Clinical risk measure for variability analysis and real-time prevention of hyper/hypo-glycaemic episodes from continuous glucose monitoring time-series

Stefania Guerra

Department of Information Engineering, University of Padova

Supervisor Prof. Giovanni Sparacino

A thesis submitted for the degree of

PhilosophiæDoctor (PhD)

2012 March

Contents

Glossary

AP	Artificial Pancreas
BG	Blood Glucose (Concentration)
CGM	Continuous Glucose Monitoring
CONGA	Continuous Overlapping Net Glycemic Action
CSII	Continuous Subcutaneous Insulin Infusion
CV	Coefficient of Variation
DR	Dynamic Risk
EN	Elastic Net
HbA1c	Glycated Hemoglobin
HBGI	High Blood Glucose Index
ICG	Index of Glycemic Control
IDDM	Insulin Dependent Diabetes Mellitus
IQR	Inter-Quartile-Range
LASSO	Least Absolute Shinkage and Selection Operator
LBGI	Low Blood Glucose Index
LLTR	Lower Limit of Target Range
MI	Multi-Injective
MODD	Mean of Daily Differences
NICGM	Noninvasive Continuous Glucose Monitoring
NIDDM	Non Insulin Dependent Diabetes Mellitus

OLS Ordinary Least Squares

PCA Principal Component Analysis

$\mathbf{SD_{b\ hh:mm\ //\ dm}}$ Standard Deviation between days within time points, corrected for changes in daily means

$\mathbf{SD_{b\ hh:mm}}$ Standard Deviation between days within time points

$\mathbf{SD_{dm}}$ Standard Deviation daily means

$\mathbf{SD_{hh:mm}}$ Standard Deviation between time points

$\mathbf{SD_{I}}$ Standard Deviation Interaction

$\mathbf{SD_{T}}$ Standard Deviation

$\mathbf{SD_{ws\ h}}$ Standard Deviation within series

$\mathbf{SD_{w}}$ Standard Deviation within days

SH Severe Hypoglycaemia

SMBG Self Monitoring Blood Glucose

SNR Signal to Noise Ratio

SVD Singular Value Decomposition

T1D Type 1 Diabetes

T2D Type 2 Diabetes

ULTR Upper Limit of Target Range

WESS Weighted estimates sum of squares

WRSS Weighted residuals sum of squares

Summary

Part I - Background and Aims

Diabetes is a chronic disease characterized by impairments both in the secretion/action the insulin hormone. As discussed in Chapter 1, insulin acts on the metabolism of glucose, since it enhances glucose uptake by the tissues and suppresses hepatic glucose production. Diabetes is treated with a combination of insulin infusions, drugs and physical exercise. Unfortunately, diabetes therapy is quite difficult to tune individually, and in diabetics glycemia (glucose concentration in blood, BG) often exceeds the normal range (70-180 mg/dl). Conditions of hypoglycemia (BG<70 mg/dl) and hyperglycemia (BG>180 mg/dl) are threatening for the patient in the short and long term, respectively. Severe hypoglycemia causes a suffering of the tissues, and in particular of the brain (because glucose is their most important nutrition factor), and may lead to coma. Hyperglycemia induces complicances such as retinopathies, nephropathies and various cardiovascular diseases. For these reasons, it is important to achieve tight control of the glycemia. Glycemia can be monitored in diabetic patients mainly though 2 strategies. The first is based on few measurements a day performed via fingerpricks (self monitoring blood glucose, SMBG). The second is based on the recently developed continuous glucose monitoring (CGM) devices, which allow collecting a CGM measurement every 1-5 minutes, hence allowing to track the physiological glucose dynamics with finer detail.

In order to achieve a good therapy it is crucial to be able to alert the patient promptly in order to allow time to the therapeutic actions (sugar ingestion, insulin administration) to be effective. Moreover, it is important to characterize the individual features of the patient

by analyzing the so-called glucose variability.

As discussed in Chapter 2, several indexes and transformation of the glucose scale may be used to characterize glucose variability. Along with several indexed based on statistical analysis of the glucose time-series, it has been suggested by Kovatchev (31) that the study of glucose concentration time-series should take into account that the glycemic range is asymmetric and that the distribution of glucose concentration values is skewed within the range. Kovatchev proposed a punctual transformation of the glucose scale (static risk, SR), assigning to each glucose level a specific risk value. Most indexes available in the literature for the analysis of glycemic time-series were developed based on SMBG measurements.Today, the availability of CGM signals opens the possibility to embed glucose dynamics information in the analysis of glucose time-series. In this thesis the concept of clinical risk and of glucose variability will be developed explicitly embedding the trend of the glucose signal in a new "Dynamic" Risk Function, which will be used as basis for online generation of alerts, and for the definition of new variability indexes. Such indexes will be then included in a multivariate analysis for the parsimonoius description of glucose variability.

Part II - Conceptual Design and Implementation

In this thesis the Kovatchev's risk function (Static Risk) will be further developed to be adapted and improved exploiting the natural information brought by CGM signals. In fact, these time-series are sampled more frequently, and allow an estimate of the first time derivative of the glucose signal. In Chapter 3, a Dynamic Risk (DR) function will be defined, which explicitly modulates the risk function through the information of the first-time derivative. In particular, the DR function will be structured in such a way that the actual SR associated with the glucose level is amplified or damped according to the time-derivative: if the glucose signal is heading toward a critical region (hypo or hyperglycemia) the risk is increased, while it is reduced if the glycemia is recovering to safety values. As far as the algorithms for the DR implementation are concerned, in Chapter 4, particular care will be put in the strategy to evaluate the first time derivative, since the measurement noise, always present in the CGM signals, tends to be amplified by the derivative. In this work this task will be carried out by a method based on regularized deconvolution which is suitable for online application, e.g. to be used in algorithms for the generation of hypo and hyperglycemic alerts.

Part III - Application of DR for Hypo/Hyperglycemic Alert Generation

DR is intrinsically predictive, allowing anticipating threshold crossings and hence alerting the patient with relative temporal gain to take adequate therapeutic actions. To show this, in Chapter 5the DR function will be applied both on simulated and on real CGM data collected by several type of sensors. Results on simulated and real data show that hypoglycemic events can be anticipated of 12 to 18 minutes by simply evaluating the DR of the glycemic time-series. Moreover, in Chapter 6 a hypo alert generation strategy based on the combination of a short-term prediction of the CGM signal via Kalman Filter and DR will allow anticipating the event of more than 20 minutes with relatively small number of false hypo/hyperglycemic alerts.

Part IV - Application of DR for the (Parsimonious) Description of Glucose Variability

Dozens of indexed evaluated from CGM/SMBG time series have been proposed to characterize and assess glucose variability. None of them explicitly exploit the continuous feature of CGM signals, in fact, none explicitly includes the information of time-derivative as a variability factor. Based on the Dynamic Risk Space (DRS), a phase plot where glucose trajectories (glucose value vs time-derivative) can be effectively shown to evaluate glucose variability and control, several new indexes will be developed in Chapter 7 to analyze potentially useful features of glucose signals.

Then, starting from the tens of indexes proposed in the literature and those developed in the present thesis, a multivariate technique, the Sparse Principal Component Analysis will be applied in Chapter 8 to a set of 48 indexes evaluated on a dataset of 60 CGM signals, in order to understand which is the best combination for the description of glucose variability. SPCA indicates that 5 components are relevant for the analysis of the dataset variance and that indexes explicitly considering information about the first time derivative are relevant for the description of glucose variability.

In summary, the utility of the DR function developed in this thesis is twofold:

- It allows assessing a measure of dynamic risk (function of both level and trend of the glucose signal) and results to be predictive of threshold crossings

- It allows defining new indexes for the assessment of glucose variability and quality of control which result to be complementary to indexes available in the literature.

Part I

Background and Aims

1 Monitoring & Therapy of Diabetes

1.1 Pathophysiology of Diabetes

In human beings, glucose represents the basic nutrition factor for the muscles and the only energy source for the brain. Glucose reaches the blood stream via several mechanisms (released by the intestine after a meal, or produced by the liver and, in small part, by the kidneys in fasting conditions) and is then absorbed by tissues either via hormone-mediate mechanisms (e.g. by the muscles) or via non-mediated transportation (e.g. by the brain). In particular, the metabolism of glucose in the healthy people is mainly controlled by insulin, an hormone secreted by the islets of Langerhans in the pancreatic beta-cells, which lowers glucose concentration in blood after a meal by facilitating the uptake of glucose by the muscles and by suppressing the hepatic production of glucose by the liver. If the glycaemia decreases and sufficient nutrients delivery to the tissues is not guaranteed, counter-regulatory hormones such as glucagon are secreted and stimulate the conversion of glycogen to glucose, allowing to keep the concentration of glucose in the safety range (39).
Diabetes is a chronic disease, characterized by the inability of the body to control the concentration of glucose (glycaemia) in blood. In particular, deficiency in the secretion and action of insulin represents the main cause of impaired glucose control. Although impairments in the glycaemic control may be referred to with the generic definition of Diabetes, such impairments can have very different genesis, two major pathologies are usually considered: Type 1 and Type 2 diabetes (T1D and T2D respectively).

1. **Type 1 Diabetes** also known as Insulin Dependent Diabetes Mellitus (IDDM or, formerly juvenile diabetes) is usually associated with autoimmune processes which de-

stroy the pancreatic beta-cells, although in some cases (10%) no evidence is present to confirm autoimmune attacks. Patients with T1D usually show the first symptoms in the childhood/adolescence; major symptoms are the polyuria (frequent urination), polydipsia (increased thirst), and weight loss.

2. **Type 2 Diabetes** also known as Non Insulin Dependent Diabetes Mellitus (NIDDM) usually affects elderly patients. It represents the 90% of the prevalence in the diabetic population. It is usually caused by a resistance to insulin action at a peripheral level, which results in augmented secretion. The combination of hyperglycemia (due to the resistance and the decreased uptake of glucose in the periphery) and the altered secretion causes a stressed the pancreas which eventually deteriorates until it is unable to secrete insulin.

Diabetes is taking on epidemic proportions with over 346 millions individuals affected by this disease worldwide (1 out of 20 adults, 90% of whom have Type-2 diabetes). By 2030, the number of diabetics is expected to double (66).
Tight control of the glycaemic level is of crucial importance, since conditions of hypoglycaemia (blood glucose concentration, BG<70mg/dl) and hyperglycaemia (BG>180 mg/dl) are life threatening. In particular, in condition of hypoglycaemia energy supply to the brain is not guaranteed and in the very short-term (minutes) symptoms like shivering, cold feeling and headache may occur. If the counter-regulation is not sufficient, coma may eventually occur when the glycaemia falls below the severe hypoglycaemia (SH) threshold (around 35-40 mg/dl). Prolonged exposure to high glucose concentration is also dangerous for patients in the long term, since it is the cause of complications such as nephropathies, cardiovascular diseases and retinopathies.
In the diabetic patient, the glycaemic level often exceeds the normality range since the action/secretion of insulin is lacking or missing. The conventional therapy for T1D and T2D is different. In T1D exogenous injections of insulin are needed to compensate for missing secretion from the pancreas. Before each meal, the patient decides the insulin bolus to be injected to allow the tissues to uptake the glucose that will reach the bloodstream. Such bolus is defined according to tables designed by the physician and tuned on the patient's history. Moreover, either slow-acting insulin or a continuous infusion of insulin are infused to mimic the so called insulin basal rate, which allows the body to continuously absorb the glucose which is produced mostly by the liver. For some details in insulin delivery including

references to research on the artificial pancreas, we refer the reader to Appendix A. In T2D, the therapy consists in exercise and diet. In some T2D subjects, after years of overproduction of insulin, the pancreas may cease to secrete insulin and exogenous insulin infusions become necessary (48).
Insulin dosing is a very difficult task, and often patients are not able to maintain their glucose concentration "in target" because of insulin under/overdosing. It is very important to keep the glycaemic concentration in blood monitored in order to effectively tune the insulin bolus and basal rate. Patient with diabetes are thus required to monitor their blood glucose levels frequently, as explained by the following section, where Self Monitoring Blood Glucose (SMBG) and the new Continuous Glucose Monitoring (CGM) will be described.

1.2 Technologies for Glucose Monitoring in Diabetic Patients

The most established and used technique to monitor the glycaemia, is the used of the so called Self Monitoring Blood Glucose, or SMBGs. Devices for the Self Monitoring of glucose have become available in the early seventies, and have now become a pocket tool that any diabetic uses daily.

1.2.1 Self Monitoring Blood Glucose (SMBG)

Figure 1.1: Commercially available SMBG devices. From left to right: LifeScan One Touch® Ultra ®2, Accu-Chek® Aviva, and iBGstar ™marketed by sanofi-aventis.

The term SMBG refers to measurements of capillary blood glucose taken via a finger pricks. SMBGs are point-in-time measurements, which provide accurate information (35) about the glucose concentration in blood. The measurement is usually performed with a glucometer; examples of commercially available devices are shown in Fig 1.1. While the first two are standalone devices (LifeScan One Touch® Ultra ®2 (36) and Accu-Chek® Aviva (51)) and only need to be fed with a measurement strips, the third device (iBGstar ™marketed by

sanofi-aventis (53)) can be connected to an Apple iPhone to register all the information that a patient needs, and can be interfaced with pieces of software that run on the smartphone. The measurements provided by the glucometers are sufficiently accurate to assess the glycemia in a specific moment, but unfortunately, the sampling procedure cannot be repeated more than 5-6 times a day. Indeed, the finger prick is painful for the patient, who needs to collect a drop of blood from the fingertips at each measurement. Few SMBGs give a glimpse on what the punctual glycaemia is, but do not allow capturing the glucose dynamics, which are much more complex. In order to overcome these limitations, the so-called Continuous Glucose Monitoring CGM devices were developed at the beginning of the 21st century.

1.2.2 Continuous Glucose Monitoring (CGM)

The need for continuous measurements of the glycaemia can be easily understood by analyzing Figure 1.2, where few SMBGs (red stars) are plotted against the continuous glucose (blue line) measured for two days with a CGM device. The richer information of the CGM device allows to highlight, for example, the hypoglycaemic event at minutes 1600 and 2800.

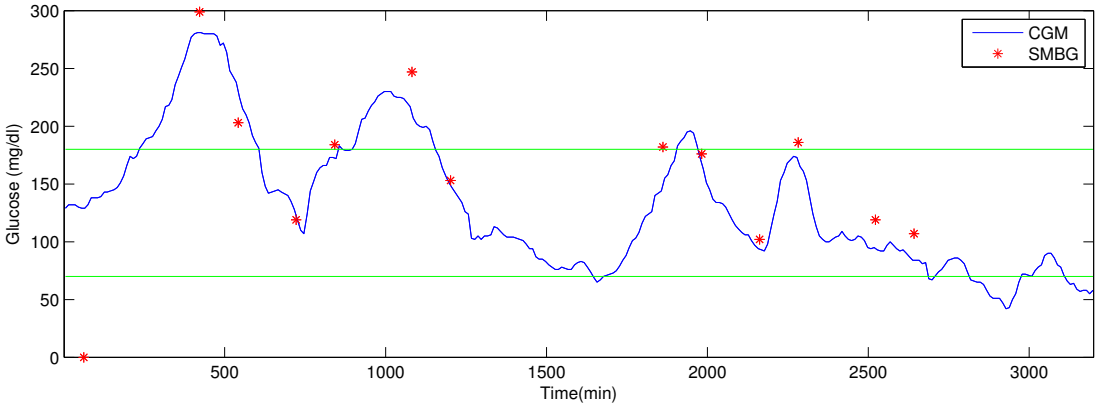

Figure 1.2: Comparison between SMBG (red stars) and CGM (blue line) measured for 2 days. Data are taken from a larger dataset collected in the 7th FP EU project DIAdvisior (1). The patient wears an Abbott Freestyle Navigator Sensor that, for this specific study, had a sampling rate of 10 minutes

Moreover, CGM can enable the possibility of developing sophisticated closed-loop control techniques to optimize insulin delivery in T1D patients (see e.g. (11), (10) and (28) for three recent reviews) or simpler glucose prediction techniques to be applied in real-time to prevent the occurrence of hypo and hyperglycaemic episodes (see (60) for a review on applications of

CGM). The greater amount of information provided by CGM also offers a completely new perspective to the analysis of the physiology and of the behavior of diabetic patients, allowing to better define the differences between subjects and the different reactions within the same subject to different stimuli.
At the present time, CGM devices can provide a glycaemic measurement every 1-10 minutes for up to 7 days. From a technological point of view it is possible to divide CGM devices in minimally or non-invasive.

1.2.2.1 Minimally Invasive CGM based on the Glucose-Oxidase Principle

Minimally invasive CGM devices measure the glycaemic level in the interstitial fluid instead of the blood compartment, as done instead by SMBG (which require direct pinching of the capillaries).

At present time the most used commercial CGM devices, also displayed in Figure 1.3, are:

- DexCom Seven Plus ™(Dexcom Inc., San Diego, CA), FDA approved (14), sampling rate 5 minutes;
- Medtronic Minimed Paradigm® Real-time Revel ™System (Medtronic Inc., Northridge, CA), FDA approved (38), sampling rate 5 minutes;
- Medtronic Guardian® Real-time System, (Medtronic Inc., Northridge, CA), FDA approved (38), sampling rate 5 minutes;
- Abbott FreeStyle Navigator ™(Abbott Laboratories, Alameda, CA), FDA approved (2), sampling time 1 minute;
- Menarini Glucoday (Menarini Diagnostic, Florence, Italy, sampling time 3 minutes).

As shown in Figure 1.3, CGM devices typically consist on two components: a wearable device, composed by a sensor placed on a micro-needle inserted in the sub-cutis and a transmitter, which usually communicates wireless with the storage component and a pocket device, which receives the data from the transmitter and allows memorizing and displaying the collected data.

Other subcutaneous continuous sensors are based on microdialysis systems, which use a fine, hollow microdialysis fibre placed subcutaneously. The probe is perfused with isotonic

fluid, from an external pool, while interstitial glucose freely diffuses into the fibre, to be then pumped out of the body to a glucose-oxidase sensor. A device exploiting this principle is the Menarini GlucoDay, produced by Menarini Diagnostics, Florence, Italy.

The crucial component is obviously the sensor, which, for all the above devices, is based on the glucose-oxydase enzyme reactions or on microdialysis structures. In presence of glycose the gluco-oxydase enzyme produces hydrogen peroxide, which is then converted in oxygen and protons with generation of current. The reaction is the following

$$\begin{aligned} \text{glucose} + O_2 &\xrightarrow{\text{glucose oxidase}} H_2O_2 + \text{gluconic acid} \\ H_2O_2 &\xrightarrow{\sim 700mV} O_2 + 2H^+ + 2e^- \end{aligned} \tag{1.1}$$

In particular the enzyme is placed at the top of a needle and protected by a membrane which is permeable to glucose. An ammeter detects the current generated by the oxidation of hydrogen peroxide at the working electrode. The measured electrical current is then translated into the glucose scale by an operation called calibration. In particular, given some SMBG reference, the current signal is corrected to match the glucose value.

Figure 1.3: Some commercial CGM devices: the Minimed Paradigm® Real-Time Revel ™System (upper left), the Abbott FreeStyle Navigator ™(upper right), the DexCom Seven Plus ™ (bottom left) and the Menarini Glucoday (bottom right).

In this thesis, data collected with the DexCom Seven Plus, the Abbott Freestyle Navigator

and the Menarini Glucoday will be used.

1.2.2.2 Other Techniques for CGM

As alternative to subcutaneous sensors based on the glucose-oxydase enzyme, other systems and prototypes have been proposed for CGM monitoring. We cite some of them for sake of completeness.

- *Iontophoresis and Sonophoresis.*
  These techniques require the to stimulate of the skin from outside with different tools, in order to extract glucose from the skin for its direct measure. The first method is based on the exctraction of glucose associated with the application of an electrical potential, causing the migration of ions from beneath the skin. In particular, sodium and chloride are pulled towards the cathode and anode respectively. The ion flow also causes neutral molecules like glucose to migrate across the skin along with the water hydrating the ions. Glucose is then detected with the enzymatic reaction reported in Eq. 1.1. The *GlucoWatch G2 Biographer* (Cygnus Inc., Redwood City, CA; not on the market anymore because it caused skin irritation in users), is an example of device which used the reverse iontophoresis. Sonophoresis uses low-frequency ultrasounds to create an array of microscopic holes on human skin, which increase its permeability, allowing glucose to trespass the skin to be directly measured. The *SonoPrep* (Echo Therapeutics Inc., Philadelphia, PA (15)) is a device which exploits this technology.

- *Micropores and Microneedles Techniques.*
  For example, micropores techniques perforate the stratum corneum without perforating the full thickness of the skin with the aid of pulsed laser or local heat. Interstitial fluid is then collected applying vacuum and a direct measure of glucose is obtained.

- *Noninvasive CGM* The needle inserted in the sub-cutis represents a nuisance for the patient, as well as the application of a current for the direct extraction of glucose from the blood stream for direct measurement. Therefore research is active in developing tools for the non-invasive detection of signals correlated with the glycaemia. Among the physical principles exploited for this scope, we can list optical techniques, e.g. based on absorption phenomena (Near InfraRed Spectroscopy, Mid InfraRed Spectroscopy), on scattering (Raman Spectroscopy, Occlusion Spectroscopy), on Optical Coherence

Tomography, on Fluorescence Technologies; Photoacoustic Spectroscopy; Impedance Spectroscopy; Electromagnetic Sensing; Thermal Emission Spectroscopy.

A general idea is to combine several of these techniques to obtain signals which are correlated to the concentration of glucose in blood (*multisensor* concept). Although non-invasive CGM are of course very attractive from a user's point of view, they do not offer the same accuracy of subcutaneous sensors yet. In particular they are difficult to calibrate, and they are not yet usable to extract reliable information on glucose dynamics (62).

In this thesis, only subcutaneous minimally invasive CGM will be considered, hence the acronym CGM will be always referred to these kind of devices.

1.3 Use of Glucose Concentration time-series

Diabetic patients who monitor themselves via SMBGs and CGM can gather a lot of information regarding their pathology. In particular, patients can exploit such information in several ways, e.g. tune the insulin boluses or to check if corrections are needed.

From the research point of view, SMBG series some insight on glucose dynamics. In fact extensive datasets of glucose monitoring via SMBGs are available and can be used, for example to study the differences among patients or to study the long term variations of glucose signals, see e.g. (61), where the relationships between glucose levels monitored via SMBGs and long term complications in diabetes are studied.

The advent of CGM devices offers an even richer insight in glucose dynamics, allowing expanding the use of monitoring signals for example for richer analysis of glucose variability in a specific subject. Also, the continuous nature of CGM makes them crucial for the optimization of insulin dosing (control) possibly with automatic delivery (artificial pancreas) and for the development of alert generation tools. In the next Chapter, we will reviews the most used algorithms and indexes proposed to assess glucose variability in diabetic patients. We will also discuss some margin of improvements which motivated the development of the present thesis.

2

Glucose Variability and Quality of Glucose Control: State of Art and Aim of the Thesis

2.1 Glucose Variability and its possible Physiological Role

Several aspects and characteristics of diabetes are enclosed in the concept of glucose variability, from the amplitude of the glucose range reached by a patient, to the repeatability in different meals or days of a specific glycemic pattern. Glycemic variability is one of the possible factors in the etiology of complications from diabetes. In a recent review article (30), an analysis on the possible relationships between glycemic variability and complications in diabetes is reported. Glucose variability has been considered as a risk factor for several issues in diabetes. In particular, as described below, many studies investigate the correlation between glycemic variability and major processes in the degeneration of tissues in diabetes.

- **Formation of Reactive Oxygen Species**. In an article by Brownlee (5) the author explains that all complications of diabetes could ultimately be explained by overproduction of the reactive free radical molecule, superoxide, generated in response to hyperglycaemia acting on cellular mitochondria. In (41) and (7) two example of studies that support the correlation between glucose variability and the oxidative stress are described. In (63) such hypothesis is not supported. Notice that the groups use different tools to measure the oxidative stress.

- **Increase in the Glycated Hemoglobin (HbA1c)** . Glucose variability may also be responsible for increase in HbA1c increase from normal to diabetic levels. While some groups report a major influence due to postprandial hyperglycemia, and only a mild

effect of basal hyperglycemia, on HbA1c concentration (at levels lower than 8-9%) (40), other groups found a correlation between HbA1c levels with glucose mean, no matter how this mean is reached (13) .

- **Microvascular Disease**. The major study investigating on the relationships between glucose varability and diabetes complication is the so called Diabetes Control and Complications Trial Research Group (*DCCT*) study (61). This study involved 1441 diabetics patients who were studied for approximatively 10 years. Although controversial, the study showed that HbA1c levels, considered to be only dependent on glucose mean, is only responsible for part of the complications. In particular the study hypothesized that big excursions in glycemic levels are more responsible for the development of complications than the average glucose level. It is still controversial which component plays a major role, since usually patients that experience big excursions are also those which have a higher glucose mean.

- **Macrovascular Risk**. The major role in complications involving the big vessels seems to be played by post-prandial hyperglycaemia. In particular, post prandial hyperglycaemia has been shown to be predictive of future cardiovascular events, even in non diabetics (4). Post prandial hyperglycaemia seems to be correlated with carotid intimal thickness (17). Other studies report a major role of the mean glucose rather than of glucose excursion (44).

The effective role of glucose variability is still controversial. There is no consensus on the influence of this feature of glucose dynamics on complication of diabetes. Also, there is no consensus on the best way of measuring and assessing glucose variability. In the next section, we will give a brief overview on the possible metrics used for such assessment.

2.2 Literature Methods to Measure Glucose Variability

A review on metrics commonly used in clinical practice for the evaluation of variability and control can be found in (52), where measures of glucose variability were divided in 4 families: *i)* methods based on standard deviations and related methods, *ii)* methods to detect excursions, *iii)* methods based on day-to-day variability and *iv)* methods based on variability during relatively short segments of the glucose time series.

2.2.1 Basic Statistical Indexes (mean and SD measures)

Statistical indexes are often used to describe the feature of glycemic signals. Clinicians use these tools because they are simple and give a rough idea of the efficiency of the therapy and some microscopic indication to tune the insulin dosing. Used metrics, with rules to compute them, are:

- Mean_T (Mean Total): Evaluates the average of all available glucose values, where T stands for Total.
- SD_T (Standard Deviation Total) : Evaluate SD of all days and all times of day, where T stands for Total..
- CV (Coefficient of Variation) : Evaluate the SD and mean on the total signal (all data), then evaluate the ratio between the two as $100 * \frac{SD_T}{mean_T}$
- SD_w (Standard Deviation within days) : Evaluate SD for all measurements in each 24-h day and then average the SD values
- $\text{SD}_{\text{hh:mm}}$ (Standard Deviation between time points) : Evaluate the average glucose for any time of day for all days, then calculate SD of this average profile versus time.
- $\text{SD}_{\text{ws h}}$(Standard Deviation within series) : Evaluate SD for any desired segment of the glucose series (e.g. intervals of 1 hour) at any possivle time, then averaged.
- SD_{dm}(Standard Deviation daily means) : Evaluate the mean glucose for each day, then calculate SD of these means
- $\text{SD}_{\text{b hh:mm}}$ (Standard Deviation between days within time points) Evaluate SD of glucose values for any specified time of day, then average these SDs
- $\text{SD}_{\text{b hh:mm // dm}}$ (Standard Deviation between days within time points, corrected for changes in daily means): same as above, but using the deviation of each observation from the mean for the same day
- SD_I (Standard Deviation of Interaction) Two-way ANOVA with replication.
- IQR (Inter-Quartile Range Measures) % of values falling between a specified range (e.g. the 10 $^{\text{th}}$, 25^{th}, 50^{th}, 75^{th}, 90^{th} percentiles)

2.2.2 Variability Measures from Glycemic Excursions

The most used index to evaluate the glycemic excursion is the MAGE (55), defined as

$$MAGE = \frac{\sum_{i=1}^{n_e} \Delta g(i)}{n_e} \tag{2.1}$$

where

- n_e: number of excursion of amplitude greater that 1 SD
- $\Delta g(i)$ i$^{\text{th}}$ glucose excursions grater than the SD_T of the whole monitoring.

The MAGE index is highly correlated with SD, so it is sometimes used as a substitute for SD (52). Moreover, the choice of the index parameters is made by the investigator, making it hard to compare results obtained on different datasets from different researchers. Also consider excursions happening on a change of day are not considered, and, most important, there is not a clear definition for the definition of what should be considered an excursion.

2.2.3 Day-to-Day Variability

A popular tool used to quantify the variability on two consecutive days is the Mean of Daily Differences (MODD), defined by Service and Nelson in (54), defined as

$$MODD = \frac{\sum_{i=1}^{s} g_{i+s} g_i}{s} \tag{2.2}$$

where

- s: is the number of samples collected in one day
- g_i: is the i$^{\text{th}}$ glucose sample of the first day
- g_{i+s}: is the sample corresponding to the i$^{\text{th}}$ sample on the next day

The peculiarity of MODD is that it was originally defined from standardized conditions, i.e. the patient was monitored for m days (typically 2) in the same conditions with invasive sampling of blood drawn at the same time of day to allow a comparison between the m days. The advent of CGM devices renders the use of MODD easier, since it is possible to compare several days monitored with frequent measurements. It can be shown that the information provided by MODD is similar to that carried by SD (52).

2.2.4 Short-Term Variability

An index called Continuous Overlapping Net Glycemic Action (CONGA), proposed by Nathan et al. in (42), evaluates a within day variability. In practice, SD is evaluated for a time-serie composed by the differences between the glycemic value and the glycemic value collected m hours later. Parameter m can be chosen by the investigator.

2.3 Concept of glucose control quality and literature indexes

Control of glucose means keeping the glycemic levels within safety regions. Safety range can be define in a very restrictive way, i.e. $80 < BG < 140$ mg/dl, although most researchers and clinicians allow a less tight definition of the hypo and hyperglycaemic threshold, by defining the euglicemic range as $70 < BG < 180$ mg/dl. In this thesis, when referring to hypo and hyperglycemic threshold we will consider the second definition. The quality of glucose control is high if a patient is able to correctly tune the carbohydrate ingestion and insulin dosing in such a way that the glycemic range stays within the safety zone with few counteractions and corrections.

The most used parameter for the evaluation of quality of glucose control is the relative time spent by the subject in different regions of the whole glycemic scale (3). For the clinician it is important to understand the percentage of time spent on target relative to the whole monitoring session, but also to distinguish cases where the percentage out of target is spent above or below the target zone. Of course a subject who spends most time in hyperglycemia needs to refine the therapy with a more intensive insulin dosing, while a subject who tends to stay in hypoglycemia for prolonged time probably needs to reduce the insulin dosing. The used indexes are hence

- Time in target: % time of the whole monitoring spent between 70 and 180 mg/dl
- Time in hypoglycemia: % time of the whole monitoring spent below 70 mg/dl
- Time in hyperglycemia: % time of the whole monitoring spent above 180 mg/dl

The main drawback of these indexes is that there is no differentiation between severe and mild episodes. Moreover if we consider to spend 20 minutes at 40 mg/dl (hypoglycemia) or at 210 mg/dl, the criticality of the two episodes is very different from a clinical perspective, since the first is a severe hypoglycemic episode, while the second is a mild hyperglycemic

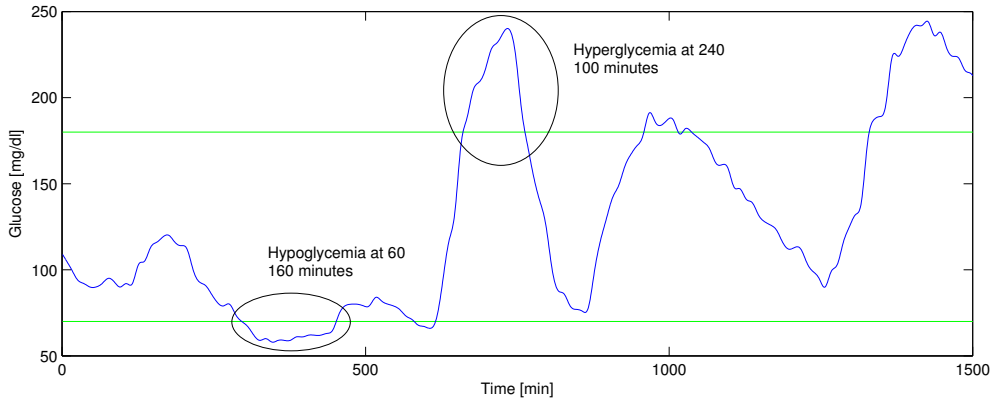

Figure 2.1: Glucose profile presenting an hypo and an hyperglycemic episode. The length of the episodes is prolonged in both cases, but the patient reaches a level of 60 mg/dl in hypo (10 mg/dl under the hypoglycemic threshold) and of 240 mg/dl in hyper (60 mg/dl above the hyperglycemic threshold). The first episode is more critical from a clinical perspective than the second despite the smaller distance from the threshold.

event. The percentage in target does not differentiate among these conditions. For example consider Figure 2.1. Here a glucose signal captured with the Abbott Freestyle Navigator is reported along with the hypoglycemic and hyperglycemic thresholds (in green, at 70 and 180 mg/dl respectively). The patient experiences a prolonged hypoglycemia at 60 mg/dl (10 mg/dl below the hypoglycemic threshold) from minute 300 to minute 460, and a prolonged hyperglycemia reaching a glucose level of 240 mg/dl (60 mg/dl above the hyperglycemic threshold) between minute 680 and minute 780. Although the hyperglycemic event seems more threatening from a numeric point of view, the prolonged hypoglycemia is clinically more severe. In statistical analysis it could be useful to explicitly highlight such clinical difference. This problem will be extensively described in Section 2.4.

Other indexes, based on combinations of mean and standard deviation of the glucose signal have been used to quantify the quality of glucose control. In particular, considering a glucose time-series g_n one can evaluate the J index, proposed by Wójcicki in (64):

$$J = 0.001 \times (mean(g) + SD(g))^2 \tag{2.3}$$

This index aims at combining in a single number the information of mean and standard deviation of the signal.

2.4 Variability and Control Literature Indexes based on Transformations of the Glycemic Scale

All the statistical and empirical indexes derived in the above section do not consider that when dealing with glucose time-series, in particular when the aim is the tuning of the therapy, it is of crucial importance to consider two main characteristics of the glycemic scale:

1. The glucose scale is not symmetric. In fact the hypo range $(20 - 70 mg/dl)$ is much narrower than the hyper range $(180 - 600 mg/dl)$. As a result, the numerical center of the scale (around 300 mg/dl) is very far from the "clinical center" (the desired target is around 110 mg/dl).

2. The distribution of glucose values is skewed in this scale. To better understand this concept, consider Fig. 2.2, which shows the glucose values (left) and first time derivative (right) distribution for 56 T1D and T2D patients monitored with the Abbott Freestyle Navigator for 1 week under the 7^{th} FP EU project DIAdvisor, during the first phase protocol "Data Acquisition Trial" (1). From the left panel, it is clear that patients spend much more time above the clinical center (around 110 mg/dl). When performing simple statistical analysis on this kind of distribution, it is clear how few, possibly severe, hypoglycemic episodes can be easily outmatched by the greater amount of data referring to (usually mild) hyperglycaemic episodes. Moreover, the threat for the patient increases much faster in the hypo region than in the hyper range.

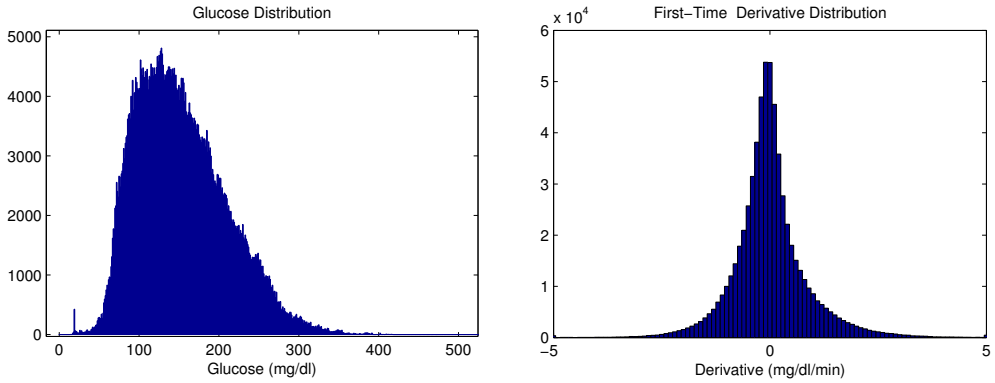

Figure 2.2: Distribution of glucose values (left panel) and First time derivatives (right panel) evaluated on a dataset of 56 Freestyle Navigator signals. Average length of the signals is 1 week.

