *et al.*, 2010] and [Naumova *et al.*, 2011].

Training and Test Data Data from the clinical part of the DAQ trial and the DIAdvisor I B and C trials were used. A number of patients participated in all three trials. Based on data completeness, six of these have been selected for this study. CGM data was used for model identification, whereas the interpolated frequent blood glucose reference measurements, see Chapter 3, were used for validation purposes.

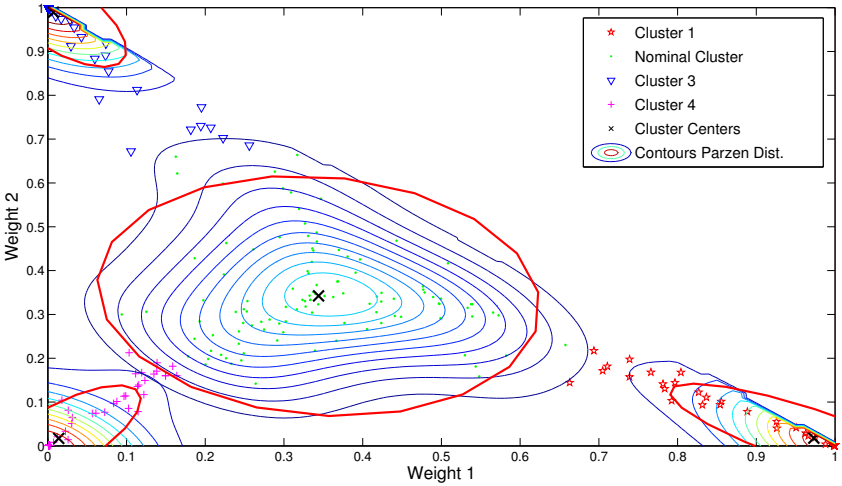

Figure 6.13 Example of distribution of weights in the training data by (6.10) and clusters given by the k-means algorithm. The red ellipses represent the fitted Gaussian covariances of each cluster (patient 0103, Trial B).

The first trial data (DAQ) was used to train the individual predictor models. The second and third trial data (DIAdvisor I.B and C) were used to train and cross-validate the SW-BMA, i.e., the SW-BMA was trained on B data and validated on C data, and vice versa.

Evaluation Criteria The prediction results were compared to the interpolated blood glucose G in terms of Clarke Grid Analysis, see Chapter 4, and the complementary Root Mean Square Error (RMSE).

Results

Training the Mode Switcher

Cluster Analysis - Finding the Modes The three predictors were used to create 40 minute ahead predictions for both training data sets $D_{T_{B(C)}}$. Using (6.10) with $N = 20$, the weights $\{\mathbf{w}_k\}_{T_{B(C)}}$ were obtained; example depicted in the (w_1, w_2) plane in Fig. 6.13. The weights received from the training are easily

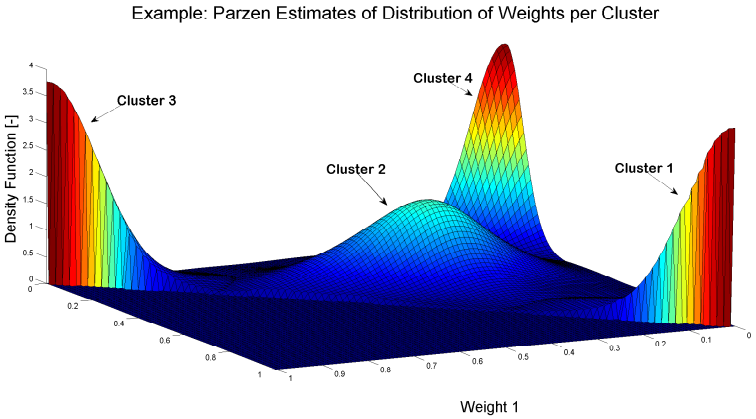

Figure 6.14 Example of estimated probability density functions for the different predictor mode clusters in the training data (patient 0103, Trial B).

visually recognized as belonging to different groups (true for all patients, not shown). Attempts were made to find clusters using a Gaussian Mixture Model (GMM) by the EM algorithm, but without viable outcome. This is not totally surprising, considering, e.g., the constraints $0 \geq w_i \geq 1$ and $\sum w = 1$. A more suitable distribution, often used as a prior for the weights in a GMM, is the Dirichlet distribution, but instead the simpler k-means algorithm was applied using four clusters (number of clusters given by visual inspection of the distribution of $\{\mathbf{w}_k\}_{T_{B(C)}}$), providing the cluster centers $\mathbf{w}_{0|\Theta_i}$.

The corresponding probability distribution for each mode $p(\mathbf{w}|\Theta_i)$, projected onto the (w_1, w_2)-plane, was estimated by Parzen window technique, and an example can be seen in Fig. 6.14. Gaussian distributions were fitted to give the covariance matrices $\mathbf{R}_{\Theta_i}$ used in (6.11).

Feature Selection The posterior mode probability $p(\Theta|D_k)$ is likely not dependent on the entire data D_k, but only a few relevant data features, possible to extract from D_k. Features related to the performance of a glucose predictor may include meal information, insulin administration, level of activity, measures of the glucose dynamics, etc. By plotting the training CGM data, colored according to the best mode at the prediction horizon retrieved by the training, interesting correlations become apparent (Fig. 6.15). The binary features in Table 6.6 were selected.

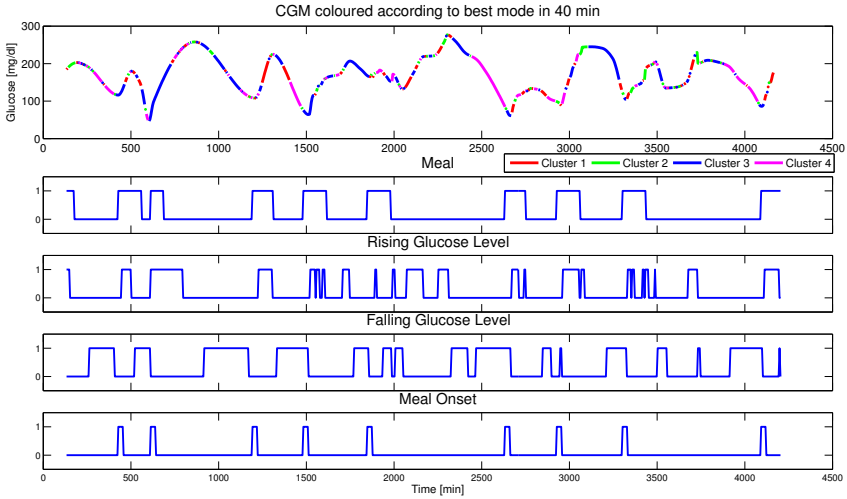

Figure 6.15 Example of CGM coloured according to best predictor mode in 40 minutes together with active features (patient 0103, Trial B).

