Università
degli Studi
di Padova

Dipartimento
di Ingegneria
dell'Informazione

Non-Invasive Continuous Glucose Monitoring: Identification of Models for Multi-Sensor Systems

School Director
Prof. Matteo Bertocco

Bioengineering Coordinator
Prof. Giovanni Sparacino

Advisor
Prof. Giovanni Sparacino

Ph.D. candidate
Mattia Zanon

Ph.D. School in
Information Engineering
XXV Series, 2013

Summary

Diabetes is a disease that undermines the normal regulation of glucose levels in the blood. In people with diabetes, the body does not secrete insulin (Type 1 diabetes) or derangements occur in both insulin secretion and action (Type 2 diabetes). In spite of the therapy, which is mainly based on controlled regimens of insulin and drug administration, diet, and physical exercise, tuned according to self-monitoring of blood glucose (SMBG) levels 3-4 times a day, blood glucose concentration often exceeds the normal range thresholds of 70-180 mg/dL. While hyperglycaemia mostly affects long-term complications (such as neuropathy, retinopathy, cardiovascular, and heart diseases), hypoglycaemia can be very dangerous in the short-term and, in the worst-case scenario, may bring the patient into hypoglycaemic coma. New scenarios in diabetes treatment have been opened in the last 15 years, when continuous glucose monitoring (CGM) sensors, able to monitor glucose concentration continuously (i.e. with a reading every 1 to 5 min) over several days, entered clinical research. CGM sensors can be used both retrospectively, e.g., to optimize the metabolic control, and in real-time applications, e.g., in the "smart" CGM sensors, able to generate alerts when glucose concentrations are predicted to exceed the normal range thresholds or in the so-called "artificial pancreas". Most CGM sensors exploit needles and are thus invasive, although minimally. In order to improve patients comfort, Non-Invasive Continuous Glucose Monitoring (NI-CGM) technologies have been widely investigated in the last years and their ability to monitor glucose changes in the human body has been demonstrated under highly controlled (e.g. in-clinic) conditions. As soon as these conditions become less favourable (e.g. in daily-life use) several problems have been experienced that can be associated with physiological and environmental perturbations. To tackle this issue, the multisensor concept received greater attention in the last few years. A multisensor consists in the embedding of sensors of different nature within the same device, allowing the measurement of endogenous (glucose, skin perfusion, sweating, movement, etc.) as well as exogenous (temperature, humidity, etc.) factors. The main glucose related signals and those measuring specific detrimental processes have to be combined through a suitable mathematical model with the final goal of estimating glucose non-invasively. White-box models, where differential equations are used to describe the internal behavior of the system, can be rarely considered to combine multisensor measurements because a physical/mechanistic model linking multisensor data to glucose is not easily available. A more viable approach considers black-box models, which do not describe the internal mechanisms of the system under study, but rather depict how the inputs (channels from the non-invasive device) determine the output (estimated glucose values) through a transfer function (which we restrict to the class

of multivariate linear models). Unfortunately, numerical problems usually arise in the identification of model parameters, since the multisensor channels are highly correlated (especially for spectroscopy based devices) and for the potentially high dimension of the measurement space.

The aim of the thesis is to investigate and evaluate different techniques usable for the identification of the multivariate linear regression models parameters linking multisensor data and glucose. In particular, the following methods are considered: Ordinary Least Squares (OLS); Partial Least Squares (PLS); the Least Absolute Shrinkage and Selection Operator (LASSO) based on ℓ_1 norm regularization; Ridge regression based on ℓ_2 norm regularization; Elastic Net (EN), based on the combination of the two previous norms. As a case study, we consider data from the Multisensor device mainly based on dielectric and optical sensors developed by Solianis Monitoring AG (Zurich, Switzerland) which partially sponsored the PhD scholarship. Solianis Monitoring AG IP portfolio is now held by Biovotion AG (Zurich, Switzerland). Forty-five recording sessions provided by Solianis Monitoring AG and collected in 6 diabetic human beings undertaken hypo and hyperglycaemic protocols performed at the University Hospital Zurich are considered. The models identified with the aforementioned techniques using a data subset are then assessed against an independent test data subset. Results show that methods controlling complexity outperform OLS during model test. In general, regularization techniques outperform PLS, especially those embedding the ℓ_1 norm (LASSO end EN), because they set many channel weights to zero thus resulting more robust to occasional spikes occurring in the Multisensor channels. In particular, the EN model results the best one, sharing both the properties of sparseness and the grouping effect induced by the ℓ_1 and ℓ_2 norms respectively. In general, results indicate that, although the performance, in terms of overall accuracy, is not yet comparable with that of SMBG enzyme-based needle sensors, the Multisensor platform combined with the Elastic-Net (EN) models is a valid tool for the real-time monitoring of glycaemic trends. An effective application concerns the complement of sparse SMBG measures with glucose trend information within the recently developed concept of dynamic risk for the correct judgment of dangerous events such as hypoglycaemia.

The body of the thesis is organized into three main parts: Part I (including Chapters 1 to 4), first gives an introduction of the diabetes disease and of the current technologies for NI-CGM (including the Multisensor device by Solianis) and then states the aims of the thesis; Part II (which includes Chapters 5 to 9), first describes some of the issues to be faced in high dimensional regression problems, and then presents OLS, PLS, LASSO, Ridge and EN using a tutorial example to highlight their advantages and drawbacks;

Finally, Part III (including Chapters 10-12), presents the case study with the data set and results. Some concluding remarks and possible future developments end the thesis. In particular, a Monte Carlo procedure to evaluate robustness of the calibration procedure for the Solianis Multisensor device is proposed, together with a new cost function to be used for identifying models.

Sommario

Il diabete è una malattia che compromette la normale regolazione dei livelli di glucosio nel sangue. Nelle persone diabetiche, il corpo non secerne insulina (diabete di tipo 1) o si verificano delle alterazioni sia nella secrezione che nell'azione dell'insulina stessa (diabete di tipo 2). La terapia si basa principalmente su somministrazione di insulina e farmaci, dieta ed esercizio fisico, modulati in base alla misurazione dei livelli di glucosio nel sangue 3-4 volte al giorno attraverso metodo finger-prick. Nonostante ciò, la concentrazione di glucosio nel sangue supera spesso le soglie di normalità di 70-180 mg/dL. Mentre l'iperglicemia implica complicanze a lungo termine (come ad esempio neuropatia, retinopatia, malattie cardiovascolari e cardiache), l'ipoglicemia può essere molto pericolosa nel breve termine e, nel peggiore dei casi, portare il paziente in coma ipoglicemico. Nuovi scenari nella cura del diabete si sono affacciati negli ultimi 10 anni, quando sensori per il monitoraggio continuo della glucemia sono entrati nella fase di sperimentazione clinica. Questi sensori sono in grado di monitorare le concentrazioni di glucosio nel sangue con una lettura ogni 1-5 minuti per diversi giorni, permettendo un analisi sia retrospettiva, ad esempio per ottimizzare il controllo metabolico, che in tempo reale, per generare avvisi quando viene predetta l'uscita dalla normale banda euglicemica, e nel cosiddetto pancreas artificiale. La maggior parte di questi sensori per il monitoraggio continuo della glicemia sono minimatmente invasivi perchè sfruttano un piccolo ago inserito sottocute. Gli ultimi anni hanno visto un crescente interesse verso tecnologie non invasive per il monitoraggio continuo della glicemia, con l'obiettivo di migliorare il comfort del paziente. La loro capacità di monitorare i cambiamenti di glucosio nel corpo umano è stata dimostrata in condizioni altamente controllate tipiche di un'infrastruttura clinica. Non appena queste condizioni diventano meno favorevoli (ad esempio durante un uso quotidiano di queste tecnologie), sorgono diversi problemi associati a perturbazioni fisiologiche ed ambientali. Per affrontare questo problema, negli ultimi anni il concetto di "multisensore" ha ottenuto un crescente interesse. Esso consiste nell'integrazione di sensori di diversa natura all'interno dello stesso dispositivo, permettendo la misurazione di fattori endogeni (glucosio, perfusione del sangue, sudorazione, movimento, ecc) ed esogeni (temperatura, umidità, ecc). I segnali maggiormente correlati con il glucosio e quelli legati agli altri processi sono combinati con un opportuno modello matematico con l'obiettivo finale di stimare la glicemia in modo non invasivo. Modelli di sistema (o a "scatola bianca"), nei quali equazioni differenziali descrivono il comportamento interno del sistema, possono essere considerati raramente. Infatti, un modello fisico/meccanicistico legante i dati misurati dal multisensore con il glucosio non è facilmente disponibile. Un differente approccio vede l'impiego di modelli di dati (o a "scatola nera") che descrivono il sistema in esame in termini di ingressi (canali misurati dal dispositivo non invasivo), uscita (valori stimati di glucosio) e funzione di trasferimento (che in questa tesi si limita alla classe dei modelli di regressione lineari multivariati). In fase di identificazione dei parametri del modello potrebbero insorgere problemi numerici legati alla collinearità tra sottoinsiemi dei canali misurati dai multisensori (in particolare per i dispositivi basati su spettroscopia) e per la dimensione potenzialmente elevata dello spazio delle misure.

L'obiettivo della tesi di dottorato è di investigare e valutare diverse tecniche per l'identificazione del modello di regressione lineare multivariata con lo scopo di stimare i livelli di glicemia non invasivamente. In particolare, i seguenti metodi sono considerati: Ordinary Least Squares (OLS), Partial Least Squares (PLS), the Least Absolute Shrinkage and Selection Operator (LASSO) basato sulla regolarizzazione con norma ℓ_1; Ridge basato sulla regolarizzazione con norma ℓ_2; Elastic-Net (EN) basato sulla combinazione delle due norme precedenti. Come caso di studio per l'applicazione delle metodologie proposte, consideriamo i dati misurati dal dispositivo multisensore, principalmente basato su sensori dielettrici ed ottici, sviluppato dall'azienda Solianis Monitoring AG (Zurigo, Svizzera), che ha parzialmente sostenuto gli oneri finanziari legati al progetto di dottorato durante il quale questa tesi è stata sviluppata. La tecnologia del multisensore e la proprietà intellettuale di Solianis sono ora detenute da Biovotion AG (Zurigo, Svizzera). Solianis Monitoring AG ha fornito quarantacinque sessioni sperimentali collezionate da 6 pazienti soggetti a protocolli ipo ed iperglicemici presso l'University Hospital Zurich. I modelli identificati con le tecniche di cui sopra, sono testati con un insieme di dati diverso da quello utilizzato per l'identificazione dei modelli stessi. I risultati dimostrano che i metodi di controllo della complessità hanno accuratezza maggiore rispetto ad OLS. In generale, le tecniche basate su regolarizzazione sono migliori rispetto a PLS. In particolare, quelle che sfruttano la norma ℓ_1 (LASSO ed EN), pongono molti coefficienti del modello a zero rendendo i profili stimati di glucosio più robusti a rumore occasionale che interessa alcuni canali del multi-sensore. In particolare, il modello EN risulta il migliore, condividendo sia le proprietà di sparsità e l'effetto raggruppamento indotte rispettivamente dalle norme ℓ_1 ed ℓ_2. In generale, i risultati indicano che, anche se le prestazioni, in termini di accuratezza dei profili di glucosio stimati, non sono ancora confrontabili con quelle dei sensori basati su aghi, la piattaforma multisensore combinata con il modello EN è un valido strumento per il monitoraggio in tempo reale dei trend glicemici. Una possibile applicazione si basa sull'utilizzo del'informazione dei trend glicemici per completare misure rade effettuate con metodi finger-prick. Sfruttando il concetto di rischio dinamico recentemente sviluppato, e' possibile dare una corretta valutazione di eventi potenzialmente pericolosi come l'ipoglicemia.

La tesi si articola in tre parti principali: Parte I (che comprende i Capitoli 1-4), fornisce inizialmente un'introduzione sul diabete, una recensione delle attuali tecnologie per il monitoraggio non-invasivo della glicemia (incluso il dispositivo multisensore di Solianis) e gli obiettivi della tesi; Parte II (che comprende i Capitoli 5-9), presenta alcune delle difficoltà affrontate quando si lavora con problemi di regressione su dati di grandi dimensioni, per poi presentare OLS, PLS, LASSO, Ridge e EN sfruttando un esempio tutorial per evidenziarne vantaggi e svantaggi. Infine, Parte III, (Capitoli 10-12) presenta il set di dati del caso di studio ed i risultati. Alcune note conclusive e possibili sviluppi futuri terminano la tesi. In particolare, vengono brevemente illustrate una metodologia basata su simulazioni Monte Carlo per valutare la robustezza della calibrazione del modello e l'utilizzo di un nuova nuova funzione obiettivo per l'identificazione dei modelli.

List of Abbreviations

WHO	World Health Organization
BGL	Blood Glucose Levels
NIR	Near InfraRed
MIR	Mid InfraRed
CGM	Continuous Glucose Monitoring
NI-CGM	Non-Invasive Continuous Glucose Monitoring
IDDM	Insulin Dependent Diabetes Mellitus
IS	Impedance Spectroscopy
DS	Dielectric Spectroscopy
LAR	Least Angle Regression
LASSO	Least Absolute Shrinkage and Selection Operator
MAD	Mean Absolute Difference
MARD	Mean Absolute Relative Difference
MSE	Mean Square Error
NI-CGM	Non Invasive Continuous Glucose Monitoring
NIDDM	Non-Insulin Dependent Diabetes Mellitus
OCT	Optical Coherence Tomography
OLS	Ordinary Least Squares
PLS	Partial Least Squares

RMSE	Root Mean Square Error
RSS	Residual Sum of Squares
ESOD	Energy of the Second Order Differences
EN	Elastic-Net
SMBG	Self-Monitoring Blood Glucose
MC	Monte Carlo

Contents

Part I

Background and Aim of the Thesis

1

Diabetes and Continuous Glucose Monitoring

According to the World Health Organization (WHO), diabetes is estimated to currently affect 347 million of people in the world and this number is expected to increase by two third in 2030 [1]. Diabetes and its complications are considered major causes of early death in most countries, with over four million deaths per year [2]. From an economic point of view, the cost of diabetes ranges from 6 to 15 % of the budget of national health systems in the EU, explaining why it is considered one of the most challenging socio-health emergencies of the 3rd millennium [3]. This chapter gives an overview of the diabetes disease and of its therapy. In this context, the importance of Continuous Glucose Monitoring (CGM) sensors is highlighted, together with a proposal of classification according to their degree of invasiveness.

1.1 The Diabetes Disease

1.1.1 The Glucose-Insulin Regulatory System

The glucose substrate represents the main source of fuel for the human body. Thanks to a complex regulatory mechanism, glucose concentration in blood of healthy subjects is tightly kept in a limited rage, i.e. 70-180 mg/dL, although it is subject to fluctuations due to utilization and production processes. Different hormones are involved in this regulation. The most important one is insulin, which is produced by the beta-cells of the pancreas, and is responsible for lowering the glucose concentrations. Insulin is also the principal control signal for conversion of glucose to glycogen for internal storage in liver [4].

As depicted in Figure 1.1, glucose is used by many organs, tissues and cells. Some, like brain or red blood cells, consume glucose continuously and independently of insulin and the interruption of this supplying may cause severe damages. For muscles, fatty tissue and liver the absorption of glucose is proportional to insulin concentration.

Glucose in blood derives both from intestinal absorption of carbohydrates and from internal production. In particular, the latter consists in the conversion to glucose of glycogen stored in the liver or in the so-called gluconeogenesis (the "re-construction" of glucose using substrate derived from glucose degradation).

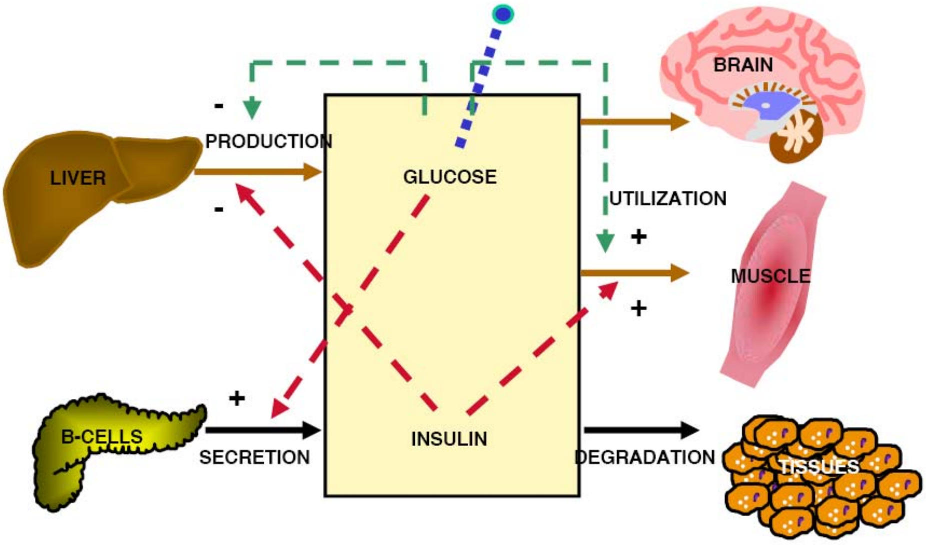


Figure 1.1: Scheme of the glucose-insulin regulatory system. Continuous arrows represent fluxes. In particular, brown ones are referred to glucose, while black ones to insulin. Dashed arrows represent the positive and negative control, indicated with "+" and "-" respectively. The green dotted arrows highlight the self-control employed by a substance, while red dotted arrows indicate the control of a substance over the other one. The blue dotted line represents the measurement site.

An increase in blood glucose concentration causes an increase in insulin secretion. Glucose and insulin concentration have the same effect on the glucose production and utilization: an increase in insulin (or glucose) concentration causes a decrease of glucose production and an increase of glucose utilization by muscle, while there is no influence on glucose utilization by brain.

1.1.2 Types of Diabetes

In people with diabetes, either the pancreas produces little or no insulin (type 1 diabetes), or the cells do not respond appropriately to the insulin that is produced (type 2 diabetes).

In particular, “*Type 1 diabetes*”, or Insulin Dependent Diabetes Mellitus (IDDM) is characterized by loss of the insulin-producing beta cells or the islets of Langherhans in the pancreas leading to insulin deficiency. In most cases, type 1 diabetes has an autoimmune origin and affects children or young adults, and in fact it is also called “juvenile diabetes”. Instead, “*Type 2 diabetes*”, or Non-Insulin Dependent Diabetes Mellitus (NIDDM), is characterized by insulin resistance which may be combined with relatively reduced insulin secretion. Insulin resistance corresponds to a loss of efficacy of insulin action, causing a reduced transport of glucose from the bloodstream into the cells. It is frequently associated with obesity and a sedentary lifestyle. Type 2 is the most common diabetes type (90% of cases) and mostly affects adult people.

1.1.3 Diabetes-Related Complications

A failure of glucose counter-regulatory system causes Blood Glucose Levels (BGL) to exceed the euglycaemic range. Hypoglycaemia and hyperglycaemia might lead to short and long term complications, respectively.

Hyperglycaemia has no immediate damaging consequence on organism, but, if this state is frequent and persist for long time, can lead to several invalidating complications. These long term complications include micro-vascular complications (involving small blood vessels) and macro-vascular complications (involving large blood vessels) [5]. The former, like neuropathy, nephropathy and retinopathy can lead to nerves damage, renal failure and blindness respectively, the latter to coronary heart disease, strokes and peripheral vascular disease. In order to prevent the onset of these complications, diabetes therapies attempt to keep BGL within the euglycemic range. This can usually be done with close dietary management, physical activity and use of appropriate medications, like insulin injections before meals. The association of faulty glucose regulatory system and neglectfully therapy could cause, principally during sleep hours and physical activity, an even more dangerous unfavorable effect, i.e. hypoglycemia (i.e. too low blood glucose level).

Hypoglycemia affects mostly the brain, given its continuous glucose demand. Therefore, when glucose levels fall, brain functions diminish and people may lose cognitive abilities and in the worst case scenario go into the so-called hypoglycaemic coma. Hypoglycemia, at the opposite of hyperglycemia, has mainly short-term effects [6] and could be classified according to the level of awareness:

- *mild hypoglycemia* (blood glucose levels between 55 and 70 mg/dL) is characterized by palpitations, extreme hunger, trembling, cold or excessive sweating and visual paleness, due to blood redirection to the vital organs and minimization of the peripheral blood circulation. In this case a small amount of carbohydrates eaten or drunk could restore normal levels;
- *moderate hypoglycemia* (between 55 and 40 mg/dL), whose symptoms include mood changes, irritability, confusion, blurred vision, weakness and drowsiness since it affects the central nervous system;
- *severe hypoglycemia* (less than 40 mg/dL) is characterized by convulsions, loss of consciousness, coma, and hypothermia. If this condition is prolonged in time could cause irreversible brain damages and heart problems, or even death. In this case, intravenous dextrose or an injection of glucagon is required.

1.1.4 Diabetes Therapies and Glucose Monitoring

In the near future, new technologies will play a crucial role in diabetes management to contrast human and socio-economical costs of this disease [7].

For type 1 diabetes, conventional therapies consist in insulin injections for compensating the lack of insulin secretion and have the goal to restore euglycaemic levels. A suitable dosage is determined using information on food intakes and current BGL. In the early stage of type 2 diabetes, a diet modification and physical exercise, associated with medications improving insulin sensitivity, may be sufficient to control glycaemic levels. If diabetes proceeds, exogenous insulin injections may be needed. In both cases, monitoring BGL is important. Indeed, several clinical studies demonstrated that long and short term complications can be reduced through a therapy based on diet, physical exercise, and drug delivery (including subcutaneous injections of exogenous insulin), tuned according to the monitoring of individual parameters [2]. The most used approach is based on the measure of glycaemia 3-4 times per day. This is referred as Self-Monitoring Blood Glucose (SMBG), i.e. the patients have to take a finger-prick blood sample on specific strips and measure BGL with a dedicated device [8]. SMBG measures are collected by the patient and then analyzed and interpreted retrospectively by the physician during periodic visits where the current therapy is revised accordingly. SMBG traces can also be analyzed retrospectively for assessing glucose variability [9]. However, a suitable time window of several months must be considered for having a reasonable number of data points.

A SMBG measure can also be used in real-time by the patient to assess the current glycaemic state. However, the sparseness of these measures does not give a complete information about the glycaemic range excursion and dynamics, leading to potentially dangerous hypo/hyper glycaemic events without any patient's awareness [10].

Self Monitoring Blood Glucose

The most common test for measuring BGL involves pricking a finger with a lancet device to obtain a small blood sample, applying a drop of blood onto a reagent test-strip, and determining the glucose concentration by inserting the strip into a measurement device. Different manufacturers use different technologies, but most systems measure an electrical characteristic proportional to the amount of glucose in the blood sample [8].

Intermittent glucose sampling can be achieved also through other physiological fluids, such as saliva, urine, sweat or tears [11]. However, in these cases, delay in the appearance of glucose in these fluids must be taken into account.

SMBG systems make a direct measure, i.e. they measure a specific property of glucose. This means that if the same property is investigated for another kind of substance, a significantly different output is produced than the one obtained from glucose. Spectral, chemical and competitive binding properties of glucose are considered to infer on blood glucose concentrations.

Direct measurements tend to be more stable than indirect ones because the signal being measured is usually unique and interferences more predictable. In fact, indirect measurements are affected by the presence of other chemicals and substances within the body that may produce the same signal, since they measure glucose effect on some secondary process [12].

Continuous Glucose Monitoring

The main drawback of SMBG is the lack of glucose measures during sleeping or daily-life activities, leading to time intervals with no informations on the glucose levels. During these intervals, dangerous hypo-/hyper-glycaemic excursions may happen without awareness for the patient. With the aim of preventing these episodes, in the last decade many devices for CGM have been developed allowing to monitor glucose fluctuations continuously with a minimum level of invasiveness.

The main advantage of CGM is the possibility to monitor BGL in a nearly continuous way, i.e. every 1 to 5 minutes, for a long period of time, i.e. 7 consecutive days. CGM time-series have been studied retrospectively for analyzing glucose variability [13, 14]. Moreover, the clinical benefit of wearing CGM devices has been demonstrated in [15, 16], showing an improvement of the glycaemic control with a decreasing of the glycated hemoglobin HbA1c (a marker of the glycamic control predictive of diabetes related complications).

More appealing are on-line applications of such technologies. In the last years, several algorithms and signal processing techniques have been developed or adapted from other fields to improve accuracy and reliability of CGM data, see [17, 18, 19]. An example is the so-called "smart" CGM architecture [20]. It consists in a cascade of independent software modules down line of the commercial CGM sensor which allow to de-noise, enhance and predict glucose levels, see e.g. [21, 22, 23, 24, 25, 26, 27, 28, 29] for examples of on-line algorithms developed for CGM. CGM are fundamental in the development of artificial pancreas, which implements a closed-loop control that has the aim to infuse the correct amount of insulin subcutaneously using a micro-infusor driven by a control algorithm, which, in turn, exploits the measurements provided by a CGM sensor as its input [30, 31, 32].

CGM are appealing for several reasons related to their degree of invasiveness and the *quasi* continuous information they provide. However, given their current performance they are still considered a complement and not a replacement of SMBG devices [33].

1.2 A Classification of Sensors for Continuous Glucose Monitoring (CGM)

CGM sensors can be classified according to: a) the kind of measure (direct or indirect); b) to the level of invasiveness; c) to the physical principle the sensor is based on. In Figure 1.2 we propose a classification scheme of existing CGM sensors according to their level of invasiveness, highlighting the physical principle or technology each sensor is based on. The following review is far from being exhaustive and a complete descriptions and reviews on the working principle, pros and cons, and future perspectives on CGM sensors can be found in [34, 35, 36, 37, 38].

1.2.1 Invasive CGM Sensors

As shown in Figure 1.2, a direct measurement of BGL could be obtained invasively by using sensors implanted into the body [39]. These sensors are extremely accurate, but given their level of invasiveness they are particularly suited for Operating Rooms and Intensive Care Units [40]. There are different technologies allowing to transduce

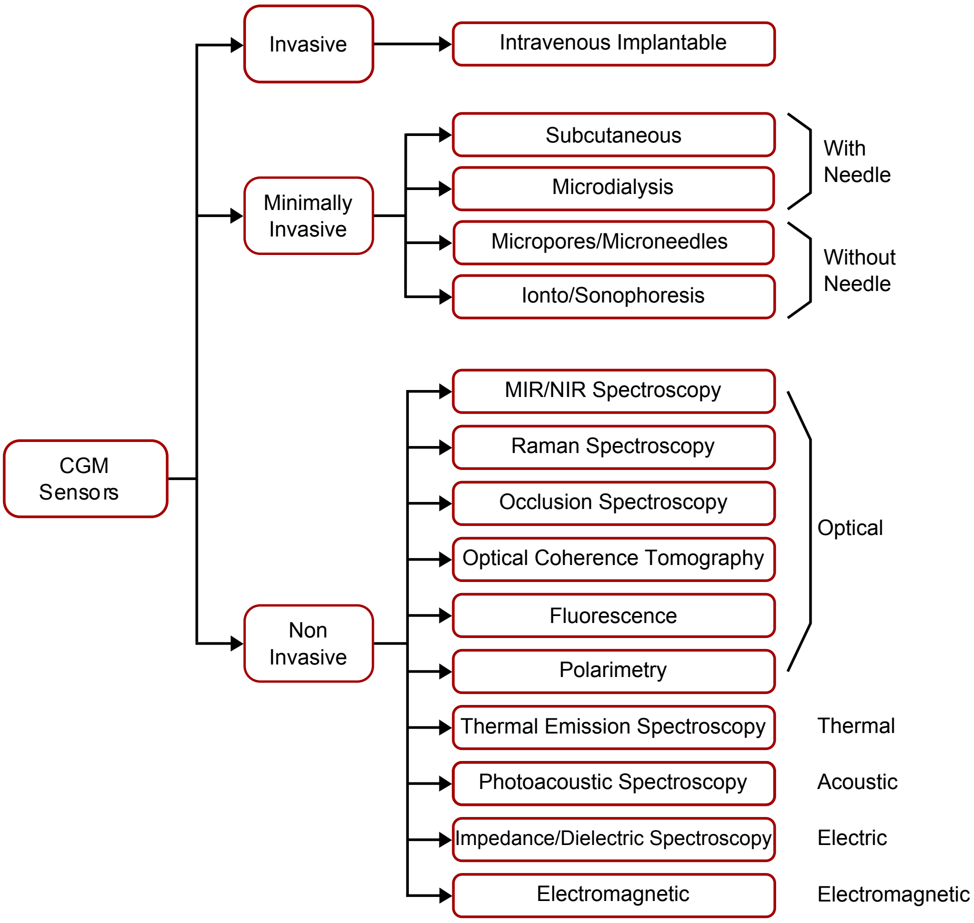


Figure 1.2: A Proposed CGM sensors classification.

glucose concentration into an electrical signal, most of them are based on glucose-oxidase principle. Other sensors are based on competitive binding of glucose with other molecules or glucose spectral properties [12].