From a clinical point of view, e.g. to evaluate the quality of control, it is important to accurately emphasize dangerous episodes which could be overlooked by straightforward application of tools like those proposed on the previous section.

Several functions have been proposed to transform the glucose scale in order to equally weight the hypo and hyperglycemic range. In sections 2.4.1 and 2.4.2 we list some of these functions, which include the so-called risk function, proposed by Kovatchev et al. in (31)which will be used in this thesis as a basis to further develop the concept of clinical risk described in Chapter 3.

2.4.1 The Kovatchev's Risk Function to Symmetrize the Glycemic Scale

A popular transformation of the glucose scale was proposed for the first time by Kovatchev and colleagues in (31). The proposed function, called *Risk Function*, was developed with the aim of symmetrizing the glucose range in order to provide a new scale such that

1. The range is centered around zero, which correspond to the clinical centre, rather than the numerical centre of the scale

2. The scale is symmetric around zero (expansion of the hypo range and compression of the hyper range)

Such a scale assigns equal weight to hypo and hyper episodes in terms of risks, i.e. a single episode of severe hypo will possibly weight more than several hours spent in moderate hyperglycemia.

Before defining the so-called risk function, the following symmetrization function

$$f(g, \alpha, \beta) = \gamma \cdot [(\ln(g)^{\alpha}) - \beta] \tag{2.4}$$

defined for g in the range [20-600] mg/dl, is introduced. Parameter γ allows the restriction of the minimal and maximal risk ($-\sqrt{10}$ and $\sqrt{10}$) at 20 and 600 mg/dl respectively. In order to obtain a symmetric scale, following constraints are imposed for the definition of parameters α and β:

$$\begin{cases} (\ln(600))^{\alpha} - \beta = -[(\ln(20))^{\alpha} - \beta] \\ (\ln(180))^{\alpha} - \beta = -[(\ln(70))^{\alpha} - \beta] \\ \gamma[(\ln(600))^{\alpha} - \beta] = -\gamma[(\ln(20))^{\alpha} - \beta] = \sqrt{10} \end{cases} \tag{2.5}$$

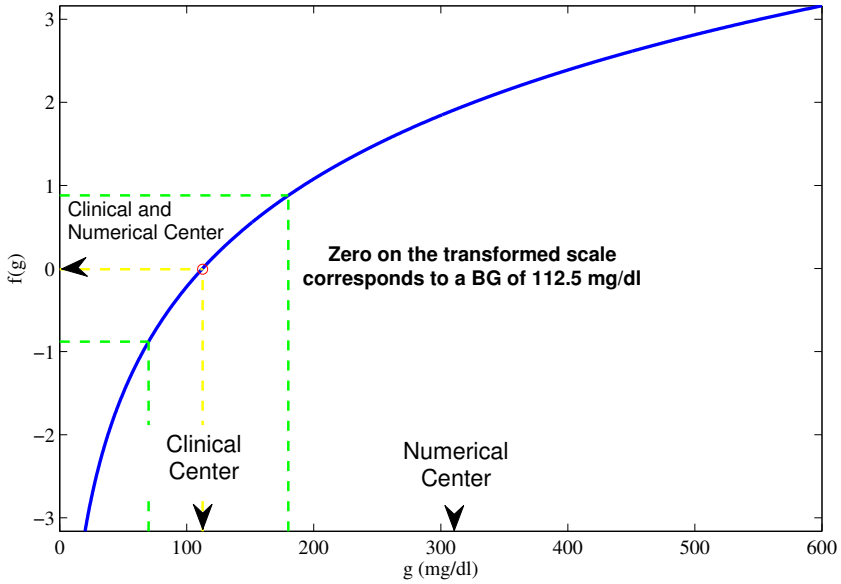

Figure 2.3: Symmetrization function by Kovatchev et al. (31) The glucose values on the x axis are matched to risk values on the y axis. Dashed lines represent the match of clinically critical values, i.e. the hypo and hyper threshold crossings.

Solving the system of Eq. 2.5, a single equation in the parameter α must be solved. Imposing $\alpha \geq 0$, the three parameters are set to $\alpha = 1.084$, $\beta = 5.381$ and $\gamma = 1.509$, assuming that glucose is expressed in mg/dl. Fig. 2.3 shows the mapping of glucose values into the simmetrized scale. Notice that the clinical centre (112.5 mg/dl) is mapped to zero-risk. The increase of risk is more rapid in hypo with respect to hyper.

From the symmetrization function, the proper risk function, plotted in Fig. 2.4 is defined as:

$$r(g) = 10 \cdot f(g)^2 \tag{2.6}$$

Notice that the function has a minimum at the clinical centre (112.5 mg/dl), which correspond to the desired target (no risk) and is maxed ($r = 100$) at 20 and 600 mg/dl. Also notice that the risk increases faster in hypo than in hyper.

From function 2.6 two functions can be defined:

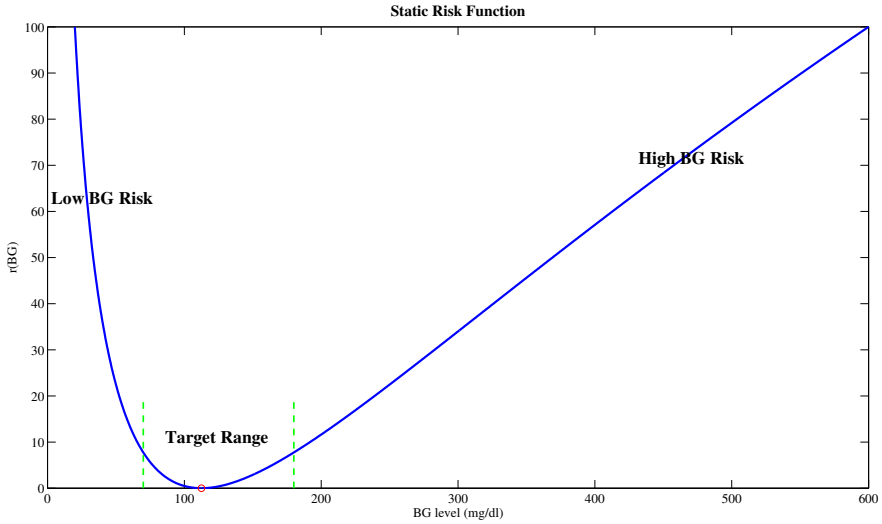

Figure 2.4: The Risk function of glucose by Kovatchev et al (31).

$$\begin{aligned} r_l(g) &= \begin{cases} r(g) & \text{if} \quad f(g) < 0 \\ 0 & \text{otherwise} \end{cases} \\ r_h(g) &= \begin{cases} r(g) & \text{if} \quad f(g) > 0 \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{2.7}$$

which are used to compute the so called Low Blood Glucose Index (LBGI) and High Blood Glucose Index (HBGI) :

$$LBGI = \frac{1}{n}\sum_{i=1}^{n} rl(g_i) \tag{2.8}$$

$$HBGI = \frac{1}{n}\sum_{i=1}^{n} rh(g_i) \tag{2.9}$$

LBGI is a quantity that increases when new samples are collected in the hypo region. The increase is faster with more severe hypos. HBGI has the same role with the hyper region. The practical use of these index was suggested by Kovatchev in (32) and (33); the LBGI and HBGI have been proven, respectively, to be predictive of severe hypoglycemia and of HbA1c concentration respectively.

2.4.2 Other Transformations of the Glucose Scale proposed in the Literature

2.4.2.1 The M_R function

A different transformation of the glucose scale, the so-called M_R function, was proposed by Wójcicki in (65)

$$M_R = 1000 \times |\log(g/R)|^3 \tag{2.10}$$

where R is a parameter that can be tuned by the investigator. This function performs a logarithmic transformation of the glucose scale, but is not able to correctly symmetrize the glycemic scale.

2.4.2.2 Index of Glycemic Control (ICG)

This transformation, defined by Rodbard in (52) represents a penalty function asymmetric with respect to the normoglycemic value of 112.5 mg/dl. It is very flexible thanks to the possibility of changing several parameters. The ICG is defined as

$$ICG = HypoIndex + HyperIndex \tag{2.11}$$

where

$$HypoIndex = \frac{\sum_{i=1}^{N}(LLTR - g(i))^b}{N \times d} \tag{2.12}$$

$$HyperIndex = \frac{\sum_{i=1}^{N}(g(i) - ULTR)^a}{N \times c} \tag{2.13}$$

with $ULTR$ (Upper Limit of Target Range) usually set to 140 mg/dl, $LLTR$ (Lower Limit of Target Range) usually set to 80 mg/dl, parameters c and d set to 30, $a \sim 1.1$ and $b \sim 1.5$. The parameters of this function can be tuned to almost match other risk function. N is the number of observed glucose values.

2.4.2.3 Glycemic Risk Assessment Diabetes Equation (GRADE)

Another transformation was proposed to symmetrize the glycemic scale by Hill et al. (27) and is defined by:

$$GRADE = \frac{\sum_{i=1}^{N} min[50, 42.5 \times \{\log_{10}(\log_{10}(\frac{g_i}{18}) + 0.16)^2\}]}{N} \tag{2.14}$$

This equation was obtained by fitting some risk scores assigned by expert clinicians to specific glucose levels.

2.4.2.4 Comparison of Transformation Functions

Figure 2.5, taken from (52) shows the transformation functions listed in Sections 2.4.1 and 2.4.2 on a semi-logarithmic scale. All the functions are based on similar concepts, and perform a transformation based on different weights of the hypo and hyper range. In this thesis we chose to consider Kovatchev's formulation of risk, since it is a popular tool with a specific mathematical definition that suits the development we need.

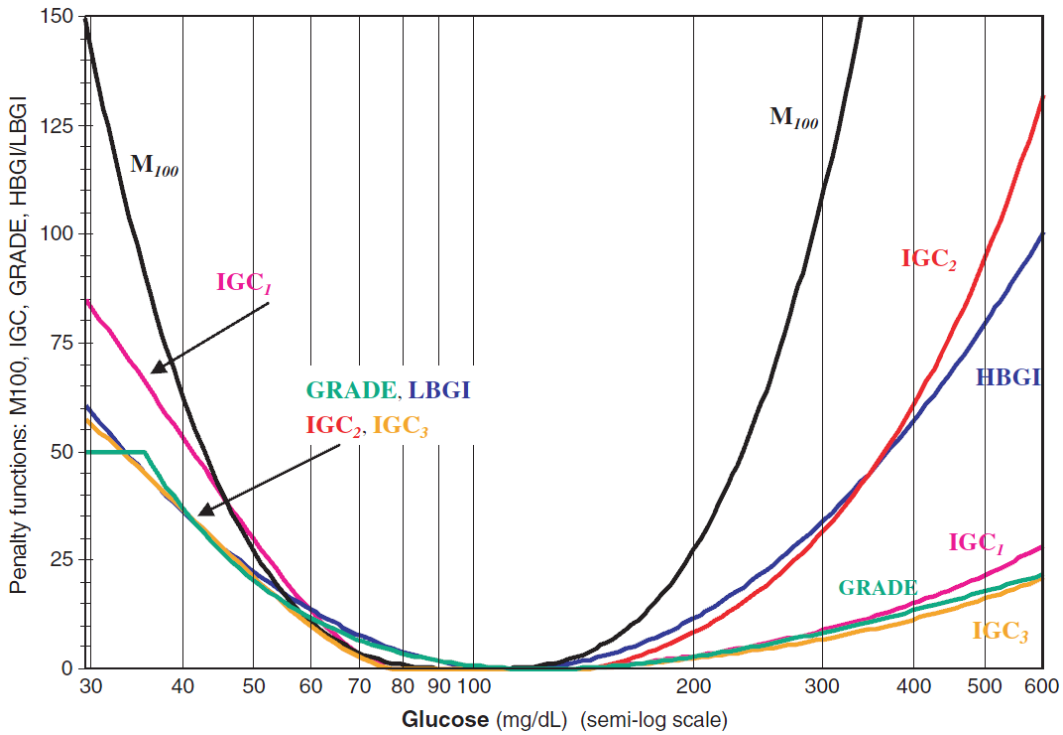

Figure 2.5: Penalty functions for the transformation of glucose scale M_{100}: black; IGC_1 (a=1.1, b=2.0): pink; HBGI e LBGI: blue; GRADE: green IGC_2 (a=1.35, b=1.9): red; IGC_3 (a=1.05, b=1.9): orange.(LLTR=80, ULTR=140, c=d=30) (taken from (52))

2.5 Limitations of Literature Measures of Glucose Variability and Control

The understanding of the role of glycemic variability on complications of diabetes is still an open field. Also, the definitions of glucose control sometimes give a partial vision of the global condition of the patient. There are some technical and practical issues in the analysis of the studies which investigate the correlations between glucose variability and complications:

1. First of all the definition of Glycemic Variability itself is not always clear. Dozens of indexes are used for the definition, and each study uses different metrics, possibly looking at only part of the whole truth of the signals.

2. Most of the used indexes were developed for SMBG references, and are intrinsically static. This means that applying such indexes to the CGM signals will provide partial summary of the whole information provided by CGM. It is important to study the effect of excursion more deeply. This can be done exploiting the information of glucose rate of change provided by CGM.

3. The clinical perspective is sometimes blurred in the simple analysis of basic statistical indexes. In particular, it is important to consider the specific characteristics of the glycaemic scale in order to have a precise idea of the so called *risk*, as will be deeply investigated in the next chapters.

2.6 Aim of the Thesis

As described in Section 2.5 the powerful information provided by CGM devices is not extensively exploited. In particular in the assessment and analysis of glucose variability, and of the clinical risk for the patient, the first time derivative of the signal is never explicitly included. The aim of this thesis is twofold:

1. To further develop the concept of clinical risk to explicitly include the time derivative as a threat factor for the patient. This task will be tackled in Chapters 3 and 4, where a formal derivation of a new "Dynamic Risk" (DR) function will be presented along with theoretical solutions to online implementation issues.

2. To provide two applications of the DR function, in particular

- The use of DR as alert generator for the prevention of hypo/hyperglycaemic events (Chapters 5 and 6)
- The development of variability and control metrics which directly include first time-derivative based indexes (Chapter 7). The importance of such indexes in the description of Glucose Variability and Control will be formalized by the use of a multivariate technique, the Sparse Principal Component Analysis (Chapter 8)

Simulated and Real Data will be used in this thesis. In particular, simulated CGM signal are obtained via smoothing of frequently sampled plasma glucose time-series available on the web (21). Real datasets consist in signals collected with different sensors (Abbott Freestyle Navigator, Menarini Glucoday, Dexcom Seven Plus) collected in several projects which involved our research group in the recent past (7th FP EU project DIAdvisor, JDRF Artificial Pancreas project).

Part II

Conceptual Design and Algorithmic Implementation of the Dynamic Risk

3

The Dynamic Risk Function

3.1 Assessment of Clinical Risk in Diabetic Patients: Role of the Glucose Trend

The availability of CGM sensors opens the possibility to embed glucose dynamics information in the evaluation of risk measures. Consider the example reported in Fig. 3.1, which displays a simulated continuous glucose profile for 1000 min. The picture highlights four particular points, labeled as A_1, A_2, B_1, B_1 together with a portion of the tangent line to the glucose profile in these points. In addition, the circle labeled as C highlights an hypo-threshold crossing event. Notably, A_1 and A_2 correspond to a glucose concentration just below the hypo threshold (65 mg/dl) but with different derivatives (-2 mg/dl/min for A_1 and +2 mg/dl/min for A_2), while B_1 and B_2 are both placed exactly on the hyper threshold (180 mg/dl with derivatives of +2 mg/dl/min and +4 mg/dl/min respectively). The Kovatchev's function (31), (33), (9), points A_1 and A_2 would be associated to the same risk value, and so would B_1 and B_2.

However, by considering the continuous glucose profile of Fig. 3.1, one easily realizes that the clinical risk associated to points A_1 and A_2 should be different. In fact, the A_1 situation is more dangerous for the patient than A_2, since in the first case the glycemia is heading deeper in the hypo region, while in the second case a recovering from the hypo region towards the normoglycemic range is happening. Similarly, B_2 situation is more dangerous than B_1, since the glycemic signal is approaching the hyperglycemic region faster. The above examples make it clear that the trend in the glycemia, which became available thanks to CGM sensors, should be considered in the risk measure.

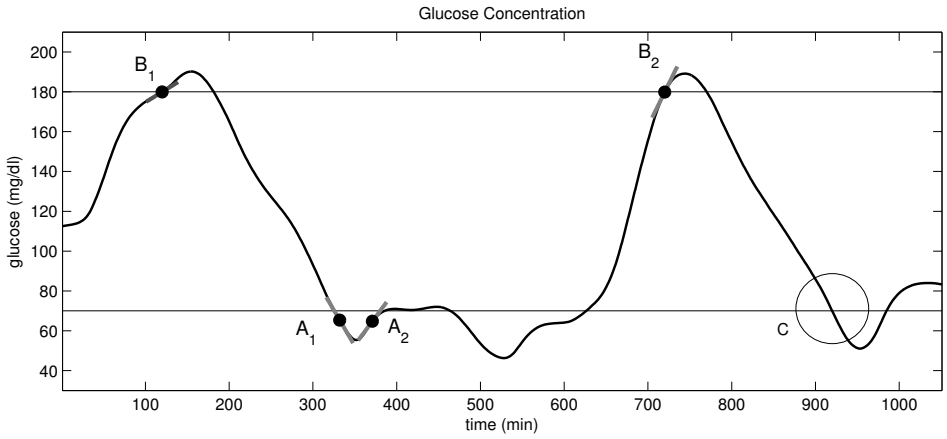

Figure 3.1: Simulated noise-free continuous glucose profile (continuous line) with hypo/hyper thresholds (horizontal lines). Four points (dots) are indicated: A_1 and A_2 (65 mg/dl, decreasing and increasing trend respectively), B_1 and B_2 (180 mg/dl, rate of change of +2mg/dl/min and +4mg/dl/min respectively) with the correspondent tangent (thin line). Circle C highlights an hypoglycemic threshold crossing event.

In this paper, starting from the original risk function proposed in (31) when only sparse SMBG data could be collected, we will define a new "dynamic risk" (DR) measure, dependent on both level and trend of actual glucose 3.2. In this chapter, a conceptual design will be presented using ideal noise free signals. When dealing with real-life signals, the computation of glucose trends in the definition of DR, stable and computationally efficient algorithms will be developed to preperly assess trends from CGM data in presence of noise, in both off-line and online situations (Sections 4.2).

3.2 Conceptual Development (Rationale and Requisites) of the Dynamic Risk (DR)

For the definition of the DR function we will exploit the two functions r_l and r_h of Eq. 2.7:

$$\begin{aligned} r_l(g) &= \begin{cases} r(g) & \text{if} \quad f(g) < 0 \\ 0 & \text{otherwise} \end{cases} \\ r_h(g) &= \begin{cases} r(g) & \text{if} \quad f(g) > 0 \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{3.1}$$

From these definitions, we obtain a new function, the so-called static risk (SR) :

$$SR(g) = r_h(g) - r_l(g) \tag{3.2}$$

As shown in Fig. 3.2 SR (black line), coincides with the standard risk function $r(g)$ of Eq. 2.6 (red dashed line), for glucose values above 112.5 mg/dl while it is a flipped version of $r(g)$ around the x-axis when the glucose value is below 112.5 mg/dl. SR(g) is hence negative for values below the clinical center, and positive otherwise.

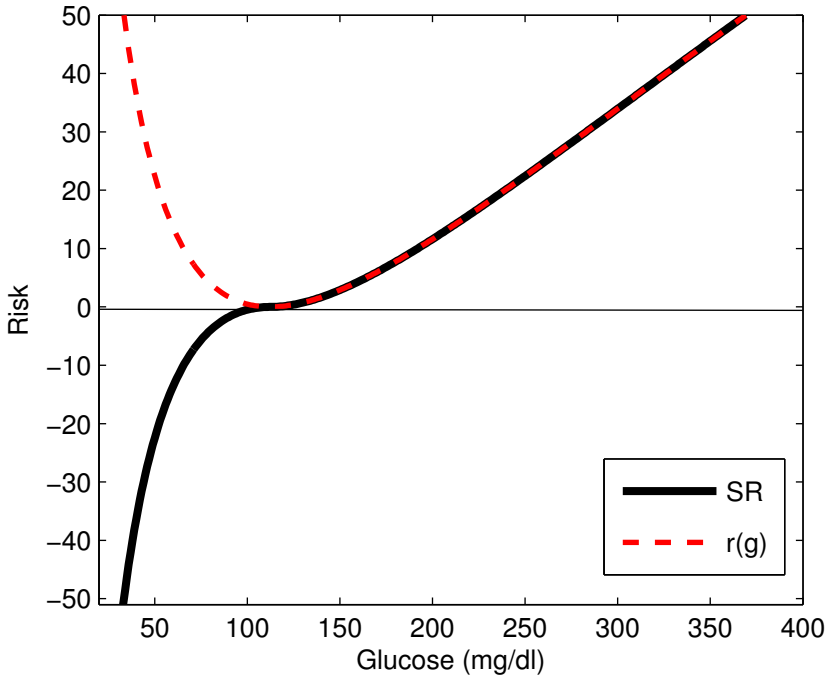

Figure 3.2: Functions $r(g)$ of Eq. 2.6 (dashed red line) and $SR(g)$ of Eq. 3.2 (black line)

The specification of the function we aim to build are easily formalized. The function DR will be a function of the actual glucose level and of the actual glucose trend, described via the first time derivative of the glucose signal itself.

$$DR\left(g, \frac{dg}{dt}\right) \tag{3.3}$$

The dependengy of g and consequently of DR from the time is not explicitly shown for simplicity. We require that in static conditions, i.e. when the derivative is zero and the glucose signal is stable, the dynamic risk equals the static risk SR, i.e.:

$$DR(g,0) = SR(g) \tag{3.4}$$

Finally, we require that the following "dynamic constraints" are satisfied:

$$\begin{cases} |DR(g,\frac{dg}{dt})| > r(g) & \text{if} \quad \frac{dg}{dt} \cdot SR(g) \geq 0 \\ |DR(g,\frac{dg}{dt})| < r(g) & \text{if} \quad \frac{dg}{dt} \cdot SR(g) < 0 \end{cases} \tag{3.5}$$

Eqs. 3.5 formally state that:

1. In case of glucose below the clinical centre, i.e. $\mathbf{SR} \leq \mathbf{0}$, and
   - negative slope, i.e. glucose entering the hypo zone, the risk is amplified (with respect to the measure given by $r(g)$)
   - positive slope, i.e. glucose exiting the hypo zone, the risk is damped ((with respect to the measure given by $r(g)$)
2. In case of glucose above 112.5 mg/dl, i.e. $\mathbf{SR} > \mathbf{0}$, and
   - positive slope, i.e. entering the hyper zone, the risk is amplified (with respect to the measure given by $r(g)$)
   - negative slope, i.e. exiting the hyper zone, the risk is damped (with respect to the measure given by $r(g)$)

In the Section 3.3, possible structures for DR is described.

3.3 Mathematical definition of DR

The general model chosen for DR, is a function such that $SR(g)$ is multiplied by an amplifier/damper function which increases or decreases the risk accordingly to the constraints of Eq. 3.5.

$$DR\left(g, \frac{dg}{dt}\right) = SR(g) \cdot F\left(\frac{dg}{dt}\right) \tag{3.6}$$

Several possibilities are open for the choice of function $F(\frac{dg}{dt})$ in Eq. 3.6.

3.3.1 Amplifier/Damper: the Exponential Structure

A first possible structure for the DR function is the following:

$$DR\left(g, \frac{dg}{dt}\right) = \begin{cases} SR(g) \cdot e^{x} & \text{if} \quad SR(g) \leq 0 \\ SR(g) \cdot e^{-x} & \text{if} \quad SR(g) < 0 \end{cases} \tag{3.7}$$

Where x is a function of t which can be computed using either $\frac{dg}{dt}$ (Section 3.3.1.1) or $\frac{dr}{dt}$ (Section 3.3.1.2) It is quite easy to show that this structure is compliant with the constraints formalized in Eq. 3.5 . Moreover it holds:

- When g is constant, $x = 0$ and $DR = SR$.
- When $x \to 0$, $e^{x} \to 0$. If the derivative sign is opposed to the actual sign of SR, DR will act as a damper. For instamce, consider a rapid recovery from hypo to normoglycemic values. In this case x is very large and positive and DR results to equal SR (negative) multiplied by a positive value. The DR becomes less negative but will not become positive. In other words this means that there will never be a positive DR risk of hyper if the glycemic value is still under the target of 112.5 mg/dl.

Fig. 3.3 shows a conceptual example of the behavior of DR and SR referring to the same signal already displayed in Fig. 3.1. The glucose signal is reported in the upper panel, along with the hypoglycemic and hyperglycemic thresholds. In the lower panel SR function (gray line) is displayed with DR (dashed black line), comuted with x in Eq. 3.7 defined as in Section 3.3.1.2. Points A_1, A_2, B_1 and B_2 are reported on the upper panel, and are linked to their transformations according to SR ($SR(A_1) = SR(A_2)$, $SR(B_1) = SR(B_2)$) and to DR ($DR(A_1)$, $DR(A_2)$, $DR(B_1)$ and $DR(B_2)$). Notice that:

- A_1 and A_2 are mapped to the same risk value by SR though, as explained before, the clinical perception of risk is different. This difference is well evidenced by DR, since $DR(A_1) > DR(A_2)$
- B_1 and B_2 are mapped to the same risk value by SR, while DR assigns higher risk to the faster approach to hyperglycemia, i.e. $DR(B_2) > DR(B_1)$.
- DR has the peculiar feature of being *intrinsically* predictive of threshold crossings. In particular, notice how around point C DR crosses the threshold 10 minutes before SR around point C. This characteristic will be used later in the present thesis both for

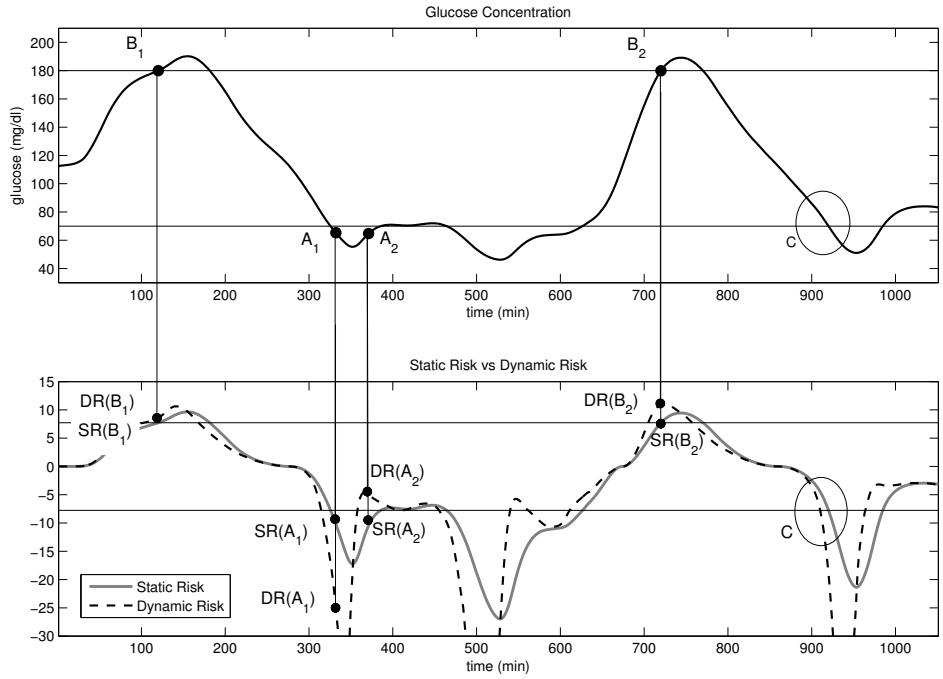

Figure 3.3: Top panel: Same simulated noise free glucose profile of Fig.3.1. Bottom panel: SR (gray) and DR (dashed black). DR and SR of the highlighted conditions on the top panel are shown as black dots in the bottom panel. The circle highlights the anticipation in threshold crossing of DR with respect to SR in event C. The horizontal lines represent the threshold crossings transformed in the risk scale

the definition of different structures Section 5.3.2, and for the generation of alerts in Capters 5 and 6.

The conceptual example reported in this section is interesting since it allows understanding the potential in DR. Some specifications need to be further defined, in particular, the choice of the derivative term x in Eq. 3.7. Two choices are possible:

1. To use the first time derivative of the glucose signal g
2. To use the first time derivative of the static risk signal $r(g)$

Details are discussed below.

3.3.1.1 Derivative term given by $\frac{dg}{dt}$

The first option for the definition of DR is to consider the following formula:

$$DR(g, \frac{dg}{dt}) = \begin{cases} SR(g) \cdot e^{\frac{dg}{dt}} & \text{if} \quad SR(g) \leq 0 \\ SR(g) \cdot e^{-\frac{dg}{dt}} & \text{if} \quad SR(g) < 0 \end{cases} \tag{3.8}$$

where $\frac{dg}{dt}$ is the time derivative of the CGM signal, which can be estimated, for example, via finite differences. This formulation is simple, but has a major drawback, since it is a symmetric function in the glucose domain, it results to be asymmetric in the risk space. This means that considering the same time derivative for two "symmetrized" glycemic values (e.g. 60 mg/dl and 210 mg/dl), the amplification would not be equal in the two regions, but it would be greater in the hyper zone. This is due to the fact that, considering equal initial value and rate, the symmetric space where the risk is evaluated is not aware of the time interval necessary for the glycemic value to reach the hypo/hyperglycemic thresholds.

3.3.1.2 Derivative term given by $\frac{dr}{dt}$

The second option to define DR, uses the first time derivative of the static risk function $r(g)$ as a penalty for the amplification/reduction of the risk. The proposed structure is the following:

$$DR(g, \frac{dg}{dt}) = \begin{cases} SR(g) \cdot e^{\frac{dr}{dt}} & \text{if} \quad SR(g) > 0 \\ SR(g) \cdot e^{-\frac{dr}{dt}} & \text{if} \quad SR(g) < 0 \end{cases} \tag{3.9}$$

where, by simple calculation on the formula 2.6, it holds:

$$\frac{dr}{dt} = \frac{dr}{dg} \cdot \frac{dg}{dt} = \left\{ 10\gamma^2 \cdot [(\ln(g))^{2\alpha-1} - \beta(\ln(g))^{\alpha-1}] \cdot 2\alpha \frac{1}{g} \right\} \cdot \frac{dg}{dt} \tag{3.10}$$

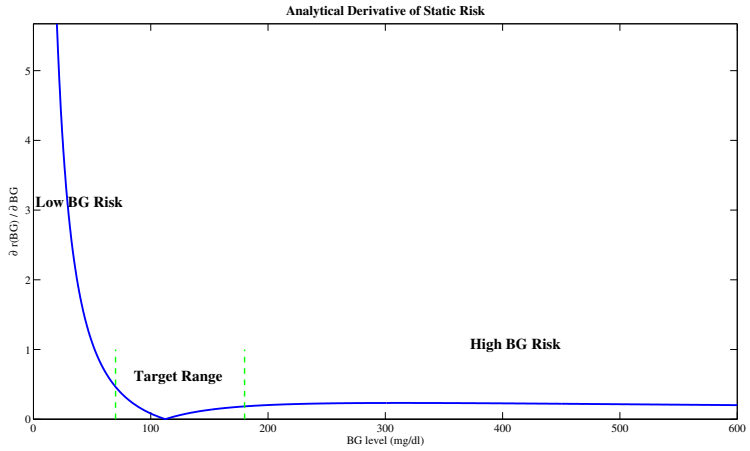

Figure 3.4: The derivative of risk with respect to glucose.

Notice how this amplification/damping term is composed by the first time derivative of the glucose signal in time (possibly the CGM signal), pre-multiplied by a term which is the analytical derivative of the risk function with respect to glucose. The derivative of risk with respect to glucose is displayed in Fig. 3.4. This term allows a greater amplification of the glucose time-derivative ($\frac{dg}{dt}$) in the hypo region, and a damping in the hyper region. In this way, the weight of the derivative is equally important in hypo and hyper, according with the whole concept of risk. To conclude with an example, consider a situation of hypo with negative time derivative (e.g. -2 mg/dl) and a condition of hyper with positive trend (e.g. +2 mg/dl) starting from the clinical center of 112 mg/dl, as shown in Fig. 3.5, top panel.

The first condition is more risky, since SR increases in magnitude much faster in the hypo region, and we would like to formalize the different clinical risk due to the derivative associated to the specific glucose value. In fact, considering the same rate of change in absolute value, the time needed to get to a severe hypo from the clinical centre is much shorter than the time needed to get to a severe hyper, as shown in Fig. 3.5. The time needed to reach the thresholds is higher for the simulation at positive rate. In the lower panel, we show different behaviors for the two options (derivative of risk, black, and derivative of glucose, blue). Notice how the first solution offers a balanced gain in hyper and hypo, while the second leads to a much earlier alert in the hyper region despite the minor criticality of the condition. The use of

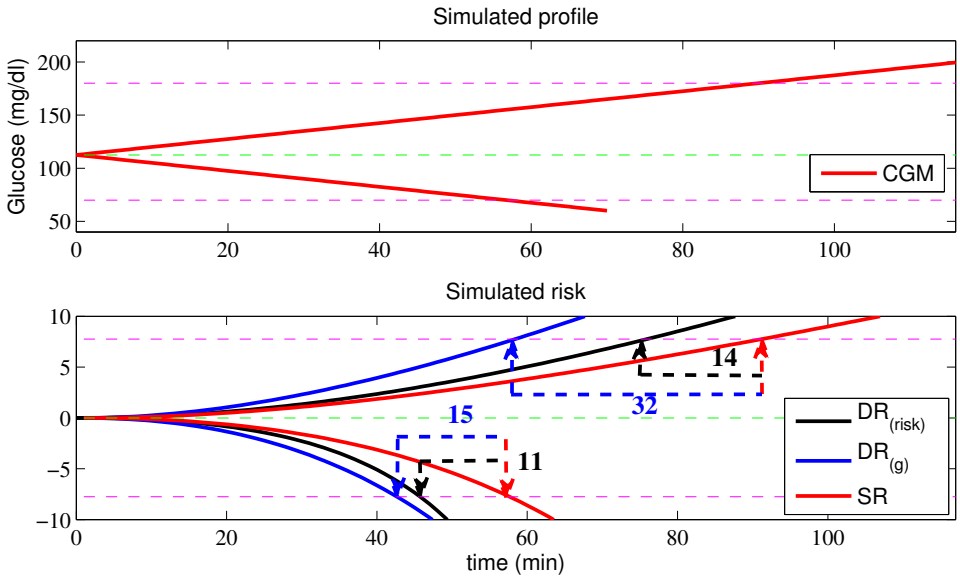

Figure 3.5: Simulation of two glucose profiles with constant rate of change of +2 mg/dl and -2mg/dl (top panel) and relative risk functions: SR (red), DR (time derivative used $\frac{dg}{dt}$, blue) and DR (time derivative used $\frac{dr}{dt}$, black)

the time derivative of SR, instead of using the simple definition of rate of change of glucose, allows us to obtain a symmetric behavior of the function in terms of clinical criticality.

The final basic structure of DR can then be summarized in

$$DR(g, \frac{dg}{dt}) = \begin{cases} SR(g) \cdot e^{\mu \frac{dr}{dt}} & \text{if} \quad SR(g) > 0 \\ SR(g) \cdot e^{-\mu \frac{dr}{dt}} & \text{if} \quad SR(g) < 0 \end{cases} \tag{3.11}$$

Notice that in the second factor of the right handside of Eq. 3.11 we have added a parameter μ at the exponential. This parameter, positive, does not change the behavior of the function in practice, but serves as a tuner for the "aggressiveness" of the risk function: the higher μ, the higher the relative weight of the role of time-derivative with respect to the role of the glucose level itself. The role of μ will be investigated more deeply in Chapter 5.