When extracting the features, meal timing and content were considered to be known 30 minutes before the meal.

From the training data, the posterior mode probabilities $p(\Theta_i|f_j)$, given each feature f_j, were determined by the ratio of active time for each mode over the time periods when each feature was present. Additionally, the overall prior $p(\Theta_i)$ was determined by the total ratio of active time per cluster over the entire test period.

The different features are overlapping, and to resolve this issue they were given different priority—only allowing the feature of highest priority, f_k^* to be present at each time step t_k. Thereafter, $p(\Theta|D_k) = p(\Theta|f_k^*)$ is determined. If no feature is active, the $p(\Theta|D_k)$ is approximated by the $p(\Theta_i)$ estimate.

Prediction Performance on Test Data

Using the estimated mode clusters $\{w_{0|i}, R_{0|i}\}, i = [1 \ldots M]$, and the estimated posteriors $p(\Theta_i|f^*)$ from Trial B (C), the ensemble machine was run on the Trial C (B) data. The parameter μ was set to 0.8 and N to 20 minutes. An example of the distribution of the weights $\mathbf{w}_k$ for the three predictors can be seen in Fig.

Table 6.6 Selected features. ε corresponds to the maximum amplitude of glucose rate-of-appearance, R_a after digesting 10 g CHO, and $\Delta G = G_k - G_{k-5}$

Feature	Threshold	Priority
Meal	$\max(Ra_k, .., Ra_{k+30}) > \varepsilon$	1
Rising G	mean $(\Delta G_{k-10}, \ldots, \Delta G_k)$ > 30 mg/(dl· h)	2
Falling G	mean $(\Delta G_{k-10}, \ldots, \Delta G_k)$ < -18 mg/(dl· h)	3
Meal and rising G	See above.	4
Meal Onset	$\max Ra(k-30, ..., k) < \varepsilon$ and $\max Ra(k, ..., k+30) > \varepsilon$	5

6.16.

Table 6.7 summarizes a comparison of predictive performance over the different patient test data sets for the RMSE evaluation criteria, and in Table 6.8 the evaluation in terms of Clarke Grid Analysis is given. The optimal switching approach, here defined as using the non-causal fitting by Eq. (6.10), is used as a measure of optimal performance of a linear combination of the different predictors.

Table 6.7 Performance evaluation for the 40 minute SW-BMA prediction compared to the optimal switching and the individual predictors. The metric is the Root Mean Square Error (RMSE), normalized against the best individual predictor $M_1 - M_3$ for each patient.

	median RMSE/RMSE$_{best}$ [min-max]	
Merging Strategy	Trial B	Trial C
SW-BMA	1.03 [0.75-1.04]	1.03 [0.94-1.05]
Optimal switching	0.97 [0.54-1.0]	0.94 [0.73-1.0]
2:nd best individual pred.	1.16 [1.09-1.27]	1.21 [1.04-1.37]
Worst individual pred.	1.44 [1.25-1.73]	1.45 [1.18-1.83]

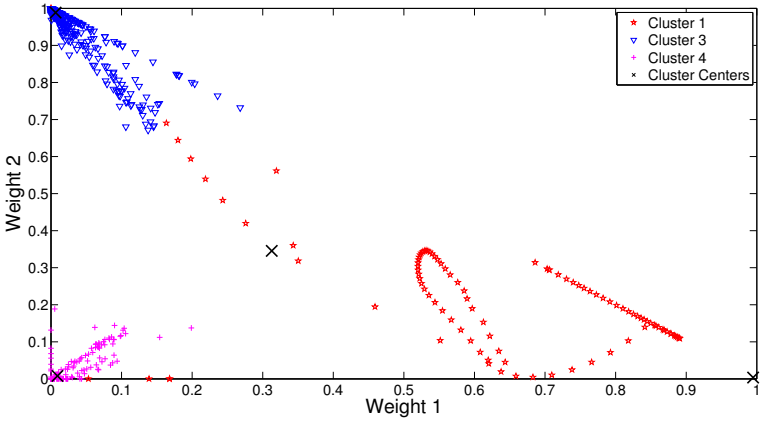

Figure 6.16 Example of the distribution of weights in the test data using the estimated clusters and feature correlations (patient 0108, Trial B).

Table 6.8 Performance evaluation for the 40 minute SW-BMA prediction compared to the optimal switching and the best individual predictor by the amount of data (%) in the acceptable A/B zones vs. the dangerous D and E zones.

Merging Strategy	Trial B			Trial C		
	A/B	D	E	A/B	D	E
SW-BMA	95.5	2.2	0	95.3	3.0	0.1
Optimal switching	96.2	1.7	0	96.9	1.3	0
Best individual pred.	94.8	2.6	0	95.0	3.4	0

6.8 Discussion

In Example I it was shown how the merged predictor could reduce the impact of the switching dynamics, resulting in performance beyond the optimal switching strategy. Example II outlined how the technique may be applied to the specific example of diabetes glucose prediction under sudden changes in the underlying physiological dynamics. Also in this example, the merged prediction turned out to be the best choice. In Example III, applying the algorithm to real-world data, the SW-BMA has, for most patients, the same RMSE and Clarke Grid per-

formance as the best individual predictor. In one case, the merged prediction clearly outperformed also the best predictor ($\mathrm{RMSE}/\mathrm{RMSE}_{\mathrm{best}} = 0.75$). However, comparison to the optimal switcher indicates that there is still further room for improvement. To fill this gap, timely switching is most important. To this purpose, features with significant correlations to mode switching, may prove useful in order to improve the likelihood that the best predictor mode is used at each time. Further research is needed to investigate and improve this aspect.