Intravenous Implantable

Glucose oxidase-based sensors technology depends on the reaction of glucose with oxygen in presence of glucose oxidase to create gluconic acid. The limitation of using this method is that the reaction requires one oxygen molecule for each glucose molecule. Since glucose is more present in the body than oxygen, the limiting reagent results to be the oxygen. For this reason, the sensor would measure oxygen levels instead of glucose levels. To avoid this problem, sensors must give oxygen an advantage over glucose, using alternative electron donors, called mediators.

The competitive binding-based sensor measures fluorescence of a binding molecule: the more glucose is bound to this molecule, the less intense is the fluorescent signal so that if glucose levels increase the measure decreases. This technique has still problems related to biocompatibility and to the risk inherent to surgical placement of these devices in blood vessels, hence it is not widely applied [41]. An additional fluorescence-based intravenous glucose sensor is presented in [42].

A new intravascular continuous glucose monitoring system is under development, using a glucose-sensitive hydrogel. When this hydrogel is bound with glucose, it changes in volume. The result is a measurable change in the hydrogel impedance that is correlated to glucose concentration. Preliminary studies have been made on a prototype of the sensor, integrated with stents as antennas for wireless data transfer from within the body [43].

1.2.2 Minimally-Invasive CGM Sensors (with needle)

There is no unanimous agreement in the literature about which kind of CGM sensors should be considered as minimally-invasive. According to the proposed sensor classification scheme, we will refer to minimally-invasive sensors (with needle), those requiring a needle inserted in the subcutis, such as subcutaneous and microdialysis (see Section 1.2.2). On the other side, minimally-invasive sensors (without needle) will be those requiring the creation of microscopic holes in the skin to perform the measurement without the need to insert needles under the skin (see Section 1.2.3).

A common limitation of all these sensors is the delay between plasma and interstitial glucose concentration. This phenomenon is due to the glucose transport from plasma to interstitium that act as a low pass filter, see [44] for details of plasma-to-interstitium glucose kinetics.

1.2.2.1 Subcutaneous Sensors

Instead of implanting the sensor into the body, a subcutaneous needle may be used to sense glucose. Usually these systems are based on enzyme electrodes, most of time exploiting a glucose-oxidase principle. Subcutaneous needles provide much more information about dynamics and glucose excursions if they are compared to a finger-prick system, providing readings every 1-5 minutes for up to 7 days. These sensors require frequent calibration to compensate drifts due to protein and cell coating of the sensor, variable tissue oxygen tension and wound response to the sensor, which alters local blood flow. From a signal processing point of view, several algorithms have been developed to deal with these calibration issues, see [18, 19] for a review. To perform periodic calibrations of the sensor, a measurement using traditional SMBG systems is usually required.

Examples of commercially available subcutaneous sensors include the FreeStyle Navigator ^{TM}(Abbott Laboratories, Alameda, CA, USA), the MiniMed Guardian Real-Time (Medtronic MiniMed, Northridge, CA, USA) and the the Dexcom® Seven® and SevenPlus® (DexCom Inc., San Diego, CA, USA) ®, to mention but a few.

Figure 1.3: *FreeStyle Navigator* CGM System[45]. Miniature electrochemical sensor placed in the subcutaneous adipose tissue (bottom left), a disposable sensor delivery unit (top right), a radiofrequency transmitter connected to the sensor (bottom right), and a hand-held receiver to display continuous glucose values.

The FreeStyle Navigator ^{TM}CGM System consists of four components (see Figure 1.3): a miniature electrochemical sensor placed in the subcutaneous adipose tissue, a disposable sensor delivery unit, a radiofrequency transmitter connected to the sensor, and a hand-held receiver to display continuous glucose values [46]. The sensor can be used for 5 days, the glucose data on the receiver are updated once a minute and include a trend arrow to indicate the direction and rate of change averaged over the preceding 15 min. The user interface of the receiver allows the threshold alarms to be set at different glucose levels. The receiver contains a built-in Free-Style blood glucose meter for calibration of the sensor as well as for confirmatory blood glucose measurements. The sensor requires four calibrations over the 5-day wearing period at 10, 12, 24, and 72 h after sensor insertion. It was approved by FDA in 2008 [47].

The DexCom ® SevenPlus® sensor consists of three parts (see Figure 1.4): a small sensor placed in the subcutaneous adipose tissue, a wireless transmitter and a receiver [48]. It performs a new measure every 5 minutes for 7 days. The receiver displays the sensor glucose value along with a graph showing glucose trend of the last 1, 3 or 9 hours. The receiver contains memory up to 30 days of continuous glucose information and

Figure 1.4: *DexCom* ® *SevenPlus*® sensor. The receiver(left) and the and transmitter (right). The subcutaneous sensor is not shown [48].

has programmable high and low glucose alerts and a non-changeable low glucose alarm set at 55 mg/dL. It must be calibrated every 12 hours. It was approved by FDA in 2009 [49]. The same company recently produced the *DexCom* ® *G4 PLATINUM*, a CGM sensor with improved performance with respect to the SevenPlus®, according to their website [48].

Figure 1.5: The *Guardian REAL-Time* [50]. REAL-Time CGM System monitor (left), the MiniLink REAL-Time Transmitter together with the glucose sensor inserted in the subcutis (right).

The Guardian Real-Time device consists of the Guardian® REAL-Time CGM System monitor (Figure 1.5, left), the MiniLink REAL-Time Transmitter (Figure 1.5, right) and of the glucose sensor inserted in the subcutis. This sensor performs a new measure every 5 minutes for 3 days [50]. The receiver contains memory up to 21 days of continuous glucose information and has alerts if a glucose level falls below or rises above preset values. It must be calibrated every 12 hours either manually or automatically via telemetry. It was approved by FDA in 2005. This sensor, integrated with an insulin delivery device composes the *MiniMed Paradigm REAL-Time* system, that was launched in 2006.

Recently, the CGM has been complemented with an insulin pump to provide the MiniMed Paradigm Real-Time Insulin Pump and Glucose Monitoring System [51].

1.2.2.2 Microdialysis

Another type of minimally invasive subcutaneous CGM sensor exploiting a needle is based on a microdialysis system, which uses a fine, hollow microdialysis fibre placed subcutaneously. This probe is perfused with isotonic fluid from an external pool, while glucose, present in the interstitial fluid, freely diffuses into the fibre, where it is pumped out of body to a glucose-oxidase based sensor [52]. The main problem related to this kind of sensor consists in modifications of chemical and physical properties of the membrane, caused by modifications in tissues characteristics such as pressure, volume, temperature and hydration. These modifications affect flow rate and composition of perfusate, which may influence glucose concentration.

The GlucoDay® by Menarini Diagnostics (Florence, Italy) is a microdialysis-based glucose monitoring system [53]. It is based on enzymatic-amperometric measurement analyzing the fluid coming from the subcutis of the abdominal region. The system comprises a walkman-size apparatus, and a sensor fibre as well as two plastic bags (one for the buffer solution, one for the waste products) as disposables. The apparatus contains also a measurement cell and a peristaltic pump. The buffer solution is pumped from a bag into the subcutaneous tissue through the microfibre and rinses the interstitial fluid, from which the measurements are obtained every 3 min and stored in memory. Data are downloaded after monitoring (maximum monitoring time, 48 h) via a serial or infrared connection to a standard PC for further analysis. It incorporates safety alarms for hypo or hyperglycaemia events and requires one daily calibration. Recently, the same company launched the GlucoMen®Day (currently waiting the CE mark), which overcomes various shortcomings of its predecessor [54]. It is smaller and more compact, and has a longer lifetime (100 hours), is more stable and embeds different algorithms for signal processing and data management [55, 56].

The SCGM 1 sensor (Roche Diagnostics, Mannheim, Germany) is also based on a microdialysis principle [57].

1.2.3 Minimally-Invasive CGM Sensors (without needle)

This section presents minimally-invasive sensors not presenting needles, but exploiting technologies for creating microscopic holes allowing glucose molecules to pass through. For this reason, this class of sensors is not regarded as fully non-invasive.

1.2.3.1 Micropores and Microneedle

Micropores techniques perforate the stratum corneum without penetrating the full thickness of the skin. A pulsed laser or the local application of heat are considered to form micropores allowing the collection of interstitial fluid applying vacuum. A measure of glucose concentration is then derived from this sample.

SpectRx is made mainly of two units. The first unit is a handheld laser, which creates micropores (size of a hair) in the stratum corneum of the skin. The interstitial fluid, containing glucose, flows through the micropores and is collected by a patch. Then, it reaches a traditional glucose sensor, which is the second unit. The meter also includes a transmitter that sends wirelessly the glucose measurements to a handheld display device [37].

Similarly capillary blood could be sampled using a hollow microneedle, which is almost sensation-less and analyses blood using an enzyme-based system.

1.2.3.2 Iontophoresis and Sonophoresis

Among minimally invasive sensors, we also include transdermal methods, which stimulate the skin exploiting different interaction fenomena in order to extract glucose from the skin micropores for its direct measures. This group comprises different techniques like reverse iontophoresis and sonophoresis [37].

Iontophoresis Principle

The first method is based on the flow of a low electrical current applied across the skin between an anode and a cathode positioned on the skin surface [38]. The application of an electrical potential causes the migration of sodium and chloride ions from beneath the skin towards the cathode and anode respectively, at rates significantly grater than passive permeability. The convective flow induced by this technique carries out neutral molecules, including glucose, along with sodium. Thus, interstitial glucose is transported across the skin towards the cathode, where it is collected and measured by a glucose oxidase-based electrode. The concentration of glucose is low so oxygen is not a limiting factor to glucose oxidase. This technique tends to generate skin irritation and cannot be used if the subject is sweating significantly; in addition it needs a long warm-up and calibration. Skin irritation may be limited by shortening the time interval of the electrical potential application. However, a minimum duration is required to get sufficient amount of glucose for measurement.

Iontophoresis-Based Sensors

The *GlucoWatch*® by *Cygnus Inc.* (approved by FDA in 2002, but withdrawn from the market in 2007) device is based on reverse iontophoresis technology [58]. It has a wrist-watch format and measures glucose through the skin using a disposable pad, which clips into the back of the meter. The pad uses an adhesive to stick to the skin allowing it to come in contact with a small electrical current, which causes the reverse iontophoresis, and then the glucose levels in the interstitial fluid can be estimated. Compared with finger-stick readings, the meter measurements have a 15-min lag time. The meter is intended for use to supplement, but not to replace, information obtained from a standard blood glucose meter. The meter has 2-3 h warm-up period, to remove the glucose on the superficial epidermis and to onset a continuous convective flow. A single-point calibration, performed using a fingerstick blood glucose measurement, accounts for variability in both biosensor sensitivity and skin permeability and is used to convert subsequent biosensor measurements into glucose readings. Afterwards, the meter provides readings every 10 min: 3 min of electrical stimulation (glucose extraction), then 7 min of glucose measurement. The meter has a memory that can store up to 8500 records and the data can be download to a PC for a subsequent analysis. An alarm also occurs in the case of a rapid change is seen in the blood sugar, in the case of sweating, and for any measurements above or below the patient's target levels. A trend indicator appears to show the direction of the blood sugar when the current measurement is more than 18 mg/dl higher or lower than the previous measurements. Event markers can be recorded for activities like meals, insulin intake and exercise. However, the meter had several limitations. In fact, the measurements could fail or be inaccurate, if the patient was sweating, or in cases of rapid temperature changes, excessive movement of the meter, or strenuous exercise. Most users reported that the electrical discharge is quite noticeable during the first use of the meter, although it becomes less noticeable on subsequent use. Moreover the disposable pad must be replaced every 12-13 h of monitoring time to ensure continued accuracy; the meter must then go through the warm-up period and calibration again. In addition, it may take more than one try to calibrate the meter, thus requiring additional finger-stick tests. Finally, the meter causes skin irritation to some extent, which limits reuse of the same site to a week or two[59, 60].

A new Reverse Iontophoresis based Glucose Monitoring Device (RIGMD) has been developed in Korea [61]. It measures a weak electric current that is dependent on glucose concentration in the interstitial fluid, by using an electrochemical enzymatic sensor located on the forearm skin. The sensor is made up of electrodes and a gelatinous material which contains glucose oxidase. A current is produced between the electrodes causing reverse iontophoresis [60].

In [62], it is described the results of preliminary experiments for the development of a mediated glucose biosensor incorporated with reverse iontophoresis function for noninvasive glucose monitoring, using an optimum combination of glucose oxidase and ferrocene.

Sonophoresis Principle

The sonophoresis based technologies use low-frequency ultrasounds to create an array of microscopic holes on human skin which increase its permeability and allow the migration of glucose contained in interstitial fluid through the skin to a glucose sensor placed in contact with the skin. Thus a direct measure is feasible [38].

Sonophoresis-Based Sensors

Echo Therapeutics produces a device based on sonophoresis technique [63]. The meter is made essentially of two units: an ultrasonic device (*SonoPrep*), coupled with the skin through an aqueous medium, which increases skin permeation, and a glucose sensor (*Symphony*), which measure glucose in the interstitial fluid reaching the sensor through the micropathways generated on the skin.

SonoPrep is an ultrasonic skin preparation generator, controlled by a microprocessor. This device delivers low-frequency ultrasound (53-56 kHz), which creates a cavitating force at the point of contact with the skin surface. This force reduces transiently the normally robust lipid barrier of normal intact skin, causing the outermost layer of skin to become increasingly conductive and permeable. Since the relationship between skin conductance and skin permeability, the active ultrasound is terminated when the skin reaches the predetermined level of permeability by continuously measuring skin conductance. This ensures that the site is properly prepared without pain, trauma (such as burn), or irritation. It is claimed that the application of the ultrasonic device for 15 s is enough to make the skin permeable for several hours (between 12 and 24 hours) [64].

Prelude SkinPrep System is a new skin preparation device under development (see Figure 1.6, right), that can be used in alternative to *SonoPrep*. The system consists of a disposable abrasive end driven by an electrical motor in a standalone hand piece. Instead of ultrasounds,*Prelude* utilizes a mechanical mean to remove stratum corneum, with the process controlled by the same conductance-based feedback mechanism used in *SonoPrep* [65]. The *Symphony* is a fully functional prototype sensor instrument designed to measure glucose through permeated skin (see Figure 1.6, left). The sensor is able to maintain reliable fluid contact with the skin through a proprietary biocompatible hydrogel, which utilizes glucose oxidase to measure glucose concentration. The sensor is

Figure 1.6: *Right:Prelude SkinPrep System. Left: Symphony* [63].

housed in a wireless transmitter, which acquires, stores, and transmits coded data to the receiver/monitor to display a reading every minute in addition to trends and alarms for excessively high and low BGL [65].

1.2.4 Non-Invasive Continuous Glucose Monitoring NI-CGM Sensors

Non-invasive Continuous Glucose Monitoring NI-CGM sensors measure glucose concentration through the skin without extracting blood or interstitial fluid or without a needle penetrating the skin for reaching these fluids. Hence, these sensors are more comfortable for the patient than the previously described sensors and do not cause displeasing physiological reactions. However, the measure is affected by different confounding factors, making more difficult to perform an accurate measurement.

NI-CGM sensors measure different physical properties of the skin and underlying tissues (optical, thermal, acoustic and electrical) which are modulated by glucose concentration changes. Given the special importance of these sensors in the present thesis, the physical principles of these sensors will be described in detail in the next chapter. For each technology, an example of its application for CGM will be presented. Particular attention will be paid to the multisensor system proposed by Solianis Monitoring AG (Zurich, Switzerland).

2

Non-Invasive Continuous Glucose Monitoring (NI-CGM) Sensors

Non Invasive Continuous Glucose Monitoring (NI-CGM) devices are appealing for obvious reasons related to patient's comfort. Even if they do not present accuracy comparable with that of subcutaneous or microdialisys-based devices yet, in the last years there has been an increasing attention concerning these non invasive technologies and several new prototypes have been designed and developed [37, 60, 66, 38]. For each of these technologies, physical principles and examples of application will be described in the following.

2.1 Physical Principles beyond NI-CGM and Prototypes

NI-CGM sensors measure glucose concentration without extracting blood or interstitial fluid or without a needle penetrating the skin for reaching these fluids. Thus, the measure is performed through the skin that is a particular multi-layer biological tissue. Consequently, to understand the characteristics of these sensors, it is convenient to have a clue of skin morphology and the non-uniform blood distribution within the layers.

2.1.1 Skin Properties

The skin is composed by several distinctive layers as illustrated in Figure 2.1. The uppermost skin layer is the *stratum corneum* of epidermis, composed of dead keratinized cells, followed by the living epidermis and the connective tissue of the dermis. The subcutaneous tissue is composed by an underlying fat layer and muscle. The dermis can be subdivided into three different layers: the upper vascular plexus, the reticular dermis and the deep vascular plexus. The epidermis does not have its own vasculature. The volume fraction occupied by blood vessels in the dermis is in the range of 1-20% and is concentrated in the upper and deep vascular plexus. Most of NI-CGM sensors, e.g. Diasensor [60], TANGTEST [68], OrSense [69], Sentris-100 [70] and other prototypes in development, are optical transducers that use light in variable frequencies to track glucose. They exploit different properties of light to interact with glucose molecules, returning a measure of some optical property proportional to glucose concentration. These optical sensors monitor glucose variations in the dermal blood; hence, the radiation needs to penetrate at least through the epidermis to reach the vascularised compartments of the dermis. Along with these optical sensors, other non-invasive approaches exploit thermal, acoustic and electrical properties. This classification follows the scheme previously proposed in Figure 1.2.

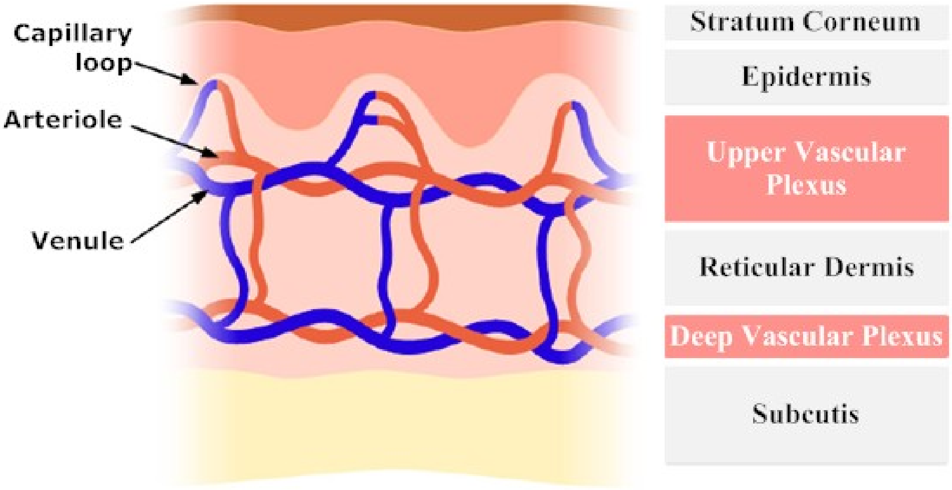


Figure 2.1: Representation of the skin layered structure highlighting the distribution of blood vasculature (left) and description of the most representative skin layers (right) [67].

2.1.2 Optical Techniques for NI-CGM

A beam of light interacts in different ways when it passes through a multilayer tissue like skin. A portion of the beam is reflected by the stratum corneum, another part is absorbed from the tissue and the remaining part is scattered (i.e. it is deviated from the straight trajectory) and diffused into a number of different directions. Figure 2.2 shows a general scheme that summarizes the different kinds of interaction of light with skin.

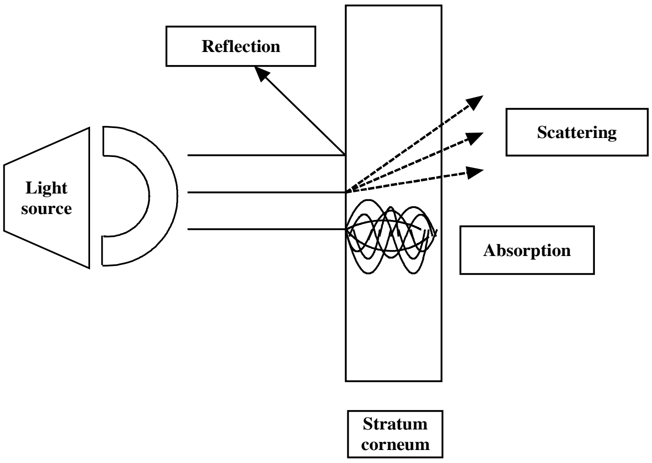


Figure 2.2: Optical properties of light utilized in glucose detection [71]. The light source (left) emits a beam of light which is partially reflected, scattered and absorbed.

Spectroscopy analyses the optical properties of light in relation to the wavelength of the radiation. Spectroscopy also provides a precise analytical method for finding the constituents (and their concentration) in materials having unknown chemical composition, since each substance exhibits characteristic spectra, which may be interpreted as the “fingerprint” of that substance. The different types of spectroscopy may be classified according to which optical properties of the light is employed.

2.1.2.1 MIR/NIR Spectroscopy

Infrared absorption spectroscopy is based on absorption phenomena: changes in glucose concentration can influence the absorption coefficient of tissues and thus the absorption bands [37].

MIR/NIR Spectroscopy Principle

In particular, the so-called Near InfraRed (NIR) spectroscopy uses light in the near infrared range (750-2000 nm). Specific spectra are chosen in order to minimize background absorption, in particular by water. The light in these wavelengths passes through the *stratum corneum* and epidermis into the subcutaneous space, allowing to measure in the deep tissues (in the range of 1 to 100 mm of depth). Perturbing factors that may interfere with glucose measurement include all the variables that influence absorption coefficient, like blood pressure, body temperature and skin hydration. Errors can also occur due to environmental variations such as changes in temperature, humidity, carbon dioxide, and atmospheric pressure. The absorption coefficient of glucose in the NIR band is low and is much smaller than that of water given the large disparity in their respective concentrations. Thus, in NIR measurements, the weak glucose spectral bands not only overlap with the stronger bands of water, but also with those of molecules such as hemoglobin, proteins, and fats [72]. Changes in glucose may affect the measurement process also in other indirect ways: for example, hyperglycaemia causes increased perfusion, which influences the spectrum and can be considered as a confounding factor. Furthermore, diabetic subjects can exhibit “thick skin” and “yellow skin” [73]. Thus, light reflected from skin of a diabetic patient may differ from that of a healthy subject at equal level of glycaemia.

In contrast to NIR, Mid InfraRed (MIR) spectroscopy utilizes light at a wavelength between 2500-10000 nm. With respect to NIR, MIR exhibits less scattering phenomena and greater absorption. Hence the tissue penetration of light in MIR can reach only the *stratum corneum*, but the glucose spectra is less perturbed from interferences from other molecules.

MIR/NIR Spectroscopy-Based Sensors

The *TANGTEST Blood Glucose Meter* seems to be based on NIR technology. This prototype measures glycemia by analyzing intensity variations in the spectrum of a weak light (about 0.1 W) transmitted through the tested hand finger (middle or index finger). In [68], the developers of the device claim that the signal noise due to other tissues is avoided by using the optical signal of pulsatile microcirculation: the signal obtained by the meter is in fact divided into a pulsatile and a direct component. The pulsatile component, which is synchronized with heart rate, is used to monitor blood glucose [74].

The *Diasensor* device is based on operates by placing the patient’s forearm on the arm tray of the meter. The dimensions of the meter are relevant compared with other meters, but it is still sufficiently compact to be used in a domiciliary environment for intermittent glucose monitoring. The blood glucose test is obtained in less than 2 minutes. However, it is not intended as a replacement for the traditional invasive blood glucose meter. It seems that the distributor was EuroSurgical Ltd., UK. However, the web site of the company does not currently mention *Diasensor*, and hence it can be speculated that it is not on sale anymore [60].

InLight Solutions is developing a device based on NIR spectroscopy and multi-variate analysis to make quantitative and qualitative measurements. Appropriate optic and software have been develop to clearly distinguish glucose molecules from water molecules. The devices are made up of three components: a light source, an optical detector, and a spectrometer. The measures are performed using the differences between the light that was sent into the skin and the light that the detector collects [75].

Other companies developing NI-CGM devices are *Pignolo Spa* developing a NIR based device, the *Glycolaser*® [76] and *MedOptix*^{TM} [77] developing a sensor based on proprietary technology at the edge between NIR and MIR spectroscopy.

2.1.2.2 Raman Spectroscopy

Raman Spectroscopy Principle

Raman spectroscopy measures the small fraction of scattered light that shows wavelengths different from the one of the exciting beam. This fraction is dependent on rotational or vibrational energy states within a molecule. Raman spectroscopy shows highly specific absorption bands and, compared with MIR and NIR spectroscopy, it has the benefit of suffering of lower interference from water. However, the Raman signal is weaker than its counterpart in other technologies due to the fact that measured photons normally have lower intensity than the original light and thus requires powerful detectors [66].

Recently, an improvement in traditional Raman spectroscopy has been proposed (surface-enhanced Raman spectroscopy), which may increase the sensitivity of the acquisition. However, it has only been tested in rats [78].

Raman Spectroscopy-Based Sensors

A prototype of sensor based on Raman spectroscopy has been described and tested by Enejder and coworkers [79]. Raman spectra were collected by means of a specially designed instrument, optimized to collect Raman light emitted from a scattering medium (tissue) with high efficiency and a diode laser as the Raman excitation source.

C8 Medisensor has recently developed a device based on Raman spectroscopy and is currently waiting for CE mark approval [80]. It implements three optical sources, accessed through an optical switch, for obtaining information about glucose as well as water and white light for normalization purposes.

2.1.2.3 Occlusion Spectroscopy

Occlusion Spectroscopy Principle

Another technique that measures scattered light is occlusion spectroscopy [71], which is based on the property of glucose to decrease the diffusion coefficient and on the enhanced transmission of light due to erythrocyte aggregation that can be reproduced in vivo by applying a pressure to the fingertip for 2-3 seconds, greater than the systolic one. One signal is collected when no pressure is applied and it is combined with the occlusion signal in order to calculate glucose concentration thanks to a specific algorithm. The advantage of this method is that it measures arterial glucose level. However, intrinsic erythrocyte aggregation and free fatty acid concentration may interfere with the measure. Calibration is needed for glucose predicting parameters using four blood glucose reference points in the first three hours, and an additional reading after 8 hours.

Occlusion Spectroscopy-Based Sensors

A Device based on this technology is the *OrSense NBM-200G* [81], which obtained the European CE mark in 2007.

Figure 2.3: *OrSense NBM-200G* [82]. On the right are visible the annular probe linked to the computation unit.

The measurement is performed using an annular probe, which is positioned on the finger's root and contains light sources, detectors and pneumatic cuffs producing oversystolic pressure to occlude blood flow. The optimization of sensitivity and specificity is achieved by the following:

- *Transmission mode.* In the transmission mode the light traverses the whole organ (finger), and the photons typically encounter many more glucose molecules along their paths than in the reflection mode. This strategy enhances the sensitivity to glucose and reduces the influence of local factors such as skin morphology and pigmentation.

- *Dynamic signal.* Occlusion spectroscopy is based on the generation of an optical signal that changes with time. The signal is induced by oversystolic occlusion at the finger's root, which causes cessation of blood flow throughout the finger. This strategy allows us to collect not only one data point per wavelength, but rather a whole function. It results in a better signal-to-noise ratio.
- *Multispectral data.* Multiple wavelengths of light sources are used. This is beneficial for specificity/selectivity, as the different behaviour of the optical signal among wavelengths allows cleaning the influence of unwanted interferences, such as the absorption of hemoglobin and changes of oxygen saturation.
- *Sophisticated algorithms.* The data are processed with sophisticated algorithms, which use only a small number of parameters, hence avoiding overfitting and false correlations [69].

2.1.2.4 Optical Coherence Tomography

Other types of NI-CGM sensors are based on Optical Coherence Tomography (OCT) that was originally developed to perform the tomographic imaging of the eye.

Optical Coherence Tomography Principle

An OCT system uses a low-power laser source, an interferometer with two arms (reference and sample) and a photodetector to measure the interferometric signal [71, 83].

The skin is irradiated with a low coherence light (light in which the emitted photons are synchronized in time and space). Backscattered radiations from tissues are combined with light returned from reference arm and the resulting interferometric signal is detected by the photodetector. Basically, it measures the delay correlation between the two original signals. Using this technique, glucose concentration in the dermis can be determined, since an increase of glucose concentration in the interstitial fluid causes an increase in the refractive index, thus determining a decrease in the refractive index mismatch.

This technique is affected by motion artifacts. In addition, while small changes in skin temperature have negligible effects, changes of several degrees have a significant influence on the signal.