3.3.2 Amplifier/Damper Hyperbolic Tangent Structure

An alternative structure exploits the hyperbolic tangent and is defined as:

$$DR(g, \frac{dg}{dt}) = \begin{cases} SR(g) \cdot \left[\delta \cdot \tanh\left(\alpha \frac{dr}{dt} + \gamma\right) + \beta\right] & \text{if} \quad SR > 0 \\ SR(g) \cdot \left[\delta \cdot \tanh\left(-\alpha \frac{dr}{dt} + \gamma\right) + \beta\right] & \text{if} \quad SR < 0 \end{cases} \tag{3.12}$$

where α, β, γ and δ are parameters whose role will be described in detail in Chapter 5. The behavior of the function is very similar to that of the exponential structure used in 3.3.1, with 4 unknown parameters to set. A comparison between the two proposed structures is shown in Figure 3.6. The picture shows the amplification factor as a function of the first time derivative for $SR \geq 0$. In case of discordance between derivative and SR, the amplification factor is in fact a damping factor (left part of the graph). If SR and the derivative have the same sign and the glycaemia is heading deeper into a dangerous zone, the right part of the graph is used, amplifying the absolute value of the risk. While the exponential structure of Eq.3.9 is unbounded in the amplification region (red line), the tanh of Eq. 3.12 structure allows setting the maximum possible amplification of risk (blue line). Moreover, it is possible to easily set the parameters in order to have a maximum damping different from zero.

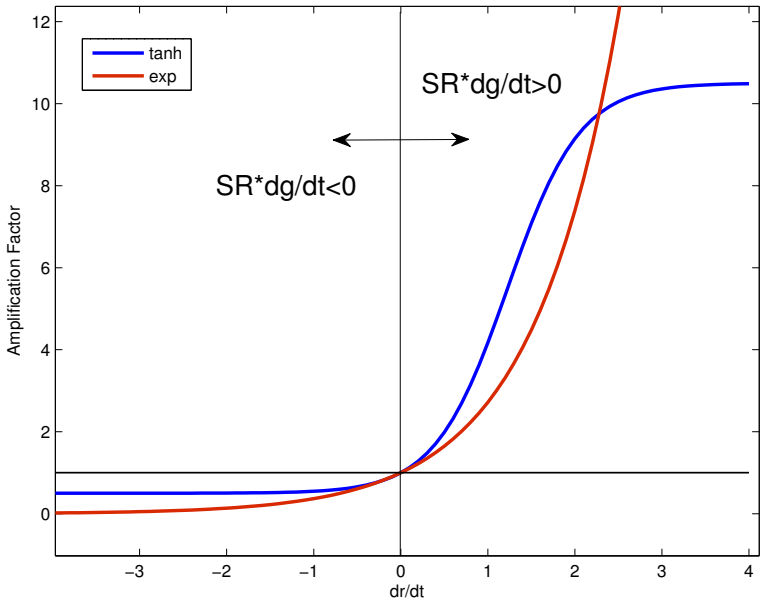

Figure 3.6: Amplification/damping factor as a function of dr/dt. the damping factor (left part of the graph) is used whenever the product of SR with dr/dt (or dg/dt) is negative, i.e. heading back to euglycemia. This factor then multiplies SR itself lowering its absolute value. Viceversa for the right part of the graph, which is the amplifier used to magnify the risk whenever the product of SR and the risk time derivative is positive (SR and time derivative have the same sign, meaning that the patient is heading deeper into threatening zones).

Notice that any function of dr/dg with this structure can fit Eq. 3.6 and can be used just like we will use Eq. 3.12 or 3.11. Therefore, the use of one structure or the other depends on

the characteristics of DR that we want to highlight.

3.4 The Concept of Dynamic Risk Space (DRS)

So far we have explored the mathematical definition of the dynamic risk. It is interesting to plot this function in a 3-dimensional graph showing how glucose levels and trend play in the definition of risk. Figure 3.7 shows DR evaluated via Eq.3.9 with $\mu = 1$ as a function of g and $\frac{dg}{dt}$. The region of hypoglycemic levels and negative time derivative are associated to negative values of risk; vice-versa, hyper regions with positive glucose time derivative are associated to positive risk values. The increase in risk is greater (in magnitude) in the hypo with respect to the hyper region. Also if we cut the graph at zero time-derivative we obtain the function SR as shown in Figure 3.2.

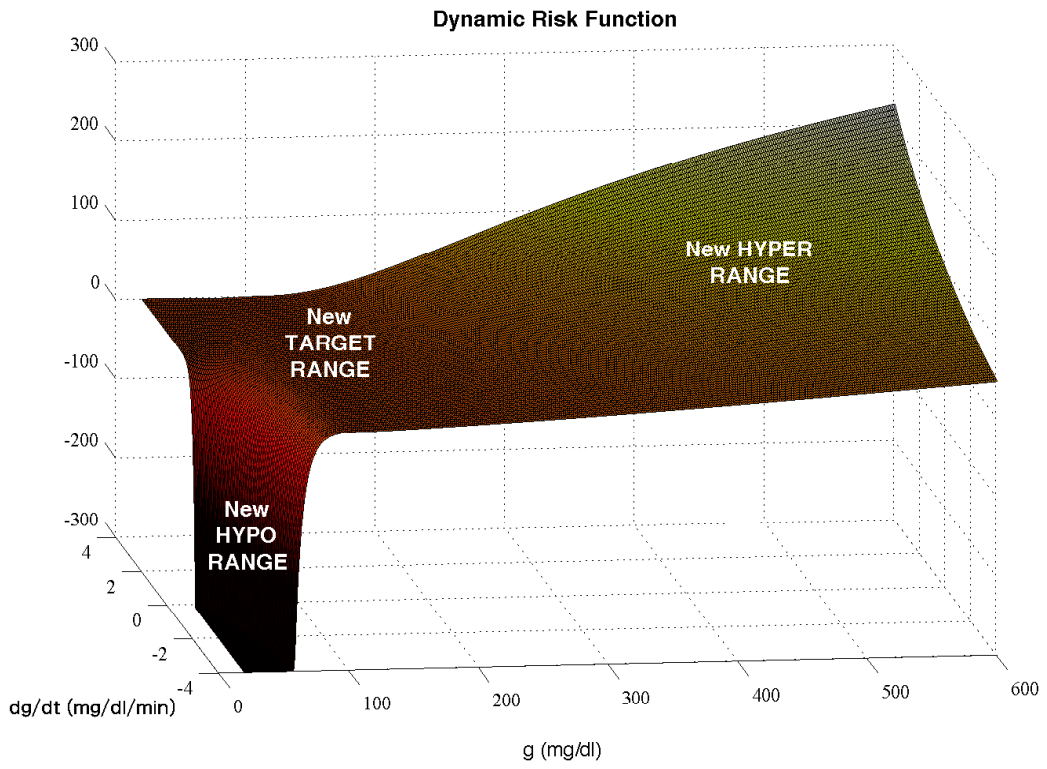

Figure 3.7: Three dimensional plot of dynamic risk as a function of g and $\frac{dg}{dt}$. Negative values correspond to hypoglycemia risk, while positive values are associated to hyperglycemic risk.

If we compress the function of Eq. 3.9 above on a bidimensional plot, we can consider

a Dynamic Risk Space (DRS) as a weighted phase plot where glucose trajectories can be displayed to extract interesting features of the patient's behavior. The DRS will be used in Chapter 7 to extract indexes for the evaluation of glucose variability. Figure 3.8 represents the absolute value of the DR as a function of glucose and its time derivative. The space is obtained using the exponential structure of Eq. 3.11, with the time derivative of risk of Eq-3.10 and $\mu = 1$.

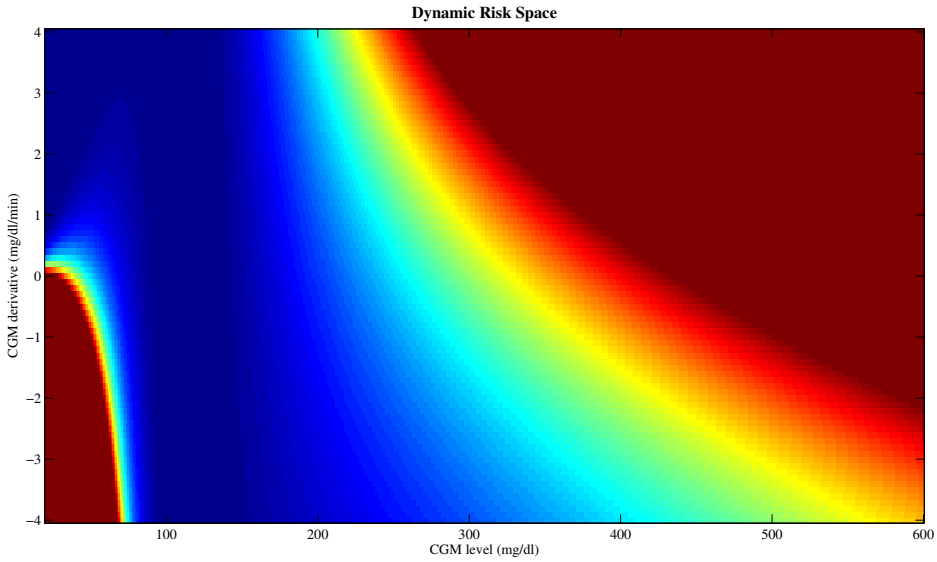

Figure 3.8: Two dimensional plot of dynamic risk as a function of glucose level and time derivative. Red corresponds to higher risk values, while blue represents the safety range.

3.5 Conclusion

In this Chapter the mathematical description of the DR and DRS was presented. This conceptual definition was presented in the ideal case of noise-free signals, to describe how the role of time derivative can be explicitly included in the clinical risk concept. In the following chapters, algorithmic issues will be presented in order to allow the online implementation of DR in real-life condition of noisy signals.

4 Algorithms for DR Implementation

4.1 Problem Formulation

Recalling again for the sake of reading the formula for the computation of the first time derivative of SR (Eq.3.10)

$$\frac{dr}{dt} = \frac{dr}{dg} \cdot \frac{dg}{dt} = \left\{ 10\gamma^2 \cdot [(\ln(g))^{2\alpha-1} - \beta(\ln(g))^{\alpha-1}] \cdot 2\alpha \frac{1}{g} \right\} \cdot \frac{dg}{dt} \tag{4.1}$$

the practical implementation of the DR function as it is defined in Eq.3.11

$$DR(g, \frac{dg}{dt}) = \begin{cases} SR(g) \cdot e^{\mu \frac{dr}{dt}} & \text{if} \quad SR(g) > 0 \\ SR(g) \cdot e^{-\mu \frac{dr}{dt}} & \text{if} \quad SR(g) < 0 \end{cases} \tag{4.2}$$

requires both CGM signal and its first time derivative.

The estimation of the first time derivative $\frac{dr}{dt}$ via finite differences is often not satisfactory and the resulting signal is noisy, since differentiation acts on the signal like a high-pass filter amplifying high frequency measurement noise. In fact, the measured CGM signal $y(t)$ can be modelled as

$$y(t) = u(t) + v(t) \tag{4.3}$$

where, at time t, $u(t)$ is the real unknown glycaemic level and $v(t)$ is measurement noise, that we can assume to be a random white noise process with zero mean and unknown variance, possibly varying with time. Several literature approaches can be used to obtain a smooth version of the signal $u(t)$, for instance Moving Average filters (25). Such filters are not adaptive, i.e. their parameters remain the same throughout the monitoring and cannot track changes in the Signal-to-Noise Ratio(SNR) of the CGM signal. Moreover such filters do not

efficiently exploit available a priori statistical information in both regularity of the signal and on measurement noise. This is why, in this thesis, we will calculate smoothed versions of glucose (risk) and (their) time derivatives by a *Bayesian* approach extensively described in (56) and (26) and, for sake of thesis completeness, summarized in Appendix B. Notably, algorithms are also suited to online applications, as will be shown in Section 4.2.4.

4.2 Computation of DR

The computation of DR requires the online estimation of a smoothed version of the CGM signal and of its derivative from noisy measures. As shown below, this can be done with two different strategies:

- Obtain a smoothed version of the CGM signal and then evaluate the first time derivative via first order finite differences.
- Consider the problem of estimating the time-derivative as a deconvolution problem. The smoothed glucose signal is obtained as a by-product of the algorithm.

4.2.1 Smoothing followed by Finite Differences

Considering y to be the measured CGM signal, and u the real glucose profile that we need to estimate and v is the superimposed noise, related by the generic model $y = Gu + v$, where $G = I_n$. The vector u can be estimated by the state-of-art technique to perform Bayesian estimation briefly described in Appendix B (by using the regularization criterion ML1 and $m = 2$).

4.2.2 Simultaneous Smoothing and Finite Differences Calculation by Deconvolution

A different approach considers two time-continuous signals $z(t)$ and $u(t)$ such that $u(t) = \dot{z}(t)$. For a generic time instant t_0 it holds:

$$z(t) = z(t_0) + \int_{t_0}^{t} u(\tau)d\tau \tag{4.4}$$

Assuming, for simplicity $t_0 = 0$ and $z_0 = 0$

$$z(t) = \int_0^t g(t-\tau)u(\tau)d\tau = g(t) \otimes u(t) \tag{4.5}$$

where

$$g(t) = 1(t) = \begin{cases} 0 & \text{if} \quad t < 0 \\ 1 & \text{if} \quad t > 0 \end{cases} \tag{4.6}$$

The problem can be discretized by integrating $g(t)$ over the time intervals where $u(t)$ is assumed to be constant. The model of data is $y = Gu + v$ where G is the following.

$$G = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \cdots & \cdots & \cdots & 1 \end{bmatrix} \tag{4.7}$$

In this way, the integral Equation 4.4 has been discretized and put in a matrix form which fits the deconvolution approach extensively explained in (12) and summarized in Appendix B. Following these state-of-art algorithms for the solution of the problem and for the definition of the optimal regularization parameters, one can obtain a smooth estimate of the first time-derivative ($\hat{u}$) and of the data ($G\hat{u}$).

4.2.3 Numerical Implementation

4.2.3.1 Efficient determination of the Regularization Parameter

In the smoothing/deconvolution algorithms, it is necessary to solve Eq. B.8, where the inversion of matrix $(G^T B^{-1} G + \gamma^o F^T F)^{-1}$ requires $O(N^3)$ operations. This computation needs to be solved several times until a parameter γ is found to satisfy the chosen convergence criterion. The whole procedure is hence computationally inefficient. It is possible to diagonalize the problem, imposing a basis change in order to invert only a diagonal matrix. This is extensively documented (12) and summarized in Appendix B.

4.2.4 Offline vs Online Implementation

In order to apply the algorithm online, it is necessary to perform the deconvolution on a sliding window of the n past samples. We recall that if we consider to use the consistency criterion summarized in the Appendix B, the noise variances σ^2 and λ^2 are estimated from the

chosen portion of data. It is hence important to choose the appropriate length of the sliding window: too many samples will not allow a good tracking of possible variations, for example, of σ^2. On the other hand, short windows do not allow to correctly capture the dynamics of the signal. For sampling time of 1 minute, a good compromise is keeping about two hours of past data (results in this thesis will be shown for this choice of the parameter). Therefore n=120 if the sampling time is 1 minute. Every time that a new glucose value sample is returned by the sensor, only the last sample of the deconvoluted profile (time derivative) and reconvoluted curve (smoothed glucose) will be saved for DR calculation. When the next sample becomes available the process is repeated with a new window which discards the oldest samples and adds the newest to the collection.

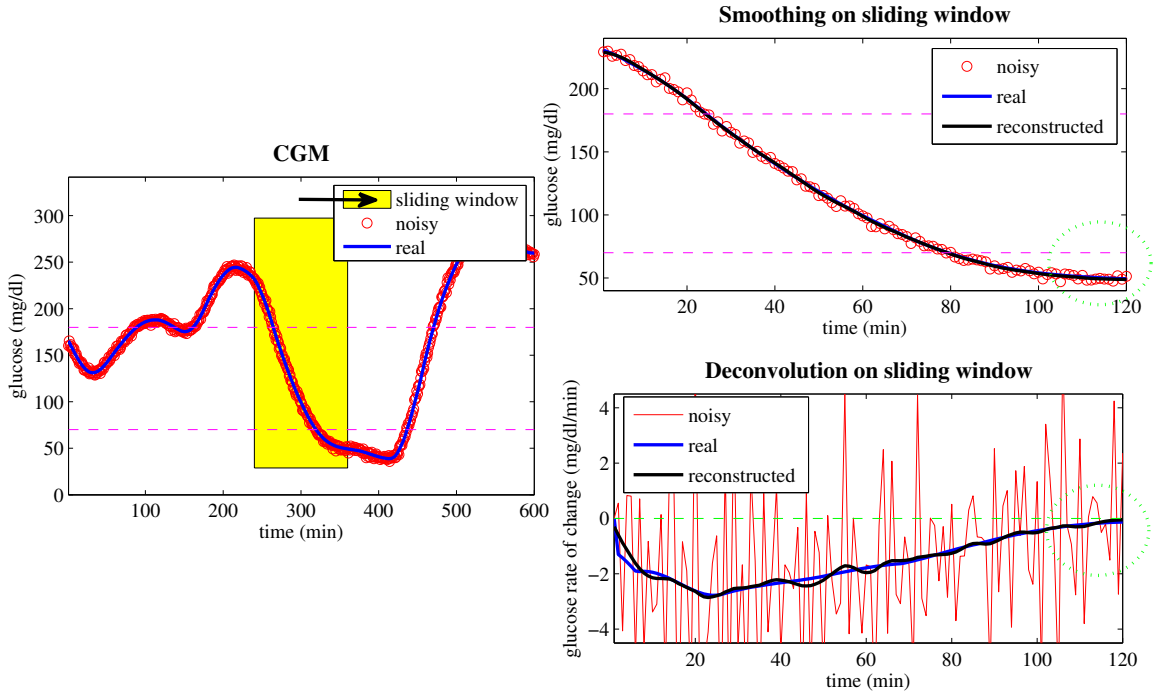

Figure 4.1: Online estimation of the smoothed CGM profile and time derivative. Left panel shows the original CGM and the portion of data available at time t (yellow box), while right panels show the results of deconvolution.

Figure 4.1 better explains this concept. The true, noise free signal (blue) is displayed with its noisy version (red). Highlighted in yellow is the portion of data that is chosen at time t, which is 120 samples long. Online, samples on the right of the yellow box are considered to be unavailable. On the right are displayed the profiles obtained by deconvolving the portion

of data in the yellow box. The top panel shows the smoothed CGM signal (black), along with the true and noisy CGM (blue and red respectively) while in the lower panel the true derivative (computed from noise free profiles, blue) is shown along with the time derivative obtained via finite differences (red) and deconvolution (black). Only the last samples (i.e. the 120^{th}) of the two curves are saved to be the smoothed CGM and its time derivative at time t. At time $t+1$ the windows shifts to the right and the process is iterated.

An important feature in smoothing, is that the uncertainty on the last component of the estimate vector usually greater than the errors on other samples. This can be appreciated in Figure 4.2, where the smoothed version of CGM and its derivative obtained via deconvolution are shown with their confidence intervals.

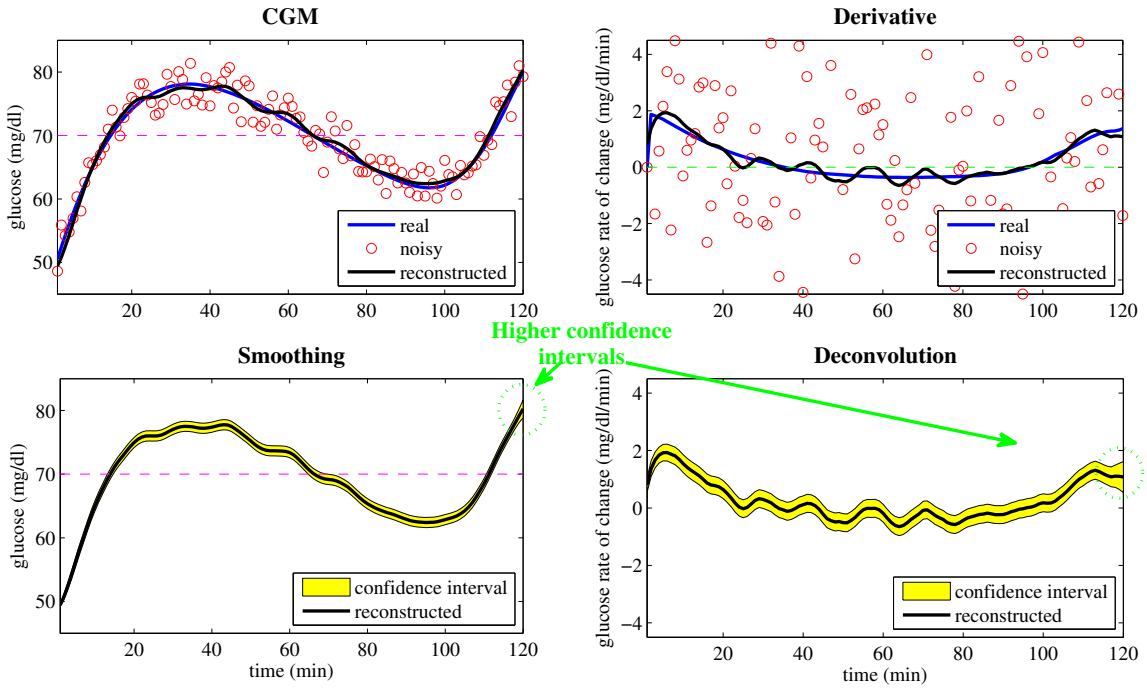

Figure 4.2: Results of deconvolution applied to the CGM signal on the upper-left panel. True CGM is displayed in blue with its noisy realization (red). The reconstructed CGM via deconvolution is shown in black. On the upper-right panel, the true time derivative (blue) is shown with its estimate provided by finite time-differences (red) and deconvolution (black). Lower panels: confidence intervals of the estimate of the smoothed CGM and of the time derivatives obtained via deconvolution.

This feature may be detrimental in our application online, since the estimate of time derivative on the last points will be more uncertain than in the rest of the winsow, propoa-

gating the uncertainty on DR. In order to reduce this effect, it is possible to force the deconvolution algorithm to consider data in the last portion of the window to have greater uncertainty, and hence to rely more on the trend (estimated based on reliable data of the first portion of the window). This can be done by using a matrix B in the equations of Appendix B which is different from the identity, for example:

$$\Delta = \begin{bmatrix} 1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & 0 & 1 & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & 0 & 2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & 0 & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & 0 & 0 & 10 \end{bmatrix} \tag{4.8}$$

The matrix of Eq. 4.8 considers the last 10 points to have greater uncertainty than the former data points. In particular the variance grows linearly on the last 10 points giving progressively less trust to the data and more on the trend when approaching the edge of the sliding window. Although this technique introduces a small delay in the estimate of the derivative, it also allows to recover a much smoother signal, which can be used to compute DR.

It is worthwhile to show some results to understand the role of the algorithm for the online estimation of time derivative.

Figures 4.3 shows a simulated signal where white random noise of variance $\sigma = 4mg^2/dl^2$ was added to simulate measurement noise (top). The middle panels represent show the time derivative computed from the noise-free signal (black line) and with two different algorithms: upper-middle panel displays in gray the first time derivative obtained via finite differences, while the mid-lower panel displays the results obtained with the deconvolution algorithm. In the bottom panel SR and DR_{exp} ($\mu = 2.2$) are shown.

Figure 4.4 is the same as Figure 4.3, where the noise signal added to the glycemic profile is a white random process of variance $\sigma = 16mg^2/dl^2$. Finally, in the lower panel, the DR of the noisy profile is shown against the static risk of the noise-free glucose profile. Notice that while SR crosses the threshold exactly when the original CGM profile does, DR anticipates such crossings. It is clear from these figures how the deconvolution approach to the estimation

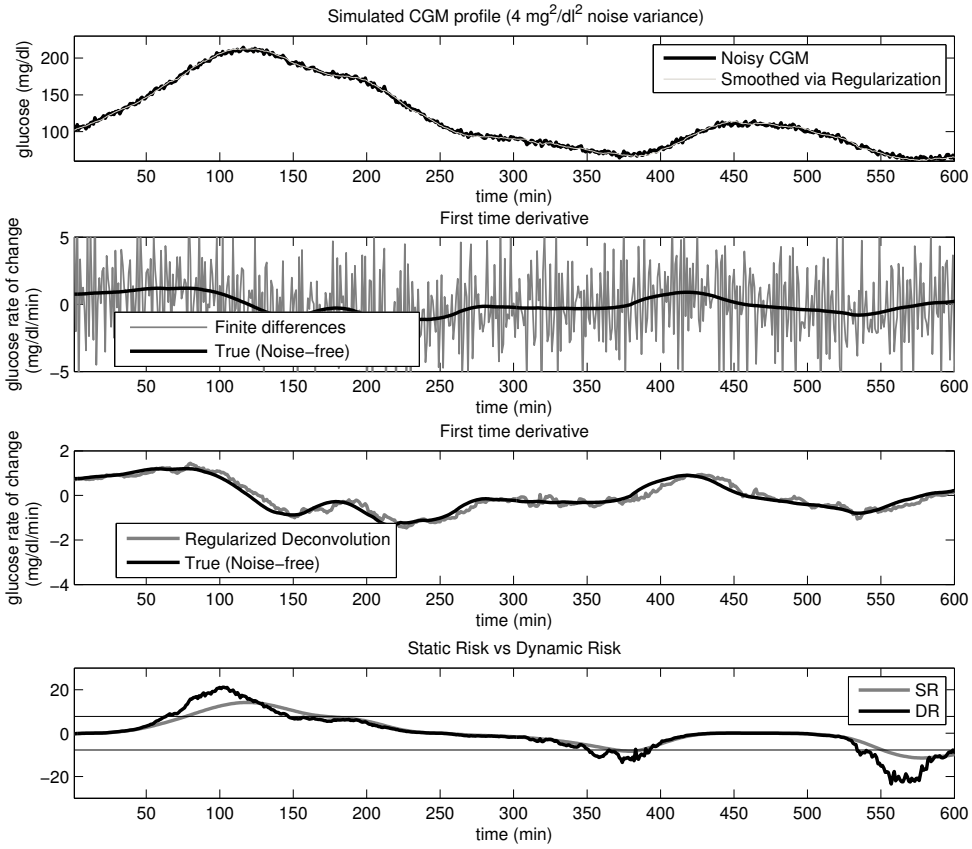

Figure 4.3: Top panel: reconstructed CGM profile (gray line) and simulated noisy profile (black line). Variance of the Gaussian noise added to the signals was $\sigma = 4mg^2/dl^2$. Middle panels: first time-derivative calculated from the noise free signal (black line) and estimated from the noisy signal via finite differences (upper middle, gray line) and regularization (lower middle, gray line). Bottom panel: SR (gray line) vs DR_{exp} (black line)

of the time derivative allows reconstructing a very smooth signal also in real time without adding a considerable amount of delay.

4.3 Conclusions

We presented an approach for the computation of the first time derivative of CGM signals which allows obtaining a reliable estimate of the first time derivative in signals affected by measurement noise. In Chapters 5 and 6 all computations are performed simulating an online conditions, and in presence of noise, the first time derivative of the glucose signals is always

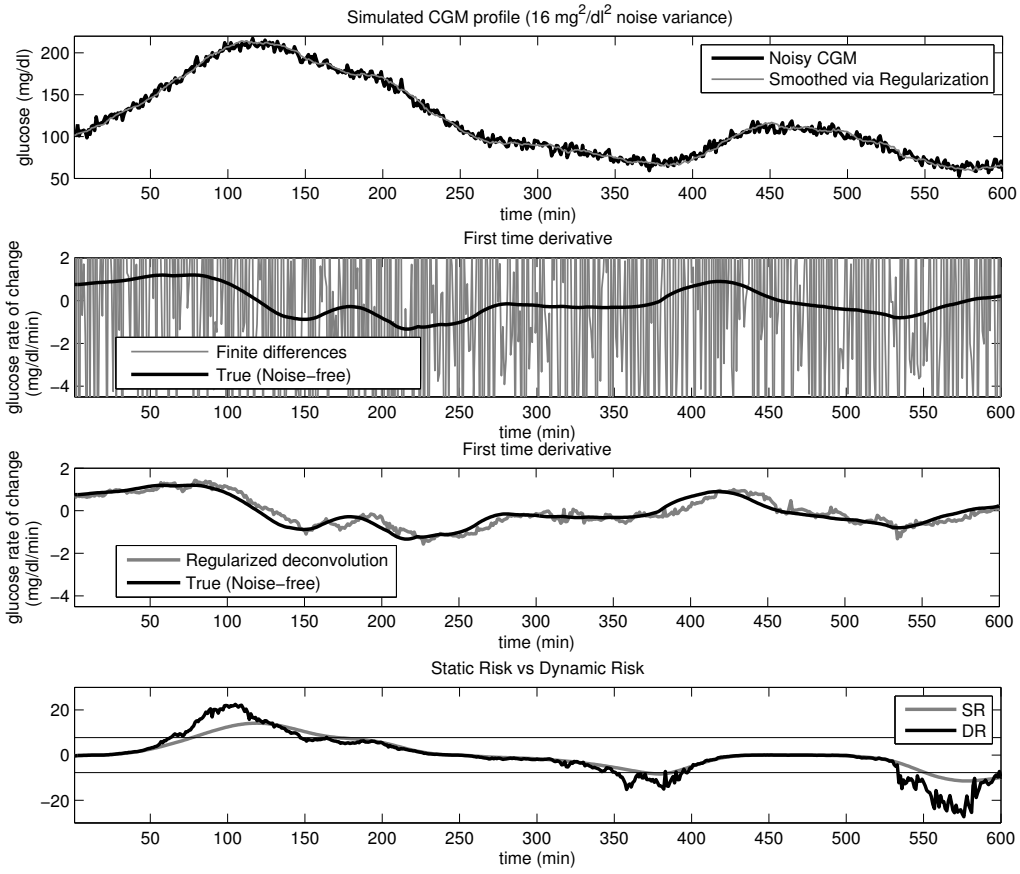

Figure 4.4: Top panel: reconstructed CGM profile (gray line) and simulated noisy profile (black line). Variance of the Gaussian noise added to the signals was $\sigma = 16mg^2/dl^2$. Middle panels: first time-derivative calculated from the noise free signal (black line) and estimated from the noisy signal via finite differences (upper middle, gray line) and regularization (lower middle, gray line). Bottom panel: SR (gray line) vs DR_{exp} (black line).

computed exploiting the deconvolution-based approach described in this chapter.

Part III

Use of Dynamic Risk for Hypo/Hyperglycemic Alert Generation

5

Application of DR for the prevention of hypo/hyperglcemic events

5.1 Prevention of Hypo/Hyperglycemic Events

Therapeutic actions, such as injection of insulin correction boluses to revert a condition of hyperglycemia or carbohydrates intake to treat hypoglycemia, cannot avoid exposure of the patients to events that can be threatening, either in the long or short term. In fact, insulin requires about an hour to be effective and to induce appreciable decrease in glucose concentration. Also, carbohydrates take time to reach the blood stream in order to compensate the effects of hypoglycaemia. In this framework, generating an alert ahead of the time could give the patient enough time for the therapeutic actions to be effective in avoiding the threats of the event itself. As anticipated in Chapter 3 (see in particular in figure 3.3), the Dynamic Risk DR can be use to predict the hypo and hyperglycemia threshold crossings ahead of time. In this Chapter we analyze the ability of DR in its online implementation (Section 4.2.4) of generating hypoglycemic alerts.

5.1.1 Generation of Hypo/Hyperglycemic Alerts in CGM Devices and Current Academic Research

The real-time detection/prevention of hypo/hyperglycemic events is a natural online application of CGM. Some of the minimally invasive CGM devices currently available in the market are provided with a visual/acoustic alert generator system that warns the patient when hypoglycemic or hyperglycemic thresholds are crossed. Often, this type of alert is based on the current glycemic value measured by the sensor (60). The FreeStyle Navigator (2) and the

MiniMed Paradigm Real-Time (38) also embed hypo/hyper alert generators based on the projection of current glucose level made by exploiting and trend.

Academic research is active in the development of methods able to generate alerts ahead of time. A detailed review can be found in (60). As suggested in (59) and (60), evaluation of alert generation must take into account the ability of anticipating threshold crossings by simoultaneously minimizing (or keeping as low as possible) the number of false alerts (i.e. alerts raised by the algorithm not followed by a real crossing of the hypo/hyperglycemic threshold).

Briefly, in Choleau et al. (8), for instance, an hypoalert is generated when the future glycemic concentration, obtained on the basis of first-order linear extrapolation of the last two/three glucose samples, is forecasted to cross the hypoglycemic threshold within 20 min. Sparacino et al. (58), demonstrated that simple prediction algorithms based on model with a reduced number of parameters, i.e. either first-order polynomial or first-order auto-regressive (AR(1)) models, with time-varying parameters identified by least squares (LS) using a fixed forgetting factor, are suitable for predicting glycemia ahead in time with a sufficient accuracy, with a PH of 30 and 45 min. Eren-Oruklu et al. (16) developed prediction algorithms based on AR(3) and ARMA(3,1) models, with time-varying parameters identified by LS, using a forgetting factor μ which could be modulated according to the glucose trend. Reifman et al.(50) proposed a predictor based on an AR(10) model, with time-invariant and subject-invariant parameters identified by regularized LS. Similarly, Gani et al. (20) developed a prediction strategy based on an AR(30) model with time-invariant parameters identified by regularized LS on pre-filtered data. Finan et al. (19) proposed a predictor based on an ARX(3) model with exogenous inputs given by ingested carbohydrates and insulin medications, both with time-invariant and time-variant parameters. Palerm and Bequette (43), after having posed the problem in a state-space setting, used the Kalman filtering methodology to predict glucose level after a given PH, using a double integrated random walk as prior for glucose dynamics. Recently, NN models have been the subject of some investigations for glucose prediction. Pérez-Gandía et al. (47) developed a feedforward NN for glucose prediction, trained and tested with 3 different PHs, i.e. 15, 30, and 45 min. More recently, Pappada et al. (46), (45) proposed a NN approach to predict glycemia with a PH of 75 min. Finally, a preliminary study carried out on a limited dataset consisting of only one patient was developed by Eskaf et al. in 2009 Inputs of their NN model include the first-order differences of the glycemic time series, and information on meals, insulin and physical exercise, extracted directly from

the blood glucose time-series, by modeling the glycemic level as a dynamic system. Another paper by Zecchin et al. (67) exploits Neural Networks incorporating a model informed with the timing and amount of meals.

5.1.2 Use of DR to generate Short-term Hypo/Hyper Alerts

The idea is to assess if and how a CGM device may take benefit from an algorithm which generates preventive hypoglycemic and hyperglycemic alerts based on the concept of current clinical risk associated to the glycemic value and its trend.

5.2 Dataset

5.2.1 Simulated Dataset

As a first step, DR was applied to simulated data, in order to tune the parameters of the function. The simulated Data Set consists of 10 frequently sampled Blood Glucose profiles (5 minutes sampling rate) available on the internet (21) were considered. The BG profiles were filtered via a smoothing algorithm and almost continuous glucose profiles were obtained (1 minute sampling rate). DR and a combination of DR with a state of art prediction algorithm will be applied on these profiles both with and without superimposed noise. Anticipation in threshold crossings and false alerts generated will be compared for the different alert generation tools.

5.2.2 Real Dataset

DR was then applied to real data, obtained with four different CGM sensors.

1. Abbott Freestyle Navigator (collected within the project DIAdvisor) Dataset 1: 10 glucose profiles, sampling time 10 minutes, average length of the signals 4297 $\pm$ 49 minutes. Number of Hypo episodes: 14, Number of Hyper episodes 72.

2. Abbott Freestyle Navigator (data taken from a previously published dataset (34)) Dataset 2: 10 glucose profiles, sampling time 1 minute, average length of the signals 1884 $\pm$ 362 minutes. Number of Hypo episodes: 8, Number of Hyper episodes 36.

3. Dexcom Seven Plus Dataset (data provided by DexCom Inc., in the framework of a development agreement with the Department of Information Engineering of the University of Padova) 1: 10 glucose profiles, sampling time 5 minutes, average length of the signals 9876 $\pm$ 172 minutes. Number of Hypo episodes: 106, Number of Hyper episodes 208.

4. Menarini Glucoday (data provided by Menarini Diagnostics, in the framework of a past research collaboration with the Department of Information Engineering of the University of Padova): 10 glucose profiles, sampling time 3 minutes, average length of the signals 2703 $\pm$ 303 minutes. Number of Hypo episodes: 45, Number of Hyper episodes 73.