6.9 Comparison to Other Merging Techniques

Compared to the strategy of pure switching between different predictors, the evaluation indicates that the proposed algorithm is more robust to sudden changes and in reducing the impact of modeling errors. Apart from that, in many applications, transition between different dynamic modes is a gradual process rather than an abrupt switch, making the pure switching assumption inappropriate. The proposed algorithm can handle such smooth transitions by slowly sliding along a trajectory in the weight plane of the different predictors, perhaps with a weaker Λ if such properties are expected. Furthermore, any type of predictor may be used, not restricting the user to a priori assumptions of the underlying process structure.

In Tagaki-Sugeno (TS) system, a technique that also gives soft switching, the underlying assumption is that the switching dynamics can be observed directly from the data. This assumption has been relaxed for the proposed algorithm extending the applicability beyond that of TS systems.

In [Raftery *et al.*, 2010], another interesting approach to online Bayesian Model Averaging is suggested for changing dynamics. In this approach, the assumed transition dynamics between the different modes is based on a Markov chain. However, in our approach no such assumptions on the underlying switching dynamics are postulated. Instead, switching is based on recent performance in regards to the applicable norm, and possibly on estimated correlations between predictor modes and features of the data stream $P(\Theta_i|D_k)$, see Eq. (6.13).

6.10 Conclusions

A novel merging mechanisms for multiple predictor has been proposed for time-varying and uncertain conditions. The approach was evaluated on both artificial and real-world data sets, incorporating modeling errors in the individual predictors, time-shifting dynamics and different cost criteria.

The results show that the merged prediction has a predictive performance in comparison with the best individual predictor in each case, and indicates that the concept may prove useful when dealing with several individual (glucose) predictors of uncertain reliability– reducing the risk associated with definite a priori model selection, or as a means to improve predictive quality if the predictions are diverse enough.

Further research will be undertaken to investigate how interesting features correlated to expected predictor mode changes should be extracted, and in regards to the possibility of making the algorithm unsupervised.

7

Conclusions and Future Work

7.1 Conclusions

This thesis has investigated prediction of glucose evolution in insulin-dependent diabetes patients. To this purpose, minimum order linear models were identified by a subspace-based method in Chapter 4 for a data set containing 47 individual patient data records from the DIAdvisor project. Model identification of diabetic data is obstructed as the two main inputs have opposing effects on the glucose, and in most cases act simultaneously. Furthermore, the a priori knowledge of the static gain of each input is not guaranteed to be incorporated into the model when using the subspace method, nor is the expected integrator-like response. For this reason, the identification results were constrained to satisfy physiologically qualitatively correct responses to the inputs, and the models were corrected in the sense that the pole closest to $z = 1$ was artificially moved to this point, with corresponding corrections of the other pole in order to minimize the disturbance of the characteristic polynomial of the model. The retrieved models were thereafter used for short-term prediction and assessed for performance. The results were compared to previous published results, developed by using other modeling and identification approaches, and proved competitive despite the low complexity. The estimated carbohydrate-to-insulin ratio, a metric often used in clinical practice to optimize the insulin therapy, were compared to the true ratio (amount of digested carbohydrates divided by total administrated bolus doses) with good

correspondence. Some population stratifications, in terms of prediction performance, were also found between the multi-dose injection and the pump patient cohorts.

Another problematic aspect to glucose dynamic modeling and identification was tackled in Chapter 5. The glucose measurements that can be retrieved in a frequent and automatic fashion, *i.e.*, the CGM measurements, are sampled in the interstitial compartment—a tissue that has a diffusion-like relationship to the compartment of primary interest—the circulatory blood system. This aspect is often overlooked in glucose modeling, but significant lagging of the glucose prediction of those models may result, as indicated by the evaluation of the lagging between the CGM signal and the corresponding reference blood glucose measurements in Chapter 3. This is unacceptable, as hypoglycaemia may quickly arise due to rapid glucose drops (see Chapter 3), and these models will, unless perfectly matched, in many cases be unable to capture these potentially dangerous situations. To overcome this deficiency, a modeling approach where the basic subspace identified model of Chapter 4 was augmented, to incorporate the dynamics responsible for the sensor delay, was developed. To prove the concept, an individual dataset with significant sensor lag, retrieved from the same data set as above, was identified in this manner, and short-term post-prandial prediction was evaluated. The results show that the lag of the glucose estimate and prediction were successfully reduced.

Diabetic glucose dynamics is known to comprise both short and more long term time-variability. Merging different diversified models may prove to be a successful approach, as a means to improve performance and robustness under such conditions. In Chapter 6, a novel merging algorithm based in a Bayesian setting was developed. The suggested method admits for soft switching and interpolation between the different models based on an evaluation of the different predictors' recent performance, using a sliding data window, and by looking for data feature identified to be correlated to switching. Different aspects of the merging approach were investigated, using two simulated data series, and the concept was thereafter successfully validated using 12 data sets from the DIAdvisor project.

7.2 Future Research

Generally, biomedical engineering has a wide spectrum of unsolved problems in both basic and applied science. This is certainly the case in diabetology where

still many low hanging fruits are ripe to pick. A list of possible direction for future research could thus be made very long, but below, the outlook has been narrowed down to a few direction for future research endeavours related to the problems addressed in this thesis.

Constrained Identification The method applied in this thesis is heuristic and does not guarantee that any model fulfils the constraints imposed. Further research into constrained identification, e.g., inspired by the possibility to incorporate constrains directly in the subspace identification routine, may be an interesting direction.

Sensor delay The concept of augmenting the model with a sensor model, describing the sensor lag, needs to be further validated by more data examples. Additionally, alternative, more complex models of the sensor dynamics, e.g., incorporating sensor drift, should be addressed.

Time-varying dynamics Time-variability is an important aspect of any time-series, with major implications to the choice of modeling and parameter estimation approach. The data sets analysed in this thesis were not long enough to evaluate long-term time-variability. However, the ambulatory data sets collected during the DAQ trial may prove useful in this regard, and could be evaluated for parameter consistency.

Population Stratification The diabetic population may be possible to stratify into smaller patient cohorts. Finding such classifications could potentially simplify parametrization of previously unmodelled patients, if model behavior could be linked to directly available, or easily measured, biomarkers. Deeper classification analysis of identified models may indicate such relationships.

Ensemble Predictor In order to detect the optimal switching point as soon as possible, the feature monitoring of the SW-BMA ensemble engine is an interesting functionality. However, finding and extracting relevant features is non-trivial and no systematic approach is utilized.