Optical Coherence Tomography-Based Sensors

The *Sentris-100* device is based on optical coherence tomography technology , exploiting infrared light to scan a cylindrical volume of skin in several steps from the skin surface down to the subcutaneous tissue. Acute changes in protein (collagen and myosin) conformation occur in response to glucose concentration changes and creates a high sensitivity in the optical coherence tomography signal; localization of signal detection to blood vessel walls minimizes any observed signal lag [60, 70].

Figure 2.4: The *Sentris-100.*

2.1.2.5 Fluorescence

Fluorescence technology has also been proposed for glucose monitoring and is based on the generation of fluorescence by human tissue when excited by lights at specific frequencies.

Fluorescence Principle

These sensors are able to measure glucose levels exploiting the dependence between fluorescence intensity and glucose concentration in the solution. Other fluorescence-based glucose sensors are based upon the affinity sensor principle, where glucose and a fluorescein-labelled analogue bind competitively with a receptor site specific for both ligands. Thus, an increase in glucose concentration causes a decrease in the binding of receptor with fluorescein-labelled analogue resulting in a decreased light emission [71].

Fluorescence-Based Sensors

A glucose-sensing contact lens has been developed using boronic acid to measure lachrymal glucose concentration [84]. The main drawback of this system is that it requires a hand-held external light source/detector. Thus, even if theoretically the lens is able to monitor continuously the glucose concentration, the information is carried out only with the detector usage.

Recently an injectable hydrogel microbeads has been developed for fluorescence-based in vivo continuous glucose monitoring. A fluorescent monomer based on diboronic acid has been developed. It enables reversible responsiveness to glucose without any reagents and enzymes. The fluorescent monomer has long, hydrophilic spacers and polymerization sites to bind flexible supports. The fluorescent monomer has sufficient intensity for in vivo transdermal monitoring; even when it is immobilized in a solid support (microbeads). Due to the virtue of their small size, the fluorescent microbeads are injectable, minimally invasive, and rapidly respond to glucose change. The microbeads have been tested with success in rats [85].

Another application of this technology is for sensing glucose from skin measurements, presenting several limitations due to epidermal tickness, skin pigmentation and other parameters [86, 87].

2.1.2.6 Polarimetry

Polarimetry is based on the optical properties of glucose, due to its chemical structure that makes glucose a chiral molecule.

Polarimetry Principle

When polarized light (light with all waves oscillating in the same plane) passes through a solution containing optically active solutes, such as chiral molecules, its polarization plane is rotated by a certain angle, which depends on solutes concentration. Measuring the rotation angle with a polarimeter allows calculating glucose concentration. This technique is sensitive to scattering properties of tissues that depolarizes light. However, skin cannot be investigated by polarimetry since it shows high scattering due in particular to the *stratum corneum*. For this reason the preferred measurement site is the eye, in particular the *aqueous humor* beneath the *cornea*, which has low scattering properties. However, this particular measurement site raise a second problem: a time delay between glucose concentration in *aqueous humor* and blood. Although polarimetry is unaffected by temperature and pH fluctuations, it suffers from motion artifacts and optical noise of other substances [66].

Polarimetry-Based Sensors

A new real-time optical polarimetric approach for glucose sensing utilizing two wavelengths is presented in [88]. Only in vitro experiments have been reported. In fact an efficient eye coupling mechanism has not been developed yet, allowing in vivo experiments on rabbits eyes.

2.1.3 Thermal Emission Spectroscopy

Thermal emission spectroscopy measures IR signals generated in the human body as a result of glucose concentrations changes. The tympanic membrane is used as measuring site, since this membrane shares the blood supply with the centre of temperature regulation in the hypothalamus [89]. Body movements and ambient temperature, also induced by pathophisiological factors, are the most significant sources of noise [90].

2.1.4 Photoacoustic Spectroscopy

Photoacoustic Spectroscopy Principle

Photoacoustic spectroscopy uses the principle that absorption of a laser light causes consequent acoustic response. Tissue is illuminated by a short laser pulse, at a specific wavelength, and the absorbed radiation causes localised heating. The small temperature increase is dependent on the specific heat capacity of the tissue irradiated. Volumetric expansion due to heating generates an ultrasound pulse, which can be detected by a microphone. Increasing tissue glucose concentrations reduce the specific heat capacity of tissue and thus increase the velocity of the generated pulse making photoacoustic spectroscopy an indirect technique for glucose estimation [37].

Besides this, the photoacoustic spectrum considered as a function of laser light wavelength mimics the absorption spectrum in clear media (i.e. optically thin) and has the advantage to present higher sensitivity in the determination of glucose, thanks to the poor photoacoustic response of water.

The main limitation of this technique is its sensitiveness to chemical interferences from some biological compounds and to physical interferences from temperature and pressure changes.

Photoacoustic Spectroscopy-Based Sensors

The *Aprise* device is based on photoacoustic technology. It exploits in fact the photoacoustic properties of the blood and tissues to estimate the prevailing glucose levels. The sensor is attached to the skin above a blood vessel, and it generates ultrasound waves by illuminating the tissue with laser pulses. Analysis of the acoustic signals provides information on the absorbance of light in the tissue at different depths, which is influenced by glucose concentration. An ultrasonic image of the optical properties of tissue directly beneath the sensor is obtained. The ultrasonic image resolves the blood vessel from the tissue layers around it, enabling separated analysis of changes in optical properties of blood and surrounding tissues [91].

2.1.5 Electromagnetic

Electromagnetic Principle

Another technique for investigating dielectric parameters of blood utilizes the electromagnetic coupling between two inductors turned around the medium under study. The coupling is influenced by variations in the dielectric parameters of the solution, which are modified by glucose. This method is based on the application of a voltage signal with proper frequency to the primary inductor and for electromagnetic coupling a signal will be produced on the secondary inductor. There exists an optimal frequency, where the sensitivity to glucose change is maximal, but it is significantly influenced by temperature. The main problem of this technique is that several other components may have an influence upon the blood dielectric parameters and not only upon glucose [37].

Electromagnetic-Based Sensors

A new electromagnetic sensor is described in [92]. Its in vitro ability to estimate variations in glucose concentration of different solutions with similarities to blood (sodium chloride and Ringer-lactate solutions) has been tested, differing though in the lack of any cellular components. The sensor was able to detect the effect of glucose variations over a wide range of concentrations.

The *Glucoband* is a non-invasive glucose monitor that uses bio-electromagnetic resonance to measure the blood glucose levels of the human body. This device is worn like a wrist watch and displays results of the test on an LCD screen. The initial measurement process takes only a few minutes. However, in the monitoring mode, measurements can be continuous. Since each concentration of glucose has its unique electromagnetic molecular self-oscillation signature-wave, the *Glucoband* perform the measure matching the self-oscillation frequencies of glucose molecules with those of hundreds of reference solutions with different levels of glucose stored in an internal database of “signatures”.

2.1.6 Impedance Spectroscopy

Impedance Spectroscopy Principle

A different kind of spectroscopy investigates the dielectric properties of a tissue using a current flow instead of a light beam. It is called Dielectric Spectroscopy (DS) or

Impedance Spectroscopy (IS). The impedance of a tissue can be obtained applying a current of known intensity and measuring the resulting current flowed through the tissue [93]. The cell membrane is semipermeable to ions, thus certain ions can pass thought it while others cannot. This makes the membrane behave like a leaky capacitor. Moreover, the intra and extra-cellular environments consists of electrolyte showing resistive properties. The impedance of a tissue can thus be decomposed into a resistive and capacitive part $Z = R + iX$, being frequency dependent, formally described in terms of a resistance R [ohm] and reactance X [ohm]. A different expression is obtained considering polar coordinates using the magnitude $|Z|$ [ohm] and phase θ [deg] according to $Z = |Z|e^{i\theta}$. Repeating IS measurements for different frequencies of the initial electrical current allows to obtain the so-called dielectric spectrum, describing the impedance as a function of frequency. Section 2.1.1 highlighted how the complex multilayer skin structure influenced optical based technologies for glucose sensing. In the same way, the tissue structure and chemical composition of a biological material in general, and of the skin in particular, may correlate with its electrical properties, thus presenting frequency dependent characteristics. In particular, low frequency and direct current must pass around the cell, in the extracellular environment, given the high capacitance of the cell membranes. On the other hand, high frequency currents penetrate through cell membranes by polarization, charging and uncharging the barrier as a capacitor. Thus, different frequency bands contain information, in terms of impedance, affected by different properties of the tissue. Figure 2.5 shows an example of the impedance

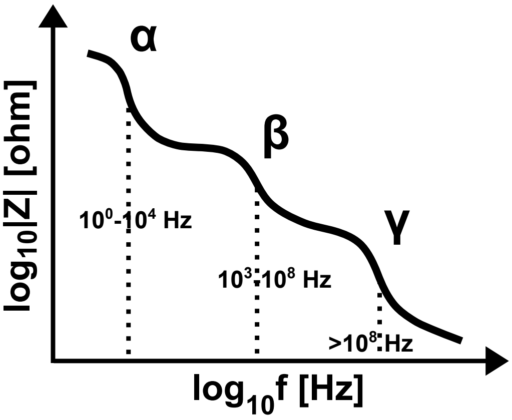


Figure 2.5: Dispersions. Magnitude of the impedance Z [ohm] as a function of the frequency of the exiting current. Adapted from [94].

magnitude as a function of the frequency [94]. The regions presenting a variation of the impedance corresponds to specific electrochemical processes, called dispersions. Four main dispersions have been identified: α, β, δ, and γ. Different mechanisms account for low frequency (α), radiofrequency (β), and microwave frequency (γ) dispersions [95]. The α-dispersion is generally considered to be associated with interfacial polarization linked with electrical double layers and surface ionic conduction effects of electrolyte at membrane boundaries. The β-dispersion has essentially two components arising from two different mechanisms: the capacitive shorting out of membrane resistances and rotational relaxations of biomolecules. Cell suspensions such as blood will typically exhibit a significant β-dispersion in the radiofrequency range between 100 kHz and 10 MHz. In addition, reorientation of free water molecules causes γ-dispersion. Water bound to protein and internal protein motion will also cause a subsidiary process, called the δ-dispersion, that is observed in the frequency region between the β and γ-dispersion.

DISPERSION	FREQUENCIES	ORIGIN
α	low(10-100Hz)	electrical double layers and electrolytes at membrane boundaries
β	radio(100kHz-10MHz)	cell suspension(blood)
δ	radio(10MHz-1GHz)	water bound to protein and internal protein motion
γ	microwave(1-100GHz)	reorientation of free water molecules

Table 2.1: Relaxation processes of biological materials.

IS-based techniques cannot measure glucose concentrations directly, since changes in glucose levels do not directly affect the dielectric properties of skin and the underlying tissue in the kHz and MHz frequency band. However, variations in plasma glucose lead to changes in the electrolyte balance in blood, cells and interstitial fluid. An increased concentration of glucose in blood involves a cellular biochemical response, which leads to changes of membrane components, nucleotide and ionic rearrangement. In particular, as a consequence of water movement, there is a decrease of sodium and an increase of potassium inside the erythrocyte. This variation of the electrolyte balance has an influence over the erythrocyte membrane potential and capacitance, which causes changes in the ac and dc conductivity and tissue permittivity that can be measured using IS [93, 96, 97]. A sensor based on IS uses electromagnetic waves in the selected frequency band that interact with the skin and the underlying tissue for monitoring these electrical properties.

Impedance Spectroscopy-Based Sensors

Several studies have been carried out and prototypes have been developed to prove the feasibility of this technology for NI-CGM. *In-vitro* studies have been performed to monitor glucose concentrations in different solutions [98], while recently a prototype has been developed embedding fringing field capacitive electrodes working in the 1-160 MHz frequency range [99].

2.2 Multisensor Approaches for NI-CGM

Multisensor Principle

NI-CGM provides reliable estimates of the glucose levels in highly restricted, i.e. in clinic conditions [100, 101, 102]. As soon as these conditions become less favourable, i.e. in daily life use, several disturbances can affect and interfere with the glucose sensing technology under use and impair the estimate of BGL. To overcome this problem, in the last years the multi-sensor approach gained larger attention [18]. It consists in the combination of some of the aforementioned NI-CGM technologies for a broader bio-physical characterization of the skin and underlying tissues. In practice, this concept translates in the embedding within the same device of different technologies able to track either properties of the skin related to glucose levels changes and environmental (temperature, humidity) and physiological (blood perfusion, body temperature) perturbations that can affect the main glucose related signals.

Multisensor-Based Sensors

An example of device based on a multi-sensor platform is the *GlucoTrack*® developed by *Integrity Applications* [103] and displayed in Figure 2.6. It resorts to a combination of thermal, acoustic and electromagnetic technologies for performing intermittent glucose monitoring, given its embedding within a ear clip that need to be wear each time a measured is performed. Data from the different sensors are processed through an algorithmic routine. Each one of the three technologies provide a signal that is converted into a glucose level with a suitable model. If at least two of the three estimated glucose levels agree then the device display the glucose value to the user, otherwise the measure is repeated.

A second example of device, at a prototype stage, is developed by Amaral and coworkers [105]. This device resorts to a combination of IS and MIR technologies combined through a suitable combination of linear and non-linear models (see Section 4.1).

In the last years, a Swiss medical company, Solianis Monitoring AG (Zurich, Switzer-

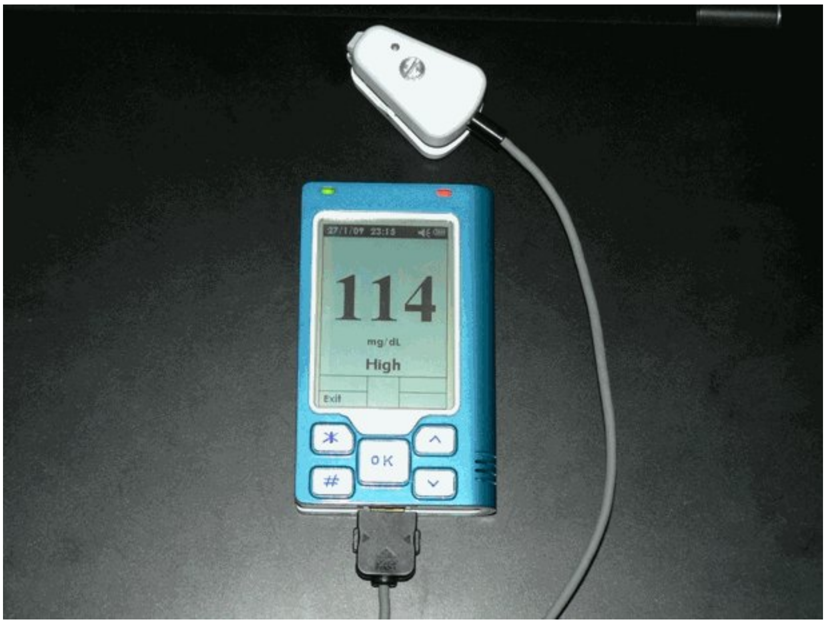


Figure 2.6: The *GlucoTrack*® consisting of the main unit (bottom) and the ear clip (top) featuring the thermal, acoustic and electromagnetic technologies [104].

land), developed a multisensor approach for NI-CGM mainly based on IS [102], whose capabilities in monitoring glucose level changes in-vivo has been recently demonstrated under clinical conditions [102]. Chapter 3 is devoted to the description of this particular device, since it provides the data that will be used in Part II of this thesis to test the proposed techniques for model identification.

3

The Multisensor Approach to CGM by Solianis Monitoring AG

This chapter will focus on the description of the multisensor approach pursued by Solianis Monitoring AG (Zurich, Switzerland), whose IP and technology have been recently acquired by Biovotion AG (Zurich, Switzerland), for providing an overview of the data that will be used in the last part of this thesis. Solianis Monitoring AG has also been partially funding the Ph.D. position during which this thesis has been developed. From this point of the thesis, we will refer to “Multisensor” (with capital M) for indicating the specific device developed by Solianis, and to “multisensor” to indicate the concept.

3.1 Description of the Solianis Multisensor

Earlier work in [106, 107] showed promising results in monitoring changes in blood glucose levels in clinical experiments in highly controlled, i.e. in clinical conditions, using IS. As soon as these conditions become less favourable, going towards a daily life use, this technique exhibits its limitations, mainly related ti deleterious effects of many perturbing factors, such as temperature fluctuations, variations of skin moisture and sweat, changes in cutaneous blood perfusion and body movements affecting the sensor-skin contact surface [108]. Consequently, all these perturbations affecting the main glucose related signals have to be identified, characterised and compensated for. As better discussed in the following, this suggested in [102] the development of a Multisensor Glucose Monitoring System, where the multisensor concept means a system that includes several sensors embedded within the same sensor substrate in contact with skin, allowing a broader bio-physical characterization of the skin and underlying tissues. The Multisensor performs continuous glucose monitoring collecting a set of signals measured through the Multisensor channels with a sampling time of 20 seconds. As shown in Figure 3.1 the Multisensor is attached to the upper arm of the patient with a flexible band and it is powered with a battery pack.

IS electrodes

As described in Section 2.1.6, changes in blood glucose levels cause dielectric changes of skin and underlying tissues within the frequency range of 0.1-100 MHz, which is measured utilizing particular capacitive fringing-field electrodes [102]. In order to achieve different penetration depths of the electromagnetic field into the various tissue layers, three electrodes with different characteristic geometries are used in the Solianis Multisensor. The interaction between an applied electromagnetic field and the skin depends not only on the frequency band, but also on the geometric properties of the electrode. The differences between the three IS electrodes consist in the distance between the active electrode and the ground potential. In particular, a distance of 0.3, 1.5 and 4 mm is associated with shallow, mid and deep penetration respectively and the sensors are referred as short, middle and long, respectively (see Figure 3.1) .

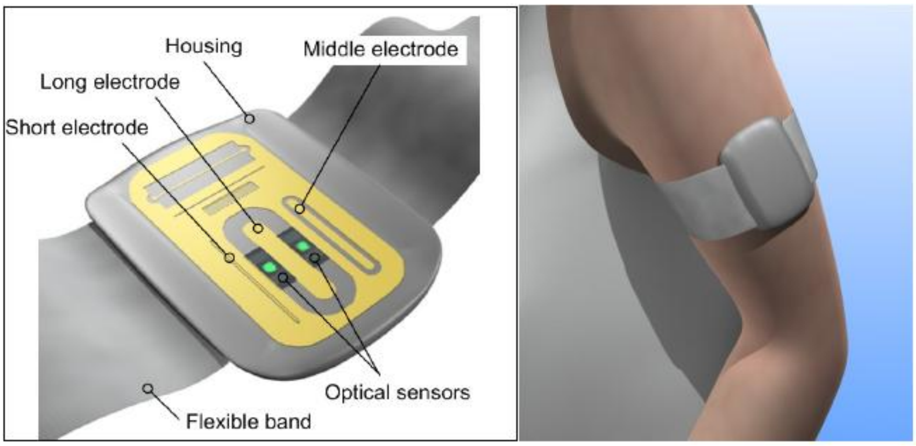


Figure 3.1: *Left*: Optical and dielectric sensors composing the Solianis Multisensor. *Right*: Solianis Multisensor attached to the upper arm with a flexible band.

The short electrode penetrates only the upper skin layers, thus it cannot yield information about glucose levels, but it may still contain information about perturbing effects related to the uppermost layers. Data from long and middle electrodes are regarded as primary signals, since they penetrate also the lower skin layers that are well micro-vascularised (see Figure 2.1) and hence particularly affected by glucose variations.

Optical sensors

As mentioned before, other sensors are used with the aim of obtaining useful information to compensate the perturbing factors: two optical sensors are embedded within the Multisensor substrate for the measurement of skin blood perfusion, which is a perturbing factor for dielectric signals [67]. Each optical sensor features 3 LEDs, located closely to each other, with the following wavelength: green (568 nm), red (660 nm) and infrared (798 nm). Light reflected back from the skin is detected by two photo-detectors (signal diodes), while the variation of emitted LEDs intensity are monitoring by two reference diodes (monitoring diodes) located near the LEDs. Simulation studies have been conducted to design the optimal position of the optical sensors within the Multisensor substrate as well as their relative distance for sampling the optimal measuring site [109].

Sweat sensors

An interdigitated electrode is used to measure the dielectric response at lower frequencies in the range of 1-200 kHz for obtaining information about sweat events. Moreover, its particular geometrical shape allow the sampling of the more superficial area of the skin. Another sensor exploits the frequencies in the range of GHz to estimate hydration levels of the underlying skin layers, since GHz excite free water molecules (see Table 2.1).

Acceleration sensors

An integrated accelerometer has the aim to monitor continuously the acceleration and the position relative to the centre of gravity of the device.

Other sensors

Finally, others sensors monitor skin and housing temperature, and ambient humidity close to the device. This is because IS data showed to be particularly sensitive to temperature fluctuations [108].

3.2 Examples of Solianis Multisensor Data

This section gives a clue of the different time-series measured from the Multisensor channels, highlighting in some cases features of the data that will influence the identification of the model in Section 3.3. Each sensor embedded on the Multisensor substrate provides its specific set of signals acquired with a sampling period of 20 seconds. To illustrate the Multisensor data, we can take advantage from the availability of reference BGL acquired in parallel with a sampling time of 10 minutes by a laboratory instruments.

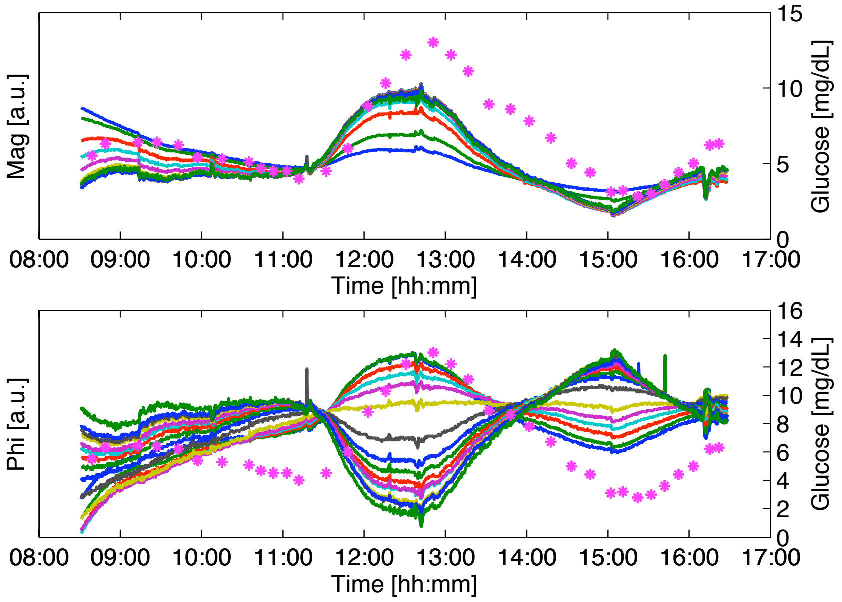


Figure 3.2: Normalized magnitude (top) and phase (bottom) impedance signals (continuous lines) from the "long" fringing field capacitive electrode vs. normalized reference BGL samples (magenta stars). Magnitude and phase at different frequencies (in the range 0.1-100 MHz) of the input current are collected with 20 sec time sampling and represented with different colors.

In Figure 3.2, representative time series, collected from the same electrode ("long" fringing field capacitive electrode) at different frequencies, are shown together with the BGL time series. In particular, the impedance at different frequencies is represented using a parametrization with magnitude (Figure 3.2, top) and phase (Figure 3.2, bottom). As shown in the top panel the magnitude signals at different frequencies are similar but not identical, thus presenting strong correlation. The same is for the phase signals, which are also correlated with the magnitude signals. Since the impedance channels, as mentioned in the previous section, contain glucose information, they are referred as the primary "glucose signals" [102].

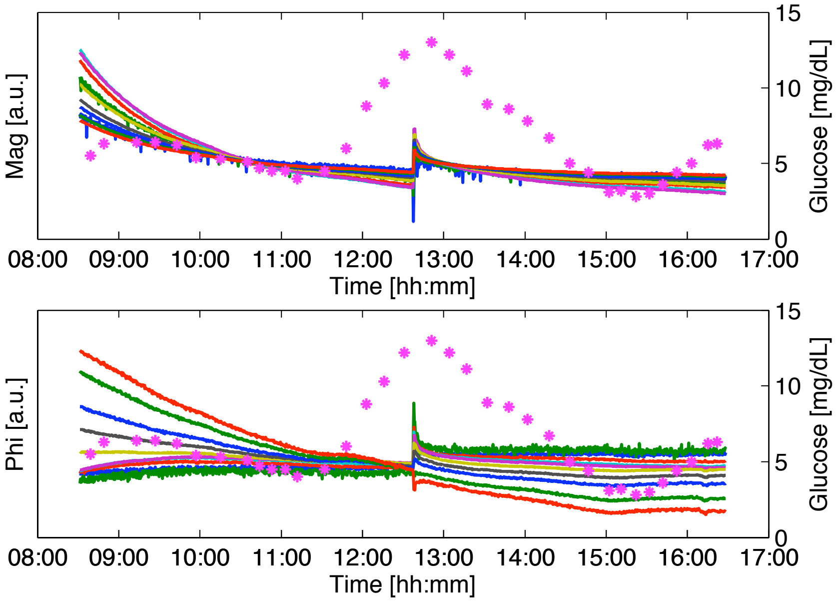


Figure 3.3: Normalized magnitude (top) and phase (bottom) impedance signals (continuous lines) from the interdigitated electrode vs. normalized reference BGL samples (magenta stars). Magnitude and phase at different frequencies of the input current are collected with 20 sec time sampling and represented with different colors.

Figure 3.3 shows the time series relative to the magnitude (top) and phase (bottom) of the impedance measured by the interdigitated electrode which is particularly sensitive to changes of the surface dielectric properties due to the creation of a saline layer after sweat events occurred. It is particularly interesting to note how these channels are particularly responsive to the on-set of the sweat event and of the following creation of the saline layer.

Figure 3.4 shows channels associated with other sensors embedded within the Solianis device. In the top panel the skin and the housing temperature are plotted together with the time-series relative to the humidity sensor, along with ambient humidity. In the bottom panel an example of optical channels is shown. Some of them seem correlated with the BGL references. However, they are noisier than impedance channels.

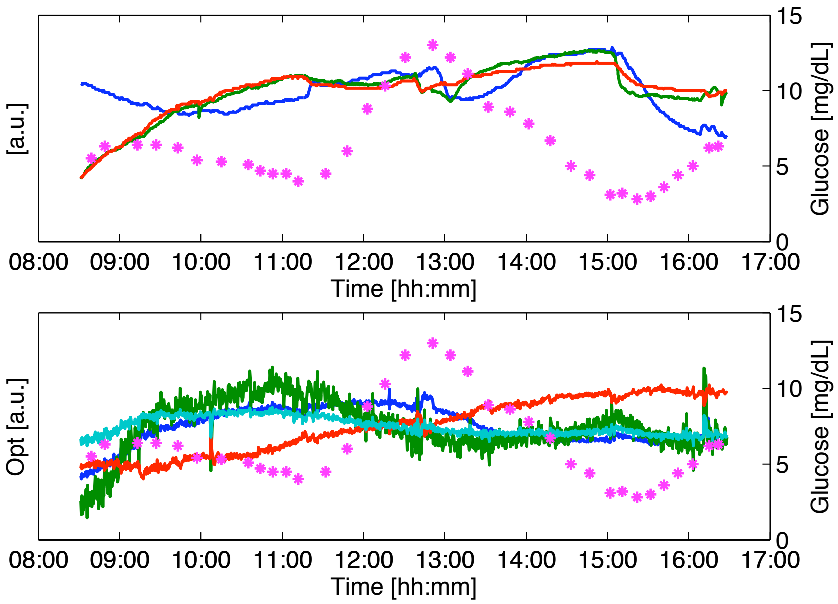


Figure 3.4: Normalized Multisensor housing temperature (green line), skin temperature (red line) and humidity (blue line)(top) and some optical (bottom) channels (continuous lines) vs. normalized reference BGL samples (magenta stars).

3.3 From Multisensor Data to Glucose: the Need of a Model

The Multisensor data described in the previous section need to properly combined in order to perform NI-CGM (see Figure 3.5). While some channels measure "glucose-related" information, others are used to characterise the perturbations affecting the primary signals, allowing for their compensation. Hence, the Multisensor signals have to be combined through a relationship, i.e. a model, for linking the measured variables with BGL.

Figure 3.5: A model (*middle*) linking Multisensor data (*left*) to BGL (*right*).

Supposing to collect the Multisensor signals at the t-th time instant whitin $\boldsymbol{X}(t)$, the unknown relationship can mathematically be expressed as:

$$BGL(t) = f(\boldsymbol{X}(t), \boldsymbol{\theta}) \tag{3.1}$$

where $f(\cdot)$ represents a generic function used to convert the signals into BGL values and $\boldsymbol{\theta}$ is an unknown vector containing the parameters that characterise this conversion.

The same problem arises for the other multisensor approaches, such as for example the one developed by Amaral and co-workers [105]. Remarkably, other research fields have to face similar challenges, mainly those related to chemometrics and dealing with spectroscopy data analysis [110, 111].