The analysis performed on the real data set were similar to those performed on the simulated dataset: alerts were generated with different tools (DR computed via exponential/hyperbolic tangent functions, state-of-the art prediction techniques) and results were compared in terms of temporal gain and number of false alerts generated.

5.3 Set-up of DR Parameters on Simulated Data

This section reports the results obtained on simulated data.

5.3.1 Tuning of parameter μ

When considering the first definition of the DR, i.e. Eq. 3.11, we needed to define what value the parameter μ should have. Figure 5.1 shows three different DR profiles (based on the same glycemic profile) evaluated for three different values of μ (continuous, dashed and dotted gray lines are evaluated with $\mu = 1$, $\mu = 2.2$ and $\mu = 4$, respectively). The higher the value of μ, the greater the amplification and damping relative to the same profile. Greater amplification of risk results in greater time gain in the threshold crossing at time $\simeq 360$ min, but, on the other hand, it increases the risk of generating false alerts. Overestimating the risk may lead to announcing a threshold crossings when in fact, the condition of the patient is still safe. Empirically, we tuned μ to the value of 2.2, since it allowed obtaining a reasonable time gain of circa 10 minutes with a relatively small number of false alerts (Results in Section 5.4).

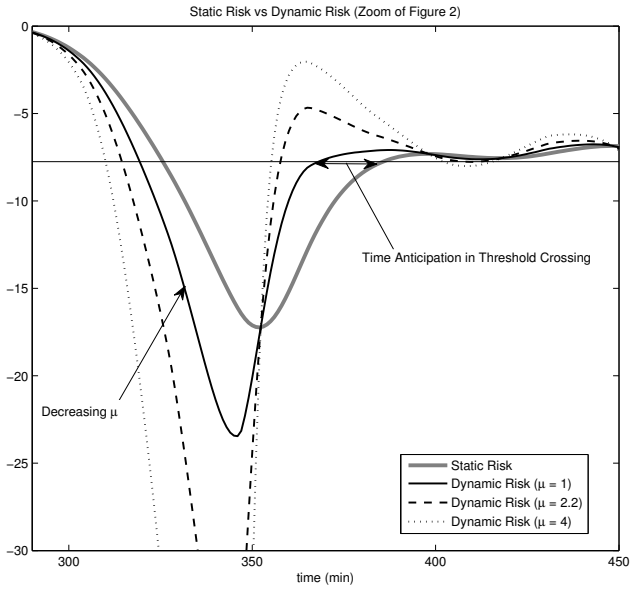

Figure 5.1: Role of μ: SR (thick solid line) vs DR for a noise-free glucose profile, where DR is calculated with μ=1, μ=2.2 and μ=4 (black solid, dashed and dotted line).

5.3.2 Assessment of Exponential-based vs. Hyperbolic Tangent-based Structure

The DR function defined as in 3.11 shows some interesting features. Being based on the exponential function it is really simple, and very rapid in changing. The conceptual example shown in Figure 3.3 demonstrates the ability of this structure to anticipate threshold crossings. This structure has nonetheless some problems. Figure 5.2 reports an example of simulated CGM profile (top panel) showing a prolonged condition of severe hypo. The signal presents a mild recovery, then a stationary period in hypo, to finally recover completely. The lower panels display SR (red) and DR (green). Notice how during the mild recovery between min 700 and min 750, DR is an attenuated version of SR, i.e. the risk is lowered and closer to zero. The damping is so strong that DR actually returns within the safety region. Unluckily, when the patient falls again a little bit deeper in hypo (minutes 800-900), DR raises a "false alert". Notice that this is a different false alert with respect to normal threshold crossings, since it starts from a clinically wrong decrease in risk.

To circumvent this problem, the hyperbolic-tangent-based structure of Eq. 3.12 was used. In particular, this structure is more flexible since it allows very easily to define two important

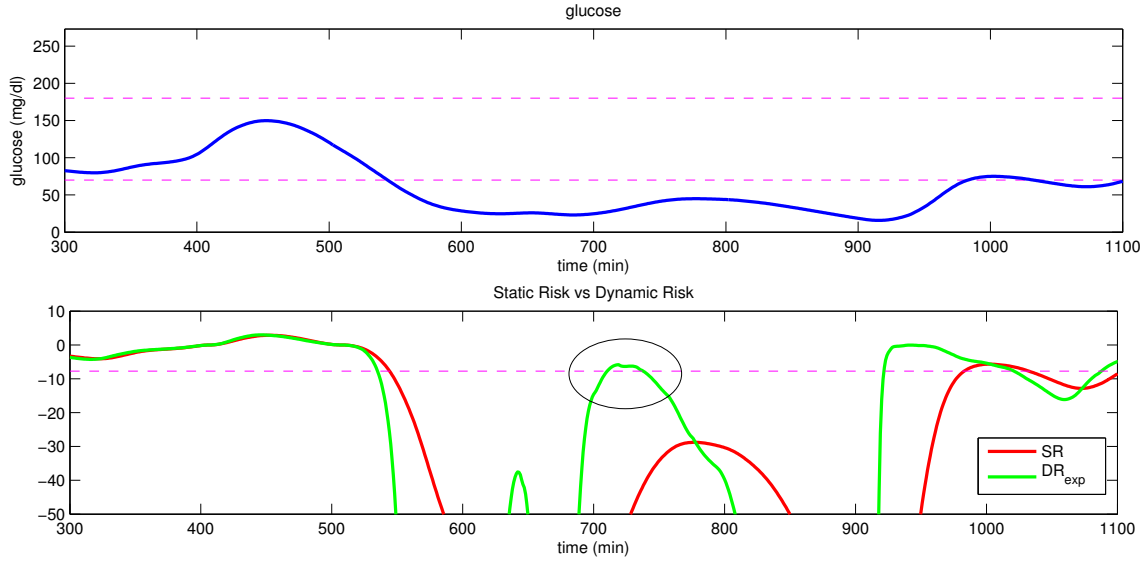

Figure 5.2: Example of failure of DR_{exp}. The damping when the subjects id heading back to euglycemia is too strong, and DR (lower panel, green) returns in safety zone very early. The derivative of the glucose signals on the way back to safety then decreases, and this results in a false positive alarm (evidenced in the circle).

features:

1. Controlled maximal damping when heading back to euglycaemia. In particular, recall Section 3.3.1. Here, the behavior of the exponential was explained when the time-derivative was in condition of very fast recovery to the euglycemic region, and it was clarified how in this case the damping reduced the DR to be exactly zero. Parameter β can be set to force the limit to when $x \to -\infty$ to be equal, for example to 0.5. This means that when dg/dt and SR have different sign, and the derivative is very high in module, SR will be reduced in magnitude no more than by its half.

2. Possibility of saturating the magnification of risk. In the exponential function, the amplification grows indefinitely with the time derivative. From a conceptual point of view this is correct, but from a practical point of view, it may be useful to assign a reference maximal possible amplification to the highest expected time derivative value with physiological meaning. In fact, knowing that variations greater than 4 or 5 mg/dl/min are not physiologically plausible and hence need to be considered spurious oscillations,

it make sense at least to set a higher limit to amplification. Parameter δ can be set to force also a maximum amplification of risk.

In this thesis we will refer to the exponential-based and hyperbolic tangent-based structures as DR_{exp} and DR_{tanh} respectively.

5.3.2.1 Parameter Choice for the DR_{tanh} Structure

As shown from Eq. 3.12, the structure based on the hyperbolic tangent has 4 parameters that need to be set:

1. Parameter γ: assuming that β and δ have been fixed, γ is set in order to fulfill the constraint that at zero time-derivative, i.e. in stationary conditions, DR equals SR. This means that for $dr/dt = 0$ the amplification factor needs to be equal to 1 (independent of α). In particular, gamma is set as:

$$\delta \tanh(\gamma) + \beta = 1 \tag{5.1}$$

   i.e.,

$$\gamma = \tanh^{-1}\left(\frac{1-\beta}{\delta}\right) \tag{5.2}$$

2. Parameter β. This parameter influences the maximum damping of DR. When dg/dt and SR have different sign (suppose $SR > 0$, and $\frac{dr}{dt} < 0$), and the derivative is very high in module the amplification factor (f) will have the form of

$$f = [\delta \lim_{x \to -\infty} \tanh(x + \gamma) + \beta] = -\delta + \beta \tag{5.3}$$

   this means that, for example to have a damping to the 75% of the risk we need to set $\beta = \delta + 0.75$

3. Parameter δ. This parameter influences the maximum amplification. In particular notice that if dr/dt and SR have the same sign (suppose $SR > 0$, and $\frac{dr}{dt} > 0$), and the derivative is very high in module the amplification factor will have the form of

$$f = [\delta \lim_{x \to \infty} \tanh(x + \gamma) + \beta] = \delta + \beta \tag{5.4}$$

if we set $\beta = \delta - y$ where with y we define the maximum damping (0.75 in the example above), and, for example, we want a maximum amplification of $h = 10$ times the risk, we have $\delta + (\delta + y) = h$, hence $\delta = (h - y)/2$. In all the discussion in the thesis we will refer to the maximum amplification parameter h as, in fact, δ. This means that when we state that parameter δ is set to 10, we mean in fact that $h = 10$, i.e. the maximum amplification is 10.

4. Parameter α. This parameter defines the rapidity of risk amplification for different time derivative values. For this parameter there are no constrains, but higher α result in faster amplification and higher "contrast" in a smaller range of derivative values. The role of α for DR_{tanh} is similar to the role of μ in DR_{exp}.

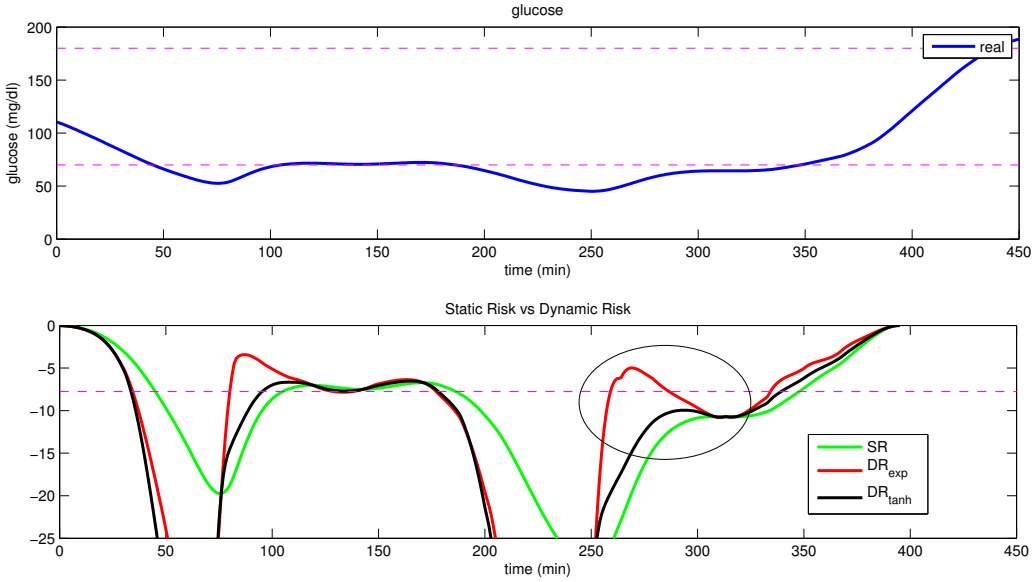

Figure 5.3: Example of profiles generated using the two presented DR structures (DR_{exp} and DR_{tanh}, red and blue respectively) vs SR (black). Notice that DR_{tanh} allows obtaining the same time gain as DR_{exp} but offers a slower return to the euglycemic range. This allows avoiding, for example, the false alert generated by DR_{exp} at minute 255.

The parameters were set in order to have a limit in the damping equal to 0.75, i.e. the risk cannot be decreased more than 25%. Maximum amplification is usually set to 10, although in Chapter 6 also other values will be used in simulation environment. Parameter γ is fixed once β and δ are defined. In this Chapter no optimization is shown for α and a fixed value of 2 is used unless differently specified. In Chapter 6 simulations will be run to investigate

the role of parameter α when using DR as an alert generator, and values between 0.5 and 5 were tested.
Figure 5.3 shows two examples of DR evaluated with DR_{exp} and DR_{tanh} structures. In the circles, critical episodes are highlighted. In particular in the bottom panel it is evident how DR_{exp} (red) causes the generation of a false alert, while the more conservative damping introduced by DR_{tanh} allows preventing the generation of such alarm.

5.4 Criteria for assessing Results on Simulated Data

Only results on hypoglycemic threshold crossings will be presented in this thesis, although the same analogous could be shown for hyperglycemic threshold crossings.
The temporal gain in threshold crossing is evaluated as the time difference between the crossing of the predictive signal at 70 mg/dl and the subsequent crossing of the static risk, which we recall, crosses the threshold simultaneously with the CGM signal. This allows comparing the condition of no prediction (alert raised when hypoglycaemia has already occurred) and the preventive alert generated by the algorithms.

A false alarm is an episode where an alert is raised by the predictive algorithm (i.e. the predicted profile crosses the 70 mg/dl threshold or the risk function falls below the static risk value corresponding to 70 mg/dl) while the glucose signal doesn't actually fall below 70 mg/dl. A snooze time of 20 or 45 minutes is set which prevents the generation of alerts right after the first alert. This makes sense in real life, where after the first alarm the patient is supposed to take actions to recover from the hypoglycemia.

5.5 Results on Simulated Data

5.5.1 Noise-free Simulated Data

Temporal gains and false positives were evaluated for the 10 simulated noise-free profiles. Figure 5.4 shows the histogram distribution (left) of time gain and false positives for DR_{exp} ($\mu = 2.2$) (top panels) and DR_{tanh} ($\alpha = 3$ and $\delta = 10$) (bottom panels). In this case the snooze time is 45 minutes, relatively long, meaning that no alerts are raised 45 minutes after the first occurance. Notice how the median gain in threshold crossing is equal to 11 minutes for both structures, but DR_{tanh} results in lower mean and median number of false alerts in

case of noise free signals. This example proves the superiority of DR_{tanh} with respect to DR_{exp} in this application.

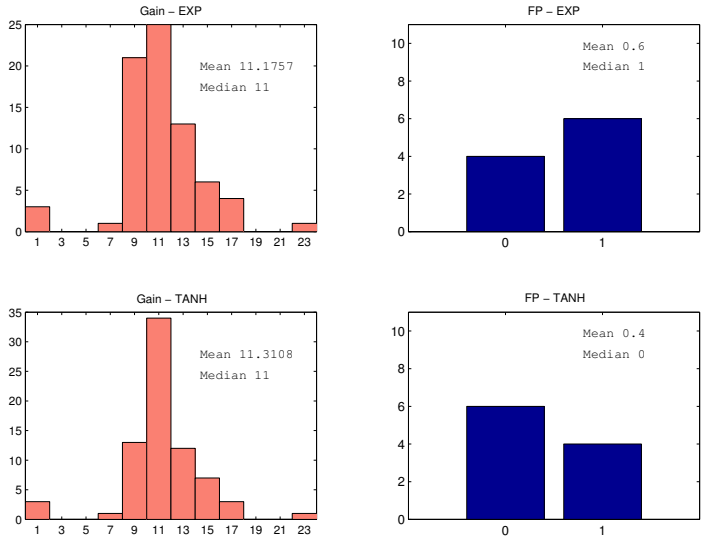

Figure 5.4: Distribution of Temporal Gain and False alerts for DR_{exp}, $\mu = 2.2$ (top panels) and DR_{tanh}, $\alpha = 3$, $\delta = 10$ (lower panels) for noise-free signals. Snooze Time was set to 45 minutes

Figure 5.5 shows the distribution of time gain and false positives for DR_{exp} ($\mu = 2.2$) and DR_{tanh} ($\alpha = 3$ and $\delta = 10$) when the snooze time is relatively short, i.e. 20 minutes. Also considering a shorter snooze time, DR_{tanh} performs better than DR_{exp} in terms of false positives, adding only one alarm in 5 out of 10 subjects, and adding no false alerts at all in the other 5. On the other hand, 7 out of 10 subjects experience one false alert with DR_{exp}. Also while the median is the same for the two structures, DR_{tanh} shows a slightly higher mean in temporal gain, which is a nice feature. A summary of the time gains in terms of mean and median and the number of false positive for the two structures on noisy data, considering again two possible implementation with snooze times of 20 and 45 minutes is shown in table 5.1.

The Table highlights that both structures allow obtaining a median temporal gain of 11 minutes, with DR_{tanh} showing a lower median number of false alerts with both snooze time settings (20 or 45 minutes).

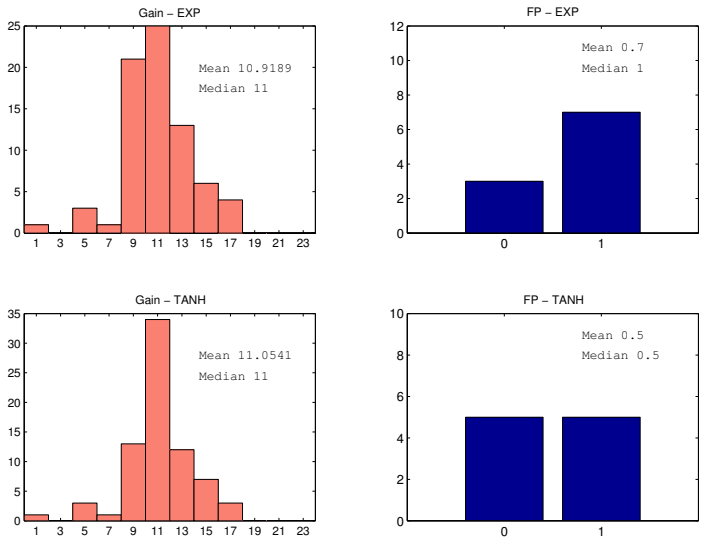

Figure 5.5: Distribution of Temporal Gain and False alerts for DR_{exp}, $\mu = 2.2$ (top panels) and DR_{tanh}, $\alpha = 3$, $\delta = 10$ (lower panels) for noise free signals. Snooze Time was set to 20 minutes

Structure	Snooze	Gain		FP	
		mean	median	mean	median
Noise free					
EXP	20	10.92	11	0.7	1
TANH	20	11.05	11	0.5	0.5
EXP	45	11.18	11	0.6	1
TANH	45	11.31	11	0.4	0

Table 5.1: Performance of DR_{exp} and DR_{tanh} in terms of temporal gain in anticipation of threshold crossing and false alerts generated for noise-free profiles. Results are summarized with the mean and median of the distribution of temporal gains and # of false positives for two different snooze-times.

5.5.2 Noisy Simulated Data

DR_{exp} and DR_{tanh} were computed for all the 10 noisy profiles. Figure 5.6 shows the distribution of time gain and false positives for DR_{exp} ($\mu = 2.2$) and DR_{tanh} ($\alpha = 3$ and $\delta = 10$) for simulated signals with zero-mean white noise with variance $\sigma^2 = 4$. In this case the snooze time is 45 minutes. With a long snooze time, DR_{exp} and DR_{tanh} perform similarly in terms of median and mean time gain, the second being slightly more reactive than the first. Median of false positives is, in this case, comparable for the two structures.

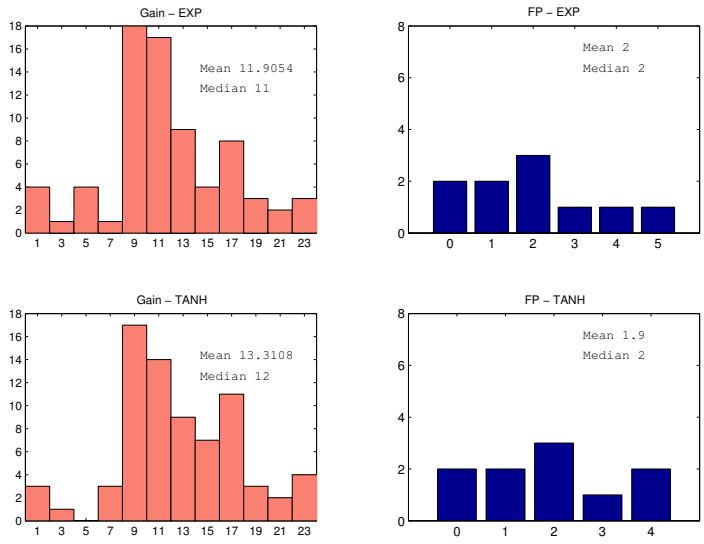

Figure 5.6: Distribution of Temporal Gain and False alerts for DR_{exp}, $\mu = 2.2$ (top panels) and DR_{tanh}, $\alpha = 3$, $\delta = 10$ (lower panels) for signals with noise variance $\sigma^2 = 4$. Snooze Time was set to 45 minutes

Figure 5.7 shows the distribution of time gain and false positives for DR_{exp} ($\mu = 2.2$) and DR_{tanh} ($\alpha = 3$ and $\delta = 10$) for the same signals but with a snooze time of 20 minutes. With a shorter snooze time, DR_{tanh} provides a slightly higher time gain than DR_{exp} with 0.5 less FA median per subject.

Finally, we summarize in table 5.2 the time gains in terms of mean and median and the number of false positive for the two structures on noisy data, considering again two possible implementation with snooze times of 20 and 45 minutes.

Structure	Snooze	Gain		FP	
		mean	median	mean	median
Noisy $\sigma^2 = 4mg^2/dl^2$					
EXP	20	11.26	11	3.6	4
TANH	20	12.15	12	3.4	3.5
EXP	45	11.18	11	2	2
TANH	45	13.31	12	1.9	2

Table 5.2: Performance of DR_{exp} and DR_{tanh} in terms of temporal gain in anticipation of threshold crossing and false alerts generated for noise-free profiles. Results are summarized with the mean and median of the distribution of temporal gains and # of false positives for two different snooze-times.

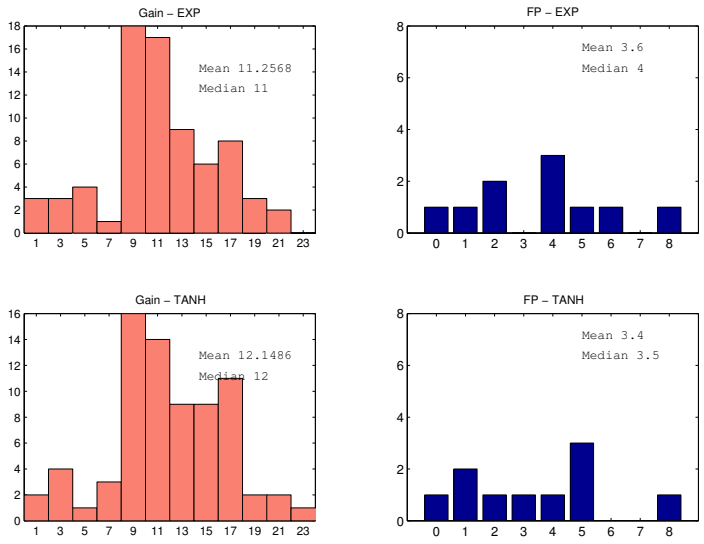

Figure 5.7: Distribution of Temporal Gain and False alerts for DR_{exp}, $\mu = 2.2$ (top panels) and DR_{tanh}, $\alpha = 3$, $\delta = 10$ (lower panels) for signals with noise variance $\sigma^2 = 4$. Snooze Time was set to 20 minutes

Also on noisy data, DR_{exp} and DR_{tanh} allow anticipating the threshold crossing of median 11 and 12 minutes, the latter being a safer algorithm with less false positive alerts than the first with a short snooze time.

5.6 Results on Real Data

Figures 5.8, 5.12, 5.14 and 5.16 show the results obtained on a representative signal from the datasets Abbott (1 minute samplig rate), Diadvisor (10 minutes sampling rate), Dexcom (5 minutes sampling rate) and Menarini (3 minutes sampling rate). The upper panels show the original signal (red), a filtered version of the signal obtained via **offline** Butterworth filter (blue - forward/backward filter) used as comparison, and the reconstructed glucose signal (black) obtained **online** via deconvolution. Lower panels show the static risk function of the filtered glucose (red), DR_{exp} and DR_{tanh} of the reconstructed glucose (green and black respectively). The filtered glucose and its static risk are only used as an offline comparison and are thought as a noise free version of the glycemia and its risk. We consider a true threshold crossing every crossing of these profiles in their respective domains. Figures 5.9, 5.13, 5.15 and 5.17 show the distributions of time gains (left panels, in minutes, considering

a snooze time of 45 minutes) and false alerts (right) for the hypoglycemic threshold crossing at 70 mg/dl. Histograms on top panels refer to the use of DR_{exp}, while lower panels refer to DR_{tanh}.

Both methods allow to achieve a temporal gain around 10 minutes on real data, with DR_{tanh} performing slightly better in terms of stability of the signal (qualitatively) and of false alerts raised. We considered a snooze time of 45 minutes to mimic real life, where a patient is likely to snooze the sensor after the first alarm in order to take the necessary therapeutic actions. The algorithm is able to work with different sensors, and at different sampling times. Different sensors have different SNR, so possibly different snooze times can be set along with different parameters of DR. So far, the tuning has been done by inspection, but future work might address the need of adapting to different SNR the parameters. More frequent sampling rate allows to better appreciate the threshold crossing gain and allow a realistic first-time derivative estimate, provided that noise is correctly dealt with.

5.7 Conclusions

DR can be used to generate alerts ahead of time thanks to its intrinsic ability to anticipate threshold crossings with a certain accuracy. The alert generated indicates that the actual time derivative and glucose levels might resolve in an hypo/hyper soon if no therapeutic actions are taken. The generation of alerts is hence clinically interesting, since it also considers explicitly the role of the trend and translates it into a risk score. In the next chapter, a refinement of the use of DR as an alert generator will be presented.

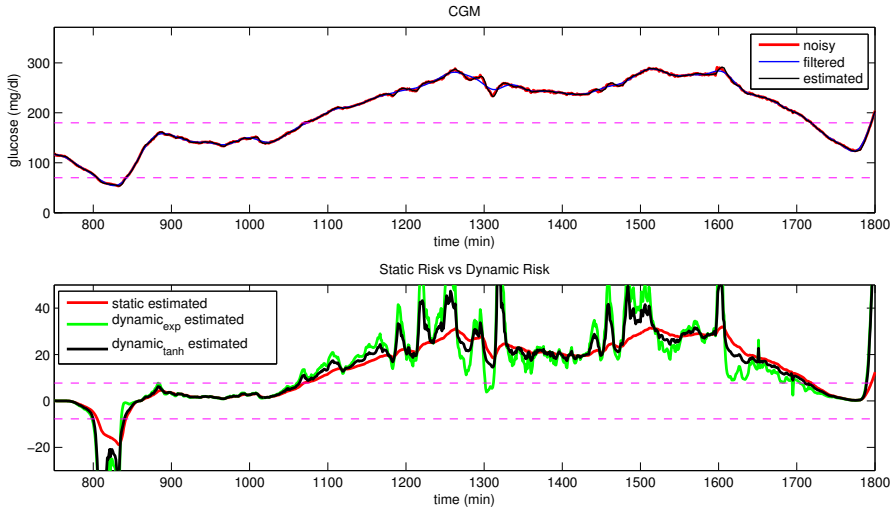

Figure 5.8: Represenatative Subject for the dataset Abbott (1 minute sampling rate). Upper panel: original signal (red), filtered version (Butterworth filter) (blue), reconstructed glucose signal (black) obtained via deconvolution. Lower panel: static risk function of the filtered glucose (red), DR_{exp} and DR_{tanh} of the reconstructed glucose (green and black respectively).

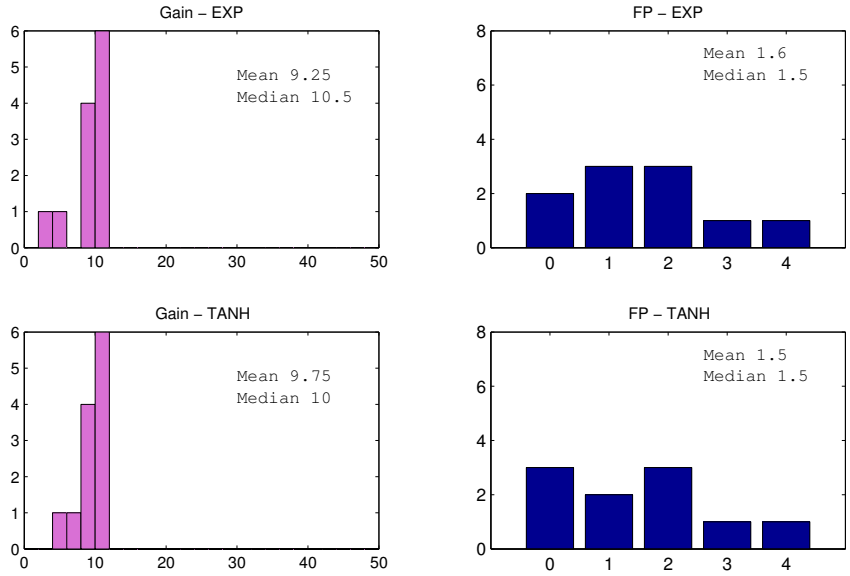

Figure 5.9: Distribution of time gains in minutes (left) and false positive (right) for the 10 subjects of the Abbott Dataset. (Sampling time 10 minutes, snooze 45 minutes. Parameters: $\mu = 2.2$ (DR_{exp}); $\delta = 10$, $\alpha = 2$ (DR_{tanh});)

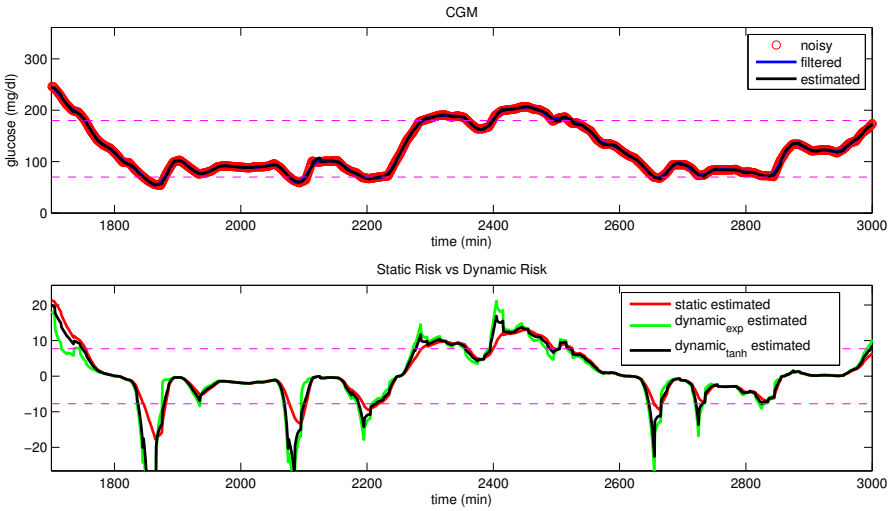

Figure 5.10: Represenatative Subject for the dataset Diadvisor (10 minutes sampling rate). Upper panel: original signal (red), filtered version (Butterworth filter) (blue), reconstructed glucose signal (black) obtained via deconvolution. Lower panel: static risk function of the filtered glucose (red), DR_{exp} and DR_{tanh} of the reconstructed glucose (green and black respectively).

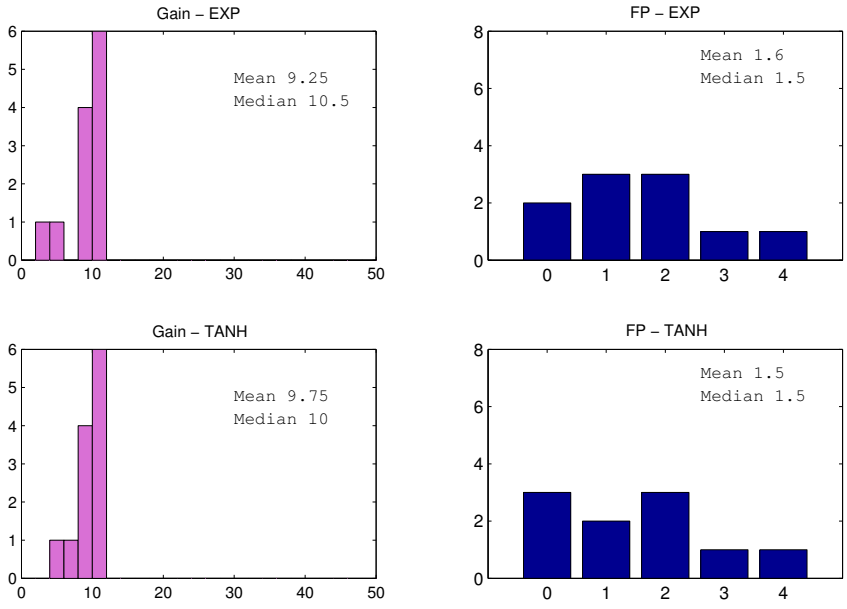

Figure 5.11: Distribution of time gains in minutes (left) and false positive (right) for the 10 subjects of the Diadvisor Dataset. (Sampling time 1 minutes, snooze 45 minutes. Parameters: $\mu = 2.2$ (DR_{exp}); $\delta = 10$, $\alpha = 2$ (DR_{tanh});)

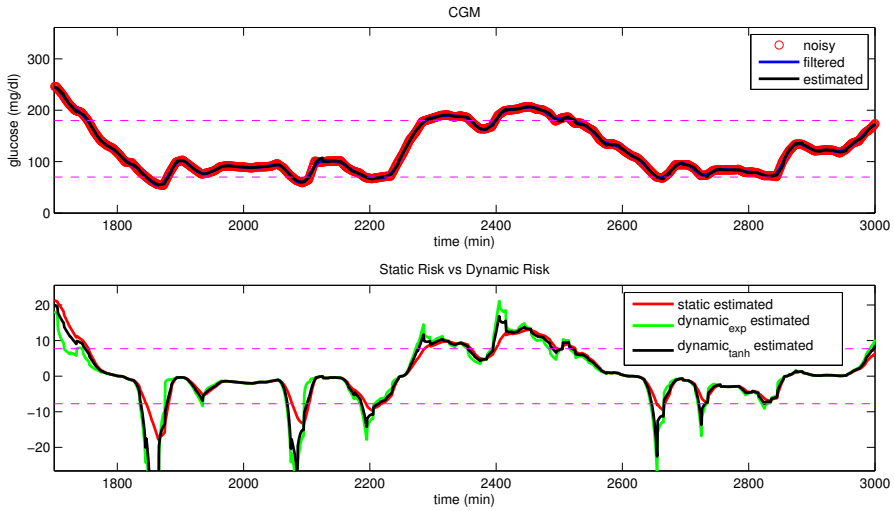

Figure 5.12: Represenatative Subject for the dataset Diadvisor (10 minutes sampling rate). Upper panel: original signal (red), filtered version (Butterworth filter) (blue), reconstructed glucose signal (black) obtained via deconvolution. Lower panel: static risk function of the filtered glucose (red), DR_{exp} and DR_{tanh} of the reconstructed glucose (green and black respectively).