8

Bibliography

Abbott Freestyle Navigator (2012): http://www.abbottdiabetescare.co.uk/your-products/freestyle-navigator.

Abdul-Rasoul, M., H. Habib, and M. Al-Khouly (2006): "'The honeymoon phase' in children with type 1 diabetes mellitus: frequency, duration, and influential factors." *Pediatric diabetes*, **7:2**, pp. 101–7.

Ackerman, E., L. C. Gatewood, J. W. Rosevear, and G. D. Molnar (1965): "Model studies of blood-glucose regulation." *Bulletin of Mathematical Biophysics*, **27:Special Issue**, pp. 21–37.

Agar, B., M. Eren, and A. Cinar (2005): "Glucosim: Educational software for virtual experiments with patients with type 1 diabetes." In *Proc. 2005 Annual Int. Conf. of the IEEE Eng. in Med. and Biol. (EMBC2005)*, pp. 845 –848.

Alessandri, A., M. Baglietto, and G. Battistelli (2005): "Receding-horizon estimation for switching discrete-time linear systems." *Automatic Control, IEEE Trans. on*, **50:11**, pp. 1736 – 1748.

Apidra™ (2012): http://www.apidra.com/.

Arenas-Garcia, J., M. Martinez-Ramòn, A. Navia-Vazquez, and A. R. Figueiras-Vidal (2006): "Plant identification via adaptive combination of transversal filters." *Signal Processing*, **86:9**, pp. 2430 – 2438. Special Section: Signal Processing in UWB Communications.

Balakrishnan, N. P., G. P. Rangaiah, and L. Samavedham (2011): "Review and Analysis of Blood Glucose (BG) Models for Type 1 Diabetic Patients." *Ind. & Eng. Chem. Research*, **50:21**, pp. 12041–12066.

Bequette, B. W. (2004): “Optimal estimation applications to continuous glucose monitoring.” In *Proc. 2004 American Control Conf. (ACC2004)*, vol. 1, pp. 958–962.

Bequette, B. W. (2012): “Challenges and recent progress in the development of a closed-loop artificial pancreas.” *Annual Reviews in Control*, October.

Bergman, R. N. and C. Cobelli (1980): “Minimal modeling, partition analysis, and the estimation of insulin sensitivity.” *Federation Proceedings*, **39:1**, pp. pp. 110–115.

Bishop, C. M. (2006): *Pattern Recognition and Machine Learning*. Springer, Secaucus, NJ, USA.

Bolie, V. W. (1961): “Coefficients of normal blood glucose regulation.” *J. of Appl. Phys.*, **16:5**, pp. 783–788.

Bolin, K., C. Gip, A.-C. Mörk, and B. Lindgren (2009): “Diabetes, healthcare cost and loss of productivity in Sweden 1987 and 2005–a register-based approach.” *Diabetic medicine : a journal of the British Diabetic Association*, **26:9**, pp. 928–34.

Boyne, M. S., D. M. Silver, J. Kaplan, and C. D. Saudek (2003): “Timing of changes in interstitial and venous blood glucose measured with a continuous subcutaneous glucose sensor.” *Diabetes*, **52**, pp. 2790–2794.

Brange, J. and A. Vølund (1999): “Insulin analogs with improved pharmocokinetic profiles.” *Advanced Drug Delivery Reviews*, **35**, pp. pp. 307–335.

Breiman, L. (1996): “Bagging predictors.” *Machine Learning*, **24:2**, pp. 123–140.

Bremer, T. and D. A. Gough (1999): “Is Blood Glucose Predictable From Previous Values? A Solicitation for Data.” *Diabetes*, **48**, March, pp. pp. 445–451.

Cescon, M. (2011): “Linear modeling and prediction in diabetes physiology.” Licentiate Thesis ISRN LUTFD2/TFRT--3250--SE. Department of Automatic Control, Lund University, Sweden.

Cescon, M., F. Ståhl, and R. Johansson (2009): “Subspace-based model identification of diabetic blood glucose dynamics.” In *Proc. 15th IFAC Symp. on System Ident. (SYSID2009), July 6 - 8, 2009, Saint-Malo, France.*

Chan, A., L. Heinemann, and S. Anderson (2010): "Nonlinear Metabolic Effect of Insulin across the Blood Glucose Range in Patients with Type 1 Diabetes Mellitus." *Journal of Diabetes Science and Technology*, **4:4**, pp. 873–881.

Chetty, V. T., A. Almulla, A. Odueyungbo, and L. Thabane (2008): "The effect of continuous subcutaneous glucose monitoring (CGMS) versus intermittent whole blood finger-stick glucose monitoring (SBGM) on hemoglobin A1c (HBA1c) levels in Type I diabetic patients: a systematic review." *Diabetes research and clinical practice*, **81:1**, pp. 79–87.

Chis, O.-T., J. R. Banga, and E. Balsa-Canto (2011): "Structural Identifiability of Systems Biology Models : A Critical Comparison of Methods." *PLoS ONE*, **6:11**.

Christen, U., C. Bender, and M. G. von Herrath (2012): "Infection as a cause of type 1 diabetes?" *Current opinion in rheumatology*, **24:4**, pp. 417–23.

Clarke, W. and B. Kovatchev (2009): "Statistical tools to analyze continuous glucose monitor data." *Diabetes Technol. & Ther.*, **11:1**, pp. S45–S54.

Clarke, W. L., D. Cox, L. A. Gonder-Frederick, W. Carter, and S. L. Pohl (1987): "Evaluating clinical accuracy of systems for self-monitoring of blood glucose." *Diabetes Care*, **10**, pp. 622–628.

Cobelli, C., E. Renard, and B. Kovatchev (2011): "Artificial Pancreas: Past, Present, Future." *Diabetes*, **60:11**, pp. 2672–2682.

Daily Dose (2012): `http://www.insulution.com/sv/insulution/`.

Dalla Man, C., M. Camilleri, and C. Cobelli (2006): "A system model of oral glucose absorption: Validation on gold standard data." *IEEE Trans. Biomed. Eng*, **53:12**, pp. 2472–2478.