3.4 Concluding Remarks

In the last years, several NI-CGM technologies have been investigated. As also depicted in Chapter 2, many sensors based on a wide range of technologies have been proposed to the market, but none of them really achieved accuracy close to that of minimally invasive, needle-based, sensors. This is because all the presented non-invasive technologies are affected by many environmental and physiological interferences affecting glucose readings. In order to measure and compensate for these detrimental effects, the multisensor approach consists in the embedding of different sensors into the same device for a complete bio-physical characterization of the skin and underlying tissues.

4

Open Problems with Model Identification in Multisensor Approaches and Aim of the Thesis

4.1 Problem Statement

A class of candidate models to be used in Figure 3.5 and described by eq. (3.1) is tha of white-box models, where differential equations are used to describe the physical relationships between measured channels and glucose variations. These kind of models have been widely exploited for modeling physiological processes with the purpose of improving physiological knowledge [112], extract clinical information [113] as well as for control purposes [114, 31]. However, in our case, a mechanicistic description for linking glucose variations with physical quantities measured with the Multisensor is not yet available. For this reason, we resort to a black-box strategy, where the system is described in terms of its inputs (Multisensor channels), outputs (glucose) and the type of model/relationship linking the two. For this reason we must define the structure (e.g. static or dynamic) and the form (e.g. linear or not) of the function $f(\cdot)$ of Figure 3.5 and eq. (3.1) [115].

The model considered throughout this thesis is a static multivariate linear regression model, formally described as:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\beta}_0 \tag{4.1}$$

where, y is the output variable, i.e. glucose, matrix $\boldsymbol{X}$ collects the data measured from the Multisensor device and $\boldsymbol{\beta}$ is the parameter vector that linearly combines the Multisensor channels in order to give an estimate of the glucose levels apart from an off-set pr basal value given by $\boldsymbol{\beta_0}$. To better highlight the notation used, the unknown variable to be predicted is called *output* or *target*, while the measured variables are called *inputs*, or *regressors* (because they contain the information for the regression model) or *predictors* (because they are used to predict the output). Thus, the aim of regression is to build and identify a prediction model. This model can then be used for estimating glucose for, new, "unseen" data. In the general case, the target variable consists of a multi-dimensional vector. However, since our purpose is to estimate glucose, throughout this thesis, only the case of a single output variable will be considered. Hence, the output will be represented by a column vector $\boldsymbol{y}$ of dimension $N \times 1$, where N is the number of available samples. In symbols:

$$\boldsymbol{y} = [y_1 \quad y_2 \quad \dots \quad y_N]^T \tag{4.2}$$

where y_i denotes the i-th sample of the reference.

The input variables are contained in the matrix $\boldsymbol{X}$ of dimension $N \times p$, where the element x_{ij} represents the i-th sample of the j-th variable.

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \\ x_{N1} & x_{N2} & & x_{Np} \end{bmatrix} \tag{4.3}$$

While each row of the matrix $\boldsymbol{X}$ contains the set of p variables corresponding to the measured Multisensor channels relative to the same i-th time instant (represented by the row vector $\boldsymbol{X}_{ip}$ $(1 \times p)$), each column contains the N samples of the j-th variables (symbolized using the column vector $\boldsymbol{X}_{jN}$ $(N \times 1)$). Hence, while subscript $i \in [1, 2 \dots N]$ indicates the sample, subscript $j \in [1, 2 \dots p]$ identifies the variable. To distinguish for example, $\boldsymbol{X}_1$ the set of p variables at the first time instant from $\boldsymbol{X}_1$ the N samples of the first variable, a second subscript is added, indicating the dimension of the vector.

$$\boldsymbol{X} = \underset{\substack{\Downarrow \\ \boldsymbol{X}_{jN}}}{\left[\begin{array}{c} x_{11} \quad \dots|\overline{x_{1j}}|\dots \quad x_{1p} \\ \vdots \quad\quad |\vdots| \quad\quad \vdots \\ \boxed{x_{i1} \quad \dots|x_{ij}|\dots \quad x_{ip}} \\ \vdots \quad\quad |\vdots| \quad\quad \vdots \\ x_{N1} \quad \dots|\underline{x_{Nj}}|\dots \quad x_{Np} \end{array}\right]} \Rightarrow \boldsymbol{X}_{ip} \tag{4.4}$$

The aim of regression is to find an estimate $\hat{\boldsymbol{\beta}}$ of the unknown coefficients $\boldsymbol{\beta}$, given the knowledge of the reference vector $\boldsymbol{y}$ and of the coupled inputs collected in $\boldsymbol{X}$ from the so called *identification data set*. After $\hat{\boldsymbol{\beta}}$ is determined, it can be used to calculate the correspondent model prediction of the target $\hat{\boldsymbol{y}}$ prospectively also on different Multisensor data than that used in the model identification stage (*test data set*).

4.1.1 Open Problems

The principal Multisensor signals are those that mostly contain the information about glucose fluctuations. However, in the everyday life use of the device, these signals are affected by different perturbing factors (temperature fluctuations, skin moisture, sweat, blood perfusion, ...). The multi-sensor concept derives from the necessity of compensating these perturbations, measuring a high number of channels presenting correlation between subsets of them. Indeed, spectroscopy data present very similar values at close frequencies. Thus the difficulties in identifying $\boldsymbol{\beta}$ in eq. (4.1) are primarily due to:

- High dimension of the measurement space;
- Correlation between subset of Multisensor channels.

These two characteristics of the recorded data make the matrix $\boldsymbol{X}$ rank deficient with numerical issues arising for the identification of $\boldsymbol{\beta}$ because the problem is ill-conditioned.

4.2 Aim of the Thesis and Outline

The aim of thesis concerns the investigation and assessment of different techniques for the identification of a multivariate linear regression model for tracking glucose levels changes non-invasively. We will consider Ordinary Least Squares (OLS), Partial Least

Squares (PLS), the regularization technique based on ℓ_1 norm, namely Least Absolute Shrinkage and Selection Operator (LASSO), and a technique based on the ℓ_2 norm, i.e. Ridge regression, and a technique based on their combination, i.e. EN regression. While our implementation will be focused on the particular Solianis Multisensor platform, the considered methodologies can have a much wider field of applicability and can be used to model other multisensor data for NI-CGM as well as for data analysis in chemometrics and related disciplines where PLS represents the current state-of-art.

The remainder of the thesis is organized into two blocks. In Part II, we explain rationale of each method, pros and cons, by exploiting a tutorial example. Part III illustrates the results obtained from the application of the five considered model identification techniques to data recorded through the Solianis Multisensor device. Finally, some hints for future developments is given regarding a Monte Carlo based methodology for assessing the robustness of the individual calibration parameter (calculated each time the Multisensor is worn) against a very common source of disturbance in daily-life conditions, i.e. sweat.

Part II

Techniques for Identification of Multivariate Models

5

Criteria for Model Identification and Model Test

5.1 Issues of High-Dimensional Regression

This section describes, using the notation of the previous chapter, the issues typically faced for regression problems when are used high dimensional datasets. Most of the material originates from the book written by Hastie et al. [116] and is reported here for sake of completeness.

It could seem reasonable that, if the identification set is large enough, it would be easy to generalize data behaviour and identify a good prediction model. However this it not true dealing with high-dimensional data because the large number of correlated predictors exacerbate the need of available data for identification of the models. This is exactly what happens in our case study where we will have to deal with more than 150 input variables (the Multisensor measured signals). In presence of high-dimensional datasets, the algorithms for solving regression problems suffer from the so called curse of dimensionality [117].

5.1.1 Curse of Dimensionality

Consider a p-dimensional unit hypercube and suppose the N regressors samples to be uniformly distributed in it. The fraction of samples included in a hypercube with side

$r(<1)$ is:

$$frac = r^p$$

Extracting the side of the hypercube as a function of the desired fraction and the dimension p, one gets:

$$r = frac^{1/p}$$

Hence, for example, to include 10% of the samples, we need a hypercube with side 0.1 for p=1, and a hypercube with side 0.8 for p=10. The different curves plotted in Figure 5.1 show the side of the hypercube as a function of the fraction of included samples for different values of dimension p.

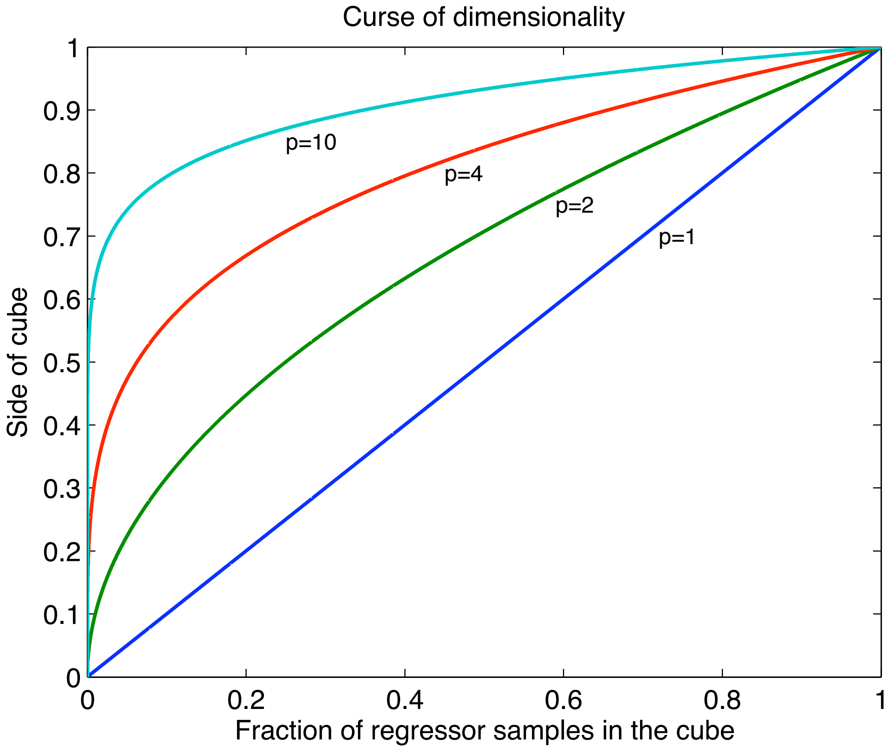


Figure 5.1: side of the cube as a function of the fraction of included samples for different values of dimension p. Adapted from [116].

As shown in Figure 5.1, the hypercube side needed for a given fraction increases even more as the dimension p increases. Hence, as the number of the regressors increases, it becomes more difficult to generalize data behaviour. In fact, the samples are more distant to each other and, in particular, they tend to be close to an edge of the sampling area, because the prediction is much more difficult near the edges of the training sample.

One can also formulate the problem considering that the sampling density is proportional to $N^{1/p}$. In high dimensional datasets, all feasible training samples sparsely populate the input space. In fact, the density rapidly decreases to zero as p increases.

5.1.2 Overfitting

In addition, high-dimensional regression algorithms have to deal with overfitting, namely, the risk of fitting a predictive model not only to the information yielded by identification data but also to noise, with considerable limitations on the generalization properties of the model. To overcome these problems, our attention will be focused on techniques based on two strategies: a) *dimensionality reduction*, which uses $M(\leq p)$ new regressors calculated from a linear combination of the original ones, or b) *regularisation*, putting a price on the values of the unknown coefficients $\boldsymbol{\beta}$ of model (4.1).

An example of regression technique using dimensionality reduction is Partial Least Squares (PLS), which will be described in detail in Chapter 6, while an example of regression technique using regularisation is Least Absolute Shrinkage and Selection Operator (LASSO), which will be discussed in Chapter 7. Both these methods require the setting of one parameter related to the model complexity (i.e. describing the new dimensionality M in PLS and the amount of regularisation in LASSO) as illustrated below.

5.2 Criteria for Selection of Model Complexity

Model complexity should be selected such that the best performance of the chosen identification method is achieved when the target variable is estimated from “unseen” data, i.e. data not used during the model identification stage. The main issue relies on the right way to evaluate the performance for selection of the model complexity.

5.2.1 The Bias-Variance *Dilemma*

Considering the identification set, it seems reasonable to assume that if model complexity increases the model will better describe the target. Hence, the Residual Sum of Squares (RSS) on identification data (describing the distance between the reference $\boldsymbol{y}$ and its model prediction $\hat{\boldsymbol{y}}$ will decrease as model complexity increases.

$$\mathrm{RSS} = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 = \|\boldsymbol{y} - \hat{\boldsymbol{y}}\| \tag{5.1}$$

This is a key aspect of the so called "internal validation". RSS of eq. (5.1) is expected to have a monotonic decreasing behaviour as model complexity increases (see Figure 5.2). This means that we cannot use RSS to determine model complexity, since we can always obtain, for sufficiently complex models, null residuals. The so-determined model usually fails in predicting new data, different from those of the identification set. In fact, a too complex model normally fits the reference data but also the noise (overfitting) and is thus not able to generalize the data behaviour properly.

As a consequence, the performance of the identification method has to be determined using independent test data. Suppose the measurement $\boldsymbol{y}$ to be modeled as a combination of a deterministic part $f(\boldsymbol{X})$ ($\boldsymbol{X}$ is the matrix collecting the inputs or regressors) and a random part $\boldsymbol{\epsilon}$ due to the noise (zero mean, constant variance):

$$\boldsymbol{y} = f(\boldsymbol{X}) + \boldsymbol{\epsilon} \qquad\qquad \boldsymbol{\epsilon} \sim N(0, \sigma^2) \tag{5.2}$$

The Mean Square Error (MSE) can be defined for measuring errors between the true value $\boldsymbol{y}_{true}$ and the model prediction $\hat{\boldsymbol{y}}$:

$$\text{MSE}(\hat{\boldsymbol{y}}) = E[\|\hat{\boldsymbol{y}} - \boldsymbol{y}_{true}\|_2^2] \tag{5.3}$$

Equation (5.3) can be divided in two terms, one representing the estimation variance and the other the bias (difference between the expected value of the estimation and the true value $\boldsymbol{y}_{true}$):

$$\text{MSE}(\hat{\boldsymbol{y}}) = trace(Var(\hat{\boldsymbol{y}})) + \|\text{Bias}(\hat{\boldsymbol{y}})\|_2^2 \tag{5.4}$$

The proof of (5.4) can be given in term of its scalar versions since $MSE = \sum_{i=1}^{N} E[(\hat{y}_i - y_{i,true})^2]$.

$$\begin{aligned}
E[(\hat{y}_i - y_{i,true})^2] &= E[(\hat{y}_i - E[\hat{y}_i]) + (E[\hat{y}_i] - y_{i,true})^2] \\
&= E[(\hat{y}_i - E[\hat{y}_i])^2 + (E[\hat{y}_i] - y_{i,true})^2 + (\hat{y}_i - E[\hat{y}_i])(E[\hat{y}_i] - y_{i,true})] \\
&= Var(\hat{y}_i) + Bias(\hat{y}_i)^2 + E[(\hat{y}_i - E[\hat{y}_i])(E[\hat{y}_i] - y_{i,true})] \\
&= Var(\hat{y}_i) + Bias(\hat{y}_i)^2 + E[\hat{y}_i E[\hat{y}_i] - (E[\hat{y}_i])^2 - y_{i,true}\hat{y}_i + E[\hat{y}_i]y_{i,true}] \\
&= Var(\hat{y}_i) + Bias(\hat{y}_i)^2 + (E[\hat{y}_i])^2 - (E[\hat{y}_i])^2 + y_{i,true}E[\hat{y}_i] - y_{i,true}E[\hat{y}_i] \\
&= Var(\hat{y}_i) + Bias(\hat{y}_i)^2
\end{aligned}$$

Generally, the variance term increases as model complexity gets higher. This can be explained observing that the more complex the model is, the more is the adherence to the data and thus the sensitivity of the estimated model parameters to the particular realization used to identify them. On the other hand, the bias term decreases as model complexity increases. As a consequence, even if estimates are influenced by noise, the effects of the bias term tends to be eliminated by averaging different estimates. Summarizing, the training error tends to decrease when model complexity is increased.

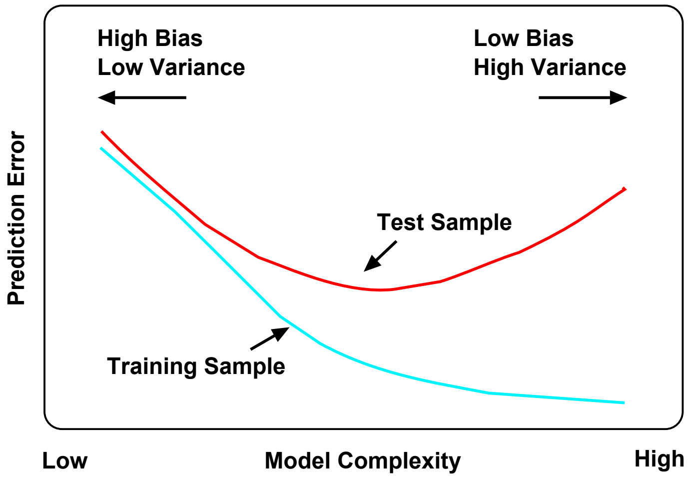


Figure 5.2: Test and training error as a function of model complexity [116]. The training error curve (blue line) and the test error curve (red line).

If the model overfits the data (too high complexity), it will not generalize well and the estimates will have too high variance. On the other side, if the model is not complex enough, it may underfit the data and have large bias. This brief discussion highlights the *dilemma* of fixing the bias-variance tradeoff and suggests that model complexity should be chosen in such a way to minimize the error on independent test data. As shown in Figure 5.2, the prediction error has a monotonic decreasing behaviour as model complexity increases, when calculated on the training set (blue curve). Hence, it can not be used to select the correct amount of model complexity. In Figure 5.2 the prediction error behaviour when calculated on the test set is also plotted (red curve). Usually, it has concave behaviour, due to the bias-variance trade-off. In this case, the curve minimum can be used to fix most reasonable model complexity. In the next subsection a method to construct the test error curve is described.

5.2.2 The Cross-Validation Principle

As far as we observed that the identification data set is not useful to select the model complexity, another set of data has to be considered (test set). As a consequence, before describing how to calculate the prediction error curve on the test set, we have to discuss how to handle the available data.

In a data-rich situation, the best way to split the available dataset is in three parts: a training set, a validation set and a test set. The training set is used to fit the model, the validation set is used to select the complexity parameter and the test set is used for assessing the generalization error of the final chosen model (Section 5.2). However, if the data are scarce (as in our case), the previous approach is not applicable.

K-fold cross-validation is a method to estimate test error, using the training set. In particular, K-fold cross-validation splits the data into K parts of approximately equal size. Iteratively, one part is left aside to calculate the test error (using MSE), while the other $K-1$ parts are used to identify the coefficients of the model. In this way a test error upon each K-th part is calculated and, averaging these values, an estimation of the test error is obtained.

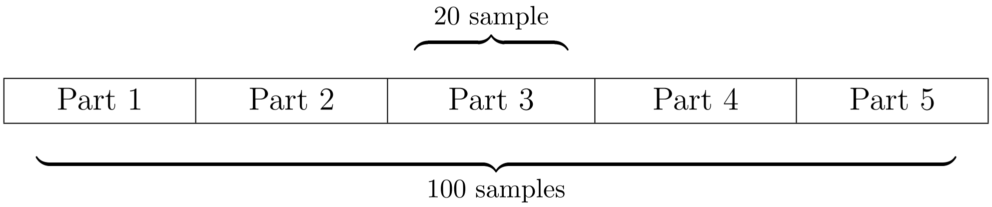


Figure 5.3: Example of dataset division for 5-fold cross-validation.

For example, suppose that a training set of 100 samples is available and that we want to perform 5-fold cross-validation. The 100 samples are randomly and equally divided in 5 parts, each of about 20 samples as shown in Figure 5.3.

At the first iteration, part 2-3-4-5 of the training set are used to estimate the coefficients of the model, obtaining $\hat{\boldsymbol{\beta}}^{-1}$, where the superscript indicates the part that was not used in the identification procedure. The estimated coefficients $\hat{\boldsymbol{\beta}}^{-1}$ are used to predict the reference of part 1 ($\boldsymbol{y}_1$) from the inputs variable of part 1 ($\boldsymbol{X}_1$):

$$\hat{\boldsymbol{y}}_1 = \boldsymbol{X}_1\hat{\boldsymbol{\beta}}^{-1} \tag{5.5}$$

The RSS is then used to calculate the test error on part 1, where the residuals denote the distance between the model predictions $\hat{\boldsymbol{y}}_1$ and the available reference points $\boldsymbol{y}_1$:

$$\mathrm{RSS}_1 = \sum_{i=1}^{N_1} (\hat{\boldsymbol{y}}_{i1} - \boldsymbol{y}_{i1})^2 = \|\hat{\boldsymbol{y}}_1 - \boldsymbol{y}_1\|^2 \tag{5.6}$$

where N_1 is the number of samples included in part 1.

At the second iteration, part 2 is left aside to calculate the RSS_2, using the coefficients estimated from part 1-3-4-5. Similarly, the procedure is iterated for other three times in order to calculate RSS_3, RSS_4 and RSS_5. These five values of RSS are then averaged in order to estimate the test error.

$$\mathrm{E}_{test} = \overline{\mathrm{RSS}} = \frac{\sum_{i=1}^{5} \mathrm{RSS}_i}{5} \tag{5.7}$$

The whole procedure is repeated for different values of the complexity parameter in order to estimate the test error as a function of the model complexity (see Figure 5.2). Usually, this function has a minimum corresponding to the most reasonable bias-variance trade-off.

Cross-validation, averaging the RSS calculated on different datasets, allows also to estimate the confidence interval for the estimated test error. Using the previous example, considering a 5-fold cross-validation procedure, the confidence interval for a given model complexity can be calculated as follows:

$$\mathrm{SD} = \sqrt{\sum_{i=1}^{5} (\mathrm{RSS}_i - \overline{\mathrm{RSS}})^2 \Big/ 5} \tag{5.8}$$

As a consequence, instead of choosing the complexity parameter at the minimum of the test error function, usually “one-standard error” rule is used to choose the model. This criterion consists in choosing the most parsimonious model whose error is no more than one standard error above the error of the best model. The model chosen according to this rule is represented by the green dashed line in Figure 5.4, corresponding to model complexity 7. A different strategy for choosing the complexity parameter is to identify its value in correspondence of a significant change of slope of the error curve. In Figure 5.4 this point corresponds to 4, and, as frequently happen, it does not coincide with the previous. This rule of thumb allows to obtain a more parsimonious model than the “minimum of the test error plus its standard deviation”, with advantages in generalization performance on new test data.

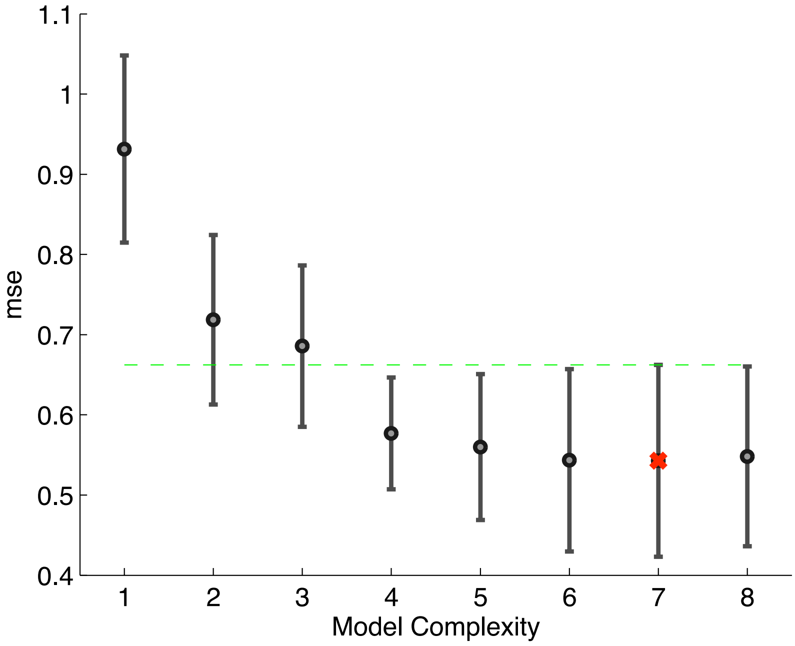


Figure 5.4: Example of test error curve in K-fold cross-validation, where the error s for each model (with different complexity) are provided by its mean and standard deviation. Red star is the minimum of the test error. Green dashed line represent the error of the best model.

5.3 Models Test

In this section, how to describe the performance of the selected model will be presented.

5.3.1 Principles for Model Test

In the previous section we described how the identification data set is used in cross-validation to choose model complexity. Once model complexity is determined the coefficients of the model can be estimated from the whole identification data set using different techniques. For instance, OLS (Chapter 6), PLS (Chapter 7) and the regularization based techniques (Chapter 8), namely, LASSO, Ridge and EN, are considered. The further step is to determine which identification method best suits for our particular problem. As a consequence, some indicators have to be defined to evaluate model performance on a test set of data. Since the error estimated from data used to identify the coefficients of the model tends to underestimate the real error, the test set must be composed by unseen data, i.e. data that are not used in cross-validation procedure nor in the identification procedure. Hence, this procedure is often called “external validation”.

Formally, in external validation, the coefficients of the linear model estimated from the identification data set $\hat{\boldsymbol{\beta}}_{train}$ are used to predict the target of the test data $\boldsymbol{y}_{test}$:

$$\hat{\boldsymbol{y}}_{test} = \boldsymbol{X}_{test}\hat{\boldsymbol{\beta}}_{id} \tag{5.9}$$

the subscript “id” denotes what is calculated from the identification data set, while the subscript “test” is appended to test set quantities through the equations. Therefore, $\boldsymbol{X}_{test}$ is the matrix collecting test data.

To quantify the prediction quality, different indicators can be considered. In particular we introduce two groups of indicators: the first aims at quantifying point accuracy of the estimated glucose profiles; the second group includes two indicators widely used within the diabetes community to judge the clinical accuracy of CGM devices. All these indexes are formally defined in the following section.

5.3.2 Indicators for Point Accuracy

The indicators defined before can be used to evaluate the performance of the identified model on unseen data (i.e. when the test data set is considered). Hence, they allow the comparison between different models [10, 118].

MSE was defined as a stochastic quantity in (5.3). However, a realization can be observed as normalized distance between prediction $\hat{\boldsymbol{y}}$ and reference data $\boldsymbol{y}$:

$$\text{MSE} = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \Big/ N \tag{5.10}$$

Root Mean Square Error (RMSE) is the square root of (5.10) and thus has the same units as the quantity being estimated.

Mean Absolute Difference (MAD) is defined as follows:

$$\text{MAD} = \sum_{i=1}^{N} |y_i - \hat{y}_i| \Big/ N \tag{5.11}$$

which differs from (5.10) since, instead of summing the squares of the differences, their absolute values are summed up.

Mean Absolute Relative Difference (MARD) is the same as (5.11), but it is an absolute indicator, since every difference $(y_i - \hat{y}_i)$ is divided for the reference value y_i:

$$\text{MARD} = \sum_{i=1}^{N} \left| \frac{y_i - \hat{y}_i}{\hat{y}_i} \right| \Big/ N \tag{5.12}$$

While these three key indicators are based only upon the distance between the test reference data $\boldsymbol{y}$ and its prediction $\hat{\boldsymbol{y}}$, others like for example R^2 measures how much the prediction is a good approximation of the reference variation.

The Pearson correlation coefficient R measures the linear dependence between two variables, representing the test reference $\boldsymbol{y}$ and prediction $\hat{\boldsymbol{y}}$. The general formula for its calculation is:

$$\mathrm{R} = \frac{N\sum_{i=1}^{N}\hat{y}_i y_i - \sum_{i=1}^{N}\hat{y}_i \sum_{i=1}^{N} y_i}{\sqrt{N\sum_{i=1}^{N}\hat{y}_i{}^2 - \left(\sum_{i=1}^{N}\hat{y}_i\right)^2}\sqrt{N\sum_{i=1}^{N}y_i{}^2 - \left(\sum_{i=1}^{N}y_i\right)^2}} \tag{5.13}$$

The correlation coefficient R ranges from -1 to +1 included. A value of +1 or -1 implies a linear relationship between the two variables. In the case R equals +1 it means that if $\boldsymbol{y}$ increases, $\hat{y_i}$ increases too (correlation); in the case R equals -1 a decrease in $\hat{y_i}$ will correspond to an increase of $\boldsymbol{y}$ (anticorrelation). A value of 0 implies that there is no correlation between the variables.

The square of correlation coefficient, R^2, ranges from 0 to +1. Hence, it does not distinguish negative from positive correlation. This indicator turns out to be useful when interested to the connection between the variables and not to the sign of the relation.

A key mathematical property of the correlation coefficient is that it is invariant to changes in location and scale, i.e. if one of the variables is transformed linearly as $a + bx$ (with a and b constants) the correlation coefficient does not change its value. This can be useful to determine if the prediction $\hat{\boldsymbol{y}}$ has the same fluctuations of the reference $\boldsymbol{y}$, without having the same scale. In this case R^2 would assume a high value (good correlation), even if the distance between the reference and test sample is high, causing bad values for RMSE, MAD or MARD.