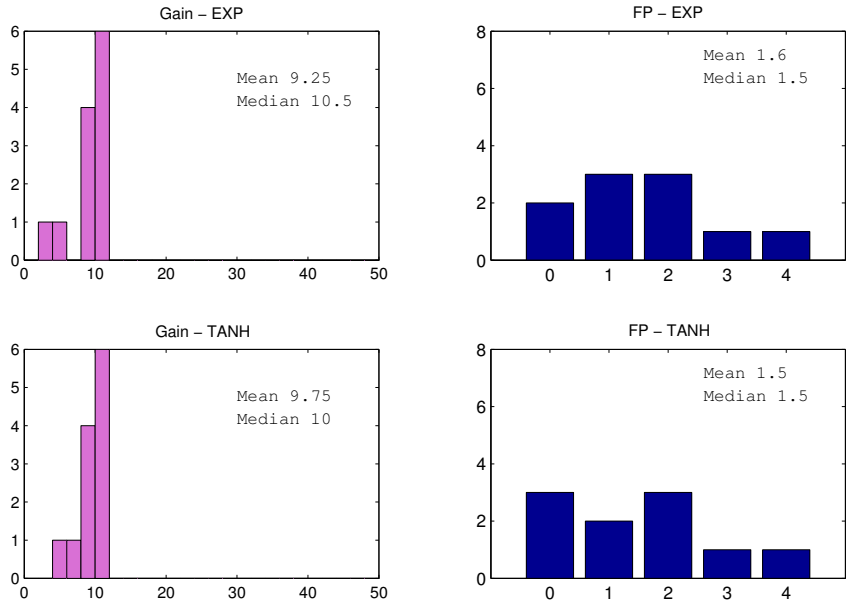

Figure 5.13: Distribution of time gains in minutes (left) and false positive (right) for the 10 subjects of the Diadvisor Dataset. (Sampling time 1 minutes, snooze 45 minutes. Parameters: $\mu = 2.2$ (DR_{exp}); $\delta = 10$, $\alpha = 2$ (DR_{tanh});)

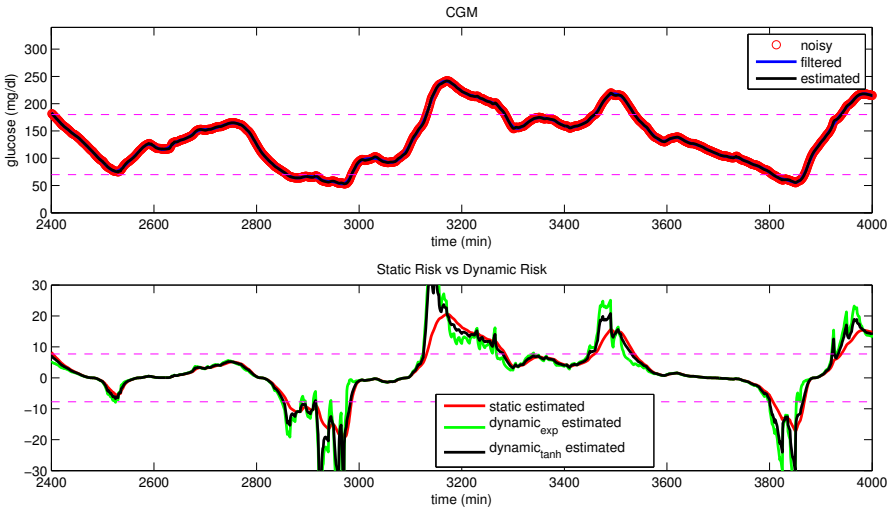

Figure 5.14: Represenatative Subject for the dataset Dexcom (5 minutes sampling rate). Upper panel: original signal (red), filtered version (Butterworth filter) (blue), reconstructed glucose signal (black) obtained via deconvolution. Lower panel: static risk function of the filtered glucose (red), DR_{exp} and DR_{tanh} of the reconstructed glucose (green and black respectively).

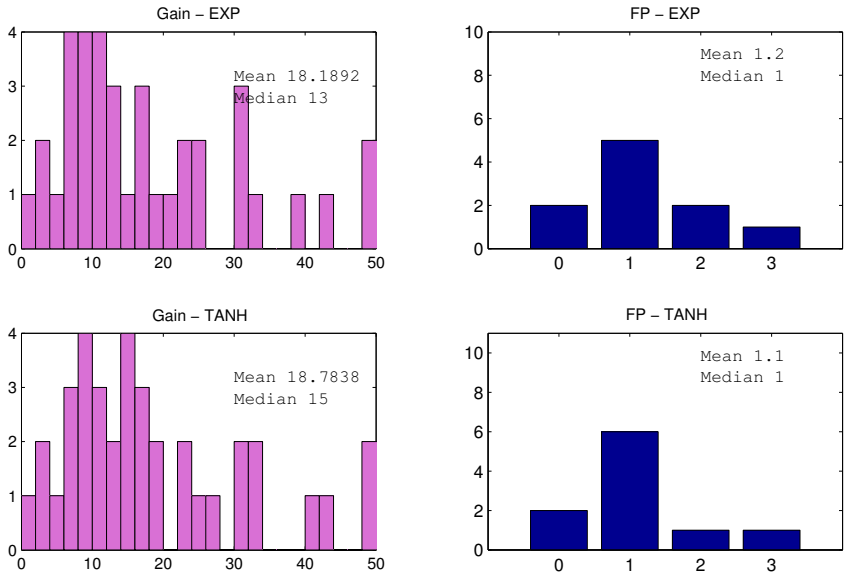

Figure 5.15: Distribution of time gains in minutes (left) and false positive (right) for the 10 subjects of the Dexcom Dataset. (Sampling time 5 minutes, snooze 45 minutes. Parameters: $\mu = 2.2$ (DR_{exp}); $\delta = 10$, $\alpha = 2$ (DR_{tanh}))

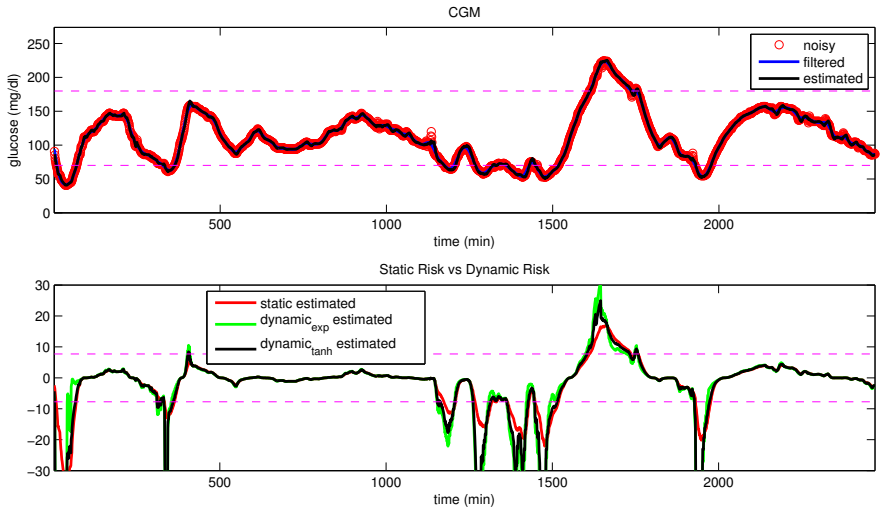

Figure 5.16: Represenatative Subject for the dataset Menarini (3 minutes sampling rate). Upper panel: original signal (red), filtered version (Butterworth filter) (blue), reconstructed glucose signal (black) obtained via deconvolution. Lower panel: static risk function of the filtered glucose (red), DR_{exp} and DR_{tanh} of the reconstructed glucose (green and black respectively).

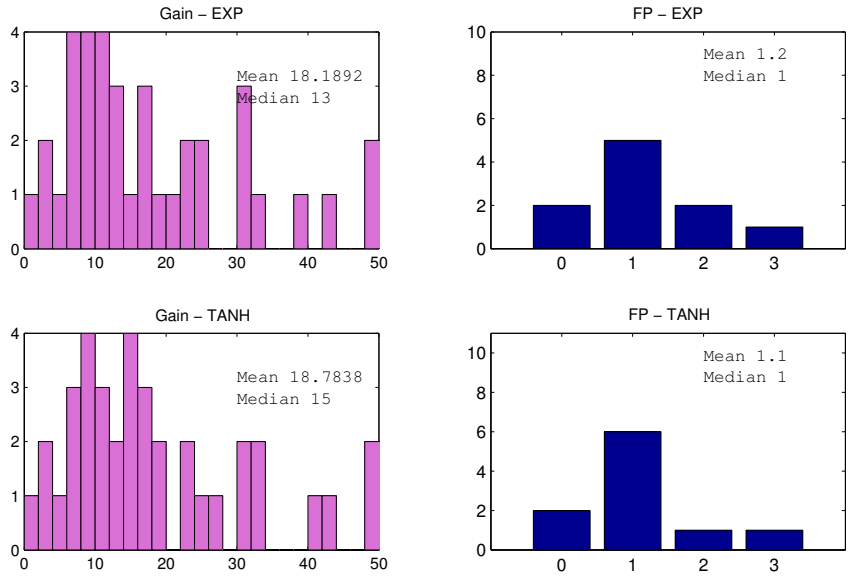

Figure 5.17: Distribution of time gains in minutes (left) and false positive (right) for the 10 subjects of the Menarini Dataset. (Sampling time 3 minutes, snooze 45 minutes. Parameters: $\mu = 2.2$ (DR_{exp}); $\delta = 10$, $\alpha = 2$ (DR_{tanh});)

5. APPLICATION OF DR FOR THE PREVENTION OF HYPO/HYPERGLCEMIC EVENTS

6

Prevention of hypo/hyperglycemic events through on-line calculation of DR: further improvement by use in combination with Kalman Filter

In this Chapter we propose to use the DR function combined with a short-term prediction based on the Kalman Filter. This will further increase the anticipation in hypo/hyper threshold crossings to 20, or more, minutes.

6.1 DR combined with Kalman Filter based Prediction

The algorithm used to obtain a short-term prediction of glucose is presented in Appendix C. It is a development of a denoising algorithm (57) which implements a Kalman Filter coupled with a Bayesian smoothing criterion to estimate the unknown filter parameters. In this section we discuss the possibility of combining the prediction performed via Kalman Filter with DR. In particular, we apply DR with the implementation based on the tanh function to the predicted profile obtained via KF. Different PH and parameters will be tested in simulation to tune the best combination of prediction and risk amplification.

6.1.1 Alert Generation

For each signal, three alarm generation tools were considered:

1. Short-term prediction per se: for this algorithm the only parameter which can be set by the user if the Prediction Horizon. We refer to profiles obtained via simple prediction as SR_p, i.e. static risk of the Short-Term Predicted Glucose, since we know that whenever the predicted signal crosses the threshold, its SR will also cross the threshold.

2. Dynamic Risk per se: in this case, parameters α and δ will be set. This signal will be referred to as DR, i.e. the Dynamic Risk of the (Original) Glucose Profile.

3. Combination of Short-term prediction and DR evaluation: in this case we considered several combinations of the parameters of the single algorithms. Signals obtained with this combination will be referred to as DR_p, i.e. the Dynamic Risk of the Short-Term Predicted Glucose profile.

4. All the above signals will be compared to SR, i.e. the static risk of the original glucose profile which, we recall, crosses the threshold exactly at the same moment as the original glucose profile.

6.2 Criteria for Assessing the Results

The evaluation of the algorithms will be performed for hypoglycemic threshold crossings. The temporal gain in threshold crossing is evaluated as the time difference between the crossing of the predictive signal at 70 mg/dl and the subsequent crossing of the static risk, which crosses the threshold simultaneously with the CGM signal. This allows comparing the condition of no prediction (alert raised when hypoglycaemia has already occurred) and the preventive alert generated by the algorithms.

The number of false positives will also be evaluated. We consider a false alarm an episode where an alert is raised by the predictive algorithm (i.e. the predicted profile crosses the 70 mg/dl threshold or the risk function falls below the static risk value corresponding to 70 mg/dl) while the CGM will not actually fall below 70 mg/dl. Moreover a snooze time of 45 minutes is set which prevents the generation of alerts right after the first alert.

6.3 Simulation Study

6.3.1 Data Generation

We consider 10 simulated signals of Section 5.2. Two simulations were run. First we used the noise-free signals and tested the method in optimal conditions. Then we added zero-mean white noise with different variance σ^2. The signals were undersampled to a 5 minutes rate grid, to simulate, for example, the performances on a device like the Dexcom Seven Plus. The average length of the signals is 5653.5 ± 60.2 min. The total number of true hypoglycemic

events is 86, and the number of alerts raised by the noise-free simulated profiles alone is, in fact, 86.

Simulations were run for every possible combination of parameters where

- $PH = \{5, 10, 15, 20\}$ minutes
- δ set to have maximum amplification of $\{1, 2, 3, 4, 5, 10, \cdots, 60\}$
- $\alpha = 0.6, 0.8, 1, \cdots, 5.4$

for a total number of 1600 combinations of parameters.

6.3.2 Results on Noise Free Signals

The performance of the prediction algorithm based on KF alone is independent of the value of α and δ and is useful to understand what is the starting point to develop the combined algorithm KF plus DR.

Figures 6.1 shows a representative simulated noise free glucose profile vs the predictions obtained via KF with PH= 5, 10, 15, 20 and 25 minutes. Notice that greater PH result in less stable and more oscillating predictions.

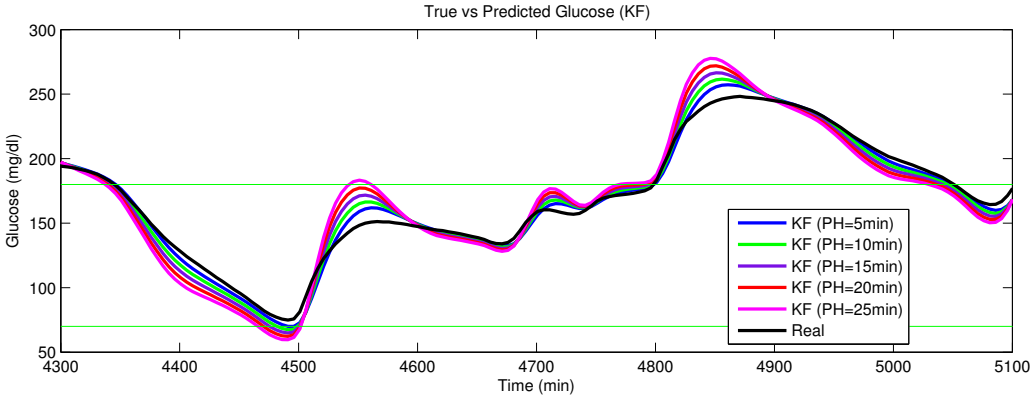

Figure 6.1: Example of simulated glucose (black) vs predicted glucose via Kalman Filter with PH=5, 10, 15, 20, 25 minutes.

Table 6.1 shows the temporal gain obtained on noise free profiles by using the KF predictive algorithm alone.

The role of DR alone as an alert generator has been shown previously in this chapter. In this particular setting, we used the structure based on the tanh function, which depends on four parameters:

Noise-Free Data				
PH	Gain (mean)	Gain (median)	FP(mean)	FP(median)
5	7.15	5	0.9	1
10	12.26	10	1.7	1.5
15	16.57	15	2.6	3
20	19.59	20	3.5	3.5
25	22.15	20	4.1	4

Table 6.1: Performance of the short-term prediction algorithm alone in terms of temporal gain in anticipation of threshold crossing and false alerts generated for noise free profiles. Results are summarized with the mean and median of the distribution of temporal gains and # of false positives for different prediction horizions (PH).

- α, determines the effective amplification.
- β determines the maximum damping. It was arbitrarily set in order to have maximum damping of 75% since this allows to have a smooth recovery to the zero risk zone when heading back to euglycemia.
- γ as shown before can be computed once the other parameters are set.
- δ defines when a saturation of the risk will occur. δ was set in order to have maximum amplification of 10. In this chatpter, δ set to 10 means that δ was set in order to obtain maximum amplification factor equals to ten.

The role of α can be appreciated in Figure 6.2, where SR for a representative simulated noise free glucose profile, against DR_{tanh} evaluated for parameter α equal to 1.5, 2 and 2.5. Greater α result in greater time gain.

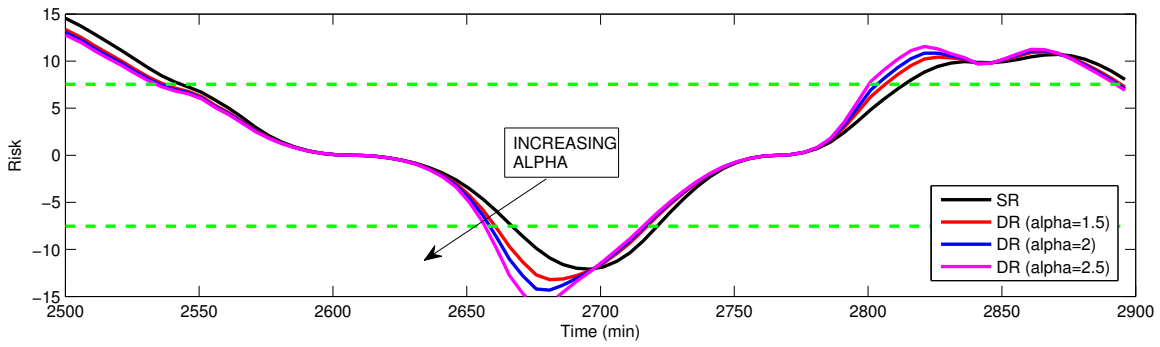

Figure 6.2: Example of SR (black) vs DR_{tanh} for α=1.5, 2, 2.5.

The role of the maximum amplification is investigated in figure 6.3 where the mean temporal gain and number of false alerts is shown for the simulated noise free profiles elaborated with DR alone for different values of α. Different values for the amplification factor were considered from 1 to 60. Basically, higher δ results in higher temporal gain, though the increase in the gain itself seems to be mainly driven by the value of α. In particular, for amplification factors ≥ 10 the family of curves collapses and the only increase in temporal gain is given by the choice of α. For this reason, in the following discussion, the maximum amplification factor will be always considered fixed and equal to 10, and the tuning of the DR component will be mainly focused in setting the best α. Only for the summary we will look for the best combinations of parameters where a different δ could provide slightly better performances.

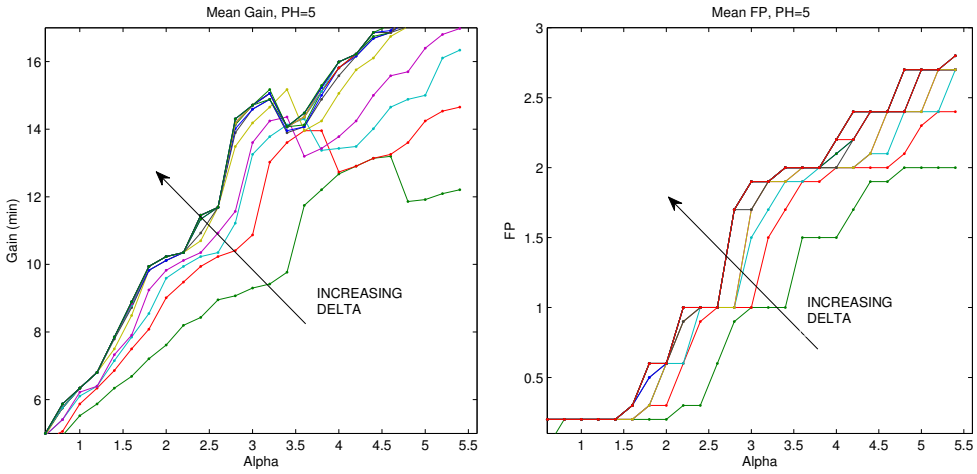

Figure 6.3: Family of curves describing the temporal gain and number of false alerts generated by DR alone as a function of α. Each curve represents the DR evaluated for a different value of δ. Notice that no significant increase is due to a specific choice of δ and the choice of α seems most crucial.

The combination of DR and the prediction algorithm based on KF can combine the nice features of the two tools. Fig 6.4 shows an example of profile obtained with combination of DR_G ($\alpha = 2$, max amplification = 10) and KF (PH=10 min). In particular, in the upper panel, a portion of the original glucose signal (blue) is displayed along with the prediction obtained via KF with PH equal to 10 minutes (red). The green lines represent the 70 and 180 mg/dl thresholds. In the lower panel we report all the predictions transformed in the risk space. SR (blue) represents the static risk of the original glucose profile, and crosses the green thresholds exactly at the same time as the original signal crosses the original thresholds.

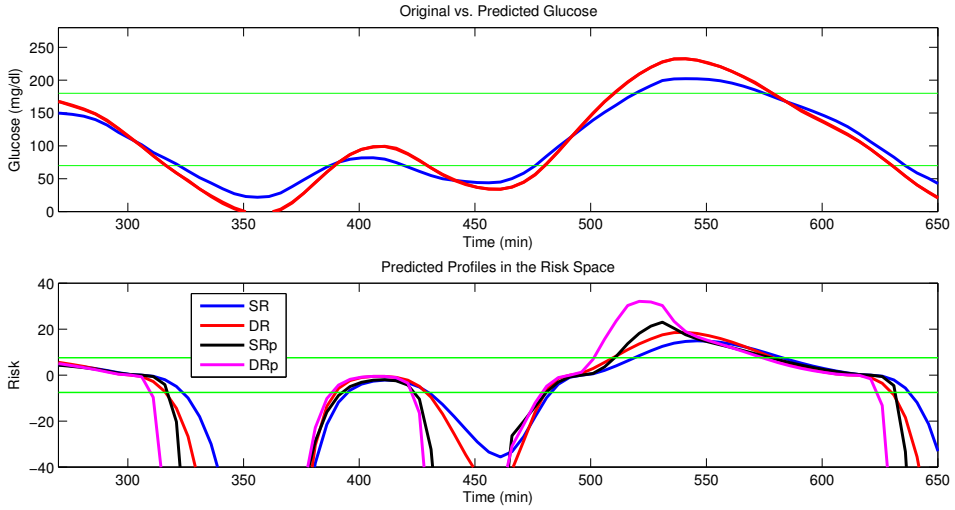

Figure 6.4: Example on simulated, noise-free signals. Top panel: Original and predicted glucose (KF with PH=10). Lower panel: SR (of the original glucose), DR (of original glucose, implemented with DR_{tanh}), DRp (of the predicted glucose).

SR_p (black) is the static risk of the predicted profile, and crosses the thresholds at the same time as the predicted glucose profile crosses the original thresholds. The DR profile (DR, red) is the dynamic risk of the original glucose signal. The magenta profile (DRp) is the dynamic risk of the predicted profile. This profile anticipates the threshold crossings with much higher than the KF predicted and the DR profiles alone. In this section we will also compare the combination of DR and KF (PH=10) with the simple use of KF with longer PH (e.g. 20 minutes).

Different PH in the evaluation of KF result in different performances of the combined algorithm DRKF. Figure 6.5 we report the mean temporal gains and the mean number of false positive alerts obtained with the combination of KF with $PH = \{5, 10, 15, 20\}$ (blue, red, black and magenta profiles respectively) with DR as a function of parameter α ($\delta = 10$).

Qualitative analysis of Figure 6.5 indicates that higher α values result in high temporal gain and, as one would expect, higher number of false alerts. In order to better understand the relative role of DR and the predictive part of the algorithm, we first observe, from Figure 6.6, the distribution of temporal gains and of false alerts generated by DR alone (yellow), by prediction alone (salmon) and by the combination of the two (pink). For this plot, maximum amplification, α and PH are set to 10, 2.2 and 10(min) respectively. The distributions are skewed, and the median of threshold crossings could be a better measurement to assess

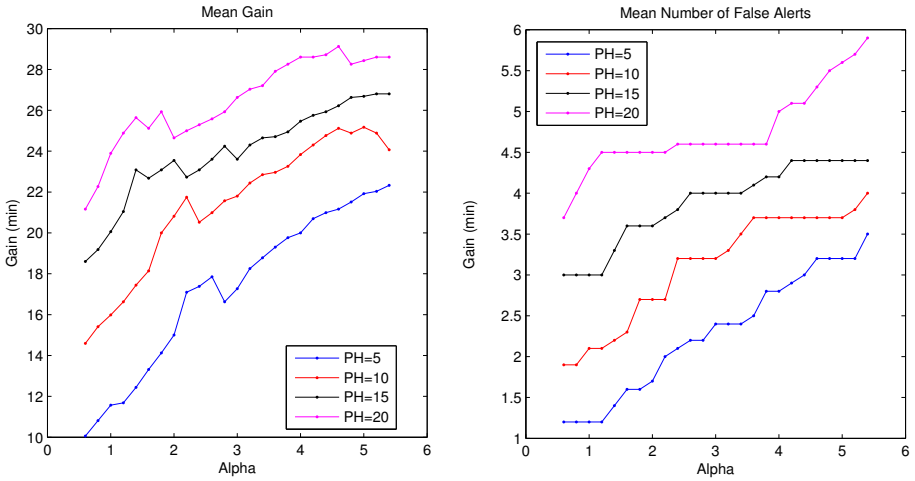

Figure 6.5: Mean Gain and # of FP for Noise free profiles: Combination of KF with $PH = \{5, 10, 15, 20\}$ (blue, red, black and magenta profiles respectively) with DR as a function of parameter α ($\delta = 10$)

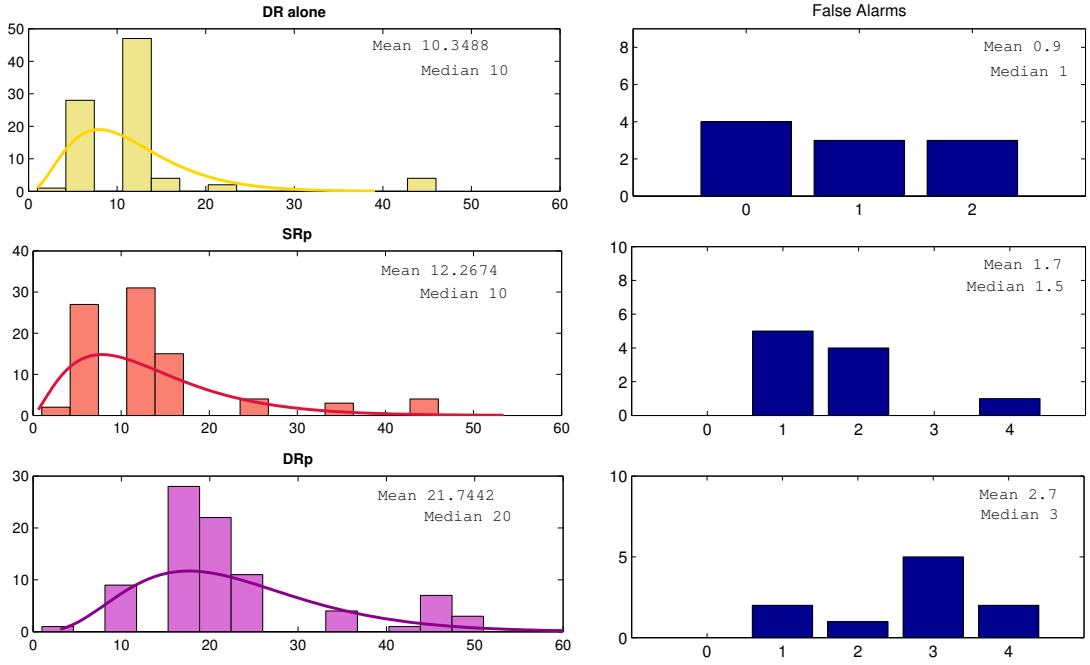

Figure 6.6: Distribution of time gains of the hypoglycemic alerts (left) and distribution of number of false alerts generated per patient (right). Results are shown for DR alone (top), Prediction alone (middle), combination of DR and KF (bottom). δ, α and PH are set to 10, 2.2 and 10(min), respectively.

the performances of the three tools. The combination of KF and DR allows shifting the distribution to the right, i.e. the average anticipation is increased.

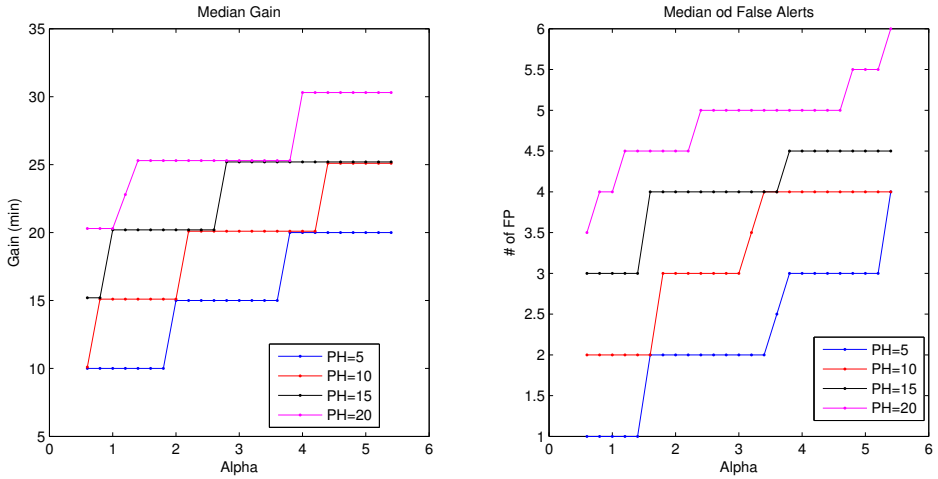

Figure 6.7: Median Gain and # of FP for Noise free profiles: Combination of KF with $PH = \{5, 10, 15, 20\}$ (blue, red, black and magenta profiles respectively) with DR as a function of parameter α ($\delta = 10$)

Figure 6.7 reports the median gain and number of false alerts for different α and PH.

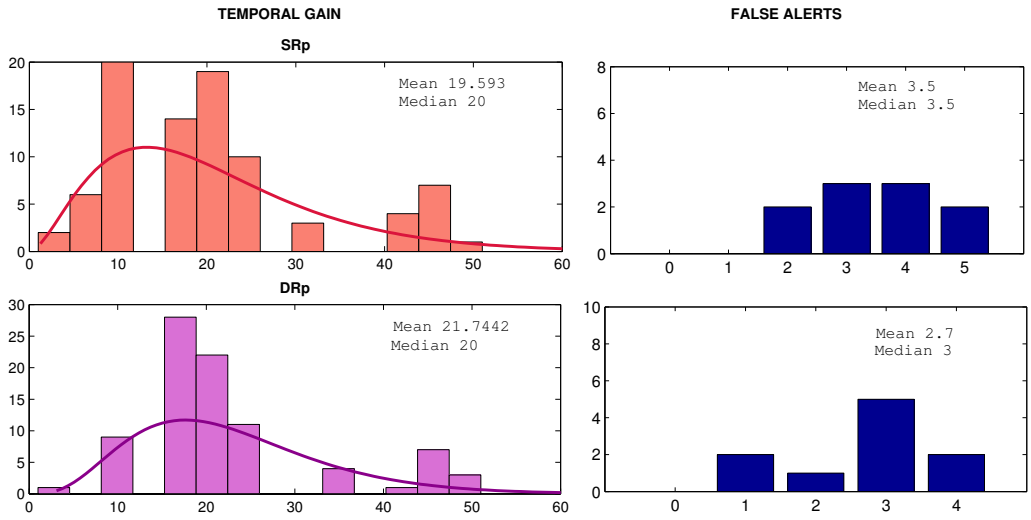

Figure 6.8: Distribution of time gains of the hypoglycemic alerts (left) and distribution of number of false alerts generated per patient (right). Results are shown for prediction alone with PH=20 (top), combination of DR and KF with δ, α and PH equal 10, 2.2 and 10(min), respectively (bottom).

It is interesting to better understand the relative role of the predictive part of the algorithm and of the DR. Figure 6.8 shows in the top panels the distribution of temporal gains obtained with a mid-term prediction (PH=20 minutes, KF alone) and in the lower panels the results obtained with the combination of DR with a short term prediction with PH=10 minutes. Notice that the combined effect of DR and KF yields better results than a longer PH by itself: despite the same median anticipation (20 minutes) in threshold crossings, the combined algorithm results in a lower number of false alerts both in mean and median.

6.3.3 Results on Noisy Signals

The same tests of section 6.3.2 were run on simulated signals, with addition of white zero-mean gaussian noise with variance $4mg^2/dl^2$ and $16mg^2/dl^2$. Figure 6.9 shows the same portion of data of Figure 6.4, where noise has been added. From the estimated profiles, no significant degradation of the performances of the combined algorithm seems visible with respect to the noise-free condition.

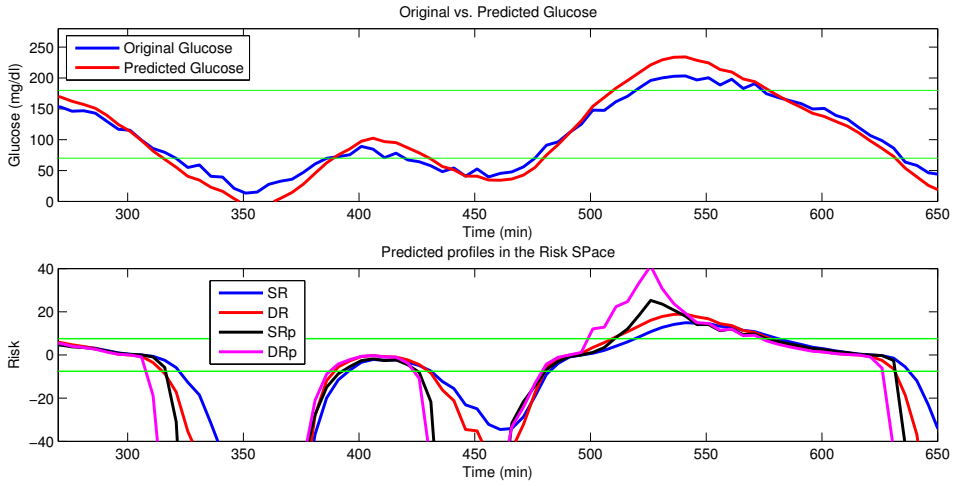

Figure 6.9: Example on simulated noisy signals ($\sigma^2 = 16$). Top panel: Original and predicted glucose (KF with PH=10). Lower panel: SR (of the original glucose), DR (of original glucose, implemented with DR_{tanh}), DRp (of the predicted glucose).

Table 6.2 shows the mean and median temporal gain and number of false alerts for the KF prediction algorithm alone for different PH. Results are comparable with the noise-free condition in terms of temporal gain. This is probably due also to the fact that the snooze time allows to raise an alert at the first threshold crossing and then silences all other alerts for

45 minutes. Notice that this is a very favorable condition. Indeed if two threshold crossings of the predicted profile were present before a true crossing of the reference, in absence of snooze time the first would be treated as a false alert, and only the second would be treated as a true alert (with shorter gain).

In any case, the number of false alerts is increased, as expected, for noisy profiles.

Noisy Data ($\sigma^2 = 16$)				
PH	Gain (mean)	Gain (median)	FP(mean)	FP(median)
5	11.3	5	1.3	1.5
10	15	10	2.2	2
15	18.95	15	3	3
20	22.35	20	4.1	3.5
25	23.90	20	4.9	4.5

Table 6.2: Performance of the short-term prediction algorithm alone in terms of temporal gain in anticipation of threshold crossing and false alerts generated for noisy profiles. Results are summarized with the mean and median of the distribution of temporal gains and # of false positives for different prediction horizions (PH).

Also for simulated noisy signals we summarize in Figures 6.10 and 6.11 the mean and median temporal gains as a function of the only parameter α. Different prediction horizons are displayed in different colors: $PH = \{5, 10, 15, 20\}$ (blue, red, black and magenta profiles respectively).