Dalla Man, C., D. M. Raimondo, R. A. Rizza, and C. Cobelli (2007a): "GIM , Simulation Software of Meal Glucose Insulin Model." *J. of Diabetes Sci. and Technol*, **1:3**, pp. 1–8.

Dalla Man, C., R. A. Rizza, and C. Cobelli (2007b): "Meal simulation model of the glucose-insulin system." *IEEE Trans. Biomed. Eng*, **54:10**, pp. 1740–1749.

Daskalaki, E., A. Prountzou, P. Diem, and S. G. Mougiakakou (2012): "Real-time adaptive models for the personalized prediction of glycemic profile in type 1 diabetes patients." *Diabetes Technol. & Ther.*, **14:2**, pp. 168–74.

Dassau, E., F. Cameron, B. W. Bequette, H. Zisser, L. Jovanovič, H. P. Chase, D. M. Wilson, B. A. Buckingham, and F. J. Doyle (2010): "Real-Time Hypoglycemia Prediction Suite Using Continuous Glucose Monitoring." *Diabetes Care*, **33:6**, pp. 1249–1254.

Davidson, P. C., H. R. Hebblewhite, R. D. Steed, and B. W. Bode (2008): "Analysis of guidelines for basal-bolus insulin dosing: basal insulin, correction factor, and carbohydrate-to-insulin ratio." *Endocrine Practice*, **14:9**, pp. 1095–1101.

Dexcom Seven Plus (2012): `http://www.dexcom.com/seven-plus`.

DIAdvisor (2012): `http://www.diadvisor.eu`.

Docherty, P. D., J. G. Chase, T. F. Lotz, and T. Desaive (2011): "A graphical method for practical and informative identifiability analyses of physiological models: a case study of insulin kinetics and sensitivity." *Biomed. Eng. Online*, **10:1**, p. 39.

Dye, L., M. Mansfield, N. Laisikiewicz, L. Mahawish, R. Schnell, D. Talbot, H. Chauhan, F. Croden, and C. Lawton (2010): "Correspondance of continuous interstitial glucose measurement against arterialised and capillary glucose following an oral glucose tolerance test in healthy voulenteers." *British J. of Nutrition*, **103**, pp. 134–140.

Elliott, G., C. W. Granger, and A. Timmermann, Eds. (2006): *Handbook of Economic Forecasting*, chapter 10. Forecast Combinations. Elsevier.

Elton, E. J., M. J. Gruber, and M. W. Padberg (1976): "Simple criteria for optimal portfolio selection." *J. of Finance*, **31:5**, pp. 1341–1357.

Eren-Oruklu, M., A. Cinar, and L. Quinn (2010): "Hypoglycemia prediction with subject-specific recursive time-series models." *Journal of diabetes science and technology*, **4:1**, pp. 25–33.

Eren-Oruklu, M., A. Cinar, L. Quinn, and D. Smith (2009a): "Adaptive control strategy for regulation of blood glucose levels in patients with type 1 diabetes." *J. of Proc. Cont.*, **19:8**, pp. 1333–1346.

Eren-Oruklu, M., A. Cinar, L. Quinn, and D. Smith (2009b): "Estimation of future glucose concentrations with subject-specific recursive linear models." *Diabetes Technol. & Ther.*, **11:4**, pp. 243–53.

Estrada, G., H. Kirchsteiger, L. del Re, and E. Renard (2010): "Innovative approach for online prediction of blood glucose profile in type 1 diabetes patients." In *American Control Conf. (ACC2010), 2010*, pp. 2015 –2020.

Facchinetti, A., G. Sparacino, and C. Cobelli (2007): "Reconstruction of glucose in plasma from interstitial fluid continuous glucose monitoring data: Role of sensor calibration." *J. of Diabetes Sci. and Technol*, **5**, pp. 617–623.

Farmer, T. G., T. F. Edgar, and N. A. Peppas (2009): "Effectiveness of Intravenous Infusion Algorithms for Glucose Control in Diabetic Patients Using Different Simulation Models." *Ind. & Eng. Chem. Research*, **48:9**, pp. 4402–4414.

Finan, D. A., F. J. Doyle, C. C. Palerm, W. C. Bevier, H. C. Zisser, L. Jovanovic, and D. E. Seborg (2009a): "Experimental evaluation of a recursive model identification technique for type 1 diabetes." *J. of Diabetes Sci. and Technol*, **3:5**, pp. 1192–202.

Finan, D. A., C. Palerm, F. J. Doyle, D. E. Seborg, H. Zisser, W. C. Bevier, and L. Jovanovic (2009b): "Effect of input excitation on the quality of empirical dynamic models for type 1 diabetes." *AIChE Journal*, **55:5**, pp. 1135–1146.

Freckmann, G., A. Baumstark, N. Jendrike, E. Zschornack, S. Kocher, J. Tschiananga, F. Heister, and C. Haug (2010): "System Accuracy Evaluation of 27 Blood Glucose Monitoring Systems According to DIN EN ISO 15197." *Diabetes Technol. & Therap.*, **12:3**, pp. 221–231.

Galvanin, F., M. Barolo, and F. Bezzo (2011): "Online redesign of clinical tests for the identification of type 1 diabetes models in the presence of continuous glucose monitoring systems." *Proc. of the 18th IFAC World Congress*, pp. 8328–8333.

Galvanin, F., M. Barolo, S. Macchietto, and F. Bezzo (2009): "Optimal Design of Clinical Tests for the Identification of Physiological Models of Type 1 Diabetes Mellitus." *Ind. & Eng. Chem. Research*, **48:4**, pp. 1989–2002.

Gani, A., A. V. Gribok, Y. Lu, W. K. Ward, R. A. Vigersky, and J. Reifman (2010): "Universal glucose models for predicting subcutaneous glucose concentration in humans." *Trans. Info. Tech. Biomed*, **14:1**, pp. 157–165.

Gani, A., A. V. Gribok, S. Rajaraman, W. K. Ward, and J. Reifman (2009): "Predicting Subcutaneous Glucose Concentration in Humans : Data-Driven Glucose Modeling." *IEEE Trans. Biomed. Eng.*, **56:2**, pp. 246–254.

Georga, E. I., V. C. Protopappas, and D. I. Fotiadis (2011): “Glucose Prediction in Type 1 and Type 2 Diabetic Patients Using Data Driven Techniques.” In Funatsu, Ed., *Knowledge-Oriented Applications in Data Mining*, chapter 17. InTech.