Finally, a measure to quantify the smoothness of the estimated glucose profiles by the different models is considered. In analogy to idea exploited in the context of regularization [119], the Energy of the Second Order Differences (ESOD) is considered, which is defined here as the energy of the second order differences of the estimated glucose profiles normalized by the energy of the second order differences of the reference BGL values in the same experimental sessions:

$$ESOD = \frac{\sum_{i=1}^{N}\Delta^2(\hat{y}_i)^2}{\sum_{i=1}^{N}\Delta^2(y_i)^2} \tag{5.14}$$

5.3.3 Indicators for Clinical Accuracy

While the indicators defined in the previous section are suited to give an indication about the point accuracy of the estimated glucose profiles, they lack in providing suitable information about the clinical information carried by the CGM traces. In order to fulfill this gap, the so called Clarke Error Grid has been extensively used among the diabetes community, initially to measure accuracy of SMBG devices and then of CGM devices.

The Clarke Error Grid shows the scatter plot of reference BGL *versus* the BGL value estimated by the device under test [120]. The plot area is broken down into five main regions as it can be seen in Figure 5.5 (left):

- *Region A* includes values within 20% of the reference;
- *Region B* includes values outside the 20% but not leading to inappropriate treatment;
- *Region C* contains points leading to unnecessary treatment;
- *Region D* contains points indicating a potentially dangerous failure in detecting hypo or hyper-glyceamia;
- *Region E* contains points that would lead to a hypo-treatment when the patient is actually in hyper-glycaemia and *viceversa.*

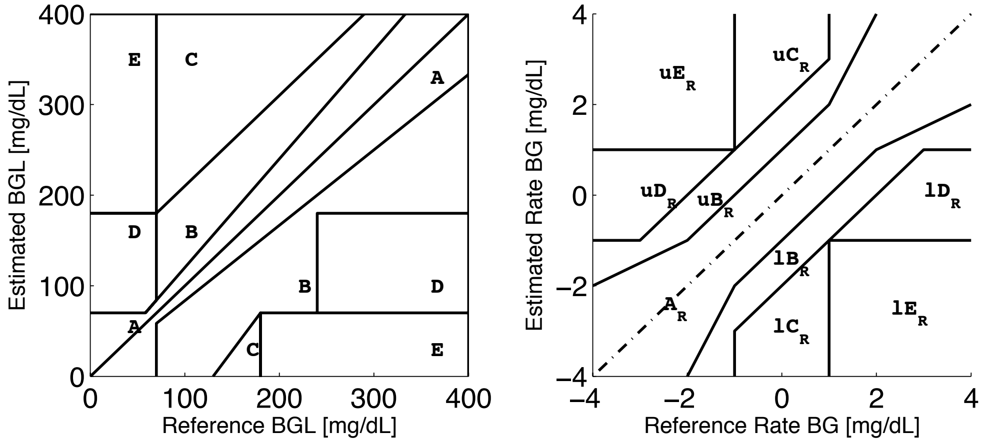


Figure 5.5: Clarke Error Grid (*left*), and Rate Error Grid (*right*).

Thus, a clinically accurate sensor should provide most of the points within the A+B zone with few, or ideally none, in the C/D/E zones. Current accuracy of minimally-invasive CGM devices show a range of values between 84.4 and 98.9 of points in the A+B zones, with point accuracy values for MARD in the range 10.3-21.5. CGM devices, measuring BGL every 1 to 5 minutes, provide information also on the trend of the glucose signal,i.e. stable, rising or falling glycaemia. To evaluate the accuracy in estimating glucose trends, the so called Rate Error Grid has been developed [121]. This grid is based on the same concept of the Clarke grid. The area is broken down into regions indicating clinically relevant information about the glucose trends estimated from the device under test. The Rate Error Grid focuses on the clinical implications of measurement errors by addressing the question of what type of clinical outcome might occur if the patient took action based on BGL rate of change.

5.4 Concluding Remarks

This chapter presented an introduction to the regression problem, with consideration that algorithms dealing with high-dimensional data suffer from the curse of dimensionality and overfitting. A general introduction to the methods trying to solve these problems was presented. As these algorithms usually require the setting of a parameter to adjust the model complexity, a commonly used procedure for such a scope was illustrated using K-fold cross-validation.

Finally, at the end of the Chapter, some indicators for the performance comparison of different models were presented. While, by visual inspection of the estimated profiles versus reference data, one could only qualitatively guess which model has the best performance, the indicators presented in Section 5.3.2 and Section 5.3.3 will allow a quantitative assessment of how much a method works better than the others in identifying linear models for regression. Further metrics are available in the literature, such as for example for evaluating the accuracy of prediction algorithms [122], but we believe those reported in this chapter are exhaustive for describing accuracy of estimated glucose profiles.

These procedures will be used in this thesis to evaluate the performance of the regression methods, presented in Chapters (6-8), when applied to the Solianis Multisensor data (Chapters 10-11).

6

Ordinary Least Squares (OLS)

The most easy and well-known method for finding an estimate of the parameter vector of the multivariate linear regression model defined in eq. (4.1) $\boldsymbol{\beta} = [\beta_0, \beta_1 \ldots, \beta_p]$, given the reference vector $\boldsymbol{y}$ and the corresponding inputs $\boldsymbol{X}$, is Ordinary Least Squares (OLS). OLS makes no assumption about the validity of the model, but simply finds the best set of parameters $\boldsymbol{\beta}$ by adjusting them in order to maximize the adhesion between the model predictions and the reference data. This chapter will present the characterization of the OLS identification procedure in a general framework. Then, with the support of a simple tutorial example (Chapter 9), advantages and drawbacks of OLS will be shown. Finally, in Chapter 11 the technique will be applied to model NI-CGM Multisensor data.

6.1 Mathematical Definition

OLS determines the estimate $\hat{\boldsymbol{\beta}}$ by minimizing the Residual Sum of Squares (RSS), where the residuals denote the distance between the model predictions (4.1) and the available reference points y_i:

$$\mathrm{RSS}(\boldsymbol{\beta}) = \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 \tag{6.1}$$

that can be written in matrix form as:

$$\mathrm{RSS}(\boldsymbol{\beta}) = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \tag{6.2}$$

where $\boldsymbol{X}$ is the matrix collecting the input data. It is easy to see that RSS is a quadratic function of the unknown parameter vector $\boldsymbol{\beta}$. Minimizing RSS in (6.2) can thus be done by setting to zero the first derivative of (6.2) with respect to $\boldsymbol{\beta}$:

$$\frac{\partial \mathrm{RSS}}{\partial \boldsymbol{\beta}} = -2\boldsymbol{X}^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \tag{6.3}$$

$$\boldsymbol{X}^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) = 0 \tag{6.4}$$

The matrix equation (6.4) collects the so-called *normal equations.* If the matrix $\boldsymbol{X}^T\boldsymbol{X}$ is not singular, a closed formula for the solution $\hat{\boldsymbol{\beta}}$ can be obtained as:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} \tag{6.5}$$

The estimated parameter vector $\hat{\boldsymbol{\beta}}$ could then be placed into (4.1) to obtain an estimate of the target $\hat{\boldsymbol{y}}$, the so-called "model prediction":

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} \tag{6.6}$$

As shown in box Algorithm 1, once the model parameters $\hat{\boldsymbol{\beta}}$ are estimated from the identification set, the linear model of eq. (4.1) can thus be used to predict unseen data through a linear combination of the inputs. The derivation of the solution is computed assuming a uniform precision of the reference data y_i, thus no weighting matrix is introduced.

```
load $X, y$ {load data and standardize}

standardize X,y

$\hat{\beta} \leftarrow \mathrm{inv}(X^TX)X^Ty;$
(or using QR decomposition)
$\hat{\beta} \leftarrow X\backslash y;$

$\hat{y} \leftarrow X\hat{\beta}$
```

Algorithm 1: OLS pseudocode.

6.2 Properties of OLS

A brief overview of the statistical and geometrical properties of the OLS estimator will be given in this section.

6.2.1 Statistical Properties

Suppose the measurement model to be a combination of a deterministic part (linear combination of regressors) and a random part (stationary, zero mean and constant variance ϵ_i affecting each measure y_i):

$$\begin{aligned} y_i &= \boldsymbol{X}_i\boldsymbol{\beta} + \epsilon_i \qquad\qquad \boldsymbol{\epsilon} \sim N(0, \sigma^2) \\ y_i &= y_{i,true} + \epsilon_i \end{aligned} \tag{6.7}$$

The Mean Square Error (MSE) of the estimate $\hat{\boldsymbol{y}}$ of the true value $\boldsymbol{y}_{true}$ is:

$$\text{MSE}(\hat{\boldsymbol{y}}) = E[\|\hat{\boldsymbol{y}} - \boldsymbol{y}_{true}\|_2^2] \tag{6.8}$$

Equation (6.8) can be divided in two terms, one representing the estimation error variance and the other the bias (difference between the expected value of the estimate and the true value $\boldsymbol{y}_{true}$):

$$\text{MSE}(\hat{\boldsymbol{y}}) = trace(Var(\hat{\boldsymbol{y}})) + \|\text{Bias}(\hat{\boldsymbol{y}})\|_2^2 \tag{6.9}$$

(see Section 5.2.1. The Gauss-Markov theorem [116] tells us that the OLS estimator of $\boldsymbol{\beta}$ has the smallest error variance among all linear unbiased estimators, namely it presents the lowest possible MSE.

However, it may well exist a biased estimator with smaller MSE. Since this estimator is biased it must have a very small variance in order to have smaller MSE than OLS (that is unbiased) as we will show in Chapter 7 and 8. Methods that shrink (or set) to zero some of the components of $\boldsymbol{\beta}$ may result in a biased estimate but with lower variance than the OLS estimator.

6.2.2 Geometrical Properties

OLS has a geometrical interpretation which is illustrated by means of Figure 6.1, that represents the simple case of two different input variables, $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$.

The input vectors $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, define a vector space S (yellow) while the target vector is represented by $\boldsymbol{y}$. Using the linear model, the estimation $\hat{\boldsymbol{y}}$ could be any linear

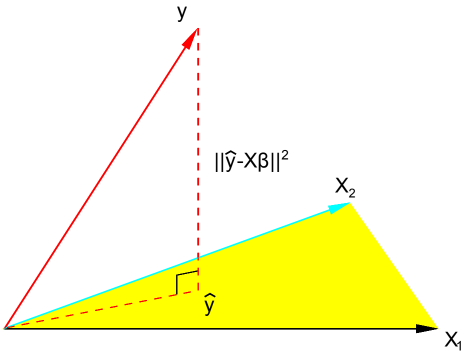


Figure 6.1: geometrical interpretation of OLS. Target vector $\boldsymbol{y}$, estimation of target vector $\hat{\boldsymbol{y}}$, input vectors $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ and in yellow the vector space S generated by the two vector. Adapted from [116].

combination of the inputs $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$. For this reason the estimate could lie anywhere in the bi-dimensional subspace S and the RSS represents the squared euclidean distance between the reference $\boldsymbol{y}$ and its estimation $\hat{\boldsymbol{y}}$. Since OLS adjusts the parameters $\boldsymbol{\beta}$ of the linear model to minimize the RSS, the OLS model prediction $\hat{\boldsymbol{y}}$ is the particular vector lying in the subspace S, which is the closest as possible to the reference $\boldsymbol{y}$. For this reason, $\hat{\boldsymbol{y}}$ corresponds to the orthogonal projection of $\boldsymbol{y}$ onto the subspace S, which is described mathematically by:

$$\boldsymbol{X}^T(\boldsymbol{y} - \hat{\boldsymbol{y}}) = 0 \tag{6.10}$$

Eq. (6.10) represents the orthogonality condition for the vector $(\boldsymbol{y} - \hat{\boldsymbol{y}})$ with respect to the subspace S defined by $\boldsymbol{X}$.

By substituting (6.6) into (6.10), one gets:

$$\boldsymbol{X}^T(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) = 0 \tag{6.11}$$

which corresponds to (6.4) and is solved by the OLS estimate.

6.2.3 Singularity Condition and Solution by QR Decomposition

If the regressors $\boldsymbol{X}_{jN}$ are not linearly independent, $\boldsymbol{X}^T\boldsymbol{X}$ is singular and can not be inverted to calculate the parameters in (6.5), yielding to a not uniquely defined $\hat{\boldsymbol{\beta}}$. However, the multiple solutions are still the projection of $\boldsymbol{y}$ onto the column space of $\boldsymbol{X}$, though there are more ways to express this projection, as there are more ways to define the subspace S.

The linear dependency of the columns of $\boldsymbol{X}$ is a consequence that one or more inputs $\boldsymbol{X}_{jN}$ present redundant information. If a couple of columns are nearly to be linearly dependent, the correlation between the two variables is high and the matrix $\boldsymbol{X}$ is not full rank. The problem of inverting $\boldsymbol{X}^T\boldsymbol{X}$ is thus ill-conditioned, leading to low accuracy of the estimated vector $\hat{\boldsymbol{\beta}}$. A typical solution for this problem is dropping redundant columns in $\boldsymbol{X}$. Other methods, as those described in the next chapters of the present thesis, provide a regularization term to cope with this low rank issue.

The most common method to recode redundant columns is the QR decomposition of $\boldsymbol{X}$:

$$\boldsymbol{X} = \boldsymbol{Q}\boldsymbol{R} \tag{6.12}$$

where $\boldsymbol{Q}$ is an orthogonal matrix ($\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}$) of dimension $(N \times p)$, while $\boldsymbol{R}$ is an upper triangular matrix of dimension $(p \times p)$. Without going into the details, these matrices are obtained by recursive orthogonalisation of the inputs, leading to an orthonormal basis for the column space of $\boldsymbol{X}$.

The QR decomposition is used to transform model (4.1) in a simpler, more stable triangular system. From (6.4) we have:

$$\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}^T\boldsymbol{y} \tag{6.13}$$

then, substituting (6.12) in (6.13) we get:

$$\begin{aligned} \boldsymbol{R}^T \underbrace{\boldsymbol{Q}^T\boldsymbol{Q}}_{\boldsymbol{I}} \boldsymbol{R}\boldsymbol{\beta} &= \boldsymbol{R}^T\boldsymbol{Q}^T\boldsymbol{y} \\ \boldsymbol{R}\boldsymbol{\beta} &= \boldsymbol{Q}^T\boldsymbol{y} \end{aligned} \tag{6.14}$$

Using QR decomposition the OLS solution is given by:

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \boldsymbol{R}^{-1}\boldsymbol{Q}^T\boldsymbol{y} \\ \hat{\boldsymbol{y}} &= \boldsymbol{Q}\boldsymbol{Q}^T\boldsymbol{y} \end{aligned} \tag{6.15}$$

The number of the estimated coefficients that are not zero is equal to the rank of matrix $\boldsymbol{X}$ and the solution coincide to (6.5) and (6.6) if $\boldsymbol{X}$ has full column rank.

6.3 Concluding Remarks

OLS is the most popular estimation method for linear regression models. The OLS solution is mathematically achieved by minimizing the residual sum of squares. This loss function has a quadratic form that allows to calculate the solution in a closed form in a very efficient way.

All these advantages make OLS an attractive estimator for linear models. However, it can lead to unsatisfactory results in several cases. First of all, the solution can not be calculated, or could be calculated only with a small precision, when there is a strong correlation between two, or more, inputs variables. In this case, the most common solution is to remove the redundant variables. In addition, it may happen that a coefficient associated with a variable results very large, while another coefficient (associated with a variable correlated with the previous one) compensates it in the opposite direction (canceling the first variable's effect). As a consequence, the information carried by one variable is deleted by the other.

7

Partial Least Squares (PLS)

As said in Chapter 4, algorithms for solving linear regression problems generally suffer from overfitting when they deal with high-dimensional datasets. This is the case of the OLS method described in Chapter 6.

In the following, we will present the PLS method and in Chapter 9, advantages and drawbacks of PLS will be shown with a simple tutorial example. Finally, in Chapter 11, PLS will be compared to the other identification techniques.

7.1 Mathematical Definition

In order to deal with overfitting, PLS regression technique discussed in this chapter resorts to dimensionality reduction, i.e. it uses M $(\leq p\ )$ new regressors $\boldsymbol{z}_k$, calculated from a linear combination of the p original ones, to model the target $\boldsymbol{y}$ $(N \times 1)$ as:

$$\hat{\boldsymbol{y}} = \boldsymbol{Z}\boldsymbol{\theta} + \epsilon \tag{7.1}$$

where $\boldsymbol{Z}$ is a $(N \times M)$ matrix, whose columns contain the so-called "latent variables" $\boldsymbol{z}_k$, $\boldsymbol{\theta}$ is the M dimensional vector of the related coefficients (which have to be estimated along with the new regressors $\boldsymbol{z}_k$(and ϵ is the error term $(N \times 1)$.

7.1.1 Derivation of the PLS estimator

Part of this material can be referred to [116]. Consider an identification set consisting of a reference vector $\boldsymbol{y}$ $(N \times 1)$, containing N samples of the target, and the corresponding input matrix $\boldsymbol{X}$ $(N \times p)$, whose rows represent the input variables $\boldsymbol{X}_{ip}$, while each column $\boldsymbol{X}_{jN}$ contains all the samples referred to the j-th variable (see Section 4.1).

Since PLS is not scale invariant, i.e. the estimates depend on the scaling of the inputs, before starting the construction of the M new regressors $\boldsymbol{z}_1$, $\boldsymbol{z}_2$,…, $\boldsymbol{z}_M$, the input variables $\boldsymbol{X}_{jN}$ have to be normalized, i.e. zero mean and unitary variance. To avoid the introduction of a new symbol below, we assume that each input variable $\boldsymbol{X}_{jN}$ is normalized.

As mentioned before, PLS iteratively constructs a set of linear combinations of the inputs, using both $\boldsymbol{X}$ and $\boldsymbol{y}$. For this construction, the original inputs $\boldsymbol{X}_{jN}$ are weighted according to their univariate effect on $\boldsymbol{y}$.

Since PLS is an iterative procedure in which the input variables $\boldsymbol{X}_{jN}$ are updated at every iteration, it is useful to add a superscript to the notation indicating the iteration number. Hence, $\boldsymbol{X}_{jN}^{(k)}$ represent the j-th input variables at the k-th iteration and $\boldsymbol{X}_{jN}^{(0)}$ correspond to the original input variables $\boldsymbol{X}_{jN}$. The same superscript is added to the estimated target variable $\hat{\boldsymbol{y}}$, as it is also updated at every iteration. In particular, at first, $\hat{\boldsymbol{y}}$ equals the mean of the reference, represented using $\bar{y}$ $(\hat{\boldsymbol{y}}^{(0)} = \bar{y})$. Then, the estimate $\hat{\boldsymbol{y}}$ is adjusted during each iteration, in which a new direction $\boldsymbol{z}_k$ is constructed.

PLS begins by computing the correlation $\hat{\varphi}_{1j}$ between the current input variables $\boldsymbol{X}_{jN}^{(0)}$ and the reference $\boldsymbol{y}$:

$$\hat{\varphi}_{1j} = \boldsymbol{X}_{jN}^{(0)}\boldsymbol{y} \tag{7.2}$$

where, in the left side, the first value of the subscript of $\hat{\varphi}$ indicates the iteration, while the second identifies the j-th variable.

Each current input variable $\boldsymbol{X}_{jN}^{(0)}$ is weighted by its corresponding correlation $\hat{\varphi}_{1j}$ in (7.2) to construct the first "derived" input $\boldsymbol{z}_1$ (N x 1):

$$\boldsymbol{z}_1 = \sum_{j=1}^{p} \hat{\varphi}_{1j}\boldsymbol{X}_{jN}^{(0)} \tag{7.3}$$

where $\boldsymbol{z}_1$ is called the first partial least squares direction. Subsequently, the reference $\boldsymbol{y}$ is regressed on $\boldsymbol{z}_1$, obtaining the scalar coefficient $\hat{\theta}_1$:

$$\hat{\theta}_1 = \frac{\boldsymbol{z}_1^T\boldsymbol{y}}{\boldsymbol{z}_1^T\boldsymbol{z}_1} \tag{7.4}$$

which is the OLS solution to the regression problem where $\boldsymbol{y}$ is the reference and $\boldsymbol{z}_1$ is the (only) input variable (compare eq. (7.4) with eq. (6.5)).

The coefficient $\hat{\theta}_1$ in (7.4) is used as the multiplier of $\boldsymbol{z}_1$ in (7.3) to update the reference estimate $\hat{\boldsymbol{y}}$:

$$\hat{\boldsymbol{y}}^{(1)} = \hat{\boldsymbol{y}}^{(0)} + \hat{\theta}_1 \boldsymbol{z}_1 \tag{7.5}$$

Using the coefficient $\hat{\theta}_1$, each current input variables $\boldsymbol{x}_{jN}^{(0)}$ is orthogonalized with respect to $\boldsymbol{z}_1$, i.e. its contribution to $\boldsymbol{z}_1$ is subtracted from it:

$$\boldsymbol{X}_{jN}^{(1)} = \boldsymbol{X}_{jN}^{(0)} - \gamma_j \boldsymbol{z}_1 \quad \text{where} \quad \gamma_j = \frac{\boldsymbol{z}_1^T \boldsymbol{X}_{jN}^{(0)}}{\boldsymbol{z}_1^T \boldsymbol{z}_1} \tag{7.6}$$

Then, the process continues until $M \leq p$ directions have been obtained.

Since the $\boldsymbol{z}_k$'s, with $k = 1, 2, \ldots, M$, are linear in the original inputs (see eq. (7.3) and (7.6)), the reference estimate after M steps, $\hat{\boldsymbol{y}}^{(M)}$, can be also computed as:

$$\hat{\boldsymbol{y}}^{(M)} = \boldsymbol{X} \hat{\boldsymbol{\beta}}^{PLS} \tag{7.7}$$

recovering the coefficients $\hat{\boldsymbol{\beta}}^{PLS}$ from the sequence of PLS transformation.

As for OLS, once the coefficients $\hat{\boldsymbol{\beta}}^{PLS}$ are estimated from the training set, they can be used in the linear model to predict unseen data through a linear combination of the inputs. It is worth noting that, if $M = p$ (i.e. the number of the PLS directions $\boldsymbol{z}_k$ equals the number of the original input $\boldsymbol{X}_{jN}$), the PLS solution is equivalent to the that of OLS.

7.1.2 Alternative implementation of PLS

Other algorithms have been developed allowing a direct estimation of the coefficients $\hat{\boldsymbol{\beta}}^{PLS}$. Without going into details, it is worth mentioning the SIMPLS algorithm [123], whose pseudo code for its derivation is depicted within box Algorithm 2, based on the input approximation using score and loading matrices:

$$\boldsymbol{X} = \boldsymbol{Z} \boldsymbol{X}_l^T + \boldsymbol{E} \tag{7.8}$$

In this case, $\boldsymbol{Z}$ is the $(N \times M)$ matrix of the M extracted score vectors (PLS directions $\boldsymbol{z}_k$), the $(p \times M)$ matrix $\boldsymbol{X}_l$ represents the matrix of loadings and $\boldsymbol{E}$ the matrix of residuals. The approximation of the target is like in (7.1):

$$\boldsymbol{y} = \boldsymbol{Z} \boldsymbol{\theta} + \boldsymbol{e} \tag{7.9}$$

```
load $X, y$ {load data and standardize}

standardize X,y

$\hat{y}^{(0)} \leftarrow 0$ {initialization}
$X^{(0)} \leftarrow X$

for $k = 1$ to $M$ do

    $\hat{\varphi}_{kj} \leftarrow X_{jN}^{(k-1)T} y$

    $z_k \leftarrow \sum_{j=1}^{p} \hat{\varphi}_{kj} X_{jN}^{(k-1)T}$
    $\hat{\theta}_k \leftarrow \frac{z_k^T y}{z_k^T z_k}$
    $\hat{y}^{(k)} \leftarrow \hat{y}^{(k-1)} + \hat{\theta}_k z_k$

    $\gamma_j \leftarrow \frac{z_k^T X_{jN}^{(k-1)}}{z_k^T z_k}$
    $X_{jN}^{(k+1)} \leftarrow X_{jN}^{(k)} - \gamma_j z_k$
```

Algorithm 2: PLS pseudocode.

The key of this algorithm is that it directly estimates a matrix of weights $\boldsymbol{W}$, representing the relationship between the PLS direction in $\boldsymbol{Z}$ with the original matrix $\boldsymbol{X}$:

$$\boldsymbol{X}\boldsymbol{W} = \boldsymbol{Z} \tag{7.10}$$

Then, substituting (7.10) into (7.9), one gets:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{W}\boldsymbol{\theta} + \boldsymbol{e} \tag{7.11}$$

the approximation of the reference $\boldsymbol{y}$ is directly related to the original inputs $\boldsymbol{X}$. Hence, ignoring the contribution of the residual matrix $\boldsymbol{e}$, the PLS reference estimate $\hat{\boldsymbol{y}}$ is obtained as:

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\boldsymbol{W}\boldsymbol{\theta} \tag{7.12}$$

By comparing (7.12) with (7.7), one gets:

$$\hat{\boldsymbol{\beta}}^{PLS} = \boldsymbol{W}\boldsymbol{\theta} \tag{7.13}$$

Hence, the matrix of weight $\boldsymbol{W}$ allows to calculate directly the estimation of the PLS coefficients $\hat{\boldsymbol{\beta}}^{PLS}$, without recovering them from the sequence of PLS transformation by a back tracking. In fact, $\boldsymbol{W}$ describes how to combine the coefficients of the new regressors $\boldsymbol{z}_k$, contained in the matrix $\boldsymbol{\theta}$.

7.2 Properties of PLS

7.2.1 Statistical Properties

It can be shown that PLS seeks directions that have high variance and high correlation with the response variable. Hence, the k-th PLS direction solves the problem:

$$\max_{\alpha} corr^2(\boldsymbol{y}, \boldsymbol{X}\alpha) var(\boldsymbol{X}\alpha) \tag{7.14}$$

with the two constraints:

$$\|\alpha\| = 1 \tag{7.15}$$

$$\alpha \boldsymbol{S} \hat{\varphi}_l = 0 \quad \text{with} \quad l = 1, 2, \ldots, k-1 \tag{7.16}$$

where $\boldsymbol{S}$ is the sample covariance matrix of $\boldsymbol{X}_{jN}$. The condition (7.16) ensures that the next direction $\boldsymbol{z}_k$ is uncorrelated with all the previous ones.

From (7.14), it can be observed that the first chosen PLS direction $\boldsymbol{z}_1$ coincides with the particular vector that lies in the $\boldsymbol{X}$ space, represented using S, and makes a compromise between its variation and its correlation with the response $\boldsymbol{y}$. Similarly, from (7.6) we can notice that the next space $S^{(1)}$, spanned by the updated input variables $\boldsymbol{X}_{jN}^{(1)}$, is the subspace of S orthogonal to the first PLS direction $\boldsymbol{z}_1$. As before, the second PLS direction $\boldsymbol{z}_2$ is that maximising the (7.14) and lying in this subspace $S^{(1)}$. Successive directions $\boldsymbol{z}_k$'s are calculated in a similar manner, with the residual subspace $S^{(k-1)}$ determined by removing from the space S, the space defined by the previous PLS directions.

7.2.2 Geometrical Properties

Figure 7.1 shows the geometry of PLS. As mentioned in Section 6.2.2, the OLS estimates $\hat{y}_{ols}$ showed as red dashed line in Figure 7.1, is the one minimizing the RSS while the first principal component indicating the direction of maximum variance of the data in $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ is indicated by the green dashed line togheter with the ellipses indicating directions of the variance of the data. The PLS solution is a trade-off between OLS and the principal components, represented as the value on the ellipse upon which OLS has the longest projection.

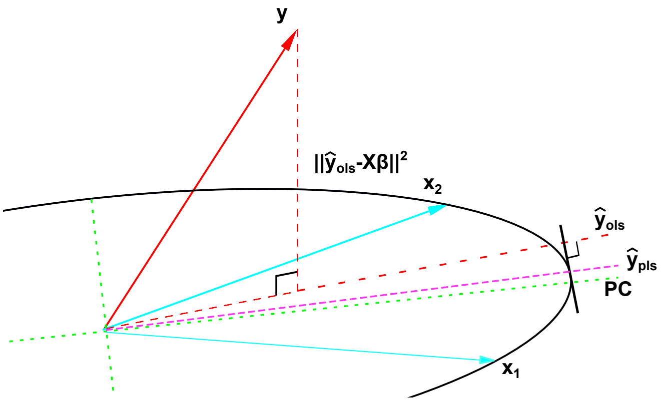


Figure 7.1: Geometrical interpretation of PLS. Target vector $\boldsymbol{y}$, estimation of target vector by OLS $\hat{\boldsymbol{y}}_{\boldsymbol{ols}}$ (red dashed line), input vectors $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$. The green dashed line represent the first principal component and the magenta line the estimation by PLS.