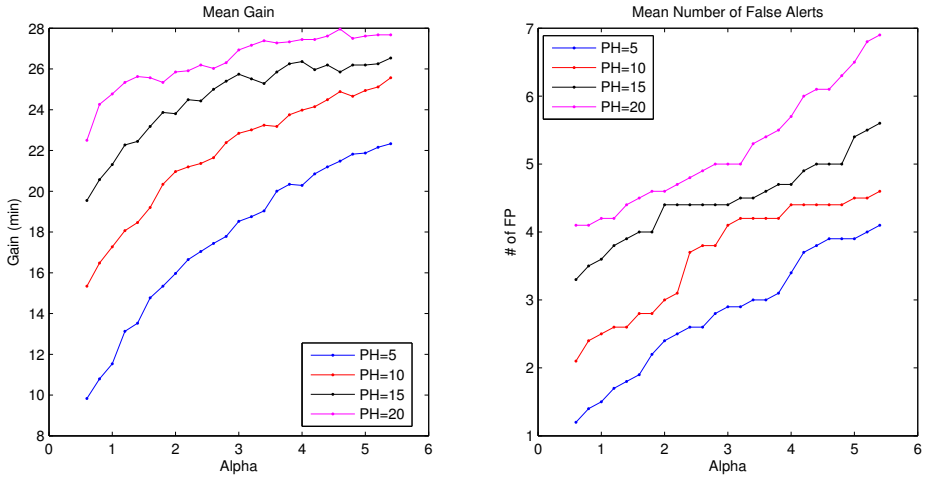

Figure 6.10: Mean Gain and # of FP for Noisy profiles: Combination of KF with $PH = \{5, 10, 15, 20\}$ (blue, red, black and magenta profiles respectively) with DR as a function of parameter α ($\delta = 10$)

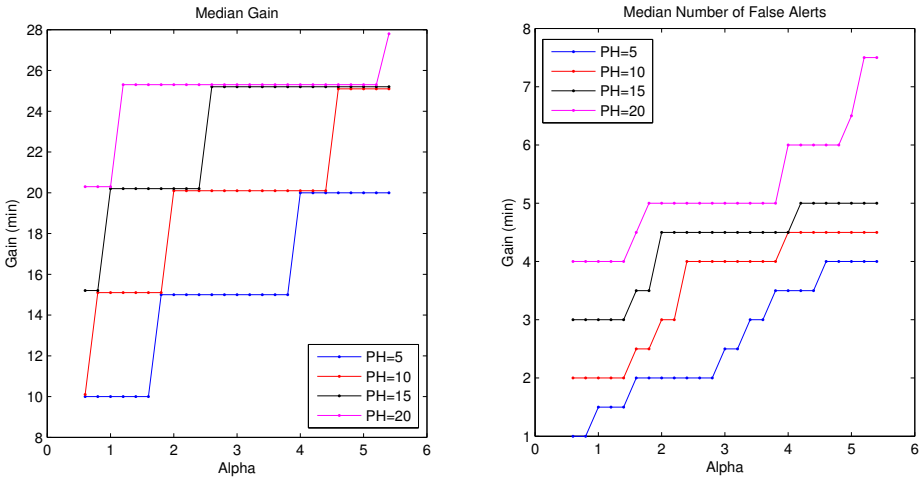

Figure 6.11: Median Gain and # of FP for Noisy profiles: Combination of KF with $PH = \{5, 10, 15, 20\}$ (blue, red, black and magenta profiles respectively) with DR as a function of parameter α ($\delta = 10$)

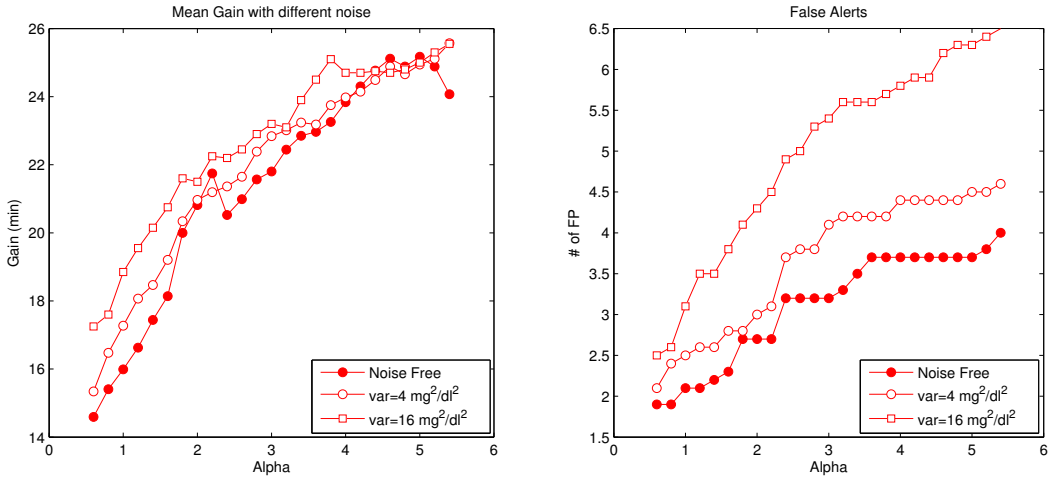

Figure 6.12: Performances with different noise levels: $\sigma^2 = 0mg^2/dl^2$ (full dots), $\sigma^2 = 4mg^2/dl^2$ (white dots) and $\sigma^2 = 16mg^2/dl^2$ (white squares). Left panel displays the mean temporal gain in the threshold crossings. Right panel shows the mean number of false alerts generated by the combination of Short Term Prediction (PH=10 min) and DR ($\delta = 10$) when α varies.

In order to better investigate the role of noise in the evaluation of the algorithm, consider Figure 6.12, where the mean temporal gains and number of false alerts are shown for a $PH = 10$ minutes and $\delta = 10$; different level of noise are represented as follows: $\sigma^2 = 0mg^2/dl^2$ (full dots), $\sigma^2 = 4mg^2/dl^2$ (white dots) and $\sigma^2 = 16mg^2/dl^2$ (white squares).

6.3.4 Lessons from Simulated Data

Simulations demonstrate that the combined use of DR on predicted profiles obtained, for example, with KF iterating the predictive step to 10 minutes, is an efficient tool to generate alerts. Table 6.3 summarizes some performances obtained by the three algorithms for a comparison in case of noise-free data. The main interesting features of the new combined algorithm can be summarized by stating that:

- DRKF performs better than the single algorithms used as standalone tools (if parameters are not adjusted), since the amplification of prediction allows increasing the time gain. This results of course in a higher number of False Alerts.
- Considering a fixed time gain of 20 minutes circa, DRKF allows lowering the mean and median number of false alerts with respect mid-term prediction by KF (PH=20) minutes (compare lines 2 and 5 of Table 6.3).
- Allowing a higher median in the number of false alerts, specific combinations of parameters result in a temporal gain 5 minutes higher than the standard mid-term prediction.

Algorithm	Gain		FP	
	mean	median	mean	median
Noise free				
KF (PH=10)	12.26	10	1.7	1.5
KF (PH=20)	19.59	20	3.5	3.5
DR ($\alpha = 2.4$, $\delta = 20$)	14.6	10	2.5	2
DR ($\alpha = 5$, $\delta = 10$)	17.2	15	2.7	2.5
DRKF ($\alpha = 2.2$, $\delta = 35$, PH=10)	21.92	15	2.7	3
DRKF ($\alpha = 5$, $\delta = 10$ PH=10)	25.17	25	3.7	4

Table 6.3: Performance of the three algorithms in terms of temporal gain in anticipation of threshold crossing and false alerts generated for noisy profiles. Results are summarized with the mean and median of the distribution of temporal gains and # of false positives.

Results obtained with this new technique are promising. Different short-term predictors could also be used, especially if are able to provide a relatively stable (non oscillating) prediction. Future work will be focused on evaluating possible different prediction strategies and on an ad-hoc tuning of the DR parameters for the specific predictor.

6.4 Real Data Study

6.4.1 Data

The combined algorithm of short term prediction via Kalman filter associated with DR_{tanh} was applied to 10 real CGM signals collected with the Dexcom Seven Plus sensor on diabetic subjects. Patients were following their therapies in their normal life. Average length of the monitorings was 9362± 739 minutes.

6.4.2 Results

Alert generation was assessed with the same criteria seen in the previous Sections, with snooze time set at 45 minutes. Parameters for DR_{tanh} were δ =10, and α =2.5 and Figure 6.13 displays the results on a representative subject: top panel shows the original vs predicted glucose via KF (blue circles and red line respectively) whereas the lower panel shows the computed risk signals (SR, SRp, DR and DRp in blue, black red and magenta respectively).

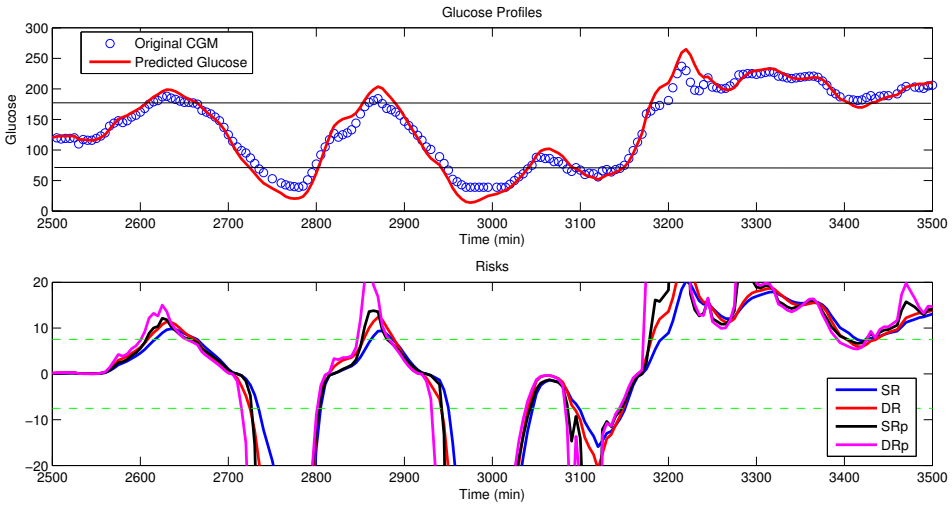

Figure 6.13: Top panel: representative glucose profile (blue circles) and predicted glucose via KF (red line). Bottom panels: Static Risk of the original and predicted glucose (blue and black respectively) and Dynamic Risk of the original and predicted glucose (red and magenta respectively).

Figure 6.14 shows the histograms with distribution of time gains (left panels) and false positives (right panels) for the three algorithms: DR alone (top), KF alone (middle) and the combination of the two (bottom). The anticipation in threshold crossings provided by

the three predictive risk signals is good, with the combined algorithm being better than the other two. On average, DR alone achieves a mean temporal gain of 13.34 minutes (median 10 minutes) with mean number of false alerts of 2.9 (median 3). Results for SRp in terms of temporal gain show a mean of 12.80 minutes (median 10 minutes) with 3.5 average false positives per day (median 3.5). Comparison of the two standalone tools on real data highlights a slight better performance of DR with respect to KF both in terms of temporal gain and of false alerts. The combination of the two algorithms results in a relatively high temporal gain (19,36 minutes on average, with median value of 15 minutes). Although the number of false positives seems rather high, it is important to recall that the real signals considered for this analysis are long (around week on average), so we can see that on average only one false alert per day is raised.

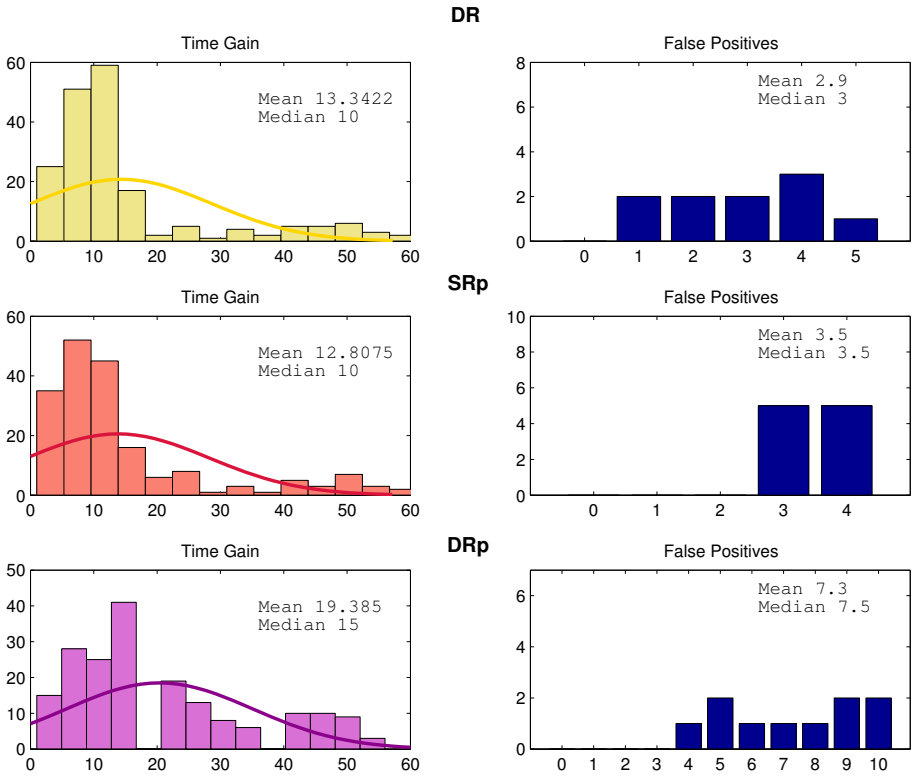

Figure 6.14: Distribution of time gains of the hypoglycemic alerts (left) and distribution of number of false alerts generated per patient (right). Results are shown for DR alone with (top), Prediction alone (middle), combination of DR and KF (bottom). Maximum amplification, α and PH are set to 10, 2.2 and 10(min), respectively.

6.5 Conclusions

The Dynamic Risk can be employed for the generation of preventive alerts. The possible use of DR as a generator of alerts per se has been published in (24), while preliminary results on the combination of DR with short-term prediction have been published in (23). The results shown in this Chapter were obtained by exploiting a rather simple algorithm for the short-term prediction of glucose. Any other predictive algorithm can be used in combination with DR. In fact, DR can be simply computed from any glucose signal, and only enhances the anticipation of threshold crossings in case of high first time derivative. Future work will be focused on the combination of DR with other prediction algorithms and on more extensive comparison of the performances of the different techniques. Moreover, a provisional US Patent has been filed concerning the possible use of DR for the generation of hypo and hyperglycemic alerts and the use of a DRS based color map to highlight the monitors of CGM devices when the patient is approaching a risky condition (22).

6. PREVENTION OF HYPO/HYPERGLYCEMIC EVENTS THROUGH ON-LINE CALCULATION OF DR: FURTHER IMPROVEMENT BY USE IN COMBINATION WITH KALMAN FILTER

Part IV

Use of Dynamic Risk for the (Parsimonious) Description of Glucose Variability

7

Possible use of DR in the assessment of Glucose Variability and definition of new indexes

In Chapter 2 the concepts of glucose variability and glucose control were introduced along with some of the most used metrics that are employed to quantify them. The main observation that emerged is the fact that none of the proposed indexes explicitly considers the glucose time derivative. This is probably due to the fact that CGM devices, which allow the estimation of the time-derivative thanks to their almost continuous nature, have only become available in the last decade. In this chapter we report some examples (sometimes conceptual) to highlight the importance of quantifying the glucose rate of change in the evaluation of the efficiency of glucose therapy. Moreover we will introduce some indexes developed on the Dynamic Risk Space (DRS) introduced in 3.4 to describe the glucose signals from the risk space perspective.

7.1 Added Value of the Time-Derivative in assessing Glucose Variability

7.1.1 Simulated Data (Conceptual Examples)

Figure 7.1 shows three sinusoids, S1, S2 and S3, with same frequency, same mean, and different amplitude (and thus different standard deviation). If, conceptually, we pretend that S1, S2 and S3 represent glucose, we note, all the sinusoids can be built to stay perfectly "in target". In this condition, all the indexes based only on mean, percentage in target, risk scores (Eq. 2.4, 2.6 in Chapter 2), cannot distinguish between the three signals. Nonetheless, it could be important to do so, since greater amplitude in the sinusoid (blue signal, S2) is a condition similar to unstable glycemia, and hence implies a higher difficulty in tuning the

therapy. In this case, standard deviation measures (such as those defined in 2.2.1) in this case succeed in characterizing these signals. Also, combined indexes of mean and standard deviation like J (Eq. 2.3), CV and M_R (Eqs. 2.10) could be useful in this condition.

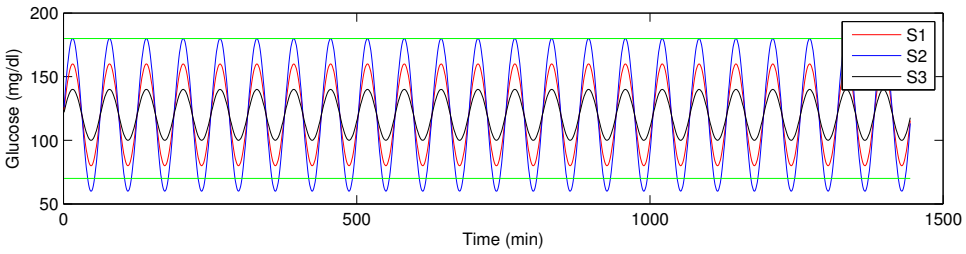

Figure 7.1: Example of three signals S1, S2 and S3 with same mean, same % in target very similar risk scores. SD in this case is able to distinguish between the 3 signals.

An example of signals where standard deviation measures fail in characterizing different glucose behaviors is illustrated in Fig. 7.2 where three sinusoids with same amplitude/SD, same frequency, and different mean are shown. In this case, SD-based measures fail in characterizing these behaviors, while mean and mean-SD combination based indexes are able to distinguish between the three signals. It is important to notice that SD is not able to distinguish between the red signal (high hypo risk) and the black signal (high hyper risk).

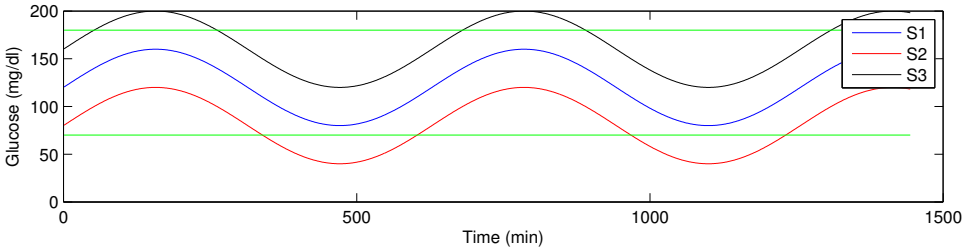

Figure 7.2: Example of three conceptual glucose signals S1, S2 and S3 with same SD but different means.

Figure 7.3 shows another example of three sinusoids with the same % in target and same frequency. The red and black signals have small amplitude (SD) with different means (the first is shifted towards hypo, the second towards hyperglycemia). The blue signal has greater amplitude and the risk of falling to hypo as the red one and the same risk of hyper as the black one. In this case the use of the maximum and minimum of the time-series (max and min) to characterize the signals would be suboptimal, since the signals cannot be distinguished by the use of only one of these two metrics.

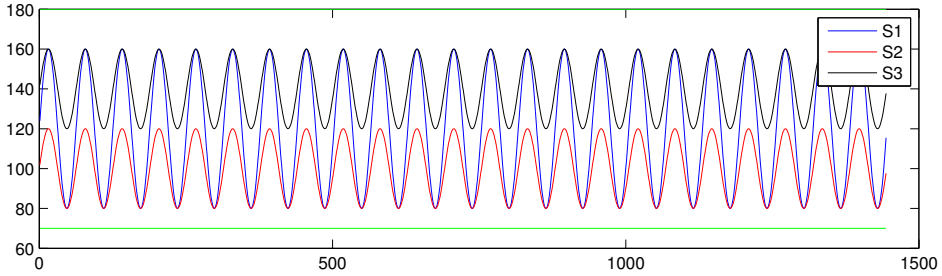

Figure 7.3: Example of three conceptual glucose signals S1, S2 and S3 with same % in target, same SD (black and red), same minimum (blue and red) and same maximum (blue and black).

Finally, we report in Figure 7.3 an insightful example, with three sinusoids with same mean, same max and min, same amplitude but different frequency. In this case, mean, SD, combinations of mean and SD, Time in target, max, min would fail in distinguishing the three sinusoids. Moreover a slight difference between penalty scores can be evaluated due to the different time spent at different levels (the distribution of glucose values is broader for the signal at low frequency, and more spiky for the faster signal). If these were true glycaemic signals, we could assume that the slower signal (blue sinusoid) represents a subject which is more easily controlled, since there are no abrupt changes in the signal, while the red signal is symptomatic of a glycaemia that is frequently corrected and very oscillating. This can be due to patient's characteristics, or to a very aggressive therapy which needs frequent corrections for preventing both hypos and hypers. This is of course a conceptual example, but can highlight the benefit that one would have by explicitly including the information of the time derivative in the variability and control analysis. This could be achieved by defining new indexes from DR, as done below in Section 7.2.

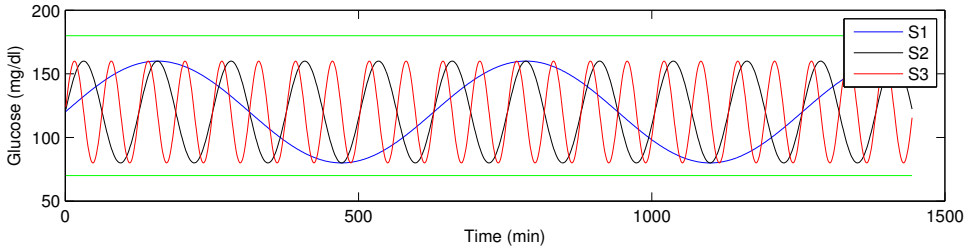

Figure 7.4: Example of three conceptual glucose signals S1, S2 and S3 with the same % in target, same mean, same SD but different frequency and hence different distribution of time derivatives.

7.1.2 Examples from Real Data

Just as shown for the conceptual examples of section 7.1.1, real glucose signals monitored with the CGM devices have different characteristics in terms of variability and control. Figures 7.6 and 7.5 show two representative glucose profiles (monitored for three days with the Abbott Freestyle Navigator device (1)). The behavior of the two subjects is comparable in terms of mean and amplitude of oscillations, but in the first case (Fig. 7.6) the patient #1 has relatively fast glucose excursions and, possibly, more corrective actions, while in the second case glucose in patient #2 is oscillating less rapidly. Indexes based on DR may help the differentiation between Subject #1 and Subject #2.

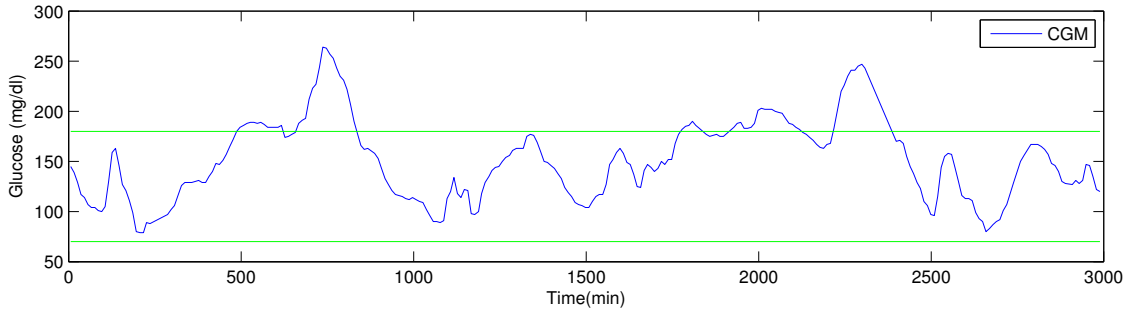

Figure 7.5: Glucose profile (three days of monitoring, Abbott Freestyle Navigator) of a representative subject.

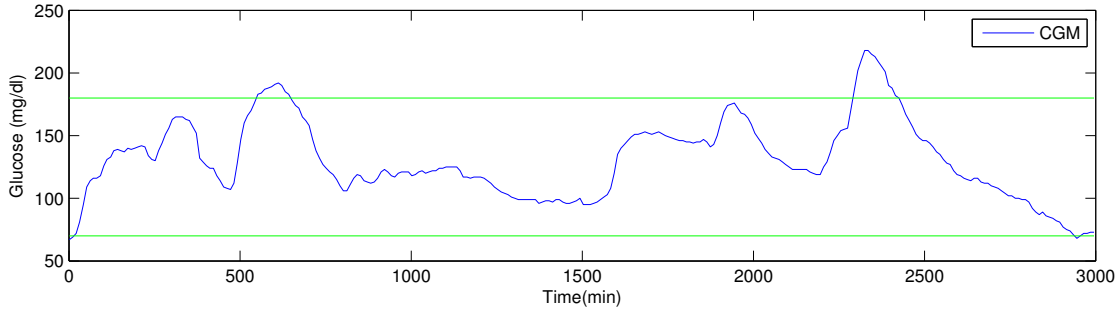

Figure 7.6: Glucose profile (three days of monitoring, Abbott Freestyle Navigator) of a representative subject.

By considering another pair of representative subjects, monitored via frequently sampled blood glucose (5 minutes rate (21), comparable to a CGM profile) and displayed in Figure 7.7 (top panels), it is interesting to see how glucose trajectories can be displayed in the DRS.

DRS (lower panels), where the color scale has been changed in order to have safety regions displayed in green and dangerous regions displayed in red.

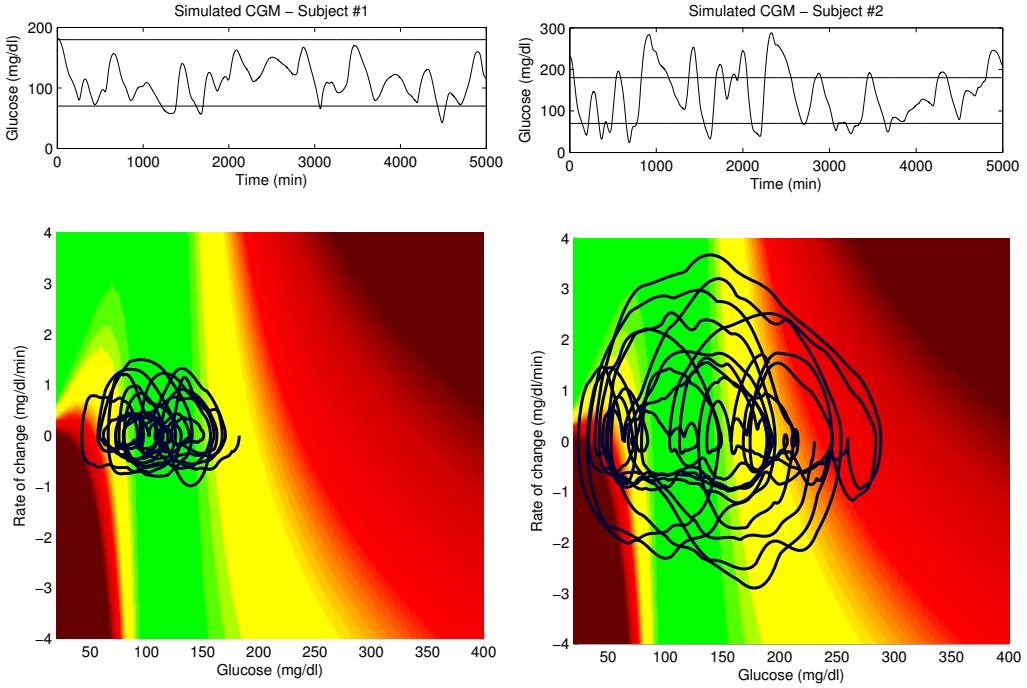

Figure 7.7: Two representative glucose profiles (top panels) in the Dynamic Risk Space (lower panels).

The left panels refer to a very well controlled patient (#1), while the panels on the right refer to a patient (#2) who often experiences hypo and hyperglycemic events. From the upper panels the difference between the two subjects is visible, although it does not appear to be substantial. On the other hand, if one analyses the trajectories of these signals in the DRS it is apparent how in the first case the trajectory is compact within the safety zone. Time derivative range is narrow, with few excursions with time derivative higher than $\pm 2 mg/dl$. In the second case (right panels), the trajectory is spread on the DRS both through the glucose axis and on the time-derivative axis.

The display of glucose trajectories in the DRS gives an immediate idea of the tightness of glucose control of a specific patient (particularly efficient when adding the colored scale of Fig. 7.7). It is also interesting to add value to this kind of analysis by developing new indexes able to capture the characteristics of the trajectories in the DRS. In this thesis we

show some possible indexes developed to capture such features. Being this a new approach to the analysis of glucose signals, several other strategies may be employed to characterize glucose variability from the DRS. We will focus on some general techniques that have been previously used in other biomedical engineering fields. For example, some ideas have been used in the posturography analysis, where the trajectory of the barycenter of a standing person (stabilogram) is tracked and analyzed from measures of pressure, see (49) for details.

7.2 New DR-based Indexes

In this section we will introduce two classes of indexes that may be used for the characterization of the signals in the DRS. The first class is a family of geometric measures that characterize the trajectory path and its barycenter, possibly weighted by the DR value of each point. The second class refers to frequency measures of the signals.

7.2.1 Geometric Measures

The first measures that we can derive on the trajectories are their barycenter and the mean distances of their points from it. In the following formulae, we will consider x and y to be the axis relative to the glucose level (g) and time derivative (dg/dt) respectively, while N will represent the total number of points in the trajectory.

- **RD(Resultant Distance)**: represents the time series of the distances of the points of the trajectories from its barycenter:

$$RD(n) = (x(n)^2 + y(n)^2)^{1/2} \text{ for } n = 1 \dots N \tag{7.1}$$

where

$$x(n) = (x_o(n) - \overline{x}) \tag{7.2}$$

$$\overline{x} = \frac{1}{N}\sum_{n=1}^{N} x_o(n) \tag{7.3}$$

$$y(n) = (y_o(n) - \overline{y}) \tag{7.4}$$

$$\overline{y} = \frac{1}{N}\sum_{n=1}^{N} y_o(n) \tag{7.5}$$

In Eq. 7.2, $x_o(n)$ is the time-serie of the distances of each point from the center (origin) while $x(n)$ is the time-series of the distances from the barycenter of the data along the x axis (similarly for $y_o(n)$ and $y(n)$ along the y-axis).

- **MDIST (Mean Distance)**: It is the mean value of the RD time-series and it represents the average distance of the points of the trajectories from the barycenter of the trajectory itself.

$$MDIST = \frac{1}{N}\sum_{n=1}^{N} RD(n) \tag{7.6}$$

- **RDIST (Root Mean Squared Distance)**: It is the root mean squared value of the RD time-series and it represents the average distance of the points of the trajectories from the barycenter.

$$RDIST = \left[\frac{1}{N}\sum_{n=1}^{N} RD(n)^2\right]^{1/2} \tag{7.7}$$

- **RDIST$_x$ (Root Mean Squared Distance along x (or y) axis)** : it represents the root mean squared distance of the distribution of x coordinate of the points in the trajectory from the x-coordinate of the barycenter.

$$RDIST_x = s_x = \left[\frac{1}{N}\sum_{n=1}^{N} x(n)^2\right]^{1/2} \tag{7.8}$$

- **SP (Sway path)**: total length of the trajectory, approximated by the sum of the distances of two consecutive points.

$$SP = \sum_{n=1}^{N-1} \left[(x(n+1) - x(n))^2 + (y(n+1) - y(n))^2\right]^{1/2} \tag{7.9}$$

- **SP$_x$ (Sway Path along x (or y) axis)**

$$SP_x = \sum_{n=1}^{N-1} |(x(n+1) - x(n))| \tag{7.10}$$

7.2.2 Ellipse-based Measures

The trajectories of glucose signal in the DRS often resemble an ellipse. In particular it is possible to define the 95% confidence ellipse as described in (49) for the trajectories of the barycenter of standing people. This confidence ellipse represents the curve containing with probability 95% the points of a bidimensional distribution, given in our specific condition by the couples of points (g and $\frac{dg}{dt}$). An example of such an ellipse is shown in Figure 7.8. In this graph, the DRS is reported with the glucose trajectory (black) displayed in the lower panel of the figure. The signal is one representative subject from the simulated dataset described in 5.2. In the upper panel, the blue ellipse comprises the area where 95% of the glucose values is expected to fall for this specific subject, while the blue whiskers represent the hypoglycemic and hyperglycemic centers, i.e. the center of distributions of values in hypoglycemia and hyperglycemia respectively.

The following indexes can be defined by using the 95% confidence ellipse:

AREA-CE Area of the ellipse that with probability of 95% includes the trajectory of the distances from the barycenter, assuming that the distances are normally distributed.

The 95% confidence ellipse is the 95% bivariate confidence ellipse, which is expected to enclose approximatively 95% of the points of the trajectory. Assuming that the number of points in the trajectory is large (i.e. $(n-1)/(n-2) \approx 1$), the major a radius and minor b radius of the 95% confidence ellipse can be computed as follows,

$$\begin{aligned} a &= [F_{.05[2,n-2]}(\sigma_x^2 + \sigma_y^2 + D)]^{1/2} \\ b &= [F_{.05[2,n-2]}(\sigma_x^2 + \sigma_y^2 - D)]^{1/2} \end{aligned} \tag{7.11}$$

In Eq.7.11, σ_x and σy are the standard deviations along the x and y axis and σ_{xy} is their covariance:

$$\sigma_{xy} = \frac{1}{N}\sum_{n=1}^{N} x(n)y(n) \tag{7.12}$$

Moreover, F is the F statistic at 95% confidence level for a bivariate distribution with n data points. For a large sample size it holds:

$$F_{.05[2,\infty]} \sim 3 \text{ if n points } >120 \tag{7.13}$$

D is defined as:

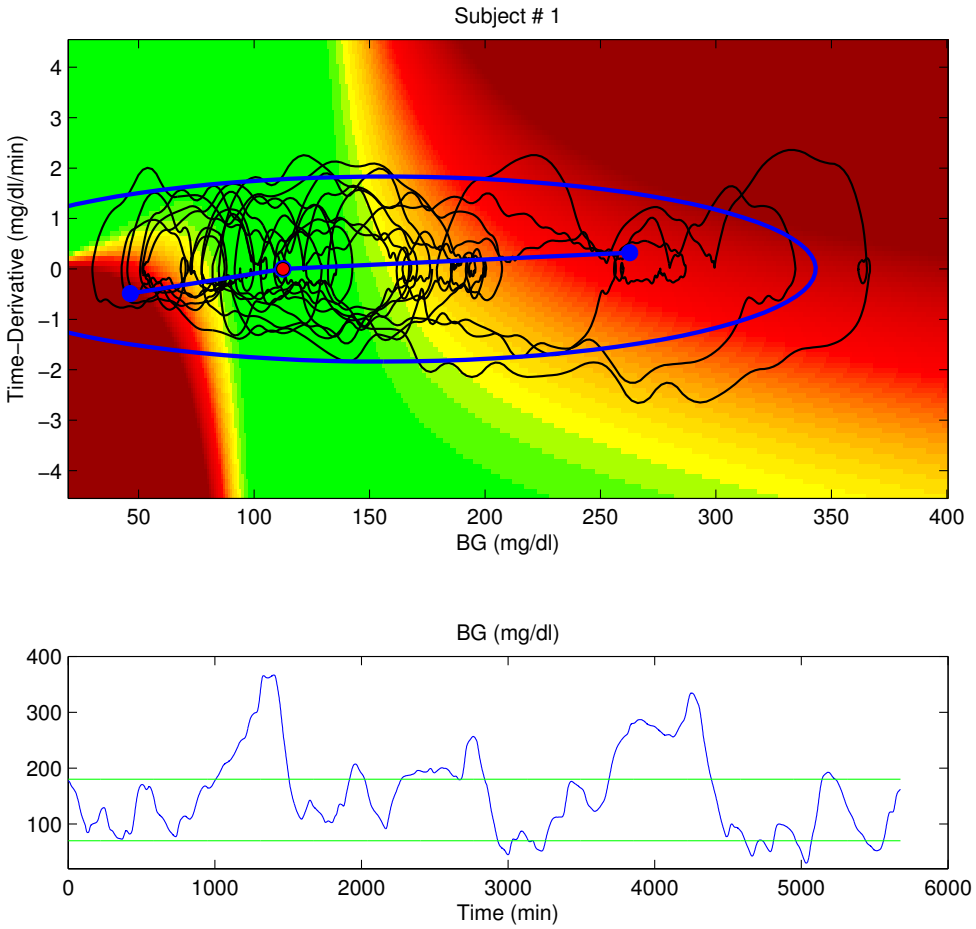

Figure 7.8: A representative glucose profile (lower panel panels) in the Dynamic Risk Space (upper panels). In blue, the 95% confidence ellipse is highlighted.

$$D = [(\sigma_x^2 + \sigma_y^2) - 4(\sigma_x^2\sigma_y^2 - \sigma_{xy}^2)] \tag{7.14}$$

Finally, the area of the ellipse can be evaluated as:

$$Area = \pi ab = 2\pi F_{.5[2,n-2]}[\sigma_x^2\sigma_y^2 - \sigma_{xy}^2]^{1/2} \tag{7.15}$$

AREA-CC is the area of the circle that with probability of 95% includes the trajectory of the distances from the barycenter, assuming that the distances are normally distributed:

$$AREA_{CC} = \pi(MDIST + z_{0.5}\sigma_{RD})^2 \tag{7.16}$$

where $z_{0.5}$, the statistic at the 95% confidence level is 1.645 and σ_{RD} is the standard deviation of the y time-series:

$$\sigma_{RD} = (RDIST^2 + MDIST^2)^{1/2} \tag{7.17}$$

It is possible to weight the barycenter of the trajectory, in order to center the ellipse according to the DR. In particular, one way to exploit the DR in this condition is using the following values for the coordinates of the center of the ellipse:

$$\begin{aligned} x_w &= \frac{\sum_{i=1}^{n} g(i)DR(i)}{n} \\ y_w &= \frac{\sum_{i=1}^{n} dg(i)DR(i)}{n} \end{aligned} \tag{7.18}$$

7.2.3 Frequency Analysis Measures

Another way to include information about the dynamics of the signals is to consider the frequency analysis. This kind of analysis is complementary to the time-domain analysis and helps highlighting specific features of the signals. Frequency analysis allows to implicitly take the time-derivative into account, since it provides information on the rapidity of the signals. In this thesis we considered specific measures on the glucose signal spectra obtained via Fast Fourier Transform (FFT). Denoting by $G(m)$ the power spectrum of the glycemic signal, we will compute following indexed:

- **TP (Total Power)**: Area under the curve $G(m)$:

$$TP = \sum_{m=1}^{N} G(m) \tag{7.19}$$

- f_{50} **(50% power frequency)**: Frequency under which is enclosed 50% of the power of the spectrum (median frequency):

$$f_{50} = \alpha\Delta f \tag{7.20}$$

  where α is the smallest integer for which

$$\textstyle\sum_{m=1}^{\alpha} G(m) \geq 0.5 \cdot TP \tag{7.21}$$

Similarly one can find f_{95} which is the frequency below which 95% of the total power is found.