Gustafsson, F. (2000): *Adaptive Filtering and Change Detection*. John Wiley & Sons, Hoboken, New Jersey, USA.

Hanas, R. and G. John (2010): “2010 consensus statement on the worldwide standardization of the hemoglobin A1C measurement.” *Diabetes care*, **33:8**, pp. 1903–4.

HemoCue Glucose 201+ Analyzer (2012): `http://www.hemocue.com/`.

Henriksson, F., C. D. Agardh, C. Berne, J. Bolinder, F. Lönnqvist, P. Stenström, C. G. Ostenson, and B. Jönsson (2000): “Direct medical costs for patients with type 2 diabetes in Sweden.” *Journal of internal medicine*, **248:5**, pp. 387–96.

Hildebrandt, P., K. Birch, L. Sestoft, and a. Vø lund (1984): “Dose-dependent subcutaneous absorption of porcine, bovine and human NPH insulins.” *Acta medica Scandinavica*, **215:1**, pp. 69–73.

Hirata, Y. and Y. Uchigata (1994): “Insulin autoimmune syndrome in Japan.” *Diabetes research and clinical practice*, **24 Suppl**, October, pp. S153–7.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999): “Bayesian model averging: A tutorial.” *Statistical Science*, **14:4**, pp. 382–417.

Home, P. D. (1997): “Insulin Therapy.” *International Textbook of Diabetes Mellitus*, pp. 899–928.

Hönes, J., P. Müller, and N. Surridge (2008): “The Technology Behind Glucose Meters: Test Strips.” *Diabetes Technol. & Therap.*, **10:s1**, pp. S–10–S–26.

Hovorka, R., V. Canonico, L. J. Chassin, U. Haueter, M. Massi-Benedetti, M. O. Federici, T. R. Pieber, H. C. Schaller, L. Schaupp, T. Vering, and M. E. Wilinska (2004): “Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes.” *Physiol. Meas.*, **25:4**, pp. 905–920.

Hovorka, R., L. J. Chassin, M. Ellmerer, J. Plank, and M. E. Wilinska (2008): “A simulation model of glucose regulation in the critically ill.” *Physiological measurement*, **29:8**, pp. 959–78.

Humalog^{TM} (2012): `http://www.humalog.com/`.

International Diabetes Federation (2012): `http://www.idf.org/fact-sheets`.

Johansson, R. (2009): *System Modeling & Identification*. KFS AB.

Kahn, C. R. (1997): "Insulin Receptors and Insulin Signaling in Normal and Disease States." *International Textbook of Diabetes Mellitus*, pp. 437–467.

Keenan, D. B., J. J. Mastrototaro, G. Voskanyan, and G. M. Steil (2009): "Delays in minimally invasive continuous glucose monitoring devices : A review of current technology." *J. Diabetes Sci. and Technol*, **3:5**, pp. 1207–1214.

Kirchsteiger, H., G. C. Estrada, S. Pölzer, E. Renard, and L. Re (2011): "Estimating Interval Process Models for Type 1 Diabetes for Robust Control Design." In *IFAC World Congress 2011*, pp. 11761–11766.

Kolter, J. Z. and M. A. Maloof (2003): "Dynamic weighted majority: A new ensemble method for tracking concept drift." *Data Mining, IEEE Int. Conf. on*, **0**, pp. 123–130.

Kovatchev, B., C. Breton, C. Dalla Man, and C. Cobelli (2008): "In silico model and computer simulation environment approximating the human glucose/insulin utilization." Technical Report. Food and Drug Administration Master File MAF 1521.

Kovatchev, B., M. Straume, D. Cox, and L. Farhy (2000): "Risk analysis of blood glucose data: A quantitative approach to optimizing the control of insulin dependent diabetes." *J. of Theor. Med*, **3**, pp. 1–10.

Kovatchev, B. P., D. J. Cox, L. A. Gonder-Frederick, and W. L. Clarke (1997): "Symmetrization of the Blood Glucose Measurement Scale and Its Applications." *Diabetes Care*, **20**, pp. pp. 1655–1658.

Kovatchev, B. P., C. King, M. Breton, S. Anderson, and W. Clarke (2006): "Clinical assessment and mathematical modeling of the accuracy of continuous glucose sensors (cgs)." In *Proc. IEEE EMBS 28th Annual Int. Conf. (EMBC2006)*, pp. 71–74.

Kovatchev, B. P., D. Shields, and M. Breton (2009): "Graphical evaluation of continuous glucose sensing time lag." *Diabetes Technol. & Ther.*, **11:3**, pp. 139–143.

Landin-Olsson, M. (2002): "Latent Autoimmune Diabetes in Adults." *Annals of the New York Academy of Sciences*, **958:1**, pp. 112–116.

Lantus™ (2012): `http://www.lantus.com/`.

Leal, Y., W. Garcia-Gabin, and J. Bondia (2010): "Real-time glucose estimation algorithm for continuous glucose monitoring using autoregressive models." *J. of Diabetes Sci. and Technol.*, **4:2**, pp. 391–403.

Lee, H., B. A. Buckingham, D. M. Wilson, and B. W. Bequette (2009): "A closed-loop artificial pancreas using model predictive control and a sliding meal size estimator." *J. of Diabetes Sci. and Technol.*, **3:5**, pp. 1082–90.

Lehmann, E., I. Hermanyi, and T. Deutsch (1994): "Retrospective validation of a physiological model of glucose-insulin interaction in type 1 diabetes mellitus." *Med. Eng. & Phys.*, **16:4**, pp. 351–352.

Lehmann, E. D. (1994): "AIDA : an interactive diabetes advisor." *Computer Methods and Programs in Biomedicine*, **2607:93**, pp. 183–203.

Lehmann, E. D. and T. Deutsch (1992): "A physiological model of glucose-insulin interaction in type 1 diabetes mellitus." *J. of Biomed. Eng.*, **14**, may, pp. pp. 235–242.

Levemir™ (2012): `http://www.levemir.com/`.

Lindholm, A., L. B. Jensen, P. D. Home, P. Raskin, B. O. Boehm, and J. Rastam (2002): "Immune Responses to Insulin Aspart and Biphasic Insulin Aspart in People With Type 1 and Type 2 Diabetes." *Diabetes Care*, **25:5**, pp. 876–882.