7.3 Concluding Remarks

PLS is a regression technique based on dimensionality reduction, which uses M new regressors, called PLS directions or latent variables, calculated from a linear combination of the original input variables depending on their univariate influence on the target. The PLS solution is iteratively obtained and at each iteration a new PLS direction is estimated.

This technique for estimating linear models tries to avoid the OLS problem of overfitting, building orthogonal PLS direction. A further feature of the PLS directions is that they are estimated maximizing both their variance and the correlation with the reference. In this way, the PLS directions try to include the informative components of the original inputs, considering also their relationship with the reference. This may be an advantage, since, as noticed from the examples, much less PLS directions are sufficient to obtain similar or even better performance than OLS. PLS will be tested on the tutorial example of Chapter 9 in order to give a general flavour of its features with respect to the other techniques. Finally, PLS will be applied in Chapter 11 to NI-CGM Multisensor data.

8

Regularization-Based Techniques: LASSO, Ridge Regression and Elastic-Net (EN)

After having presented OLS and PLS, in this chapter we will cover regression techniques that estimate the parameters of the multivariate linear model exploiting regularization. As shown below in detail, these methods add a further term to the RSS cost function in order to penalize complex models and avoid overfitting.

8.1 General Mathematical Definition

According to eq. (6.5) of the OLS estimation, the unknown coefficients of the linear regression model of eq. (4.1) can be identified minimizing the RSS($\boldsymbol{\beta}$). To reduce the risk of overfitting and numerical problems for the estimation of $\boldsymbol{\beta}$, regularization based techniques add a further term $F(\boldsymbol{\beta})$ to the cost function, tipically putting a price on $\boldsymbol{\beta}$ in order to discourage coefficients to become, in absolute value, too large, as it may happen with OLS (see tutorial example of Chapter 9). Hence, the function to minimize turns into:

$$L(\boldsymbol{\beta}, \lambda) = RSS(\boldsymbol{\beta}) + F(\boldsymbol{\beta}, \lambda) \tag{8.1}$$

and the estimated coefficients are obtained as:

$$\hat{\boldsymbol{\beta}}^{REG} = \arg\min_{\beta}(RSS(\boldsymbol{\beta}) + F(\boldsymbol{\beta}, \lambda)) \tag{8.2}$$

As we will discuss in detail in the following sections, the term $F(\boldsymbol{\beta}, \lambda)$ can incorporate ℓ_1 norm (Least Absolute Shrinkage and Selection Operator (LASSO), Section 8.2), ℓ_2 norm (Ridge regression, Section 8.3)) or a combination of the two (Elastic-Net, EN, Section 8.4), whose effects are controlled by the scalar λ [116]. The parameter λ can be thought as a parameter controlling the complexity of the model, since it prevents the model coefficients from becoming too large. According to the regularization form of the penalty term, different features will be induced on the estimated parameter vector $\hat{\beta}$.

8.2 l_1 Norm Regularization (LASSO Regression)

The LASSO solution is found as:

$$\hat{\boldsymbol{\beta}}^{lasso} = \arg\min_{\beta}\left(RSS(\boldsymbol{\beta}) + \lambda\sum_{j=1}^{p}|\beta_j|\right) \tag{8.3}$$

where, in the cost function, the coefficients of the multivariate model are penalized by considering the sum of their absolute values ($\lambda \geq 0$). Using eq. (6.1), eq. (8.3) becomes:

$$\hat{\boldsymbol{\beta}}^{lasso} = \arg\min_{\beta}\left(\sum_{i=1}^{N}\left(y_i - \beta_0 - \sum_{j=1}^{p}X_{ij}\beta_j\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j|\right) \tag{8.4}$$

By using Lagrangian multipliers [124], it can be shown that an equivalent way to write problem (8.4) is as follows:

$$\begin{aligned}\hat{\boldsymbol{\beta}}^{lasso} &= \arg\min_{\beta}\sum_{i=1}^{N}\left(y_i - \beta_0 - \sum_{j=1}^{p}X_{ij}\beta_j\right)^2\\ \text{subject to} \quad &\sum_{j=1}^{p}|\beta_j| \leq t\end{aligned} \tag{8.5}$$

where t is proportional to λ. Because of the nature of the constrains, making t sufficiently small will cause some of the coefficients to be exactly zero, leading to a sparse solution.

Unfortunately, eq. (8.4) is not differentiable when $\boldsymbol{\beta}$ contains zero values. Hence, a solution of (8.4) in closed form is not available and iterative methods are needed to compute an approximated solution. As a consequence, for computing the LASSO solution, a wide variety of approaches have been proposed in the literature to solve such a problem. In the next section, some algorithms for computing LASSO solution in an efficient way will be briefly listed; in Section 8.2.2 particular attention will be given to the Least Angle Regression (LAR) algorithm that will then be used to analyze the tutorial example data in Chapter 9 and the Multisensor data in Chapter 11.

8.2.1 Numerical Methods for Computing LASSO Estimates

This subsection gives a brief overview of the numerical methods most currently used in the literature for computing the LASSO solution. Then, in Section 8.2.2, we will describe a modification of the LAR procedure for the LASSO implementation along with its interpretation.

As mentioned above, a closed form solution for estimating the LASSO model is not available, thus iterative techniques have to be considered based on Newton's method [125]. These methods update the vector of coefficients $\boldsymbol{\beta}$ at each iteration using a descent direction of the form:

$$\boldsymbol{\beta}_{k+1} \leftarrow \boldsymbol{\beta}_k - \alpha \boldsymbol{\nabla} L(\boldsymbol{\beta}_k)/\boldsymbol{\nabla}^2 L(\boldsymbol{\beta}_k) \tag{8.6}$$

where the subscript k indicates the iteration.

Since the gradient $\boldsymbol{\nabla} L(\boldsymbol{\beta}_k)$ does not exist if some coefficients β_i are zero, different strategies were proposed to solve this problem.

Sub-gradients based algorithms use sub-gradients of the function at non-differentiable points [125] and can be classified in three different strategies, according to which variables are optimized at every iteration: coordinate descent methods [126, 127], that optimize over one variable at a time, active set methods [128, 129, 130], that optimize all the non-zero variables at every iteration and orthant-wise descent methods [131], that are similar to the previous but adds two projection operators.

Unconstrained approximation methods replace the minimization function $L(\boldsymbol{\beta})$ with a twice differentiable surrogate objective function, whose minimizer is sufficiently close to the minimizer of $L(\boldsymbol{\beta})$. The main advantage of this approach is that, since the replaced function is twice differentiable, we can directly apply an unconstrained optimization method to minimize the function. See for example [132, 133, 134] where the L1-norm constrained is replaced with the multi-quadratic functions.

Constrained optimization methods re-formulate problem (8.4) as a differentiable one with constraints. In this case, each variable β_i is represented as the sum of two variables:

$$\beta_i = \beta_i^+ - \beta_i^- \tag{8.7}$$

where $\beta_i^+ \geq 0$ and $\beta_i^- \geq 0$. In this formulation the absolute value function becomes:

$$|\beta_i| = \beta_i^+ + \beta_i^- \tag{8.8}$$

An obvious drawback of this approach is that it doubles the number of variables in the optimization problem. Different methods are based on this approach, for instance: log-barrier [135], interior-point [136], projected Newton [137] and two-metric projection [138].

8.2.2 LAR Method for Computing LASSO Solution

LAR is an iterative method intimately connected with LASSO. In fact it provides an extremely efficient algorithm for computing the entire LASSO path, i.e. the behaviour of the coefficients $\boldsymbol{\beta}$ for different values of the complexity parameter.

8.2.2.1 The LAR procedure

The LAR algorithm has been developed as a model selection algorithm [139]. It is useful to define the active set $\mathcal{A}_k$ (of dimension m) as the set of the non-zero coefficients at the k-th step. When $\mathcal{A}_k$ is used as a subscript for a matrix or a vector, it selects the values connected to the active variables at the k-th step. Hence, $\boldsymbol{X}_{\mathcal{A}_k}$ is the sub-matrix of $\boldsymbol{X}$ composed by the active variables and $\boldsymbol{\beta}_{\mathcal{A}_k}$ is the coefficient vector for these variables. To simplify the notation, the subscript k will be dropped, if it is clear that we are referring to the k-th step.

The LAR solution is computed following these steps:

1. set all the coefficients β_i to zero;

2. choose the variable $\boldsymbol{X}_{jN}$ most correlated with the reference $\boldsymbol{y}$;

3. move the correspondent coefficient β_j from zero towards its OLS value β_j^{ols} (in this way the correlation of the variable $\boldsymbol{X}_{jN}$ with the current residual $\boldsymbol{r} = \boldsymbol{y} - \boldsymbol{X}_{jN}\beta_j$ decreases);

4. continue the process until another variable $\boldsymbol{X}_{lN}$ has as much correlation with the current residual as $\boldsymbol{X}_{jN}$ has;

5. add variable $\boldsymbol{X}_{lN}$ to the active set $\mathcal{A}_k$;

6. move the coefficients $\boldsymbol{\beta}_{\mathcal{A}_k}$ towards their OLS values, in such a way that their correlation with the current residual $\boldsymbol{r} = \boldsymbol{y} - \boldsymbol{X}_{\mathcal{A}_k}\boldsymbol{\beta}_{\mathcal{A}_k}$ continues to be the same;

7. repeat steps 4-6 until $\mathcal{A}_k$ has reached the desired dimension or until all the variables have been included to $\mathcal{A}_k$ (in this case the OLS solution is obtained).

Figure 8.1 shows an example of the progression of the absolute correlations during each step of the LAR procedure. The labels at the top of the plot indicate which variable enters the active set at each step.

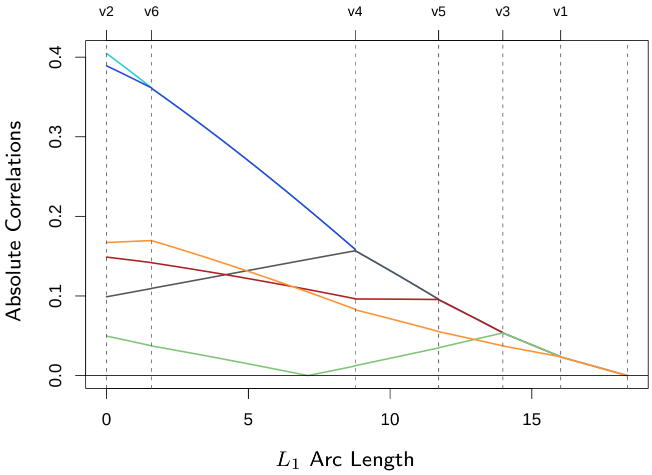


Figure 8.1: Progression of the absolute correlations during each step of the LAR procedure [116].

By construction, the coefficients β_js in the LAR algorithm change in a piecewise linear fashion. Note that we do not need to take small steps and re-check the correlation in step 4. In fact, using the knowledge of the covariance of the predictors and the piecewise linearity of the algorithm, the exact step length can be calculated at the beginning of each step.

8.2.2.2 The LAR Implementation

Having introduced the guidelines of the LAR algorithm, we can now go into its mathematical details. First of all, let us define some useful notation. $\boldsymbol{X}_{s\mathcal{A}}$ is the same as $\boldsymbol{X}_{\mathcal{A}_k}$, but each regressor is multiplied by the sign s_j of its correlation with the current residual $\boldsymbol{r}$:

$$\boldsymbol{X}_{s\mathcal{A}} = \begin{bmatrix} \dots & s_j \boldsymbol{X}_{jN} & \dots \end{bmatrix} \tag{8.9}$$

where $\boldsymbol{X}_{jN} \in \mathcal{A}_k$. For simplicity, let's define $\boldsymbol{G}_{\mathcal{A}}$ $(m \times m)$ as:

$$\boldsymbol{G}_{\mathcal{A}} = \boldsymbol{X}_{s\mathcal{A}}^T \boldsymbol{X}_{s\mathcal{A}} \tag{8.10}$$

and the scalar $A_{\mathcal{A}}$ as:

$$A_{\mathcal{A}} = \left(\mathbf{1}_{\mathcal{A}}^T \mathbf{G}_{\mathcal{A}} \mathbf{1}_{\mathcal{A}}\right)^{-1/2} \tag{8.11}$$

where $\mathbf{1}_{\mathcal{A}}$ (m x 1) is a column vector of ones.

Since the LAR procedure is not scale invariant, data have to be normalized before starting the iterative procedure. Hence, the initial target estimation $\hat{\boldsymbol{y}}_0$ is set to zero. Let $\hat{\boldsymbol{y}}_k$ the current target estimation at the k-th step, the current correlation $\boldsymbol{c}$ ($m \times 1$) of the predictors with the current residual can be written as:

$$\boldsymbol{c} = \boldsymbol{X}^T(\boldsymbol{y} - \hat{\boldsymbol{y}}_k) \tag{8.12}$$

The current active set $\mathcal{A}_k$ includes all the variables, whose absolute correlation correspond to the maximum of all the absolute correlations C_{max}:

$$\mathcal{A}_k = \{j : |c_j| = C_{max}\} \quad \text{where} \quad C_{max} = \max_j \{|c_j|\} \tag{8.13}$$

The solution at the next step is updated as follows:

$$\hat{\boldsymbol{y}}_{k+1} = \hat{\boldsymbol{y}}_k + \gamma \boldsymbol{u}_{\mathcal{A}} \tag{8.14}$$

where $\boldsymbol{u}_{\mathcal{A}}$ is a versor ($\|\boldsymbol{u}_{\mathcal{A}}\| = 1$) defining the direction to which the current target estimation $\hat{\boldsymbol{y}}_k$ is moved. This direction is calculated in such a way that the correlation of each active variables with the current residual vector equals the correlation of the other active variables. The versor $\boldsymbol{u}_{\mathcal{A}}$ is calculated as follows:

$$\boldsymbol{u}_{\mathcal{A}} = \boldsymbol{X}_{s\mathcal{A}} \boldsymbol{w}_{\mathcal{A}} \quad \text{where} \quad \boldsymbol{w}_{\mathcal{A}} = A_{\mathcal{A}} \boldsymbol{G}_{\mathcal{A}}^{-1} \mathbf{1}_{\mathcal{A}} \quad (m \text{ x } 1) \tag{8.15}$$

and, since it is an equiangular vector, it enjoys this property:

$$\boldsymbol{X}_{s\mathcal{A}}^T \boldsymbol{u}_{\mathcal{A}} = A_{\mathcal{A}} \mathbf{1}_{\mathcal{A}} \tag{8.16}$$

Instead, the coefficients are updated as follows:

$$\hat{\boldsymbol{\beta}}_{k+1} = \hat{\boldsymbol{\beta}}_k + \gamma \boldsymbol{d}_{\mathcal{A}} \tag{8.17}$$

where $\boldsymbol{d}_{\mathcal{A}}$ (m x 1) is the vector equaling $s_j w_{\mathcal{A}j}$ for $j \in \mathcal{A}_k$ (note the connection with the versor $\boldsymbol{u}_{\mathcal{A}}$ in (8.15)) and zero elsewhere.

As said before, γ can be exactly computed as to update the variables to the point in which another variable enters the active set. In particular, γ is calculated as follows:

$$\gamma = \min_{j \in \mathcal{A}^c}{}^{+} \left\{ \frac{C_{max} - c_j}{A_{\mathcal{A}} - a_j}, \frac{C_{max} + c_j}{A_{\mathcal{A}} + a_j} \right\} \quad \text{where} \quad a_j = \boldsymbol{X}_{jN}^T \boldsymbol{u}_{\mathcal{A}} \tag{8.18}$$

where $\min^{+}$ indicates the minimum between the positive values, being $\gamma > 0$.

The explanation of (8.18) is obtained by comparing the current correlation of a variable that is not in the active set with the correlation of the active variables. In particular, the current correlation of the j-th variable is:

$$c_j(\gamma) = \boldsymbol{X}_{jN}^T (\boldsymbol{y} - \hat{\boldsymbol{y}}_{k+1}) \tag{8.19}$$

then, substituting (8.14) in (8.19) one gets:

$$c_j(\gamma) = \boldsymbol{X}_{jN}^T (\boldsymbol{y} - \hat{\boldsymbol{y}}_k - \gamma \boldsymbol{u}_{\mathcal{A}}) \tag{8.20}$$

which using (8.12) and (8.18), becomes:

$$c_j(\gamma) = c_j - \gamma a_j \tag{8.21}$$

If the absolute value of (8.21) is referred to an active set variable, using (8.13) and (8.16), it becomes:

$$|c_j(\gamma)| = C_{max} - \gamma A_{\mathcal{A}} \tag{8.22}$$

then, equalling (8.21) with (8.22) one gets:

$$\left\{ \begin{array}{rcl} C_{max} - \gamma A_{\mathcal{A}} & = & c_j - \gamma a_j \\ -C_{max} + \gamma A_{\mathcal{A}} & = & c_j - \gamma a_j \end{array} \right. \tag{8.23}$$

Solving the set of equations in (8.23) for γ, one obtains the values of γ for which the correlation of a variable that is not in the active set equals the correlation of the active variables. Since we search the minimum positive value of γ, corresponding to the step of the first non active variable equalling the correlation of the active ones, we finally get the (8.18).

8.2.2.3 LAR vs. LASSO

In Figure 8.2 the coefficient profiles are plotted as model complexity increases for both LAR (left) and LASSO (right). It can be noticed that the profiles are similar to each other, except when a non-zero variable hits zero (highlighted by a red circle in Figure 8.2).

In fact, a small modification in LAR procedure allows implementing the LASSO path. The modification is the following: if a non-zero coefficient hits zero the corresponding variable is dropped from the active set and the current joint least squares direction recomputed. Below we explain why LAR and LASSO are so similar.

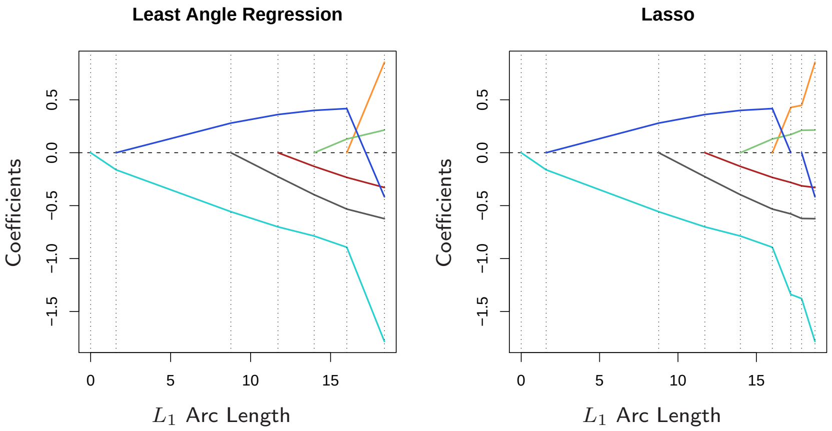


Figure 8.2: *Left*: LAR coefficients profile as the model complexity increases. *Right*: LASSO coefficients profile as the model complexity increases [116].

The correlation of an active set variable with the current residual can be expressed as:

$$\boldsymbol{X}_{jp}^T(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}) = \gamma s_j \quad \forall j \in \mathcal{A}_k \tag{8.24}$$

where $s_j \in \{-1, 1\}$ indicates the sign of the correlation and γ is the absolute value of the correlation.

Since the non-active variables are less correlated to the current residual than the active variables, we can write:

$$\left|\boldsymbol{X}_{lp}^T(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})\right| \leq \gamma \quad \forall k \notin \mathcal{A}_k \tag{8.25}$$

The LASSO minimisation function:

$$L(\boldsymbol{\beta}) = \frac{1}{2}\left\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\right\|^2 + \lambda\left|\boldsymbol{\beta}\right| \tag{8.26}$$

is differentiable for the active variables. For these variables the stationarity conditions (first derivative set to zero) are:

$$\boldsymbol{X}_{jp}^T(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}) = \lambda sgn(\beta_j) \quad \forall j \in \mathcal{A}_k \tag{8.27}$$

which corresponds to (8.24) if the sign of the correlation s_j matches the sign of the coefficients β_j. That is why the LAR algorithm and the LASSO start to differ when an active coefficient passes through zero. The LASSO condition (8.26) is violated for that variable, which is, thus, kicked out of the active set.

Finally, the stationarity conditions for the non-active variables are:

$$\left|\boldsymbol{X}_{lp}^T(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})\right| \le \gamma \quad \forall k \notin \mathcal{A}_k \tag{8.28}$$

which corresponds to the LAR equation (8.25).

8.2.2.4 LASSO Implementation by LAR modification

```
load $X, y$
normalize $X, y$

$\hat{y}_0 \leftarrow 0$;
$\hat{\beta}_0 \leftarrow 0$;
$c \leftarrow X^T y$;
$C_{max} \leftarrow \max(c)$;
$j \leftarrow \text{find}(c = C_{max})$;
$A \leftarrow x_j$

while active variables$< p$ do
  $a = X^T u_A$
  $\gamma = \min^+_{l \in A^c} \left\{ \frac{C_{max} - c_l}{A_A - a_l}, \frac{C_{max} + c_l}{A_A + a_l} \right\}$
  (associated with $X_l$)
  $\tilde{\gamma} = \min_j(-\hat{\beta}_k / d_A)$

  if $\tilde{\gamma} < \gamma$ then
    $\gamma = \tilde{\gamma}$;

  $\hat{y}_{k+1} = \hat{y}_k + \gamma u_A$
  $\hat{\beta}_{k+1} = \hat{\beta}_k + \gamma d_A$
  $C_{max} = C_{max} - \gamma A_A$

  if $\tilde{\gamma} < \gamma$ then
    drop $X_j$ from $A$
  $A \leftarrow X_l$
  update $u_A, d_A$ and $A_A$
  $c = X^T(y - \hat{y}_{k+1})$;
```

Algorithm 3: LASSO pseudocode.

The only modification of the LAR procedure needed for implementing LASSO is a check of the γ value calculated in (8.18) [139]. In fact, we have to make sure that during the LAR step none of the coefficients $\boldsymbol{\beta}$ changes its sign. In particular, starting from the updating of the coefficients in (8.18), here reported:

$$\hat{\boldsymbol{\beta}}_{k+1} = \hat{\boldsymbol{\beta}}_k + \gamma \boldsymbol{d}_{\mathcal{A}}$$

a β_j will change sign at:

$$\gamma_j = -\frac{\hat{\beta}_j}{d_j} \tag{8.29}$$

The first change occurs at:

$$\tilde{\gamma} = \min_{\gamma_j > 0} \{\gamma_j\} \tag{8.30}$$

corresponding to the j-th variable.

Hence, if $\tilde{\gamma} > \gamma$ calculated in (8.18), no sign change will occur and the LAR step does not violate any LASSO condition. Contrarily, if in (8.18) $\tilde{\gamma} \leq \gamma$, the updated coefficients $\hat{\boldsymbol{\beta}}_{k+1}$ cannot be a LASSO solution. To avoid this, the LAR step is not completed, but it is stopped at $\tilde{\gamma} = \gamma$. Then, the $\tilde{j}$-th variable is removed from the active set and a new equiangular direction in (8.15) is calculated.

The LASSO path can be estimated using the LAR modification. It can be implemented by the pseudo-code in the box Algorithm 3 (the updates of $\boldsymbol{u}_{\mathcal{A}}$, $\boldsymbol{d}_{\mathcal{A}}$ and $A_{\mathcal{A}}$ have not been reported).

8.2.3 Properties of LASSO

8.2.3.1 Geometrical Properties

As for OLS in Chapter 6, we now consider the case of two different input variables $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ [139], as can be seen from Figure 8.3. LAR builds up the estimates in successive steps, each step adding one variable to the model, according to the value of its correlation with the target variable. In the case of two input variables, the current correlations $\boldsymbol{c}$ depend only on the projection $\bar{\boldsymbol{y}}$ of $\boldsymbol{y}$ into the plane spanned by $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$:

$$\boldsymbol{c} = \boldsymbol{X}^T \boldsymbol{y} = \boldsymbol{X}^T \bar{\boldsymbol{y}} \tag{8.31}$$

As shown in Figure 8.3, $\bar{\boldsymbol{y}}$ makes a smaller angle with $\boldsymbol{X}_1$ than with $\boldsymbol{X}_2$, that corresponds to a greater correlation with $\boldsymbol{X}_1$ than with $\boldsymbol{X}_2$. Hence, the variable $\boldsymbol{X}_1$ enters the active set (step 2) and the solution moves in direction of $\boldsymbol{X}_1$, indicated in Figure 8.3 by the equiangular unit vector $\boldsymbol{u}_1$ (step 3-eq. (8.15)). Representing the moving solution of this first iteration with $\vec{\boldsymbol{y}}_1$, the current correlations $\boldsymbol{c}$ with the current residual becomes:

$$\boldsymbol{c} = \boldsymbol{X}^T(\bar{\boldsymbol{y}} - \vec{\boldsymbol{y}}_1) \tag{8.32}$$

From the Figure 8.3, we can see that the correlation of $\boldsymbol{X}_1$ with the current residual decreases. This process stops when the current residual is equally correlated with $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ (step 4), that happens when the residual vector $(\bar{\boldsymbol{y}} - \vec{\boldsymbol{y}}_1)$ bisects the angle between $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$. Hence, the variable $\boldsymbol{X}_2$ is added to the active set (step 5). Now the solution moves in such a direction as to keep equal the two correlations (step 6). This direction is represented in Figure 8.3 by the equiangular unit vector $\boldsymbol{u}_2$ (eq. (8.15)), that corresponds to the bisector of the two vectors $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$. In this case all the variables were added to the active set; hence, at the next iteration, the OLS solution is reached. Note that the OLS solution corresponds to $\bar{\boldsymbol{y}}$ (Section 6.2.2). In the general case, subsequent iterations are taken along equiangular vectors, generalizing the concept of the bisector $\boldsymbol{u}_2$.

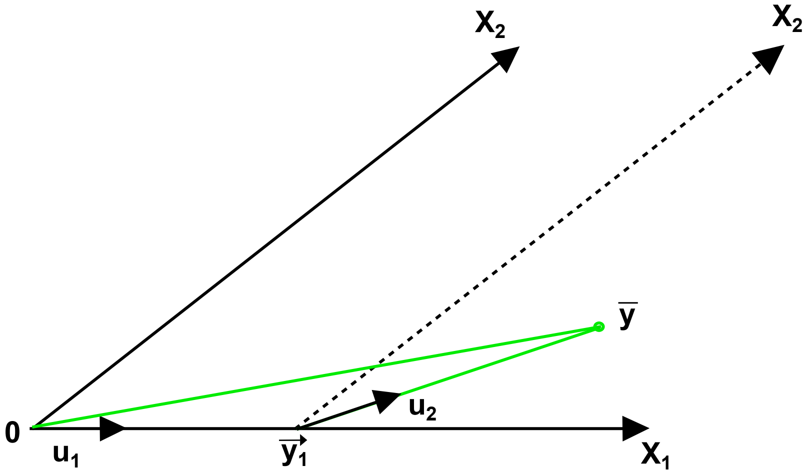


Figure 8.3: Geometrical interpretation of LASSO solution using LAR modification. Projection of the target vector $\bar{\boldsymbol{y}}$, input vectors $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$. Versor $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ indicating the equiangular vectors. Adapted from [139].

8.2.3.2 Sparse Solution

As said in the Section 8.2, the regularisation term added to RSS yielded to a sparse solution. In this Section it will be described the reason why such a constraint lead to a sparse solution, using, for simplicity, the same example of two input variables $\boldsymbol{X}_1$ and

$\boldsymbol{X}_2$.

From (8.5) the constraint region defined by LASSO is:

$$|\beta_1| + |\beta_2| \leq t \tag{8.33}$$

which is represented by a diamond area in the Cartesian space of the coefficients (blue region in Figure 8.4). As a consequence, all the possible solutions of LASSO lie in this region.

Plotting in the same Cartesian space the OLS solution ($\hat{\boldsymbol{\beta}}$ in Figure 8.4), we can see how the OLS estimates, minimizing the RSS, fall in the center of the elliptical contours which represent the RSS behaviour for different estimates of $\boldsymbol{\beta}$.

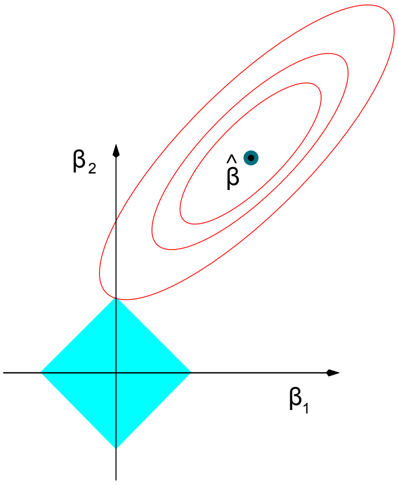


Figure 8.4: Interpretation of the sparse solution of LASSO. $\hat{\boldsymbol{\beta}}$ represents the OLS solution, the red ellipses are the contours of the residual sum of squares and the blue areas correspond to the constraint region $|\beta_1| + |\beta_2| \leq t$ (taken from [116]).