To better understand and quantify the importance of time-derivative when assessing the signals behavior, Table 7.1 reports some of the indexes evaluated on CGM signals of the two representative subjects of Figures 7.5 and 7.6. Notice how the subjects present similar mean, SD and similar percentage of time spent in each region (hypo/hyper/euglycemia). Nonetheless the behavior assessed considering the whole 7-days monitoring highlights that the first patients presents faster oscillations than the second. This is summarized in the DR-based indexes reported in the shaded rows of the Table.

Index	Subj # 1	Subj # 2
SD_T	50.00	45.62
CV	34.42	30.27
$Mean_T$	145.27	150.70
Mr	12.19	12.32
J	38.13	38.54
igc	1.96	1.58
ADRR	43.92	35.48
GRADE	6.67	7.84
MAGE	79.89	77.89
% hypo	1.93	0.87
% hyper	15.97	20.05
% eu	82.10	79.08
min	50.00	65.00
max	285.00	272.00
RMS	9.22	8.34
path	1003.61	621.32
$path_{dg}$	54.46	27.82
Area EC	106.02	58.72

Table 7.1: Computed values for some of the proposed indexes evaluated on the two representative Subjects of Figures 7.5 and 7.6. The shaded rows highlight the indexes which are able to clearly highlight the different behavior of the two patients (rapidly vs slowly oscillating signals).

7.3 Next Steps

In this Chapter we introduced new DR-based indexes on the glucose profile which also consider the time-derivative of the signal. It is important to understand the role of such indexes with respect to the other indexes already introduced in Chapter 2 and taken from the literature.

7. POSSIBLE USE OF DR IN THE ASSESSMENT OF GLUCOSE VARIABILITY AND DEFINITION OF NEW INDEXES

In Chapter 8 the problem of selecting indexes or combination of indexes for the description of a specific dataset will be introduced.

8

Multivariate Analysis of Glucose variability Indexes via SparsePCA and Parsimonious Description

Several indexes can be used to describe different features of glucose signals. Some of them are available form the literature (Chapter 2), while some others have been defined in this thesis from the DR function (Chapter 7). Tables 8.1, 8.2 and 8.3 summarize those indexes divided in three classes, i.e. glucose variability, glucose control and DRS. Obviously it is expected that these indexes are, in part, correlated and redundant. It is therefore important to understand how many of them are indeed necessary to describe the features of glucose behavior in a given subject. In this Chapter we demonstrate that considering the time-derivative as a potential variability factor is important from a statistical point of view and that DR-based indexes add valuable information in describing and characterizing glucose behavior in different patients.

8.1 Problem Statement

Assume having a dataset of n glucose signals relative to n different subjects. For each signal evaluate m indexes and summarize such information in matrix X $(n \times m)$. We want to describe the variability of this set of indexes evaluated on the dataset, with fewer indexes p, $(p < m)$or combination of indexes. At the same time we want to find the most parsimonius combination of the p indexes able to capture the maximum possible variance of the m original indexes.

The PCA is a popular tool used to achieve the goal of understanding the real dimensionality of a dataset . This tool has been deeply studied and is consolidated in the statistical practice, so, for sake of brevity, an introduction to PCA is reported in Appendix D.

		Glucose Variability
SD/MEAN	Mean	Average of all available Glucose values
	SD_T	SD of all data from all days and all times of day
	SD_w	Evaluate SD for all measurements in each 24-h day, then average these values
	$SD_{hh:mm}$	Evaluate the average glucose profile by averaging the glucose value of the same time of day for all available days. Then evaluate SD of the average profile
	SD_{wsh}	SD for a fixed number of consecutive glucose values (e.g. 1 hours) evaluated at all possible times of day and then averaged
	SD_{dm}	Evaluate mean glucose for all days, then evaluate the SD of these values
	$SD_{bhh:mm}$	Evaluate the SD of glucose values for any specified time of day, then average these values
	$SD_{bhh:mm//dm}$	As above, considering deviations from the mean for the same day
	SD_I	Obtained using two-way analysis of variance with replication
	CV	$100 * \frac{SD_T}{mean_T}$
	M_R	$mean(1000 \cdot \lvert \log_{10}(Glucose/100)\rvert^3)$
Excursions	MAGE	Average Amplitude of upstrokes or downstrokes with magnitude greater that 1SD
Day-to-Day	MODD	Mean difference between glucose values obtained at the same time of day on two consecutive days under standardized conditions
	$MODD_d$	Mean of differences in glucose between each value and the value exactly $d \times 24$ hours later
Short-Term	CONGA	

Table 8.1: Literature Indexes for the definition of Glucose Variability

8.2 PCA as a Regression Problem

In order to be able to move further from the standard PCA approach, it is important to pose the PCA problem as a regression problem. This will be helpful to define new strategies for the definition of a new base with useful features for the reduction of the dimension of the

Glucose Control		
Target	% in target	70(80)-140(180) mg/dl
	% above target	≥140(or 180/200/250) mg/dl
	% below target	≤70(or 50/80) mg/dl
SD/MEAN	J	$.001 \times (mean_T + SD_T)^2$
	MAGE	As defined in Table
	SD_T	As defined in Table
Transformation of glucose scale	ICG	Sum of Hyperglycaemia and Hypoglycaemia Index
	Hyper Index	Weighted average of hyperglycaemic values
	Hypo Index	Weighted average of hypoglycaemic values
	ADRR	Average Daily Risk Range average sum of HBGI for maximum glucose and LBGI for minimum glucose for each day
	LBGI	Low Blood Glucose Index - Average of Kovatchev's risk values in hypo
	HBGI	High Blood Glucose Index - Average of Kovatchev's risk values in hyper
	GRADE	Average of transformed glucose values using Hill's transformation

Table 8.2: Literature Indexes for the definition of Glucose Control

DR and Frequency-based Indexes		
Areas	Ellipse Area	Weighted average of hyperglycaemic values
	Circle Area	Weighted average of hypoglycaemic values
Trajectory path	Sway Path	Total length of the path of the trajectory in the DRS
	Glucose path	Length of the path of the trajectory along the x-axis in the DRS
	Derivative Path	Length of the path of the trajectory along the y-axis in the DRS
Distance Measures	MDIST	Average distance of the points of the trajectory from its barycenter
	RDIST	Root mean squared distance of the points of the trajectory from its barycenter along x or y axis
	Barycenter	Weighted or non weighted coordinates of the barycenter of the trajectory

Table 8.3: Indexes based on the Dynamic Risk Space

original problem (68).

Considering the SVD computation of PCA, we know that in

$$X = U\Sigma V^T \tag{8.1}$$

matrix U represents the principal components of X, and V represents the loadings matrix. Simple manipulations of Eq 8.1 yield the following relation:

$$XV = U\Sigma \tag{8.2}$$

Denoting $U\Sigma = Y$ and $V = B$, we can consider the PCA problem as a regression problem. In fact, principal components are linear combination of the m variables whose realizations are reported in X, and the vectors y_i of matrix Y can be written as

$$y_i = X\beta \tag{8.3}$$

where β are the column vectors of B. For each PC p_i, it is possible to evaluate the loadings β that allow the transformation from X to Y. The problem of the estimation of β can be tackled in different ways, that will be explained in the following sections.

The easiest way to compute the loadings β, given matrix Y is to performa the linear regression of Eq. 8.4:

$$\hat{\beta} = \min_{\beta} ||y_i - X\beta||^2 \tag{8.4}$$

PCA as a regression problem indicates a major drawback. While it is possible to cut some of the components that do not carry information, it is true that the chosen principal components are linear combinations of all the vectors (indexes) of the original basis. This means that we can describe a dataset by combining the original indexes (m) with proper loadings. It is sometimes desirable not only to achieve the dimensionality reduction, but also to reduce the size of explicitly used variables. In the specific problem of dimensionality reduction for glucose variability, this means that only certain indexes contribute in the linear combination that define the principal components with coefficients different from zero. In order to solve this problem, several approaches have been tested, which consider the PCA as a regression problem and make use of regularization techniques to solve the optimization problem. In the following section, two regularization operators are briefly presented.

8.2.1 Regularized Regression Problem

The Least Absolute Shrinkage and Selection Operator (LASSO) is an algorithm for the smart solution of regression problems. In particular, the minimization of the cost function in the regression problem is handled adding a regularization term to the function that needs to be minimized. Briefly, instead of simply minimizing the Residual Sum of Squares

$$RSS(\beta) = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{8.5}$$

one can decide to minimize a different function, which includes a term $F(\beta)$

$$L(\beta) = RSS(\beta) + F(\beta) \tag{8.6}$$

and the estimated coefficient vectors becomes

$$\hat{\beta}^{REG} = \arg\min_{\beta} \left(RSS(\beta) + F(\beta) \right) \tag{8.7}$$

The result is that some coefficients (associated to less informative variables) shrink to zero. In particular, the regularization term in the LASSO consists in the sum of coefficients absolute values:

$$\hat{\beta}^{LASSO} = \arg\min_{\beta} \left(\left|\left| Y - \sum_{j=1}^{p} X_j \beta_j \right|\right|^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right) \tag{8.8}$$

where $\lambda \geq 0$ is a complexity parameter that controls the amount of shrinkage and hence the number of coefficients that will be put to zero.

Another possible optimization is the so called ridge-estimate:

$$\hat{\beta}^{RIDGE} = \arg\min_{\beta} \left(\left|\left| Y - \sum_{j=1}^{p} X_j \beta_j \right|\right|^2 + \lambda \sum_{j=1}^{p} ||\beta_j||^2 \right) \tag{8.9}$$

which will be used in the following sections for a different exact derivation of PCA.

A similar behavior is given by the use of the (naive) Elastic Net operator (68), which achieves similar results by shrinking some of the coefficients to zero by exploiting a different minimization function which includes a quadratic term of the coefficients.

$$\hat{\beta}^{EN} = \arg\min_{\beta} \left(\left\| Y - \sum_{j=1}^{p} X_j \beta_j \right\|^2 + \lambda_2 \sum_{j=1}^{p} ||\beta_j||^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j| \right) \tag{8.10}$$

8.2.2 LASSO and Elastic Net for PCA

Considering the problem of shrinking some of the coefficient (related to less informative variables) to zero, one can think of exploiting the two above presented operators in the regression step of PCA. Some strategies propose to force to zero all variables whose contribution (coefficient) to the linear combination is smaller than a certain threshold. This approach has been proven to be misleading (6). Alternative solutions to straightforward PCA were proposed. For example, McCabe (37) proposed a method to find a subset of principal variables. Notice that in both LASSO and Elastic Net, sparse coefficients are a direct consequence of the L_1 penalty, not depending on the squared error loss function. Joliffe and Uddin proposed a methodology called the ScotLASS (29), which directly puts the L_1 constraint in PCA to get sparse loadings. The major problem in this algorithm is the choice of the parameter which defines when the coefficients will be shrink to zero. In the following section a brief review of the so-called Sparse Principal Component Analysis proposed by Zou et al. (68) is presented.

8.3 Sparse Principal Component Analysis

A new method, called Sparse Principal Component Analysis was introduced by Zou et al in (68) in 2004. The method was developed in order to obtain a tool such that

1. without any sparsity constraint, the method reduces to PCA
2. it can be computed efficiently
3. it should avoid mis-identifying important variables.

The method uses the LASSO or Elastic Net regression of Section8.2.1 to produce modified principal components with sparse loadings. The first exploratory example shown in (68) is a two steps approach, which first requires to perform a regular PCA and then use a regression step to find sparse loadings. The regression step is the following:

$$\hat{\beta} = \arg\min_{\beta} ||Y_i - X\beta||^2 + \lambda_1 |\beta| \tag{8.11}$$

where Y_i is the i^{th} PC and the cost function is, in this case, the same as Eq 8.10, i.e. regularized using the Elastic Net approach.

Notice that different algorithms may be employed to obtain sparse loadings, the two-steps approach being the simplest one. Since to our experience the regularization approach has never been tested on glucose variability data, this will be the approach that we will use in the discussion of the results. Further studies will focus on testing different algorithms.

We report here an example of different,self-contained regression-type approach presented in (68). Denoting with X_i the i^{th} row vector of X, for any $\lambda > 0$ it can be proved that

$$(\hat{\alpha}, \hat{\beta}) = \arg\min_{\alpha,\beta} \sum_{i=1}^{n} \left|X_i - \alpha\beta^T X_1\right|^2 + \lambda|\beta|^2 \tag{8.12}$$

$$\text{subject to } |\alpha|^2 = 1 \tag{8.13}$$

subject to then $\hat{\beta} \propto V_i$ (recall that $X\hat{V}_i$ is the i^{th} approximated principal component). In (68) the procedure is shown to generalize Eq. 8.12 to all the principal component. The interesting feature of this step is that it proves that exact PCA can be obtained while relaxing the restriction $\beta = \alpha$ and adding a ridge penalty term (see Eq.8.9) to the simple regression procedure. To obtain sparse loadings, the LASSO penalty is added into the generalization of Eq. 8.12, and the following optimization problem is considered:

$$(\hat{\alpha}, \hat{\beta}) = \arg\min_{\alpha,\beta} \sum_{i=1}^{n} ||X_i - \alpha\beta^T X_i||^2 + \lambda \sum_{j=1}^{k} ||\beta j||^2 + \lambda \sum_{j=1}^{k} \lambda_{1,j}|\beta_j| \tag{8.14}$$

subject to $\alpha\alpha^T = I_k$. Eq. 8.14 is called the SPCA criterion. In its general formulation, the same λ is used for all the k components, while different $\lambda_{1,j}$ are allowed for penalizing the loadings of different principal components. This formulation of PCA with sparse loadings is more flexible than the ScotLASS and can be implemented with efficient algorithms.

8.4 Dataset and Indexes Evaluated

Sixty CGM signals collected with the Abbott Freestyle Navigator (10 minutes sampling) in the framework of the DIAdvisor project (1) were used for this study. CGM was recorded for 7 days (3 in hospital, plus 4 at home) in diabetics subjects in three clinical centers (Department of Clinical and Experimental Medicine at University Padova, the Centre Hospitalier Universitaire at Montpellier and the Institut Klinick a Experimentln Medicny in Prague). While in

hospital, subjects received 3 controlled meals per day and mostly rested during the day in the hospital. At home, patients lived their normal life. Sensors were calibrated accordingly with manufacturer's instructions.

On the 60 CGM signals, 26 literature indexes for the evaluation of glucose variability and control were computed. Also, 22 new DRS-based indexes were calculated. The detailed List of Indexes evaluated is reported in Tables 8.1, 8.2 and 8.1.

Since the analysis is retrospective the signals, originally sampled at a 10 minutes sampling rate were undertaken a bayesian smoothing procedure with a 1 minute spread virtual grid with the same algorithm presented in Appendix B.

8.5 Preliminary Analysis: Standard PCA

PCA was performed on matrix X, which is an $n \times m$ (60×48) matrix having on each of the 60 rows the 48 indexes evaluated on the i$^{\text{th}}$ subject. From the PCA we obtain two matrixes:

- β $(m \times m)$: matrix of the coefficients, i.e. the rotation matrix
- Z $(n \times m)$: the score matrix, representing the data matrix rotated along the new direction of the Principal Components. It is obtained by pre-multiplying X by B.

Matrix β has on the i^{th} column the coefficients for the linear combination of the indexes which allows to obtain the i^{th} principal component. It can be used to evaluate the relative weight of the different indexes in the specific principal component. The values σ_i are related to the variance of the dataset, and the less informative ones (on the lower right part of matrix Σ, associated to the right columns of matrix U of Eq 8.2) can be forced to zero. At this point, one may reconstruct the so called scores (matrix Z) and evaluate its variance. If we use all the principal components, the transformation is a simple rotation and the variance of the scores will remain unchanged. if we use less principal components to reconstruct the data, the Z matrix will be able to explain only part of the total variance in the original dataset, since the information carried by the less informative principal component has been lost. Table 8.4 shows a summary of the variances explained by using the first k principal components (first column). In particular, the second column reports the absolute variance of the k^{th} component, the third column reports the % of variability explained by each principal component, while the last column shows the cumulative variance obtained by using the first k PC (results are shown for the first 20 components out of 48). The shaded area shoes the results for the

5 most significative components. The total variance explained by truncating the principal components matrix U to the 5th column is 87.94 % of the total variance evaluated on the original matrix X.

# of principal components	Variance	%Variability	%Cumulative
1	27.3775	57.0365	57.0365
2	7.9991	16.6647	73.7013
3	2.9303	6.1047	79.8060
4	2.3521	4.9002	84.7062
5	1.5498	3.2287	87.9349
6	1.3584	2.8299	90.7649
7	1.0310	2.1479	92.9127
8	0.7577	1.5786	94.4913
9	0.7018	1.4622	95.9535
10	0.3243	0.6757	96.6292
11	0.2980	0.6209	97.2502
12	0.2613	0.5444	97.7945
13	0.1982	0.4129	98.2075
14	0.1805	0.3760	98.5835
15	0.1360	0.2833	98.8667
16	0.1009	0.2101	99.0769
17	0.0719	0.1497	99.2266
18	0.0628	0.1309	99.3575
19	0.0560	0.1167	99.4742
20	0.0494	0.1029	99.5771

Table 8.4: Cumulative variances explained by the first 20 principal components. The shaded area highlights the first 5 components, which, together allow explaining 87.9% of the total variance in the dataset.

In this thesis we consider 5 principal components, meaning that with 5 linear combinations of the original indexes, a sufficient (87%) variance of the original dataset is represented. Notice that, so far, these linear combinations may assign non-zero coefficient to each index. This means that one should first evaluate all possible indexes to combine them in 5 components that describe the specific subject. In this framework, a summary of the estimated coefficients is shown in Table 8.5.Each one of the 5 principal components is evaluated by linearly multiplying the coefficients in columns 2 to 6 with the computed index. For simplicity, only the first and last eight rows (indexes loadings) are shown in this Table.

From Table 8.5 the difficulty in the interpretation of PCA results is clear. The 5 principal components, which are basically 5 coefficients which determine the position of a certain

Index	PC1	PC2	PC3	PC4	PC5
1	-0.1583	-0.1465	0.1934	-0.0704	-0.0299
2	-0.1495	-0.1434	0.1300	-0.0778	0.0850
3	-0.1319	-0.0671	0.1582	0.0252	-0.2435
4	-0.1123	-0.0930	0.0393	-0.0655	0.2750
5	-0.1481	-0.1224	0.1814	-0.0221	-0.1942
6	-0.1413	-0.1064	0.1856	-0.0050	-0.2641
7	-0.0380	-0.2867	0.2656	-0.0148	-0.0256
8	-0.1778	0.1163	-0.0389	-0.0603	0.0209
...	...	...	...	...	...
41	0.1353	-0.1207	-0.1162	-0.0921	-0.0564
42	0.1092	-0.1442	-0.3315	-0.0071	-0.0610
43	0.1351	-0.1338	-0.2542	0.0143	-0.1268
44	0.0729	-0.1524	-0.3179	-0.2057	-0.0212
45	-0.0863	-0.2644	-0.2308	0.0163	0.0020
46	-0.1769	0.0443	-0.0602	0.1192	-0.0092
47	-0.1803	0.0995	-0.0348	-0.0471	0.0290
48	-0.1801	0.1065	-0.0444	-0.0321	0.0151

Table 8.5: Coefficients evaluated via standard PCA. Notice that all coefficients are non-zero, even if some contribute only for a very minimal part to the total variance.

subject defined through the indexes evaluated on his/her CGM signal, on a 5 dimensional space, are the combination of all the 48 indexes. No indication is given on which are the dominant indexes or combination of indexes that determine the position of the subject in the new space with reduced-dimensionality.

8.6 Analysis via Sparse PCA

As anticipated, the Sparse PCA approach can be utilized in order to have an easier interpretation of the dimensionality reduction performed by the PCA approach. Here we apply the two-steps SPCA (PCA plus LASSO regression to obtain sparse loadings) exploiting following parameters:

1. The number of principal components used was set to 5 as explained in Section 8.5.

2. The number of loadings allowed to be different from zero in the LASSO was set to 5 as well. The choice was made by analyzing the cumulative variance explained by two-steps SPCA with 5 principal components, allowing 1 to 48 indexes to be different from

zero for each component. With only 5 non-zero coefficients for each PC, we are able to explain about 70% of the variance explained by the full 5-PC analysis (all coefficients different from zero).

A portion of the resulting loading matrix is reported in Table 8.6. Notice how many loadings are shrink to zero by the SPCA algorithm.

Index	PC1	PC2	PC3	PC4	PC5
...	...	...	...	...	...
3	0	0	0	0	-0.0861
4	0	0	0	0	0.1686
...	...	...	...	...	...
6	0	0	0	0	-0.0098
7	0	-0.2919	0.4558	0	0
...	...	...	...	...	...
10	-0.3916	0	0	0	0
11	0	0	0	0.6453	0
...	...	...	...	...	...
15	0	0	0	0.1910	0
...	...	...	...	...	...
18	0	0.1094	0	0	0
19	0	0.2661	0	0	0
...	...	...	...	...	...
23	-0.2361	0	0	0	0
...	...	...	...	...	...
29	-0.0018	0	0	0	0
30	0	-0.1687	-0.2240	0 0	
...	...	...	...	...	...
32	0	0	-0.2946	0	0
33	-0.2066	0	0	0	0
34	-0.0600	0	0	0	0
...	...	...	...	...	...
36	0	0	0	-0.0702	0.4542
...	...	...	...	...	...
38	0	0	0	0	0.1133
39	0	0	0	-0.1470	0
...	...	...	...	...	...
42	0	0	-0.2774	0	0
...	...	...	...	...	...
44	0	0	-0.0260	-0.0169	0
45	0	-0.3019	0	0	0
...	...	...	...	...	...

Table 8.6: Loadings obtained via two-steps SPCA. Notice that many coefficients are forced to zero, meaning that relative index will not contribute in the evaluation of the specific Principal Component.

8.6.1 Selected Variables

Forcing sparsity in the SPCA algorithm, results, of course, in a certain loss of information. The more the zero-loadings, the higher the reduction in explained variance. This is the major

drawback in the use of this method, which, nonetheless, results in a greater readability of the loading matrix. In particular, we summarize in Table 8.7 the chosen indexes for each principal component, i.e. those indexes whose coefficient is not shrink to zero by the regularization in the regression step of PCA.

PC1	PC2	PC3	PC4	PC5
J	CV	CV	IGC	SD_{dm}
90^{th}percentile	Min	$Path_{dG}$	% Hypo	SD_{hhmm}
$Path_r$	10^{th}percentile	Sway Area	$C_{y\ ellipse}$	$SD_{bhhmm\ //\ dm}$
A_{circle}	$path_{dG}$	$f_{50\%\ dr}$	$Asse_x$	$C_{y\ ellipse}$
$A_{ellipse}$	P_{dr}	$f_{95\%\ dr}$	$f_{95\%\ dr}$	$C_{y\ weighted\ ellipse}$

Table 8.7: Selected Variables for each Principal Component

Table 8.7 allows making some quite thick statements about the analysis of the glucose signals. In particular, the first thing to be noticed is that the information about the time derivative adds relevant information when considered along with simple glucose level measurements. This can be seen by noticing that many time-derivative related indexes (e.g. the path of the time derivative, the weighted or non-weighted y-coordinate of the ellipse, the area of the ellipse) are picked in the generation of the sparse loadings. This is interesting, although quite obvious, since if no information was added by the time-derivative, these indexes would be ignored by the SPCA. Of course we are adding information to the dataset, but the important thing is that with the time-derivative we are adding relevant information in terms of dataset variability.

Also Table 8.7 shows how different kind of information is clustered in different components: PC1 carries information on the range spanned by the glucose signal in particular with respect to hyper excursions. PC2 is informative on the rapidity of the signal (path of the derivative and total power of the risk spectrum) with attention to the hypo excursions. PC3 again includes information on the derivative path and spectral informations (related to the rapidity of the signal). Finally, PC4 and PC5 collect other "average" indexes, such as the information of the coordinate location of the ellipse in the DR Space.

8.7 Classification via SPCA

The 60 CGM signals analyzed in the previous section were screened by an expert clinician at the Department of Clinical and Experimental Medicine at the University of Padova. The

clinician was asked to classify the 60 subjects accordingly to the CGM signals. Two different levels of classifications were defined. In a first level, the subjects were classified roughly in two classes, well and poorly controlled. The second level is more detailed and gives a score from 1 to 5 to the signal from very bad to optimal control.

The clinician was asked to observe the signals and was not informed with mean, SD or any other index score evaluated on the trace. The clinician had access to information of the beginning/end of the night. The main parameters in the evaluation was the qualitative estimation of the time spent in hyper, in hypo and the severity of each episode.

8.7.1 Results (Two Control Classes)

When considering the first classification (two classes) the dataset comprises:

- Good Control: 30 patients
- Poor Control: 30 patients

Sparse PCA was evaluated on the dataset, and the first two components were considered for a qualitative analysis in a two dimensional space (this is due to simplicity, although future studies will comprise statistical analysis considering more principal components if needed). Figures 8.1 and 8.2 represent the normalized values of S-PC1 and S-PC2 selected indexes in the 60 subjects.

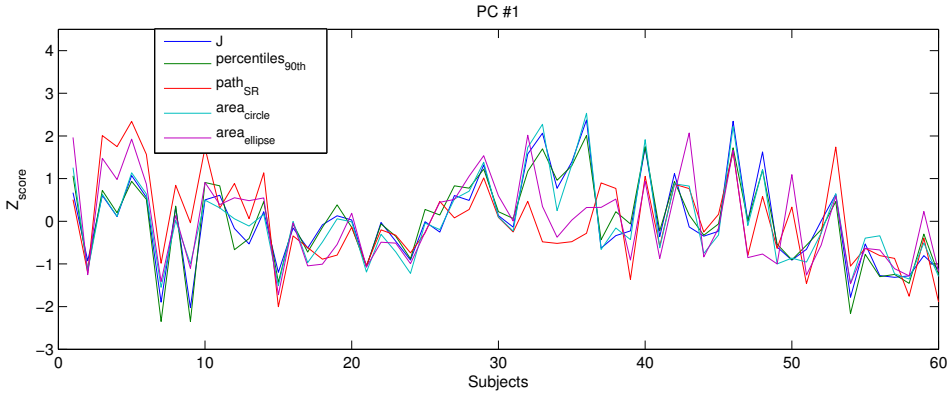

Figure 8.1: Boxplot of SPC1 and SPC2 in the two classes

Figure 8.3 displays a standard boxplot representing the values of the first and second S-PC (left and right panels respectively) for the two control classes of subjects.

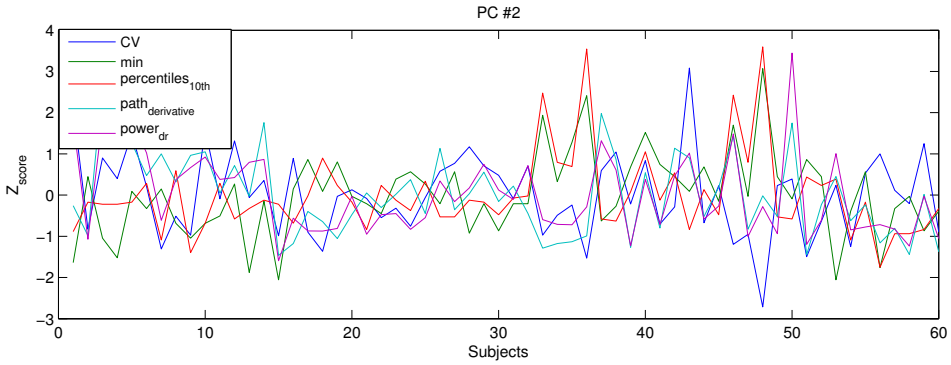

Figure 8.2: Boxplot of SPC1 and SPC2 in the two classes

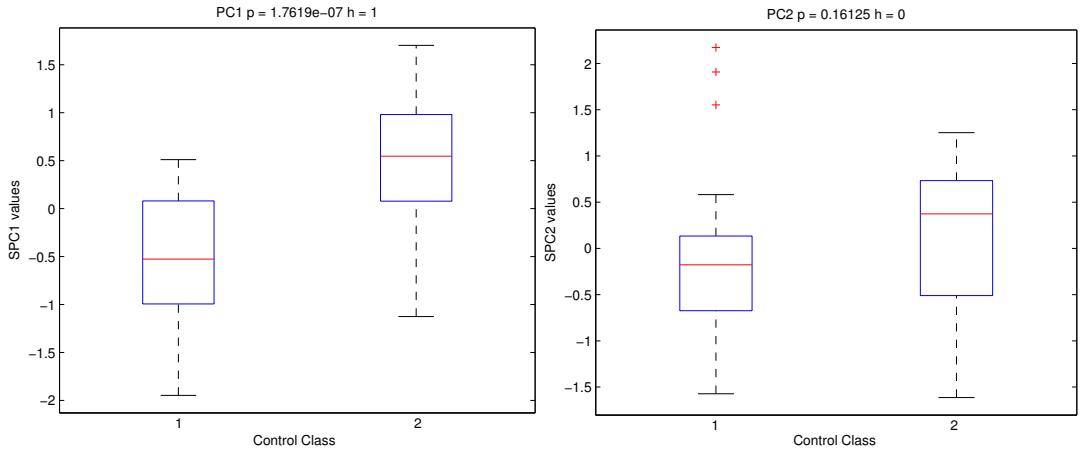

Figure 8.3: Boxplot of SPC1 and SPC2 values evaluated on the 60 subjects divided accordingly to the 2 control classes as defined by the expert clinician. Left boxes represent

Notice that the values of principal components in the two control classes are quite different for the first principal component, and the paired t-test performed to assess statistical significance of such difference results positive ($p< e^{-7}$). No significant difference was found in the value of the second S-PC in the two control classes (p=0.16).

Figure 8.4 represents the 60 subjects in the bidimensional space defined by the first two sparse principal components. The two classes result to be nicely separated in this space with well-controlled patients located on the upper-right part of the plane, and the poorly controlled patients located in the lower left part.

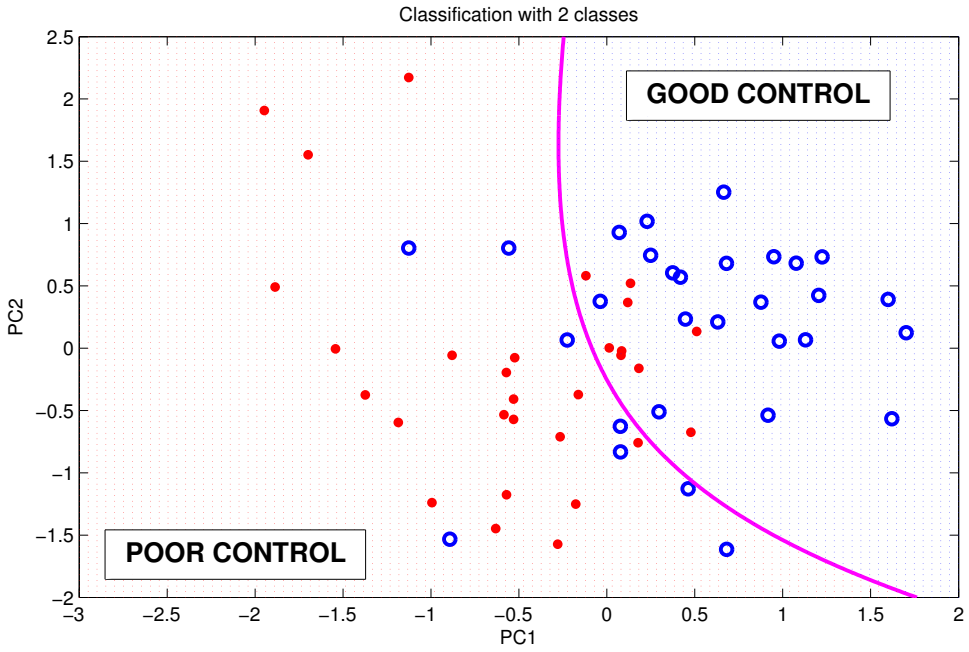

Figure 8.4: Scatterplot representing the position of the 60 subjects in the bidimensional space defined by the first two principal components. Subjects are clustered accordingly with their control class (good control in blue, poor control in red) as defined by an expert clinician.

8.7.2 Results (Five Control Classes)

When considering the second classification (five classes) the dataset comprises:

- Class 1 (very poor control): 15 patients
- Class 2 (poor control): 15 patients
- Class 3 (average control): 14 patients
- Class 4 (good control): 10 patients
- Class 5 (excellent control): 6 patients

Figure 8.5 displays a standard boxplot representing the values of the first and second sparse principal components (left and right panels respectively) for the five control classes of subjects.

A standard two-ways ANOVA was performed in order to understand if the two PC are significantly different among the control classes. The test results positive for PC1 and PC2,

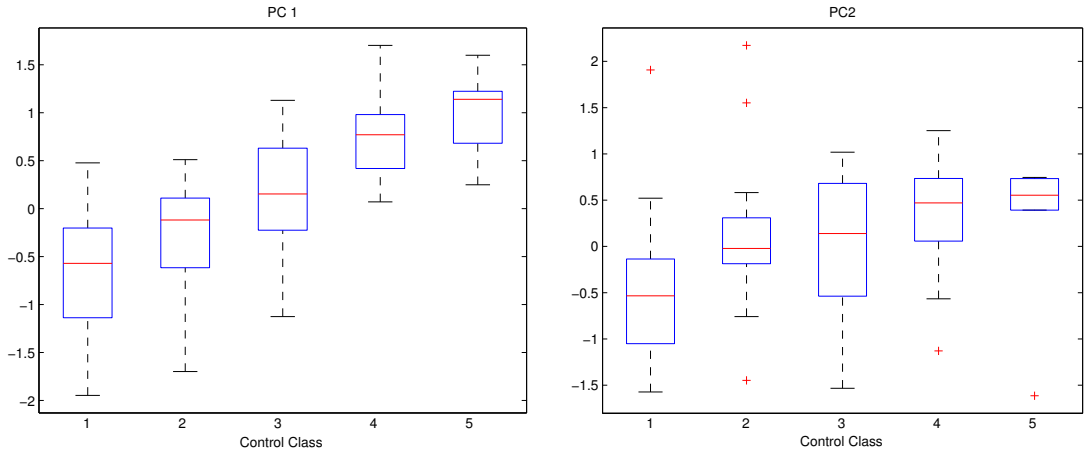

Figure 8.5: Boxplot of SPC1 and SPC2 values evaluated on the 60 subjects divided accordingly to the five control classes as defined by the expert clinician.