Littlestone, N. and M. K. Warmuth (1994): "The weighted majority algorithm." *Information and Computation*, **108:2**, pp. 212–261.

Lu, Y., S. Rajaraman, W. K. Ward, R. A. Vigersky, and J. Reifman (2011): "Predicting human subcutaneous glucose concentration in real time: a universal data-driven approach." In *Proc. 2011 Annual Int. Conf. of the IEEE Eng. in Med. and Biol. (EMBC2011)*, pp. 7945–8.

Makroglou, A., J. Li, and Y. Kuang (2006): "Mathematical models and software tools for the glucose-insulin regulatory system and diabetes: an overview." *Appl. Num. Math*, pp. 559–573.

MathWorks (2012): "Matlab™." `www.mathworks.com`.

MedTronic (2012): `http://www.medtronic-diabetes.se/`.

Mokan, M., A. Mitrakou, T. Veneman, C. Ryan, M. Korytkowski, P. Cryer, and J. Gerich (1994): "Hypoglycemia unawareness in IDDM." *Diabetes Care*, **17:12**, pp. 1397–1403.

Nauck, M. A., B. Baller, and J. J. Meier (2004): "Gastric inhibitory polypeptide and glucagon-like peptide-1 in the pathogenesis of type 2 diabetes." *Diabetes*, **53 Suppl 3:December**, pp. S190–6.

Naumova, V., S. Pereverzyev, and S. Sampath (2011): "A meta-learning approach to the regularized learning—Case study: Blood glucose prediction." Technical Report. Johann Radon Institute for Computational and Applied Mathematics (RICAM), Linz Austria.

Novolog™ (2012): `http://www.novolog.com/`.

Nucci, G. and C. Cobelli (2000): "Models of subcutaneous insulin kinetics. a critical review." *Computer Methods and Programs in Biomedicine*, **62**, pp. pp. 249–257.

Nussey, S. and S. Whitehead (2001): *Endocrinology: An Integrated Approach.* BIOS Scientific Publishers Limited.

Ohlsson, H., L. Ljung, and S. Boyd (2010): "Segmentation of ARX-models using sum-of-norms regularization." *Automatica*, **46:6**, pp. 1107–1111.

Oza, N. (2005): "Online bagging and boosting." In *Systems, Man and Cybernetics, 2005 IEEE Int. Conf. on*, vol. 3, pp. 2340 – 2345 Vol. 3.

Palerm, C. C. and B. W. Bequette (2007): "Hypoglycemia detection and prediction using continuous glucose monitoring-a study on hypoglycemic clamp data." *Journal of diabetes science and technology*, **1:5**, pp. 624–9.

Palerm, C. C., J. P. Willis, J. Desemone, and B. W. Bequette (2005): "Hypoglycemia Prediction and Detection Using Optimal Estimation." *Diabetes Technology & Therapeutics*, **7:1**, pp. 3–14.

Pappada, S. M., B. D. Cameron, P. M. Rosman, R. E. Bourey, T. J. Papadimos, W. Olorunto, and M. J. Borst (2011): "Neural network-based real-time prediction of glucose in patients with insulin-dependent diabetes." *Diabetes Technol. & Ther.*, **13:2**, pp. 135–41.

Percival, M., W. Bevier, and Y. Wang (2010): "Modeling the effects of subcutaneous insulin administration and carbohydrate consumption on blood glucose." *J. of Diabetes*, **39:3**, pp. 800–805.

Percival, M., Y. Wang, B. Grosman, E. Dassau, H. Zisser, L. Jovanovič, and F. Doyle (2011): "Development of a multi-parametric model predictive control algorithm for insulin delivery in type 1 diabetes mellitus using clinical parameters." *J. of Proc. Control*, **21:3**, pp. 391–404.

Pérez-Gandía, C., A. Facchinetti, G. Sparacino, C. Cobelli, E. J. Gómez, M. Rigla, A. D. Leiva, and M. E. Hernando (2010): "Artificial Neural Network Algorithm for Online Glucose." *Diabetes Technol. & Ther.*, **12:1**, pp. 81–88.

Perriello, G., P. De Feo, E. Torlone, C. Fanelli, F. Santeusanio, P. Brunetti, and G. B. Bolli (1991): "The dawn phenomenon in Type 1 (insulin-dependent) diabetes mellitus: magnitude, frequency, variability, and dependency on glucose counterregulation and insulin sensitivity." *Diabetologia*, **34:1**, pp. 21–28.

Plank, J., M. Bodenlenz, F. Sinner, C. Magnes, E. Görzer, W. Regittning, L. A. Endahl, E. Draeger, M. Zdravkovic, and T. Pieber (2005): "A Double-Blind, Randomized, Dose-Response Study Investigating the Pharmacodynamic and Pharmacokinetic Properties of the Long-Acting Insulin." *Diabetes Care*, **28:5**, pp. 1107–1112.

Poulsen, J., A. Avogaro, F. Chauchard, C. Cobelli, R. Johansson, L. Nita, M. Pogose, L. del Re, E. Renard, S. Sampath, F. Saudek, M. Skillen, and J. Soendergaard (2010): "A diabetes management system empowering patients to reach optimised glucose control: From monitor to advisor." In *Proc. 2010 Annual Int. Conf. of the IEEE Eng. in Med. and Biol. (EMBC2010)*, pp. 5270 – 5271.

Puckett, W. R. (1992): *Dynamic Modeling of Diabetes Mellitus.* Ph.D. thesis, University of Wisconsin-Madison.

Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Pololakowski (2005): "Using bayesian model averaging to calibrate forecast ensembles." *Monthly Weather Review*, **133**, May, pp. 1155–1174.

Raftery, A. E., M. Kárný, and P. Ettler (2010): "Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill." *Technometrics*, **52:1**, pp. 52–66.

Rebrin, K. and G. M. Steil (2000a): "Can interstitial glucose assesment replace blood glucose measurements." *Diabetes Technol. & Ther*, **2:3**, pp. 461–472.

Rebrin, K. and G. M. Steil (2000b): "Can interstitial glucose assessment replace blood glucose measurements?" *Diabetes Technol. & Therap.*, **2:3**, pp. 461–72.

Roy, A. and R. S. Parker (2007): "Dynamic Modeling of Exercise Effects on Plasma Glucose and Insulin Levels." *J. Diabetes Sci. and Technol.*, **1:3**, pp. 338–347.