The LASSO solution is the first point where the elliptical contour hits the constraint region. Since the diamond region presents corners, it is probable that the solution occurs at a corner. In this case, one coefficient is exactly zero, in particular $\boldsymbol{\beta}_1$ in Figure 8.4. In addition, when there are more predictors, the diamond becomes a rhomboid, and has many more corners and flat edges. As a consequence, there are many more opportunities for the estimated parameters to be zero.

8.3 l_2 Norm Regularization (Ridge Regression)

Ridge regression, from now on “Ridge”, is a technique for the estimation of the parameter vector $\hat{\boldsymbol{\beta}}^{ridge}$. It is defined as the value of $\hat{\boldsymbol{\beta}}$ that minimizes a cost function given by RSS plus a regularization term given by the sum of the squares of the coefficients weighted by a parameter λ controlling model complexity [116]:

$$\hat{\boldsymbol{\beta}}^{ridge} = \arg\min_{\beta} \left(\sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} X_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right). \tag{8.34}$$

Problem (8.34) can also be formulated as a constrained optimization problem, as happened for LASSO:

$$\begin{aligned} \hat{\boldsymbol{\beta}}^{ridge} &= \arg\min_{\beta} \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} X_{ij}\beta_j \right)^2 \\ \text{subject to} &\quad \sum_{j=1}^{p} \beta_j^2 \leq t \end{aligned} \tag{8.35}$$

where t is, as in eq. (8.5), inversely proportional to λ.

λ (≥ 0) is the complexity parameter that control the amount of shrinkage. The larger its value, the greater the amount of shrinkage. The problem formulated as in (8.35) makes explicit constraint on the size of the parameters. In the case of correlated variables in the linear regression model, a large positive coefficient on one variable can be canceled by a similar large negative coefficient on a correlated predictor. As happened for LASSO, imposing a size constraint on the coefficients alleviates the problem. Since Ridge regression is not equivariant under scaling of the inputs, the predictors are centered and scaled (also to be uniform with the other identification methods).

8.3.1 Definition of Ridge Regression

Equation (8.34) is continuous and derivable thus the Ridge model, $\hat{\boldsymbol{\beta}}^{ridge}$, has a closed form solution that can be obtained by setting to zero the derivative of eq. (8.34). Recalling the function $L(\lambda, \boldsymbol{\beta}) = RSS(\boldsymbol{\beta}) + \lambda \boldsymbol{\beta}^T \boldsymbol{\beta}$, we have:

$$\frac{\partial L(\lambda, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -2\boldsymbol{X}^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) + \lambda\beta \tag{8.36}$$

$$-\boldsymbol{X}^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) + \lambda\boldsymbol{\beta} = 0 \tag{8.37}$$

By rearranging 8.37, we obtain the estimate of the model parameter vector:

$$\hat{\boldsymbol{\beta}}^{ridge} = (\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{I}_{p\times p})^{-1}\boldsymbol{X}^T\boldsymbol{y} \tag{8.38}$$

where $\boldsymbol{I}_{p\times p}$ is the $p \times p$ identity matrix. The solution adds a positive constant (λ) to the diagonal of $\boldsymbol{X}^T\boldsymbol{X}$ before inversion. Thus, even if $\boldsymbol{X}^T\boldsymbol{X}$ is not full rank, the matrix in eq. (8.38) is invertible.

To estimate the complexity parameter λ, the prediction error is plotted against the degree of freedom (df), a quantity given by:

$$df(\lambda) = tr[\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{I})^{-1}\boldsymbol{X}^T] = \sum_{j=1}^{p} \frac{d_j^2}{d_j^2 + \lambda} \tag{8.39}$$

representing the effective degrees of freedom of the ridge regression fit. Usually, the degrees of freedom in a linear regression are given by the number of free parameters. However, since all the p coefficients will be non-zero, a measure of the complexity is given in term of λ through eq. (8.39), where d_j ($d_1 \geq d_2 \geq \cdots \geq d_p \geq 0$) are the singular values of $\boldsymbol{X}$.

The Ridge estimator can be implemented by the pseudo-code depicted in box Algorithm 4. In this case, a Cholesky factorization was used to invert the $(X^TX + \lambda I)$ matrix, creating an upper triangular matrix R, satisfying the equation $R^TR = X^TX + \lambda I$.

```
load $X, y$
normalize $X, y$

$\hat{\beta}^{ridge} \leftarrow \mathrm{inv}(X^TX + \lambda I)X^Ty$;
(or using a Cholesky decomposition)
$R \leftarrow chol(X^TX + \lambda I)$
$\hat{\beta}^{ridge} \leftarrow R\backslash(R^T\backslash(X^Ty))$;

$\hat{y} \leftarrow X\hat{\beta}^{ridge}$
```

Algorithm 4: Ridge pseudocode.

8.3.2 Properties of Ridge Regression

A comparison between Ridge and LASSO constraints may help to understand the features of the two methods. Referring to eq. (8.40), the constraint region defined by Ridge Regression is a disk area in the Cartesian space of the coefficients (blue region in Figure

8.5):

$$\beta_1^2 + \beta_2^2 \leq t \tag{8.40}$$

Since in this case the disk has no corners, there is a lower probability for one coefficient collected within $\hat{\boldsymbol{\beta}}^{ridge}$ to be exactly zero. Thus, the ℓ_2 norm shrink the model coefficients but does not induce sparseness properties as the ℓ_1 norm does for the LASSO.

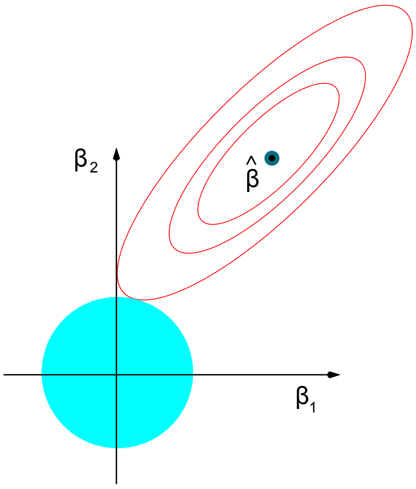


Figure 8.5: Ridge Regression regularised solution. As in 8.4, but here the blue area corresponds to the constraint $\beta_1^2 + \beta_2^2 \leq t$ (taken from [116]).

8.4 $\ell_1 + \ell_2$ Norm Regularization: Elastic-Net (EN) Regression

Considering the usual linear regression model in the form $y = \boldsymbol{X}\boldsymbol{\beta} + v$, the estimation of the parameter vector can be obtained minimizing a cost function given by RSS plus a regularization term given by the combination of the ℓ_1 and ℓ_2 norms [140].

8.4.1 Definition of Elastic-Net Regression

The term $F(\beta, \lambda)$ of the cost function in (8.1) includes two terms [140]:

$$F(\lambda_1, \lambda_2) = \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=1}^{p} \beta_j^2 \tag{8.41}$$

where λ_1 and λ_2 represent the two complexity non-negative parameters weighting the two norms. The Elastic-Net (EN) estimator is defined as:

$$\hat{\boldsymbol{\beta}}^{en} = \arg\min_{\beta} \left(\sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} X_{ij}\beta_j \right)^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=1}^{p} \beta_j^2 \right) \tag{8.42}$$

An alternative representation of the EN cost function is given by:

$$L(\boldsymbol{\beta}, \lambda, \alpha) = \frac{1}{2} \sum_{i=1}^{N} \left(y_i - \sum_{j=1}^{p} X_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} \left(\alpha|\beta_j| + \frac{1}{2}(1-\alpha)\beta_j^2 \right) \tag{8.43}$$

The cost function (8.43) introduces some scaling factors (1/2) that will be useful in the following (see Section 8.4.3.2). With this new representation, α weighs the contribution of the two norms, while λ measures the trade-off between adherence to the data and model complexity.

8.4.2 Properties of EN

In order to understand some of the EN estimator properties, it is useful to consider the geometrical interpretation of this estimator [116], as was done for LASSO and Ridge in Section 8.2 and 8.3 respectively. The constraint region defined by the combination of the ℓ_1 and ℓ_2 norms with the parameterization defined in 8.43 can be represented as a contour area, presenting sharp, thus non-differentiable corners, given by the presence of the ℓ_1 norm. This suggests a possible explanation for the properties related to the EN estimator, that share both properties of sparseness induced by the ℓ_1 norm, and a grouping effect among correlated variables given by the ℓ_2 norm.

The parameterization defined in eq. (8.43) states, even more clearly than eq. (8.42), the trade-off between Ridge and LASSO. By anticipating some properties that will be more clear analyzing the tutorial example and the application to the multi-sensor data, EN tends to average input variables highly correlated and then entering the average contribution into the model, providing a grouping effect.

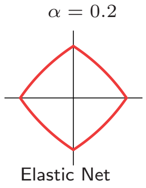


Figure 8.6: Contour of the EN penalty norm given by eq. (8.43) for $\alpha = 0.2$, presenting sharp non-differentiable corners (although not easily visible) (taken from [116]).

8.4.3 Numerical Methods for Computing EN Estimates

The EN solution can be obtained thought different approaches whose derivation depends on the form of the cost function.

8.4.3.1 LAR-EN

The LAR-EN algorithm for computing the EN solution resorts to the same algorithm proposed in Section 8.2.2 for solving the LASSO problem and is based on the cost function defined as in eq. (8.42) where λ_1 and λ_2 independently weigh the two norms. More in detail, the algorithm exploits the LAR procedure for solving the regularization problem with the ℓ_1 norm (as for the LASSO), but considers an augmented data set in order to artificially take into account the ℓ_2 norm effect. Let's consider $\alpha = \lambda_2/(\lambda_1 + \lambda_2)$, then solving eq. (8.42) is equivalent to solve the following optimization problem:

$$\begin{aligned} \hat{\boldsymbol{\beta}}^{en} &= \arg\min_{\beta} \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} X_{ij}\beta_j \right)^2 \\ \text{subject to} \quad & \alpha \sum_{j=1}^{p} \beta_j^2 + (1-\alpha) \sum_{j=1}^{p} |\beta_j| \leq t \end{aligned} \tag{8.44}$$

Eq. (8.44) is the EN penalty which is a convex combination of the LASSO and ridge penalties. For $\alpha \in [0, 1)$ the EN penalty is singular (without first derivative) at 0 and is strictly convex for all $\alpha > 0$. We define an artificial data set $(\boldsymbol{y}^*, \boldsymbol{X}^*)$ from the original one and the couple (λ_1, λ_2):

$$\boldsymbol{X^*}_{(N+p)\times p} = (1+\lambda_2)^{-1/2} \begin{pmatrix} \boldsymbol{X} \\ \sqrt{\lambda_2}\boldsymbol{I} \end{pmatrix}, \boldsymbol{y^*}_{(N+p)} = \begin{pmatrix} \boldsymbol{y} \\ 0 \end{pmatrix} \tag{8.45}$$

Let $\gamma = \lambda_1\sqrt{1+\lambda_2}$ and $\boldsymbol{\beta^*} = \sqrt{1+\lambda_2}\boldsymbol{\beta}$, the EN criterion can be written as:

$$L(\gamma, \boldsymbol{\beta^*}) = L(\gamma, \boldsymbol{\beta^*}) = \sum_{i=1}^{N+p}\left(y_i - \beta_0 - \sum_{j=1}^{p} X_{ij}\beta_j\right)^2 + \gamma\sum_{j=1}^{p}|\beta_j| \tag{8.46}$$

In this way, we have transformed the EN problem into an equivalent LASSO problem on augmented data:

$$\hat{\boldsymbol{\beta}}^* = \arg\min_{\beta^*} L(\gamma, \boldsymbol{\beta^*}) \tag{8.47}$$

with the solution to the original problem given by:

$$\hat{\boldsymbol{\beta}}^{en} = \frac{1}{\sqrt{(1+\lambda_2)}}\hat{\boldsymbol{\beta}}^* \tag{8.48}$$

Empirical evidence [140] showed that the estimator (8.48) does not perform satisfactorily unless it is close to either ridge or the LASSO. Indeed, the $\hat{\boldsymbol{\beta}}^{en}$ in eq. (8.48) is referred to as naïve EN because it performs a double shrinkage that does not help to reduce the variance much and introduces extra bias compared with pure LASSO or Ridge. This is because the naïve EN solution $\hat{\beta}$ is a two stage procedure: the Ridge regression coefficients are first obtained fixing λ_2; then, the LASSO-type problem is solved. The EN (corrected) estimate of the parameter vector $\hat{\boldsymbol{\beta}}^{en}$ is defined as $\hat{\boldsymbol{\beta}}^{en} = \sqrt{1+\lambda_2}\hat{\boldsymbol{\beta}}^*$ where $\hat{\boldsymbol{\beta}}^*$ is defined in (8.47). Rearranging and substituting in eq. (8.48) we obtain:

$$\hat{\boldsymbol{\beta}}^{en} = (1+\lambda_2)\hat{\boldsymbol{\beta}}^{old-en} \tag{8.49}$$

This scaling preserves the variable selection property of the naïve EN and is the simplest way to undo the unnecessary shrinkage mentioned above.

To calculate the LASSO-step solution for the problem (8.46), the LAR algorithm used in Section 8.2.2 can be used on an augmented data set for fixed λ_2. In particular, the step described in (8.10) now consists in calculating $\boldsymbol{G}_{\mathcal{A}_k} = \boldsymbol{X}^{*T}_{A_k}\boldsymbol{X}^*_{A_k}$ that, substituted with (8.45), becomes (at the k-th iteration):

$$\boldsymbol{G}_{\mathcal{A}} = \frac{1}{1+\lambda_2}\left(\boldsymbol{X}^T_{\mathcal{A}}\boldsymbol{X}_{\mathcal{A}} + \lambda_2\boldsymbol{I}\right) \tag{8.50}$$

8.4.3.2 Cyclical Coordinate Descent

Cyclical coordinate descent methods have been proposed several times for solving the LASSO problem [141, 125]. They belong to the family of *sub-gradients* strategies that use sub-gradients of the objective function to minimize at non-differentiable points, namely

```
load $X, y$
normalize $X, y$

$\hat{\beta}^{en}_{old} = 0$

while $true$ do
  for $k = 1, ..., p$ do
    $r_{ik} = y_i - \sum_{\substack{j=1 \\ j \neq k}}^{p} X_{ij}\tilde{\beta}_j$
    $\tilde{\beta}^*_k = \sum_{i=1}^{N} X_{ik} r_{ik}$
    $\hat{\beta}_k \leftarrow \frac{S(\tilde{\beta}^*_k, \lambda\alpha)}{1+\lambda(1-\alpha)}$

  if $\hat{\beta} = \hat{\beta}^{en}_{old}$ then
    $\hat{\beta}^{en} = \hat{\beta}$
    break
  else
    $\hat{\beta}^{en}_{old} = \hat{\beta}$

$\hat{y} \leftarrow X\hat{\beta}^{en}$
```

Algorithm 5: Pseudo-code of the cyclical coordinate descent method for computing the EN solution.

where there is a $|\beta_j|$ equal zero. The same strategy can be used to optimize the EN cost function 8.42, given the presence of the ℓ_1 term. In particular, these methods are very simple because they optimize over one variable β_j at the time, applying a soft-thresholding operator to deal with the non-differentiability points, as widely discussed by Friedman at al. [142] and Van der Kooij [143]. In practice, the following steps are repeated for $j = 1, 2, k, ..., p, 1, ...$ until convergence is reached, i.e. coefficients stabilize:

1. a coefficient of the multivariate linear regression model is chosen, for example β_k
2. the coefficient β_k is updated, with the remaining coefficients fixed

Empirical evidence showed that starting with any values for β_k, the sequence converge to the true solution [141].

The coordinate descent step for solving (8.43) is obtained computing the gradient at $\beta_k = \tilde{\beta}_k$ for $\beta_k > 0$, assuming to have estimates of $\tilde{\beta}_j$ for $j \neq k$. Thus, partially optimizing with respect to β_k:

$$\left.\frac{\partial L(\boldsymbol{\beta},\lambda,\alpha)}{\partial \beta_k}\right|_{\boldsymbol{\beta}=\tilde{\boldsymbol{\beta}}} = -\sum_{i=1}^{N} X_{ik}\left(y_i - \sum_{\substack{j=1\\ j\neq k}}^{p} X_{ij}\tilde{\beta}_j - X_{ik}\beta_k\right) + \lambda(1-\alpha)\beta_k + \lambda\alpha \tag{8.51}$$

where the sum has been broken down to isolate the contribution of the β_k coefficient. The quantity $\tilde{y}_i^{(k)} = \sum_{j\neq k} X_{ij}\tilde{\beta}_j$ is the value fitted by the model excluding the contribution of X_{ik}, so:

$$\begin{aligned} y_i - \tilde{y}_i^{(k)} &= y_i - \hat{y}_i + X_{ik}\tilde{\beta}_k \\ &= r_{ik} + X_{ik}\tilde{\beta}_k \end{aligned} \tag{8.52}$$

where $\hat{y}_i$ is the current fit of the model for observation i, and r_{ik} the current residual. The quantity $y_i - \tilde{y}_i^{(k)}$ represents the partial residual for fitting β_k. Moreover, because of the standardization of the predictors, the first term on the right hand side of (8.51) becomes:

$$\sum_{i=1}^{N} X_{ik}\left(y_i - \sum_{\substack{j=1\\ j\neq k}}^{p} X_{ij}\tilde{\beta}_j - X_{ik}\beta_k\right) = \sum_{i=1}^{N} X_{ik}(y_i - \tilde{y}_i^{(k)}) = \sum_{i=1}^{N} X_{ik} r_{ik} + \tilde{\beta}_k \tag{8.53}$$

thus, the update of the coefficient $\tilde{\beta}_k$ can be obtained setting the gradient of the cost function to zero and rearranging:

$$\tilde{\beta}_k \leftarrow \frac{S\left(\sum_{i=1}^{N} X_{ik}(y_i - \tilde{y}_i^{(k)}), \lambda\alpha\right)}{1+\lambda(1-\alpha)} \tag{8.54}$$

where the operator $S(\cdot)$ indicates the soft-thresholding operator as pointed out in [144], that showed how the solution of a problem involving ℓ_1 norm is a soft-thresholded version of the least squares estimate:

$$sign(z)(|z|-\gamma)_+ = \begin{cases} z-\gamma & \text{if} \quad z>0 \quad \text{and} \quad \gamma<|z| \\ z+\gamma & \text{if} \quad z<0 \quad \text{and} \quad \gamma<|z| \\ 0 & \text{if} \quad \gamma \geq |z| \end{cases} \tag{8.55}$$

In practice, from the k-th coefficient calculated by least-squares on the partial residuals ($\tilde{\beta}_k$), the operator $S(\cdot)$ subtracts $\lambda\alpha$ and, if it hits zero, then $\tilde{\beta}_k = 0$, otherwise the coefficient is shrunk by an amount $\lambda\alpha$.

The EN solution can be obtained implementing the algorithm described in the previous section which is based on cyclical updates of single coefficients of the parameter vector β till they stabilize or reach convergence. In the pseudo code of box Algorithm 5, the algorithmic procedure for deriving the EN solution is presented. Notice that the coordinate descent algorithm can be used for computing the LASSO solution, in the same way the LAR-EN can be used to identify the EN model. The only modification consists in the scaling factor in eq. (8.54), as discussed in [141].

8.5 Concluding Remarks

This chapter presented different techniques for the estimation of multivariate linear regression models based on regularization. In particular, the models are estimated by minimizing a cost function given by the sum of the RSS plus a term controlling complexity, penalizing complex models. This additional term can have different forms. For example, considering the ℓ_2 norm, the sum of squares of the coefficients of the parameter vector $\boldsymbol{\beta}$ will be penalized, while if the ℓ_1 norm is considered, the sum of the coefficient absolute values will be penalized. A combination of the two norms can also be used as regularization cost function, providing the so called EN model. Depending on the form of the regularization term considered, the identified model will have different features, as shown in the tutorial example of Chapter 9.

The sparseness properties of the LASSO model, identified with a modified LAR based algorithm, avoid the cancellation effect of correlated predictors occurring with OLS, by choosing only one important variable and discarding the others. Ridge regression has a closed form solution, in the opposite of LASSO, and tends to keep all the predictors but with smaller coefficients with respect to the LASSO model. Thus, similarly to what happens with PLS, the Ridge model estimates the output taking information potentially from the whole set of predictors. Finally, EN represents a trade-off between LASSO and Ridge, obtained with a combination of the two norms. This feature allows to retain the benefits derived from the use of both norms, shrinking many channel weights to zero and, at the same time, averaging those predictors with non-zero weights inducing a grouping effect.

9

Tutorial Example

While the scope of the present thesis concerns the application of the 5 parameter identification techniques discussed in Chapters 6-8 to NI-CGM Multisensor data, it is useful to first present a "tutorial" literature example that will help us to gain some confidence with the methods and highlight their and potential pros and cons.

9.1 Data Set

The data for this example is taken from a study concerning prostate cancer [145] by Stamey *et al., 1989*. They examine the correlation between height clinical measures and a target variable defined as the logarithm of the prostate-specific antigen (**lpsa**). The height variables are: log(cancer volume) (**lcavol**), log(prostate weight)(**lweight**), age, the logarithm of the amount of benign prostatic hyperplasia (**lbph**), seminal vescicle invasion (**svi**), log(capsular penetration) (**lcp**), Gleason score (**Gleason**) and percentage Gleason score 4 or 5 (**pgg45**). To model this connection, samples referred to 97 different subjects are available.

The data have been split in two data subsets: one is the identification set consisting of 67 subjects, which is used to identify the parameters of the model and the other is used to estimate the prediction error, the so called test set, consisting of 30 subjects.

With reference to the linear model of eq. (4.1), here our target vector $\boldsymbol{y}$ is the measure of “prostate-specific antigen”:

$$\boldsymbol{y} = \left.\begin{bmatrix} lpsa(1) \\ lpsa(2) \\ \vdots \\ lpsa(97) \end{bmatrix}\quad\right\}\textbf{dp}$$

while the clinical measures, defined above represent our input variables $\boldsymbol{X}_{jN}$, which are collected within the matrix $\boldsymbol{X}$:

$$\boldsymbol{X} = \left[\begin{array}{c} \boxed{lcavol(1) \quad lweight(1) \mid \overline{age(1)} \mid lbph(1) \quad \dots \quad pgg45(1)} \\ lcavol(2) \quad lweight(2) \mid age(2) \mid lbph(2) \quad \dots \quad pgg45(2) \\ \vdots \qquad\qquad \vdots \qquad \mid \quad \vdots \quad \mid \quad \vdots \qquad\qquad \vdots \\ lcavol(67) \quad lweight(67) \mid \underset{\substack{\Downarrow \\ \textbf{age}}}{\underline{age(67)}} \mid lbph(67) \quad \dots \quad pgg45(67) \end{array}\right] \begin{array}{l} \Rightarrow subject \quad 1 \\ \Rightarrow subject \quad 2 \\ \\ \Rightarrow subject \quad 67 \end{array}$$

Thus each column of the matrix $\boldsymbol{X}$ contains one of the height input variables, while the rows correspond to 67 samples referred to different subjects.

Our aim is to determine the coefficients vector $\boldsymbol{\beta}$ that describes the influence of the clinical variables upon the “prostate-specific antigen” target. Since the identification data have been standardised (mean=0 and standard deviation=1), the offset parameter β_0 can be dropped and the unknown vector $\boldsymbol{\beta}$ has dimension height:

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_8 \end{bmatrix}$$

First of all it can be useful to calculate the correlation between the different input variables.

From Table 9.1 it can be seen that **lcp** and **pgg45** are two of the most correlated variables, which is also confirmed by a visual inspection of Figure 9.1.

Before using the data for model identification and model test, a pre-processing is performed standardizing $\boldsymbol{X}$ and $\boldsymbol{y}$. This is done for allowing a direct comparison of the estimated coefficients for the different models, since the methods controlling complexity, i.e. PLS, LASSO, Ridge and EN, are not scale invariant. This introduction of the data

	lweight	**age**	**lbph**	**svi**	**lcp**	**gleason**	**pgg45**
lcavol	0.3002	0.2863	0.0632	0.5929	0.6920	0.4264	0.4832
lweight	-	0.3167	0.4370	0.1811	0.1568	0.236	0.0742
age	-	-	0.2873	0.1289	0.1730	0.3659	0.2758
lbph	-	-	-	-0.1391	-0.0885	0.033	-0.0304
svi	-	-	-	-	0.6712	0.3069	0.4814
lcp	-	-	-	-	-	0.4764	0.6625
gleason	-	-	-	-	-	-	0.7571

Table 9.1: Correlation between the different input variables, with highlighted the most elevated correlations.

set and of its features will turn out to be useful in the following sections to show pros and cons of the different identification techniques.

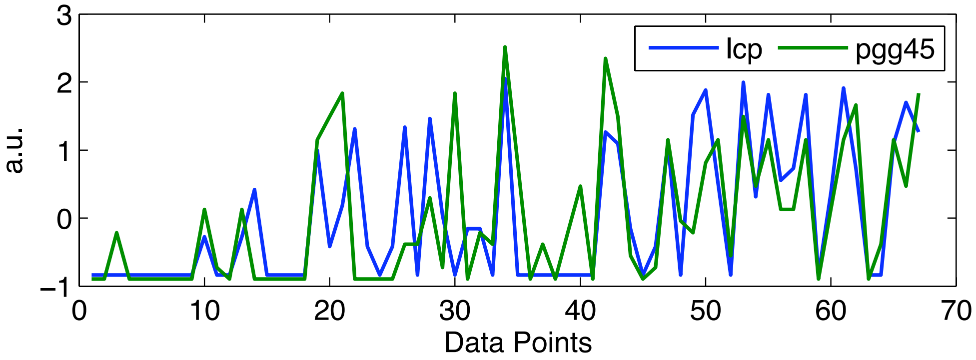


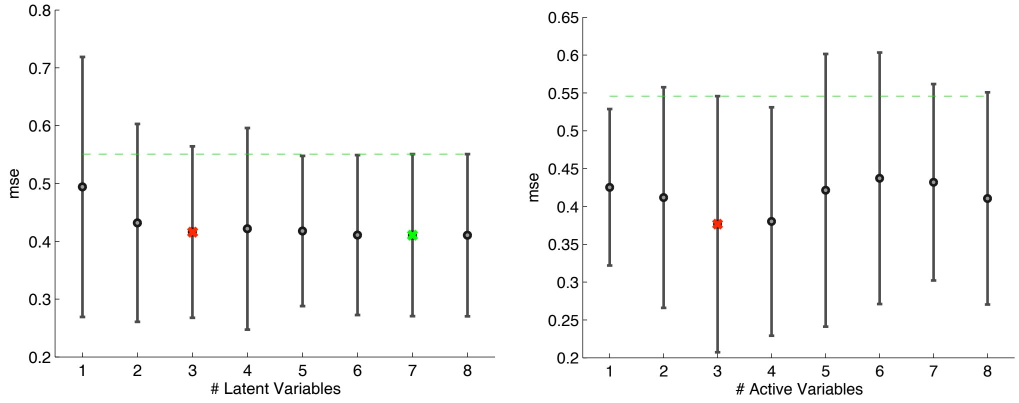
Figure 9.1: Plot of two of the most correlated variables **lcp**(blue) and **pcc45** (green).

9.2 Cross-Validation for Model Complexity Estimation

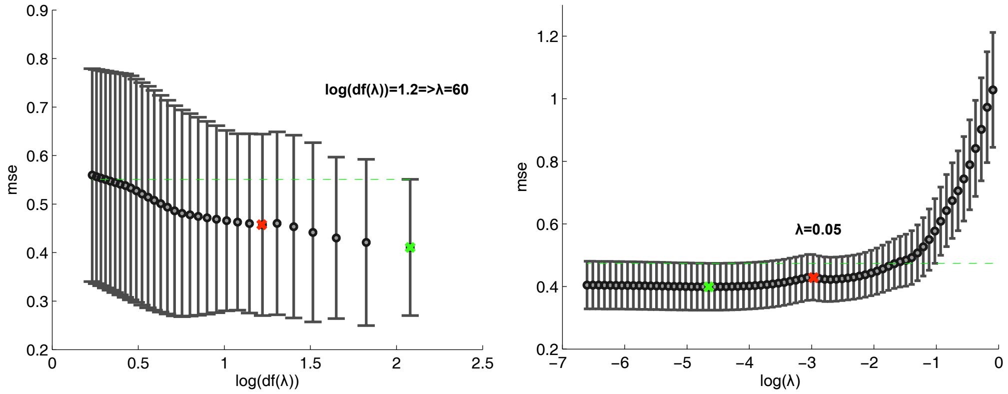
The methods controlling complexity require the estimation of the complexity parameter(s) before identifying the coefficients of the model on the identification data set. Figure 9.2 shows in each subplot the error curve for each method estimated by means of the cross-validation procedure described in Chapter 5. The test error curve is estimated using 8-fold cross-validation. The identification data are randomly split into 8 parts of approximately equal size. Iteratively, one part is left aside to calculate the test error (using MSE), while the other 7 parts are used to "estimate" the coefficients of the model. In this way a test error upon each of the 8-th parts not used for identify the models is calculated and, averaging these values, an estimation of the test error is obtained.