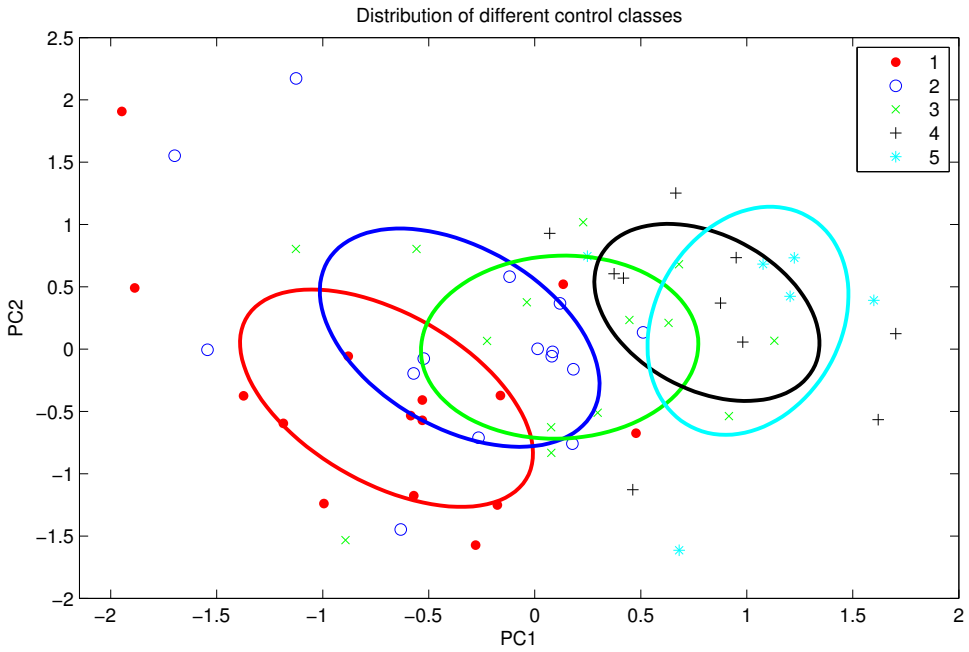

Figure 8.6: Scatterplot representing the position of the 60 subjects in the bidimensional space defined by the first two PCs. Subjects are clustered accordingly with their control class (scores from 1 to 5) as defined by an expert clinician.

meaning that indeed both components are different in the different control classes ($p< e^{-8}$ and $p< 0.02$, respectively). Figure 8.6 represents the 60 subjects in a bidimensional space

defined by the first two principal components.

The exploratory results shown in this section demonstrate that the use of dimensionality reduction techniques such as the two-steps SPCA can be a useful tool to discriminate among well-controlled and poorly controlled subjects. This approach seems promising for the understanding of glucose variability and control.

8.8 Conclusion

Although it seems quite obvious that the information of time-derivative adds value to the analysis of glucose signals, this proof-of-concept study helps the understanding of the relative importance of this information from a statistical point of view. Further studies will be focused on the analysis of clinical application of these concepts. In this Chapter we have shown a new approach to the analysis of glycemic variability and control. The major novelties are the explicit inclusion of indexes based on the DR and on the evaluation of the first-time derivative of signals in the analysis, and the quantitative, statistically meaningful framework of the Sparse Principal Component Analysis to study these data. Although this is just an exploratory study, and further studies may be done to fully understand the potential in this approach, it is nonetheless worth it underling that preliminary results indicate that the explicit use of DR based indexes adds value in the analysis of CMG data. Great benefit would come from a study on the long term complications of diabetes, where the monitoring is performed with continuous sensors rather than with SMBGs.

8. MULTIVARIATE ANALYSIS OF GLUCOSE VARIABILITY INDEXES VIA SPARSEPCA AND PARSIMONIOUS DESCRIPTION

9 Conclusions

The analysis of glucose time-series is very complex. Information about the patient health status can be inferred provided that proper analysis is performed. Historically, several interesting analysis have been performed on SMBG time-series, which provide sparse information about the glycemic level (3-4 measurements per day) hence ignoring wide fluctuations in the glycemic signal which can be dangerous in the short and in the long term.

The advent of sensors for the continuous monitoring of the glycemia, allows enhancing the analysis tools and consequently the knowledge and treatment of diabetes. In particular, the concept of clinical risk, proposed by Kovatchev et al. (31) and based on SMBG readings, can be re-interpreted in order to fully exploit the continuous information of CGM.

In this thesis, starting from a literature formulation of the concept of risk, we developed a new DR function which amplifies the clinical risk whenever the diabetic patient is receding from the safety region (Chapter 3). This function requires the time-derivative of the glucose signal to be evaluated; this computation is particularly challenging, since the derivation operation amplifies measurement noise. In Chapter 4 algorithmic solutions to this problem were proposed based on a deconvolution process.

DR has been shown to be potentially interesting for the generation of hypo/hyperglycemic alerts, both as a standalone tool (Chapter 5) and in combination with literature short-term prediction algorithms (Chapter 6). Simulated and real data were used for the assessment of the new technique. Simulations include noisy and noise-free signals that were analyzed separately. Real data were collected with three different continuous glucose sensors (Abbott Freestyle Navigator, Dexcom Seven Plus and Menarini Glucoday). Computation of DR allows generating alerts with 10 to 18 minutes ahead of time on real datasets, and up to 25 minutes

when combined with a KF based predictor, with a moderate number of false alerts generated each day. A provisional US patent has been recently deposited, which describes the possibility of exploiting DR based algorithms for the generation of alerts. Moreover, the patent describes the possibility of alerting a patient of an incoming risk of hypo/hyperglycemia through visual alerts by highlighting the monitor of CGM devices based on the DRS color map (22).

In this thesis we have also proven how retrospective analysis may greatly benefit from the exploitation of the continuous nature of CGM and the DR concept. In Chapters 7 and 8, conceptual examples of the importance of the information of time derivative. Measures derived from DR, that consider the time-derivative as a variability factor, have been developed by analyzing trajectories of glucose in the DRS. Sparse PCA approach was employed for the analysis of a dataset of 48 indexes evaluated on 60 patients monitored for 7 days with a continuous glucose sensor. Dimensionality reduction proved that the indexes based on the DRS are statistically relevant for the characterization of glycemic signals. The Sparse PCA approach allowed determining that glucose variability should not be assessed by means of only one aspect of the glucose signal, but should be a more comprehensive analysis, possibly including DR-based metrics.

Further studies will focus on refining the use of DR for the two proposed applications of alert generation and of indexes development for the description of glucose variability. In particular, comparison with other literature strategies for the generation of alerts will be necessary to better understand the applicability of the method in commercially available devices. Moreover, formal strategies for the tuning of the parameters may be developed. As far as the dimensionality reduction problem is concerned, future work will focus on the refinement of the SPCA algorithm used to compute the principal components, and on the assessment of the correlation of these components with short and long term complications of diabetes. Other studies will focus on the exploitation of the DRS to tune the therapy of the patients, for example via analysis of single-meal trajectories and personal features of patients in the DRS.

A

Some additional Details on Diabetes Therapy

Accurate monitoring of the glycaemia is of fundamental importance for diabetes control. Also, the delivery techniques of insulin and insulin analogues is at least as important as the accurate information on the variable to be controlled (the glycaemia). T1 diabetics need to inject themselves with insulin after each meal to prevent hyperglycaemia and to assure the uptake of glucose by the target tissues.

A.1 Conventional and Innovative Therapy

In the conventional therapy, patients tune the infusions based on few SMBGs measurements per day, or based on the retrospective analysis of CGM signals. Different kinds of insulin are available on the market, from slow acting insulins (e.g. Glargine, Detemir), to mid-term acting insulins (e.g. human NPH) to rapid and ultra rapid acting insulins (e.g. Lispro, marketed by Ely-Lilly as Humalog, and Aspart, marketed for example by Novo Nordisk as NovoRapid). There are two main insulin delivery techniques which require the injection in the body through a needle: the multi-injective (MI) therapy and the infusion via a Continuous Subcutaneous Insulin Pump (Continuous Subcutaneous Insulin Infusion, CSII). In the MI therapy the patient usually performs a bolus of slow acting insulin which should mimic the basal insulin secretion, plus fast acting insulin boluses at meal time to regulate the glycaemic increase due to the food ingestion. In the CSII therapy, a subcutaneous needle is inserted in the abdomen and a pump delivers insulin microboluses every 5 to 15 minutes. This allows having an almost continuous provisioning of insulin to the body. At meal time, the patient can deliver a larger bolus through the same needle, so the carbohydrates can be rapidly

absorbed by the tissues. Insulin pumps are only used in combination with ultrarapid insulins (aspart and lispro). Several studies report a better therapy efficiency patient treated with insulin pumps with respect to the conventional MI therapy.

A.2 Artificial Pancreas

In the recent past, researchers have been focusing on the problem of optimal tuning of the insulin delivery. In particular, the advent of Continuous glucose monitoring devices and of pumps able to continuously infuse insulin in the body, allowed to tackle the problem of "closing the loop". Automatic controllers are being developed exploiting the information of CGM to tune the optimal insulin bolus/basal rate or, in general, dose, to be fed continuously through the pump. Several issues are still open in this field, from the structure of the controller itself (Model Predictive Controllers, Proportional Integrative Derivative structures, bi-hormonal controllers), to open problems with CGM accuracy (calibration, poor SNR) to possible system failure detection (both in pump and CGM). The interested reader is referred to (10), and (11) for two recent review on the achievements and open problems in the technologies for diabetes.

Brief Review of the Bayesian Approach to Smoothing and Deconvolution

Consider the following general linear model:

$$y = Gu + v \tag{B.1}$$

where y and v are n-dimensional vectors, u is an N-dimensional vector, and G is an $n \times N$ matrix. In a Bayesian framework, the estimate of u given y and the model B.1, can be tackled as a minimum error variance estimate. The solution to this problem coincides with that of determining the vector $\hat{u}$ which minimizes the expected value of the quadratic norm of the estimate error:

$$E\left[\|u - \hat{u}\|^2\right]. \tag{B.2}$$

Restricting the search into the class of linear estimators, it can be demonstrated that, if u and v are uncorrelated, have zero mean and known covariance matrixes Σ_u and Σ_v respectively, the best linear estimate of u given y is the solution of the following optimization problem:

$$\min_{\hat{u}} (y - G\hat{u})^T \Sigma_v^{-1} (y - G\hat{u}) + \hat{u}^T \Sigma_u^{-1} \hat{u}, \tag{B.3}$$

Notably, the argument of the minimization problem is made of two terms:

- $(y - G\hat{u})^T \Sigma_v^{-1} (y - G\hat{u})$: it represents the adherence of the model to the data, i.e. it accounts for the posterior information yielded by the data

- $\hat{u}^T \Sigma_u^{-1} \hat{u}$: it represents the adherence of the unknown vector u to the prior information available on u itself.

Eq. B.3 has a closed form solution given by:

$$\hat{u} = (G^T \Sigma_v^{-1} G + \Sigma_u^{-1})^{-1} G^T \Sigma_v^{-1} y \tag{B.4}$$

In order to solve the minimization problem, it is necessary to have a second-order a priori statistical description of vectors u and v. For example, for specific problems, one can assume following models for the statistical a priori knowledge.

- *Second order statistical description of vector v*: covariance matrix of vector v described as

$$\Sigma_v = \sigma^2 B \tag{B.5}$$

with B known matrix and σ^2 scaling factor, unknown.

- *Second order statistical description of vector u*: for biological signals, the prior information of the signal to be estimated are qualitative, i.e. the signal is assumed to be smooth and regular. Formally, this can be described considering the components of vector $u(t)$ to be the realization of a stochastic process that is modeled by a cascade of m integrators driven by stationary white noise with zero mean and variance (unknown) λ^2 with prior covariance matrix:

$$\Sigma_u = \lambda^2 (F^T F)^{-1} \quad \text{with} \quad F = \Delta^m \tag{B.6}$$

where m represents the penalty index of the m-th order detivatives of $u(t)$ and Δ is the lower-triangular Toeplitz matrix:

$$\Delta = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & -1 & 1 & 0 \\ 0 & \cdots & \cdots & 0 & -1 & 1 \end{bmatrix} \tag{B.7}$$

higher values of m allow to have a smoother reconstruction of the signal $u(t)$.

Assuming to have a statistical description of the covariances matrixes as defined in Eq. B.5 and B.6, Eq. B.3 can be rewritten as:

$$\hat{u} = (G^T B^{-1} G + \gamma^o F^T F)^{-1} G^T B^{-1} y \tag{B.8}$$

The ratio $\gamma^o = \sigma^2/\lambda^2$ determines the best trade-off between the adherence to the prior and the data. Higher γ^o define an estimator which relies mostly on the prior, while lower γ^o values are associated to estimators which mostly rely on the data. The Bayesian framework allows to optimally weight the relative importance of a prior and of the data, since the statistical knowledge on their variance is explicitly considered in the computation of the estimate. Considering the problem defined in Eq. B.2 and its closed form solution of Eq. B.8, two quantities can be defined:

- *Weighted residuals sum of squares, WRSS*

$$WRSS = (y - G\hat{u})^T B^{-1} (y - G\hat{u}) \tag{B.9}$$

- *Weighted estimates sum of squares, WESS*

$$WESS = \hat{u}^T F^T F \hat{u} \tag{B.10}$$

In a stochastic framework, WRSS and WESS are random variables, whose realizations depend on the value of γ, since $\hat{u}$ depends on γ. It can be shown that for the optimal value γ^o following properties hold:

$$E\left[WESS(\gamma^o)\right] = \lambda^2 \cdot q(\gamma^o) \tag{B.11}$$

$$E\left[WRSS(\gamma^o)\right] = \sigma^2\{n - q(\gamma^o)\} \tag{B.12}$$

where

$$\gamma^o = trace[B^{-1/2} G (G^T B^{-1} G + \gamma^o F^T F)^{-1} G^T B^{-1/2}] \tag{B.13}$$

with $B^{-1/2}$ is a square matrix such that $B^{-1} = B^{-1/2} B^{-1/2}$. The properties defined in Eqs. B.12 and B.11 some consistency criteria con be defined for the choice of the best γ in

conditions where σ^2, λ^2 or both are unknown. In particular:

Consistency Criterion 1: If λ^2 is unknown and σ^2 is known, choose γ such that the following condition is matched:

$$WESS(\gamma) = \lambda^2 \cdot q(\gamma^o) \quad \text{with} \quad \lambda^2 = \sigma^2\gamma \tag{B.14}$$

Consistency Criterion 2 If σ^2 is unknown and λ^2 is known, choose γ such that the following condition is matched:

$$WRSS(\gamma) = \sigma^2\{n - q(\gamma^o)\} \quad \text{with} \quad \sigma^2 = \lambda^2\gamma \tag{B.15}$$

Consistency Criterion 3: When both σ^2 and λ^2 are unknown, choose γ such that the following condition is matched:

$$\frac{WRSS(\gamma)}{n - q(\gamma)} = \gamma \cdot \frac{WESS(\gamma)}{q(\gamma)} \tag{B.16}$$

then, when γ is determined from B.16, compute the posterior estimate of σ^2 from B.9 as

$$\hat{\sigma}^2 = \frac{WRSS}{n - q(\gamma)} \tag{B.17}$$

A final remark. In a Bayesian framework it is possible to compute the confidence interval of the estimate. In particular, if only gaussian variables are involved, the covariance matrix of the estimate error defined as $\tilde{u} = u - \hat{u}$ is

$$\Sigma_{\tilde{u}} = \Sigma_u - \Sigma_{uy}\Sigma_y^{-1}\Sigma_{yu} \tag{B.18}$$

which becomes (demonstration not shown)

$$\Sigma_{\tilde{u}} = (G^T\Sigma_v^{-1}G + \Sigma_u^{-1})^{-1} \tag{B.19}$$

with the model of measurement of Eq. B.1. For a more detailed description of this topic, the reader is referred to (12).

B.1 Numerical Implementation

The inversion of matrix $(G^T B^{-1} G + \gamma^o F^T F)^{-1}$, which is required for the solution of the deconvolution problem, is computationally time-consuming, since it requires $O(N^3)$ operations. Moreover, this inversion needs to be done several time in the iterative procedure for the consistent choice of the regularization parameter γ. Here we report the procedure which can be used to diagonalize the problem in order to obtain a computationally efficient solution. Given matrixes G, F and B:

- define matrix $H = B^{-1/2} G F^{-1}$, $n \times N$
- perform the Singular Value Decomposition (SVD) of H, defining the unitary matrixes U and V such that $U^T U = I_n$ and $V^T V = I_N$ and D composed by a diagonal matrix of diamension n and a zero padding matrix of $N - n$ columns.
- define the new coordinates $\xi = U^T B^{-1/2} y$, $\eta = V^T F u$ and $\epsilon = U^T B^{-1/2} v$; Eq. B.1 can be rewritten as:

$$\xi = D\eta + \varepsilon \tag{B.20}$$

  where $cov(\varepsilon) = \sigma^2 I_n$ and $cov(\eta) = \lambda^2 I_N$

- the regularized estimate of $\hat{\eta(\gamma)}$ can be obtained as

$$\hat{\eta}_i = \begin{cases} \frac{d_i}{d_i^2+\gamma}\xi_i & \text{for} \quad i = 1, \cdots, n \\ 0 & \text{for} \quad i = n+1, \cdots, N \end{cases} \tag{B.21}$$

  in $O(n)$ computations using following regularization quantities:

$$q(\gamma) = \sum_{i=1}^{n} \frac{d_i^2}{d_i^2 + \gamma} \tag{B.22}$$

$$WESS(\gamma) = \sum_{i=1}^{n} \left(\frac{d_i \xi_i}{d_i^2 + \gamma} \right)^2 \tag{B.23}$$

$$WRSS(\gamma) = \sum_{i=1}^{n} \left(\frac{\gamma \xi_i}{d_i^2 + \gamma} \right)^2 \tag{B.24}$$

- once the optimal γ has been computed, the input of the system can be obtained as $\hat{u} = F^{-1}V\hat{\eta}$ in $O(nN)$.

The only operation which is computationally difficult is the evaluation of the SVD, which can be done in $O(N^3)$. It is important to notice, that this operation is done only once for every estimate.

C

Some Details on Prediction via Kalman Filtering

Here we give a brief review of an online technique tusable to obtain a predicted profile in the short-term. The method is a development of a denoising algorithm (57). by Facchinetti et al. (18), which implements a Kalman filter coupled with a Bayesian smoothing criterion to estimate the unknown filter parameters. Briefly, let $y(t)$ be the glucose profile of Eq. 4.3. We recall that $y(t) = u(t) + v(t)$, with $u(t)$ representing the true glucose profile and $v(t)$ representing the noise component, assumed to be zero-mean gaussian noise with variance σ^2 varying in time. In order to put the estimation problem in a suitable form to be put in a Kalman Filter form, we define the following model of the signal $u(t)$:

$$u(t) = 2u(t-1) - u(t-1) + w(t) \tag{C.1}$$

where $w(t)$ is a zero-mean Gaussian noise, with variance λ^2. One can estimate $u(t)$ putting the equation in state-space form. If we allow $x_1(t) = u(t)$ and $x_2(t) = u(t-1)$ we have:

$$\begin{bmatrix} x_1(t+1) \\ x_2(t+1) \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} w(t) \tag{C.2}$$

$$y(t) = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + v(t) \tag{C.3}$$

Equation C.2 and C.3 are the process update and the measurement equations used by the Kalman filter to estimate $\hat{x}(t|t)$ which is the linear minimum variance estimate of the state vector that can be obtained from measurement $y(t)$ at time t.

The prediction via KF can be obtained considering the predictive step alone:

$$\hat{x}(t+1|t) = F\hat{x}(t|t) \tag{C.4}$$

where F is the state transition matrix and $\hat{x}(t+1|t)$ is the state estimate based only on the measurements collected until time t. By iterating the predictive step one can obtain the prediction n steps ahead, at the time $t + PH$

$$\hat{x}(t+PH|t) = F^{PH}\hat{x}(t|t) \tag{C.5}$$

A detailed description of the algorithm can be found in (18).

D

Basic Aspects of Principal Component Analysis

Principal Component Analysis (PCA) is a non-parametric multi-variate method for extracting relevant information from datasets difficult to interpret. In particular, PCA is a simple tool used to reduce the correlation of a set of p variables (collected in a matrix X). The correlation within the data collected in the matrix X can be expressed by the (mxm) covariance matrix C_X:

$$C_X = \frac{1}{n-1} X^T X \tag{D.1}$$

The diagonal terms of C_X are the variance of particular measurements types while the off-diagonal terms are the covariance between measurements types. Thus, in order to reduce the correlation of the data collected in X, we are seeking a linear transformation defined by a matrix P:

$$Y = XP \tag{D.2}$$

such that the covariance matrix C_Y of the matrix Y is diagonal. In particular, PCA finds the matrix P to re-express the data with the constraints that all basis vector $p_1, \cdots, p_m$ are orthonormal. The matrix P is thus called orthogonal and has the properties that $PP^T = I$. The matrix P can be obtained in two closely related way:

1. by exploiting the eigenvector decomposition of the covariance matrix or
2. by resorting to the Singular Values Decomposition (SVD) of the original data matrix X.

By considering point 1, we preliminary recall that a symmetric matrix A can be diagonalized by an orthogonal matrix of its eigenvectors E, such that $A = EDE^T$, and where D is diagonal. It is easy to show that the covariance matrix of Y can be rewritten as $C_Y = \frac{1}{n-1}PX^TXP^T$ by substituting D.2 in D.1. If we now select the P matrix to be a matrix of eigenvectors of C_X (thus holding $PP^T = I$) and perform some easy steps we obtain that $C_Y = D$. In this way we have diagonalized the covariance matrix of the new set of variables and have assured that these new variables are not correlated anymore (since the off-diagonal elements of C_Y are zero).

A second possibility to find a linear transformation that reduces the correlation in the X matrix finding a new set of uncorrelated variables is by exploiting the so called SVD which factorizes X as:

$$X = U\Sigma V^T \tag{D.3}$$

In D.3, V is the matrix collecting the set of orthonormal eigenvectors (with associated eigenvalues λ_i) for the symmetric matrix X^TX, Σ is the diagonal matrix collecting the so called singular values $\sigma_i = \sqrt{\lambda_i}$ and U is the orthonormal matrix collecting the vectors $u_i = 1/\sigma i X v_i$. By a simple manipulation ,D.3 can be rewritten as $XV = U\Sigma$ and it is easy to see that the change of basis that we are seeking is given by V.

The interesting feature of PCA is that the principal components are sorted in decreasing importance in terms of explained variance. In particular:

1. Principal components sequentially capture the maximum variability among the X matrix, guaranteeing minimal information loss

2. Principal components are uncorrelated, so one component can be addressed without need to refer to others.

This means that the smaller components are easily identified and may be forced to zero. In this way only some of the components of the new basis are used to describe the data, cutting those which do not add crucial information in terms of variance of the data. One of the possible criteria to decide how many informative principal components need to be kept to allow a good reconstruction of the total variability is considering a number of principal components which explains 80-90% of the total variance.

References

[1] 7th Framework European Project DIAdvisor. http://www.diadvisor.eu/.

[2] Abbott Diabetes Care. http://www.abbottdiabetescare.com/index.htm.

[3] American Diabetes Association (ADA). http://www.diabetes.org/.

[4] B. Balkau, M. Shipley, R. Jarrett, K. Pyrl, M. Pyrl, A. Forhan, and Eschwge. High blood glucose concentration is a risk factor for mortality in middle-aged nondiabetic men. 20-year follow-up in the whitehall study, the paris prospective study, and the helsinki policemen study. *Diabetes Care*, 21(3):360–367, 1998.

[5] M. Brownlee. The pathobiology of diabetic complications: a unifying mechanism. *Diabetic Medicine*, 54(6):1615–1625, 2005.

[6] J. Cadima and I. Jolliffe. Loadings and correlations in the interpretation of principal components. *Journal of Applied Statistics*, 22:203–214, 1996.

[7] A. Ceriello, K. Esposito, L. Piconi, M.A. Ihnat, J.E. Thorpe, R. Testa, M. Boemi, and D. Giugliano. Oscillating glucose is more deleterious to endothelial function and oxidative stress than mean glucose in normal and type 2 diabetic patients. *Diabetes*, 57: 1349–1354, 2008.

[8] C. Choleau, P. Dokladal, J. Klein, W. Ward, G. G. Wilson, and G. Reach. Prevention of hypoglycemia using risk assessment with a continuous glucose monitoring system. *Diabetes*, 51:32633273, 2002.

[9] W. Clarke and B. Kovatchev. Statistical tools to analyze continuous glucose monitor data. *Diabetes Technol. Ther.*, 11 Suppl 1:45–54, Jun 2009.

[10] C. Cobelli, C. Dalla Man, G. Sparacino, L. Magni, G. De Nicolao, and B. Kovatchev. Diabetes: Models, signals, and control. *IEEE Trans. Biomed. Eng.*, 2:54–96, 2009.

[11] C. Cobelli, E. Renard, and B. Kovatchev. Artificial pancreas: Past, present, future. *Diabetes*, 60(11):2672–2682, 2011.

[12] G. De Nicolao, G. Sparacino, and C. Cobelli. Nonparametric input estimation in physiological systems: problems, methods, case studies. *Automatica*, 33:851–870, 1997.

[13] R. Derr, E. Garrett, G.A. Stacy, and C.D. Saudek. Is hba(1c) affected by glycemic instability? *Diabetes Care*, 26(10):2728–2733, 2003.

[14] DexCom Inc. http://www.dexcom.com/.

[15] Echo Therapeutics Inc. http://www.echotx.com/.

[16] M. Eren-Oruklu, A. Cinar, L. Quinn, and D. Smith. Estimation of the future glucose concentrations with subject specific recursive linear models. *Diabetes Technol Ther*, 11 (4):243253, 2009.

[17] K. Esposito, M. Ciotola, D. Carleo, B. Schisano, L. Sardelli, D. Di Tommaso, L. Misso, F. Saccomanno, A. Ceriello, and D. Giugliano. Post-meal glucose peaks at home associate with carotid intima- media thickness in type 2 diabetes. *J Clin Endocrinol Metab*, 93 (4):1345–1350, 2008.

[18] A. Facchinetti, G. Sparacino, and C. Cobelli. An online self-tunable method to denoise CGM sensor data. *IEEE Trans Biomed Eng*, 57:634–641, 2010.

[19] D. Finan, F. Doyle, C. Palerm, W. Bevier, H. Zisser, L. Jovanovic, and D. Seborg. Experimental evaluation of a recursive model identification technique for type 1 diabetes. *J Diabetes Sci Technol*, 5(3):11921202, 2009.

[20] A. Gani, A. Gribok, J. Rajaraman, and J. Reifman. Predicting subcutaneous glucose concentration in humans: Data-driven glucose modeling. *IEEE Trans Biomed Eng*, 56 (2):246254, 2009.

[21] Glucosecontrol.Edu. http://glucosecontrol.ucsd.edu.

[22] S. Guerra, A. Facchinetti, M. Schiavon, G. Sparacino, and C. Cobelli. Alert system for hypo and hyperglycemia prevention based on clinical risk associated to glucose level and trend for diabetes-related applications, provisional us patent, serial no. 61/551,773.

[23] S. Guerra, A. Facchinetti, G. Sparacino, and C. Cobelli. Dynamic Risk hypoalert generation from CGM signals. In *5th conference on Advanced technologies and tretments for Diabetes*, Barcelona, Spain, Feb 2011.

[24] S. Guerra, G. Sparacino, A. Facchinetti, M. Schiavon, C. Dalla Man, and C. Cobelli. A dynamic risk measure from continuous glucose monitoring data. *Diabetes Technol. Ther.*, 13:843–852, Aug 2011.

[25] J.D. Hamilton. *Time Series Analysis*. rinceton University Press, 1994.

[26] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. Springer Series in Statistics, 2009.

[27] N.R. Hill, P.C. Hindmarsh, R.J. Stevans, I.M. Stratton, J.C. Levy, and D.R. Matthwes. A method for assessing quality of control from glucose profiles. *Diabetic Medicine*, 24: 753–758, 2007.

[28] R. Hovorka. The future of continuous glucose monitoring: closed loop. *Curr. Diabetes Rev*, 4:269279, 2008.

[29] I. Jolliffe, N. Trendafilov, and M. Uddin. A modified principal component technique based on the lasso. *Journal of Computational and Graphical Statistics*, 12:531–547, 2003.

[30] E.S. Kilpatrick. Arguments for and against the role of glucose variability in the development of diabetes complications. *J Diabetes Sci Technol.*, 3(4):649–655, 2009.

[31] B. P. Kovatchev, D. J. Cox, L. A. Gonder-Frederick, and W. Clarke. Symmetrization of the blood glucose measurement scale and its applications. *Diabetes Care*, 20:1655–1658, Nov 1997.

[32] B. P. Kovatchev, D. J. Cox, L. A. Gonder-Frederick, D. Young-Hyman, D. Schlundt, and W. Clarke. Assessment of risk for severe hypoglycemia among adults with IDDM: validation of the low blood glucose index. *Diabetes Care*, 21:1870–1875, Nov 1998.

[33] B.P. Kovatchev, M. Straume, D.J. Cox, and L.S. Fahry. Risk Analysis of Blood Glucose data: a quantitative approach to optimizing the control of insulin dependent diabetes. *J Theor. Med*, 3:1–10, 2000.

[34] B.P. Kovatchev, L.A. Gonder-Frederick, D.J. Cox, and W.L. Clarke. Evaluating the accuracy of continuous glucose-monitoring sensors: continuous glucose-error grid analysis illustrated by TheraSense Freestyle Navigator data. *Diabetes Care*, 27(8):1922–1928, 2004.

[35] C.Y. Kuo, C.T. Hsu, C.S. Ho, T.E. Su, M.H. Wu, and C.J. Wang. Accuracy and precision evaluation of seven self-monitoring blood glucose systems. *Diabetes Technol. Ther .*, 13 (5):596600, 2011.

[36] LifeScan Inc. http://www.lifescan.com/.

[37] G. McCabe. Principal Variables. *Technometrics*, 26:137–144, 1984.

[38] Medtronic Inc. http://www.medtronicdiabetes.net.

[39] S. Melmed, Polonsky K.S., P.R. Larsen, and H.M. Kronenberg. *Williams Textbook of Endocrinology*. Saunders, 2008.

[40] L. Monnier, H. Lapinski, and C. Colette. Contributions of fasting and postprandial plasma glucose increments to the overall diurnal hyperglycaemia of type 2 diabetic patients: variations with increasing levels of hba1c. *Diabetes Care*, 26(3):881–885, 2003.

[41] L. Monnier, E. Mas, C. Ginet, F. Michel, L. Villon, J.P. Cristol, and C. Colette. Activation of oxidative stress by acute glucose fluctuations compared with sustained chronic hyperglycemia in patients with type-2 diabetes. *JAMA*, 295(14):1681–1687, 2006.

[42] D.M. Nathan, J. Kuenen, R. Borg, H. Zheng, D. Shoenfeld, and R.J. Heine. A1c derived average glucose (adag) study group: translating the a1c assay into estimated average glucose values. *Diabetes Care*, 31:1473–1478, 2008.

[43] C. Palerm and W. Bequette. Hypoglycemia detection and prediction using continuous glucose monitoring - a study on hypoglycemic clamp data. *J. Diabetes Sci. Technol.*, 1: 624629, 2007.

[44] J. Pankow, D. Kwan, B. Duncan, M. Schmidt, D. Couper, S. Golden, and C.M. Ballantyne. Cardiometabolic risk in impaired fasting glucose and impaired glucose tolerance: the atherosclerosis risk in communities study. *Diabetes Care*, 30(2):325–331, 2007.

[45] S. Pappada, B. Cameron, and P. Rosman. Development of a neural network for prediction of glucose concentration in type 1 diabetes patients. *J Diabetes Sci and Technol*, 2(5): 792801, 2008.

[46] S. Pappada, B. Cameron, P. Rosman, R. Bourey, T. Papadimos, W. Oloruntu, and M. Borst. Neural network-based real-time prediction of glucose in patients with insulin-dependent diabetes. *Diabetes Technol Ther*, 13(2):135141, 2011.

[47] C. Perez-Gandia, A. Facchinetti, G. Sparacino, C. Cobelli, E. Goomez, M. Rigla, A. de Leiva, , and M. Hernando. Artificial neural network algorithm for on-line glucose prediction from continuous glucose monitoring. *Diabetes Technol Ther*, 12(1):81–88, 2010.

[48] M.A. Powers. *Handbook of diabetes medical nutrition therapy.* Jones and Bartlett Learning, 1996.

[49] T.E. Prieto, J.B. Myklebust, R.G. Hoffmann, E.G. Lovett, and B.M. Myklebust. Measures of postural steadiness: differences between healthy young and elderly adults. *IEEE Trans Biomed Eng.*, 43(9):956–966, 1996.

[50] J. Reifman, S. Rajaraman, A. Gribok, and W. Ward. Predictive monitoring for improved management of glucose levels. *J Diabetes Sci. Technol.*, 1(4):478486, 2007.

[51] Roche Diagnostics - Accu-Chek. https://www.accu-chek.com/index.html.

[52] D. Rodbard. Interpretation of continuous glucose monitoring data: Glycemic variability and quality of glycemic control. *Diabetes Technol. Ther.*, 11:55–67, 2009.

[53] sanofi-aventis iBGstar. www.bgstar.com/web/ibgstar.

[54] F.J. Service and R.L. Nelson. Characteristics of glycemic stability. *Diabetes Care*, 3: 58–62, 1980.

[55] F.J. Service, G.D. Molnar, Rosevear J.W., E. Ackerman, L.C. Gatewood, and W.F. Taylor. Mean amplitude of glycemic excursions, a measure of diabetic instability. *Diabetes*, 19:644–655, 1970.

[56] D.S. Sivia. *Data Analysis: A Bayesian Tutorial*. Oxford Science Publications, 1996.

[57] G. Sparacino, A. Facchinetti, and C. Cobelli. Method and device for processing glycemia level data by means of self-adaptive filtering, predicting the future glycemia level and generating alerts. WO 2009/136372, International Publication Date: November 12, 2009.

[58] G. Sparacino, F. Zanderigo, S. Corazza, A. Maran, A. Facchinetti, and C. Cobelli. Glucose concentration can be predicted ahead in time from continuous glucose monitoring sensor time-series. *IEEE Trans Biomed Eng*, 54(5):931937, 2007.

[59] G. Sparacino, A. Facchinetti, A. Maran, and C. Cobelli. Continuous glucose monitoring time series and hypo/hyperglycemia prevention: requirements, methods, open problems. *Curr Diabetes Rev*, 4:181–192, 2008.

[60] G. Sparacino, A. Facchinetti, and C. Cobelli. Smart continuous glucose monitoring sensors: On-line signal processing issues. *Sensors*, 10:6751–6772, 2010.

[61] The Diabetes Control and Complications Trial Research Group. The Effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin dependent diabetes mellitus. *N Engl. J. Med.*, 329:977–986, 1993.

[62] S. Vaddiraju, D.J. Burgess, I. Tomazos, F.C. Jain, and F. Papadimitrakopoulos. Technologies for continuous glucose monitoring: Current problems and future promises. *J. Diabetes Sci and Technol.*, 4(6):15401562, 2010.

[63] I. Wentholt, W. Kulik, R. Michels, J. Hoekstra, and J.H. deVries. Glucose fluctuations and activation of oxidative stress in patients with type 1 diabetes. *Diabetologia*, 51(1): 183–190, 2008.

[64] J.M. Wójcicki. Mathematical description of the glucose control in diabetes therapy. analysis of the schlichtkrull m value. *Horm. Metab. Res*, 27:1–5, 1995.

[65] J.M. Wójcicki. J-index. a new proposition of the assessment of current glucose control in diabetic patients. *Horm. Metab. Res*, 27:41–42, 1995.

[66] World Health Organization. http://www.who.int/diabetes/en/.

[67] C. Zecchin, A. Facchinetti, G. Sparacino, G. De Nicolao, and C. Cobelli. Neural network incorporating meal information improves accuracy of short-time prediction of glucose concentration. *IEEE TMBE*, In Second Revision .

[68] H. Zou, T. Hastie, and R. Tibshirani. Sparse principal component analysis. *Jcgs*, 15: 262–286, 2006.

Acknowledgements

I would like to thank my mentor and advisor, Professor Giovanni Sparacino, for the precious help and guidance that he has offered me in the last four years. Thank you for the patience, the support and for the trust you always gave me.

All my colleagues and friends PhD students and post-docs at University of Padova. Two special acknowledgments: the first to Dr. Andrea Facchinetti, my travel-mate, for the wonderful working experiences we shared (and for waiting for me every time I was held by immigration during our trips) and the second to Dr. Chiara Dalla Man, for the great scientific insights that she always shares so generously with her trainees (and for making the EU project meetings even fun). It was a pleasure and a honor to work with you. Thanks to Michele Schiavon and Mattia Zanon for the constructive work done together on the Dynamic Risk (and for making fun of me whenever I grumble).

Thank you to all those friends who crossed my path during these three years, helping me focusing also on life outside academia: to Selena and Tiziana, my angel-devil friends, to all my friends at Salsa de Calle, for reminding me that dancing is freedom and pure joy.

A special thanks of course to my family: Andrea ("Santa Pazienza") and my parents (and, well, ok you too Carlo). You gave me trust and support in so many different ways, and each little part of it helped me growing into a better person.