Sasaki, S. (2002): "Mechanism of insulin action on glucose metabolism in ruminants." *Animal Science Journal*, **73**, pp. pp. 423–433.

Shoelson, S. E. and P. A. Halban (1994): "Insulin Biosynthesis and Chemistry." *Joslin's Diabetes Mellitus*, pp. 29–55.

Siler, S. Q., R. A. Neese, M. P. Christiansen, M. K. Hellerstein, Q. Scott, and M. P. Chris (1998): "The inhibition of gluconeogenesis following alcohol in humans." *Am. J. Physiology*, **275**, pp. E897–E907.

Sorensen, J. T. (1985): *A physiologic model of glucose metabolism in man and its use to design and assess improved insulin therapies for diabetes.* PhD thesis, Massachusetts Institute of Technologhy.

Sparacino, G., F. Zanderigo, S. Corazza, A. Maran, A. Fachinetti, and C. Cobelli (2007): "Glucose concentration can be predicted ahead in time from continuous glucose monitoring sensor time-series." *IEEE Trans. Biomed. Eng*, **54:5**, pp. 931–937.

St John, A., W. A. Davis, C. P. Price, and T. M. E. Davis (2010): "The value of self-monitoring of blood glucose: a review of recent evidence." *J. of diabetes and its complications*, **24:2**, pp. 129–41.

Ståhl, F. (2003): "Diabetes Mellitus Modelling Based on Blood Glucose Measurements." Master Thesis TFRT-5703. Dept. Automatic Control, Lund University, Sweden.

Ståhl, F. and R. Johansson (2009): "Diabetes mellitus modeling and short-term prediction based on blood glucose measurements." *Math. Biosci.*, **217**, January, pp. 101–117.

Stork, A. D., H. Kemperman, D. W. Erkelens, and T. F. Veneman (2005): "Comparison of the accuracy of the Hemocue glucose analyser with the Yellow Springs Instrument glucose oxidase analyser, particularly in hypoglycemia." *Eur. J. of Endocrionol*, **153**, pp. 275–281.

Takagi, T. and M. Sugeno (1985): “Fuzzy identification of system and its applications to modelling and control.” *IEEE Trans. on Systems, Man, and Cybernetics.*, **SMC-15**, Jan-Feb, pp. 116–132.

The DIAdvisor Consortium (2012): “Final publishable summary report.” `http://cordis.europa.eu/results/home_en.html`.

The Swedish National Board of Health and Welfare (2012a): “Folkhälsorapporten 2009.” `http://www.socialstyrelsen.se/`.

The Swedish National Board of Health and Welfare (2012b): “Öppna jämförelser av hälso- och sjukvårdens kvalitet och effektivitet jämförelser mellan landsting 2009.” `http://www.socialstyrelsen.se/`.

Tonyushkina, K. and J. H. Nichols (2009): “Glucose meters: a review of technical challenges to obtaining accurate results.” *J. of diabetes Sci. and Technol.*, **3:4**, pp. 971–80.

Toumaz (2012): `http://www.toumaz.com`.

Turner, B. C., E. Jenkins, D. Kerr, R. S. Sherwin, and D. A. Cavan (2001): “The Effect of Evening Alcohol Consumption on Next-Morning Glucose Control in Type 1 Diabetes.” *Diabetes Care*, **24:11**, pp. 1888–1893.

Vaddiraju, S., D. J. Burgess, I. Tomazos, F. C. Jain, and F. Papadimitrakopoulos (2010): “Technologies for continuous glucose monitoring: current problems and future promises.” *J. of Diabetes Sci. and Technol*, **4:6**, pp. 1540–62.

Van Cauter, E., K. S. Polonsky, and A. J. Scheen (1997a): “Roles of circadian rhythmicity and sleep in human glucose regulation.” *Endocrine reviews*, **18:5**, pp. 716–38.

Van Cauter, E., K. S. Polonsky, and A. J. Scheen (1997b): “Roles of circadian rhythmicity and sleep in human glucose regulation.” *Endocrine Reviews*, **18:5**, pp. 716–38.

Van Haeften, T. W. (1989): “Clinical significance of insulin antibodies in insulin-treated diabetic patients.” *Diabetes Care*, **12:9**, pp. 641–8.

VivoMetrics (2012): `http://vivonoetics.com/products/sensors/lifeshirt/`.

Voulgari, C., S. Pagoni, S. Paximadas, and A. I. Vinik (2012): “”Brittleness” in diabetes: easier spoken than broken.” *Diabetes Technol. & Therap.*, **14:9**, pp. 835–48.

White, M. F. and C. R. Kahn (1994): "Molecular Aspects of Insulin Action." *Joslin's Diabetes Mellitus*, pp. 139–162.

Wild, S., G. Roglic, A. Green, R. Sicree, and H. King (2004): "Global Prevalence of Diabetes Estimates for the year 2000 and projections for 2030." *Diabetes Care*, **27:5**, pp. 1047–1053.

Wilinska, M. E., L. J. Chassin, H. C. Schaller, L. Schaupp, T. R. Pieber, and R. Hovorka (2005): "Insulin kinetics in type-1 diabetes: Continuous and bolus delivery of rapid acting insulin." *IEEE Trans. Biomed. Eng*, **52:1**, pp. 3–12.

Worthington, D. R. L. (1990): "The use of models in the self-management of insulin-dependent diabetes mellitus." *Computer Methods and Programs in Biomedicine*, **32**, pp. pp. 233–239.

Yellow Springs Instrument (2012): `http://www.ysilifesciences.com/`.

Zecchin, C., A. Facchinetti, G. Sparacino, G. De Nicolao, and C. Cobelli (2011): "A new neural network approach for short-term glucose prediction using continuous glucose monitoring time-series and meal information." *2011 Annual Int. Conf. of the IEEE Eng. in Med. and Biol. Soc. (EMBC2011)*, August, pp. 5653–5656.

Zecchin, C., A. Facchinetti, G. Sparacino, G. D. Nicolao, S. Member, and C. Cobelli (2012): "Neural Network Incorporating Meal Information Improves Accuracy of Short-Time Prediction of Glucose Concentration." *IEEE Trans. Biomed. Eng.*, **59:6**, pp. 1550–1560.