The model complexity is selected using the “one-standard error” rule (Section 5.2.2), which indicates the best model as the most parsimonious one, whose error is less then the minimum plus one time its standard deviation.

For PLS, the selected model correspond to the minimum value of the test error curve at seven directions. However, as mentioned in Section 5.2.2, the complexity parameter can be chosen as the one where there is a clear drop in the error curve. In our case, the value of M can be set to 3. Similar considerations can be done for LASSO, Ridge and EN.



(a) PLS (b) LASSO



(c) Ridge (d) EN

Figure 9.2: 8-fold Cross-validation curves for the choice of the most reasonable complexity parameters for PLS (a), LASSO (b), Ridge (c) and EN for $\alpha = 0.8$ (d). The MSE (mean value and one standard deviation) is represented as a function of the model complexity parameter for each method. The green cross represents the value of the complexity parameter according to the “one-standard error” rule (green dotted line), while the most reasonable complexity is chosen in correspondence of the drop of the error curve and displayed as a red cross..

In particular, for LASSO, the test error curve is plotted as a function of the number of active variables, instead of λ. However, the number of the active variables is intuitively connected to the model complexity, but also to the degrees of freedom of the model (see [146] for more details). The test error curve has a minimum in correspondence of the "one-standard error" rule, which also coincides with the point of drop of the test error curve. Hence, the finally chosen model has 4 active variables.

In subplot (c) of Figure 9.2 the test error curve for the ridge model is reported as a function of a quantity defined as the degree of freedom which is inversely related to λ. The complexity parameter should be chosen in correspondence to a value close to 1.25 of the degree of freedom on the logarithmic scale (corresponding to $\lambda = 300$), according to the "one-standard error" rule. However, this value is too large, and a more reasonable value to choose for λ is the one where the test error curve presents a drop in slope, namely, at $log(df(\lambda)) = 1.2$ corresponding to $\lambda = 60$.

As already mentioned, EN has two parameters for controlling complexity: λ is the regularization parameter weighting the trade-off between adhesion to the data (low RSS) and model complexity (discouraging complex models), while α controls the contribution of the two norms. The cross-validation procedure is used for choosing the complexity parameters. In particular, a grid of 11 equally spaced αs is considered, in the interval $(0 \div 1)$. Then, a set of λs equally spaced on the log scale is evaluated for each value of α. Thus, each cross-validation plot is inspected separately and λ will be chosen according to the specific α with the one-standard-error rule (see red cross in Figure 9.2) or as an alternative after the first drop in the error curve, namely when $log(\lambda) \approx -3$ corresponding to $\lambda = 0.05$. In this case, $\alpha = 0.8$ was chosen, since it was the one giving the lower MSE with reasonable complexity.

9.3 Model Identification

The models are identified over the same data set used for the cross-validation procedure. Table 9.2 shows the coefficients for each variable estimated with the proposed identification techniques. It can be noticed that for OLS, the contribution of the two correlated variables **lcp** and **pgg45** (see Figure 9.1) to the estimation of $\hat{\boldsymbol{\beta}}$ is not relevant, since their relative coefficients tend to compensate their effects. This phenomenon occurs when OLS deals with highly correlated variables: their relative coefficients tend to become large but with opposite signs and thus they compensate each others.

After having fixed $M = 3$, the PLS estimates can be computed. It is interesting to compare the estimated OLS coefficients $\hat{\boldsymbol{\beta}}^{ols}$ with the PLS ones $\hat{\boldsymbol{\beta}}^{pls}$, as reported in Table 9.2. The compensation effect occurring with the variables **lcp** and **pgg45** for OLS is not happening using PLS, that weights the variable **lcp**less. From Table 9.2 and Figure 9.4, we can also notice that the estimated PLS coefficients have, on average, a smaller absolute value than the OLS ones, indicating the a control of the complexity as been achieved.

As described before, the LAR procedure allows to create the entire LASSO path (see Figure 9.3), i.e. the behaviour of the coefficients $\hat{\boldsymbol{\beta}}$ as the model complexity increases. At first all the parameters $\boldsymbol{\beta}$ are set to zero and enter in the active set according to their correlation with the current residual. Notice that, at the end of the LASSO path, i.e. when the selected number of active variables is equal to the number of predictors in the matrix $\boldsymbol{X}$, $\hat{\boldsymbol{\beta}}^{lasso}$ correspond to $\hat{\boldsymbol{\beta}}^{ols}$.

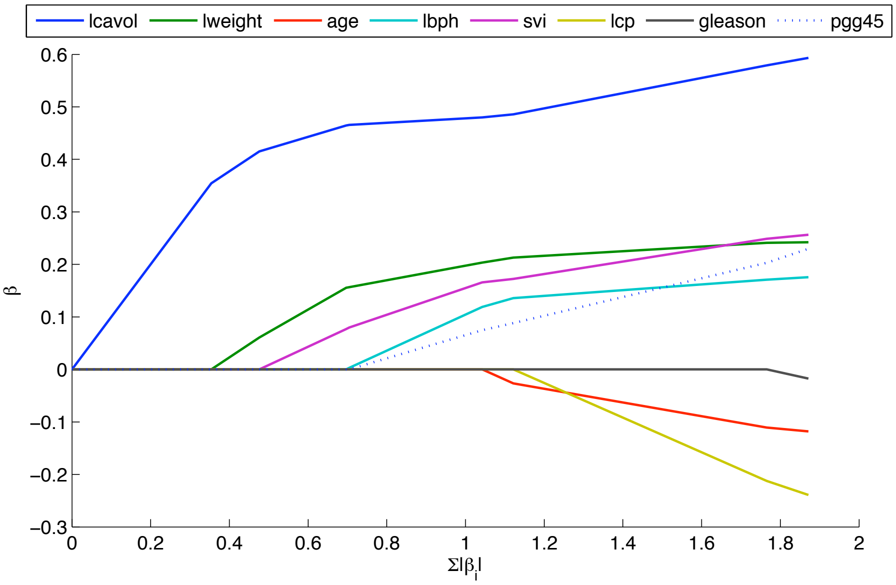


Figure 9.3: LASSO path for prostate cancer data of the "tutorial" example. The coefficients weighting the different variables (expressed in different colors) are shown as a function of the model complexity (expressed as the sum of the absolute value coefficients in the model).

Analyzing Figure 9.3 and Table 9.1 we can see the LASSO behaviour when dealing with correlated variables, e.g. **lcp** and **pgg45**. **pgg45** (blue dashed line in Figure 9.3) enters the active set before **lcp** (yellow continuous line) showing the regularization and variable selection performed by the ℓ_1 norm. As soon as the model complexity increases, **lcp** enters the active set and its coefficient becomes large until compensation of the two variables occurs.

Table 9.2 shows that Ridge estimated coefficients have sum of absolute values clearly lower than that of the coefficients obtained with OLS ($\sum |\hat{\beta}_j^{ols}| = 1.8716$ against

$\sum|\hat{\beta}_j^{ridge}| = 0.6686$ respectively). This proves the regularization performed by the ℓ_2 norm that shrunk the model coefficients. However, the ℓ_2 norm does not induce sparsity on the coefficients, thus allowing all the variables to enter the model. However, they are individually much smaller that the coefficients of the other methods (see Figure 9.4).

As it can be seen from Table 9.2, the EN model still share the property of sparseness with LASSO (induced by the ℓ_1 norm), but retains more variables (thanks to the ℓ_2 norm). Unfortunately, the grouping effect is not visible, probably because of the small number of predictors and data available.

	lcavol	**lweight**	**age**	**lbph**	**svi**	**lcp**	**gleason**	**pgg45**
OLS	0.5931	0.2423	-0.1180	0.1755	0.2563	-0.2393	-0.173	0.2296
PLS	0.4948	0.2594	-0.1520	0.1696	0.2586	-0.0306	0.0053	0.1019
LASSO	0.48	0.2034	0	0.1188	0.1658	0	0	0.0746
Ridge	0.1669	0.1156	0.0195	0.0701	0.1122	0.0688	0.422	0.0713
EN	0.4778	0.2101	-0.0185	0.1309	0.1722	0	0	0.0863

Table 9.2: Estimated coefficients of the parameter vector $\hat{\boldsymbol{\beta}}$ for OLS, PLS, LASSO, Ridge and EN.

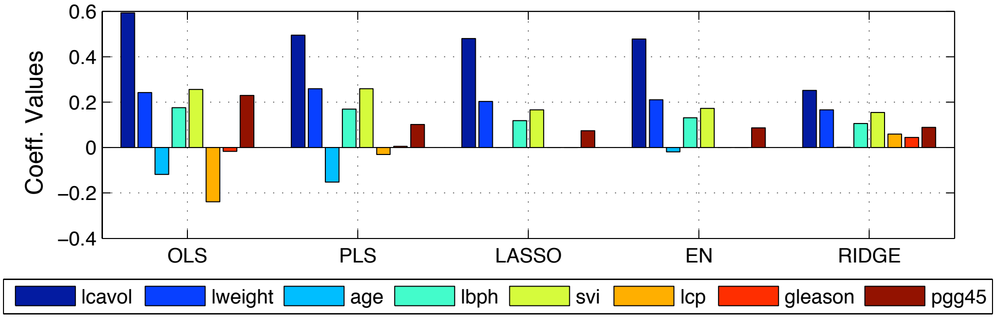


Figure 9.4: Coefficients of the multivariate linear regression model identified by the different techniques.

9.4 Model Test

As described in Chapter 3, to evaluate the performance of the two different methods, it is convenient to analyse their behaviour in predicting unseen data. Hence, the previously estimated coefficients are applied on inputs of the test set and the results are compared with the test reference. To quantify the performance of the models, the MSE indicator was computed. Table 9.3 shows, as expected, that OLS is the model with lower accuracy indicating the occurrence of overfitting.

As said before, using the OLS estimator, the coefficients of highly correlated variables tend to grow large in opposite directions compensating each others. It was the case of **lcp** and **pgg45** which are positively correlated. LASSO choses only one of the two variables, discarding the other one by shrinking its coefficient to zero. To quantify the performance of the three methods, the MSE indicator is considered, as shown in Table 9.3. Table 9.3 confirms that the estimators have similar performances.

Despite the regularization, the Ridge model is not able to generalize as well as the other models do in predicting the target variable from the input data. However, results remains comparable with those of the other models.

Although the grouping effect is not visible in this tutorial example, the combination of the two norms allows the EN model to outperform the other two models identified with a regularization technique (see Table 9.3). Probably due to the few data and predictors available, performance is not as good as that of PLS, although very close.

	MSE
OLS	0.5213
PLS	0.4284
LASSO	0.4593
Ridge	0.5257
EN	0.4583

Table 9.3: MSE indicator for OLS, PLS, LASSO, Ridge and EN on test data.

9.5 Concluding Remarks

This chapter illustrated a procedure for assessing accuracy of different identification techniques. The same logic will be used in Part III of this thesis to test the same techniques over the Multisensor data. From data pre-processing, to cross-validation for setting the more reasonable complexity parameter for each techniques, the models are finally identified and tested over an independent test data set. Of particular interest is the effects on the model coefficients induced by those techniques controlling complexity (see Figure 9.4). In particular, while retaining information from all the predictors, PLS estimates a model with visually smaller coefficients than OLS, resulting in a less complex model able of better generalization on test data (see Table 9.3). On the other side, LASSO induces a sparse model, with 3 coefficients shrunk to zero in a total of 8, obtaining good prediction performance on test data, not comparable with that of PLS, thought. The reason could be that with few data available, PLS has better prediction capabilities because it takes information from all the variables. The ℓ_2 norm induces a model (Ridge) where all the coefficients are non-zero, as can be easily seen from Figure 9.4. However, they are individually much smaller that the coefficients of the other methods. Finally, the EN model is a trade-off between the LASSO and Ridge ones, presenting 2 coefficients shrunk to zero. Unfortunately, the grouping effect representing its main feature is not fully visible in this tutorial example, but will be clear in Chapter 11 where the identification techniques will be applied with the aim of performing NI-CGM.

Part III

Case Study

10

Data Set

The present chapter illustrates the data set and the relative acquisition protocol that will be used later in Chapter 11 to assess the performance of the identification techniques in modeling multi-sensor data. Starting from this chapter, we will refer to a particular multi-sensor device, namely the Solianis Multisensor, from now on, for sake of brevity, called “Multisensor” (note the capital M).

10.1 Acquisition Protocol

Data, provided to us by Solianis Monitoring AG, were acquired during an experimental clinical study conducted at the University Hospital Zurich that included six patients with Type 1 *Diabetes Mellitus* (T1DM) (age 44 ± 16 years; body mass index BMI 24.1 ± 1.3 kg m^{-2}; duration of diabetes 27 ± 12 years; glycated hemoglobin HbA1c 7.3 ± 1.0), identified by the following labels: “AA02”, “AA03”, “AA04”, “AA05”, “AA06”, and “AA09”. Each subject performed different recording sessions in different days. Each recording session had an approximative duration of 8 hours during which plasma glucose was induced to vary according to a desired profile. In particular, glucose was loaded either orally or by intra-venous glucose administration to induce different hyper and hypoglycaemic excursions. In total, four different desired profiles were considered. These profiles are shown with different colors in Figure 10.1, where the black vertical dashed line represent the first 75 minutes of the experiment (that will be later removed from the study). The rationale of forcing glucose to mimic such a variety of profiles is to assess the ability of the "Multisensor hardware + model" system to discriminate among both different glucose rates of change and levels of glucose concentrations.

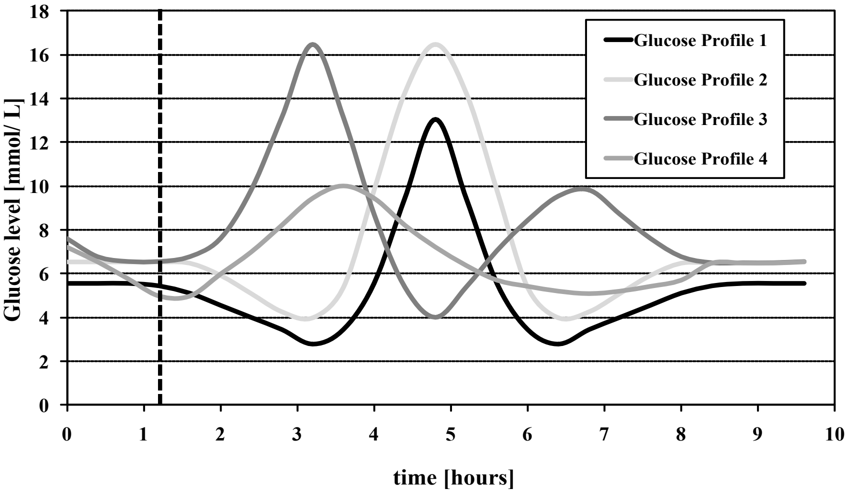


Figure 10.1: The four desired glucose profiles considered in the protocol. Time zero corresponds to intravenous insulin infusion, and the black vertical line the first 75 minutes.

The study was performed in accordance with Good Clinical Practice and the Declaration of Helsinki. All patients signed an informed consent agreement,performed the screening visit and were then enrolled in the study. After a patient arrived in the clinical study unit in the morning, blood glucose was measured and an intravenous insulin infusion was performed. Glucose was administered after a 75 min equilibration time needed for establishing euglycaemic level and to allow the skin of the subject to adjust to the application of the sensor. Multisensor data were recorded by placing the device on the right upper arm. Reference glucose values were acquired in parallel, every 10-20 min, using a HemoCue Glucose 201 Analyzer (HemoCue AG, Switzerland). On average, seven recording sessions were performed by each patient (min. 5 and max. 10). This provided a data set of 45 recording sessions available for the analysis described in the following.

As mentioned in Chapter 3, the Multisensor provides a set of measurements of different nature, mainly based on dielectric and optical sensors, for a total of more than 150 measured signals. Most of the signals come from the dielectric electrodes (see Figure 10.2), showing a high correlation and exhibiting similar but not identical behaviour.

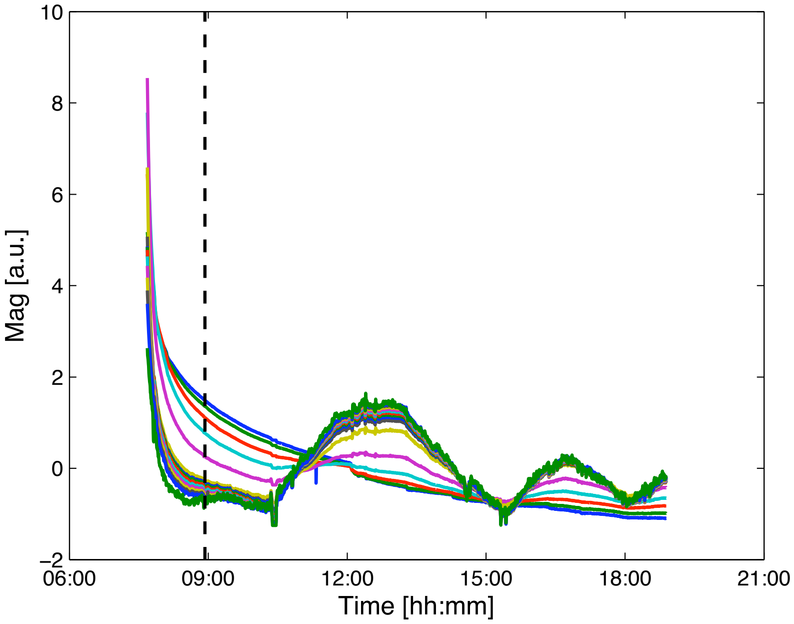


Figure 10.2: Example of IS Multisensor data. The first 75 mins (on the left of the dashed vertical line) are removed for the presence of Multisensor-skin adaptation processes and for allowing euglycaemic level to be established.

Hence, there are two important characteristics of this dataset: it is a high-dimensional dataset and there are many correlated variables. Figure 10.2 also clarifies the reason the first 75 minutes are removed: in this time interval there is a strong influence of adaptation processes due to Multisensor-skin contact.

10.2 Data Partition Between Model Identification and Model Test

As said in Chapter 3, it is a good choice to evaluate the performance of the different models using unseen data. Hence, in order to evaluate the performance in estimating glucose profiles from Multisensor data not used during the model identification stage, the data set was split into two parts, in such a way that each subject in each data subset underwent a similar number of days with a specific profile. Data subsets used in the following are:

- data subset “part 1”, consisting of 23 recording sessions;
- data subset “part 2”, consisting of 22 recording sessions.

These two data subsets will be used separately for model identification and model test, namely, if data subset "part 1" is used for model identification, data subset "part 2" is used for model test and *viceversa*. In Chapter 11 we will refer to "internal validation" results if the model is applied to the same data subset used for its identification, and "external validation" if the model is applied to a new data subset.

Notice that both the data subsets contain data recorded from different subjects, thus the identified models will have a "global" or "population" validity since they are not tailored to a specific subject. In a practical prospective this could be an appealing aspect since it could allow to use a previously identified population model for estimating glucose profiles also in subjects whose data did not participated to model building.

10.2.1 Preprocessing

Data for model identification

For each Multisensor channel, the first 75 min of each recording session are removed since this interval is dominated by an adaptation process due to the Multisensor/skin contact. Signal channels undergo a causal median filtering (window width of 5 samples) for the removal of occasional spurious spikes. Signals used for model identification are standardized to have zero mean and standard deviation one, namely, they are shifted and scaled with their own sample mean and standard deviation.

Data for model test

The first 75 min of each recording session are removed and the same causal median filtering above was applied. Then, each signal channel is shifted and scaled using sample mean and sample standard deviation of its correspondent in the identification data set. In such a way, the analysis can be considered consistent with a realistic on-line application of the models. Indeed, in a prospective use of the device, sample mean and standard deviation cannot be known in advance, and only the values estimated during the model identification stage can be used.

10.2.2 Determination of Model Complexity

While for OLS the identification data subset is used to identify the model coefficients, which are then applied on the test data subset to estimate BGL, for the techniques controlling complexity an additional step is needed before estimating $\hat{\boldsymbol{\beta}}$. In particular, the complexity parameter(s) need(s) to be fixed exploiting K-fold cross-validation over the identification data subset (see Section 5.2.2). After having estimated the model complexity, PLS, LASSO, Ridge and EN models are identified from the same identification data subset used for cross-validation and applied on the test data subset for predicting the BGL values.

10.2.3 Model Calibration

While models obtained during the model identification stage will have a “global” validity because they are obtained considering identification data subset containing data from different subjects, during the model test phase an individualized calibration step is required at the beginning of each experimental session to adjust the baseline of the estimated glucose profile by the model. Formally, such a calibration is described by equation:

$$\hat{\boldsymbol{g}}_{cal} = \boldsymbol{X}\hat{\boldsymbol{\beta}} + b \tag{10.1}$$

where $\hat{\boldsymbol{g}}_{cal}$ is the $(N \times 1)$ vector containing the calibrated glucose profile, from now on only “glucose profile”, $\boldsymbol{X}$ is the $(N \times p)$ matrix collecting Multisensor data, $\hat{\boldsymbol{\beta}}$ is the $(p \times 1)$ identified parameter vector of the multivariate linear model (no matter which of the 5 parameter identification techniques is adopted) and b is the scalar value representing the baseline glucose calibration parameter calculated exploiting a single RBG provided by a “gold standard” technique based on finger prick. This additional parameter is obtained as the difference between the estimated glucose value given by the multivariate linear model $\boldsymbol{X}_i\hat{\boldsymbol{\beta}}$ and the RBG point at the same time instant t_i:

$$b = \boldsymbol{X}_i\hat{\boldsymbol{\beta}} - RBG(t_i) \tag{10.2}$$

In practice, the glucose profile is shifted to the first RBG value available. This initial adjustment is usually performed after 75 minutes the Multisensor is placed in contact with the skin, for allowing adaptation processes related to Multisensor-skin contact to deplete, and then kept fixed for all the duration the Multisensor is worn.

11

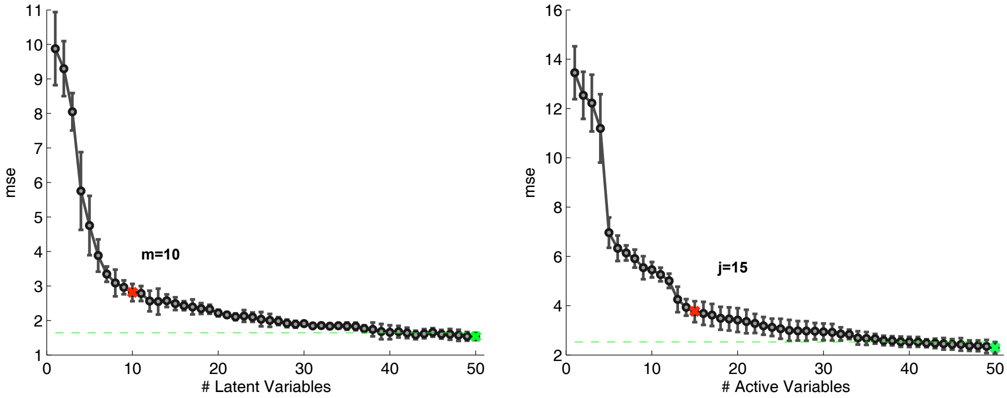
Results

As already mentioned in the previous chapter, the full dataset was split in “part 1” and “part 2”. Hereafter, if “part 1” is used for model identification, “part 2” is used for model test and *viceversa*. The identification data subset used to find the model parameter vector is also previously used to find the most reasonable complexity parameters for PLS, LASSO, Ridge and EN.

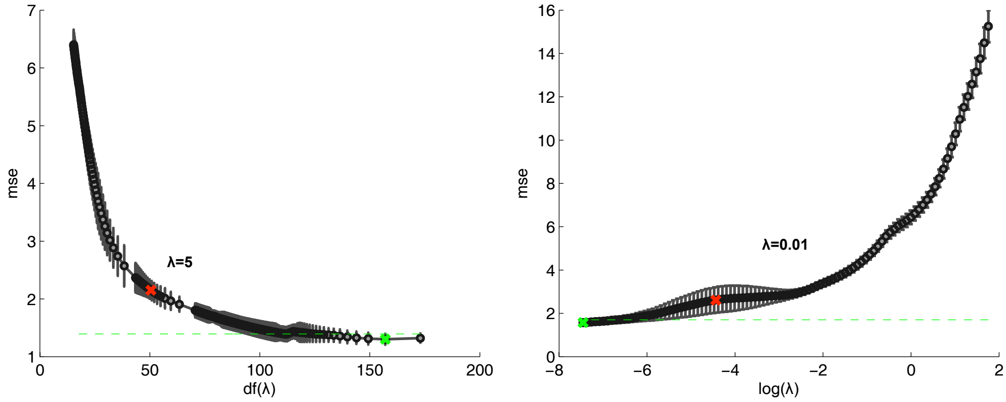
11.1 Determination of Model Complexity

The “optimal” complexity parameter values are shown in Table 11.1-11.4, for the different techniques. Their values are determined according to reasonable empirical evidence, i.e. where the cross-validation curve presents a clear drop in slope (values are reported as red crosses in Figure 11.1), rather than with the “one-standard error” rule (whose values are reported as green crosses in Figure 11.1). Figure 11.1 shows the cross-validation results when data subset “part 1” is considered for model identification, and comparable results (not shown) are obtained when data subset “part 2” is used.

The cross-validation curve in Figure 11.1 (a) shows the error curve as a function of the number of latent variables for the PLS technique. The “optimal” complexity parameter value suggested by the “one-standard error” rule, indicated with a green cross in subplot (a) at the value of $m = 50$, is likely to lead to an unnecessary too complex model. Indeed,



(a) PLS (b) LASSO



(c) Ridge (d) EN

Figure 11.1: 10-fold Cross-validation curves for the choice of the "optimal" complexity parameters for PLS (a), LASSO (b), Ridge (c) and EN for $\alpha = 0.4$ (d). The MSE (mean value and one standard deviation) is represented as a function of the model complexity parameter for each method. The green cross represents the value of the complexity parameter according to the one-standard-error rule (horizontal green dashed line), while the red crosses represent the values according to the drop in the error curve.

visual inspection of the cross-validation plot shows a clear drop of the error curve around 10. The complexity parameter for the different identification techniques suggested by the "one-standard error" rule is shown in the subplots of Figure 11.1 with green crosses, while the red crosses is the chosen value according to the drop in the error curve. This empirical consideration also drives the choice of the complexity parameter for the LASSO model, indicating a drop of the cross-validation curve around 15 (see subplot (b) in Figure 11.1). The choice of the complexity parameter for Ridge follows a similar approach. Indeed, the cross-validation curve shown in subplot (b) of Figure 11.1 has a drop when the degree of freedom, defined by eq. (8.39), is approximately 50, corresponding to $\lambda = 5$. Similarly for EN, the ending part of the drop in the error curve can be noticed for $log(\lambda) \approx -4.5$ (subplot (d) of Figure 11.1), corresponding to $\lambda = 0.01$. For EN different cross-validation curves for different values of α where examined. The most reasonable choice seemed that obtained for $\alpha = 0.4$. Indeed, this combination of complexity parameters is the one providing a good trade-off between the ℓ_1 and ℓ_2 norms allowing a reasonable complexity for the EN model to be achieved. A value of $\alpha = 0.4$ can suggest that, although it is important to shrink channel weights to zero in order to lower the probability of occasional jumps or spikes entering the model, allowing a grouping effect over correlated predictors is also important for a more robust estimation of glucose profiles.

11.2 Model Identification

	RMSE [mg/dL]	R^2	MAD [mg/dL]	MARD [%]	ESOD	EGA [%] $A+B(A)$ $C\backslash D\backslash E$	CEGA [%] $A_R+B_R(A_R)$ $C_R\backslash D_R\backslash E_R$
OLS	20.5	0.94	17	13.8	2.1	96.2(78.3)	86.9(58.7)
	(7.9)	(0.02)	(8)	(6.1)	(1.5)	0\3.8\0	3.5\6.2\3.4
PLS	39.5	0.84	33.3	28.9	1.8	88.9(39)	83.7(56.5)
(m=10)	(12.3)	(0.09)	(12.1)	(14.2)	(1.7)	4.1\5.7\1.3	7.2\5.2\3.4
LASSO	49.8	0.78	41.8	36.8	0.8	89.9(40.8)	86.9(61.1)
(j=15)	(16.7)	(0.14)	(16.5)	(20.1)	(0.7)	2.1\6.9\1.1	7.9\3.4\1.8
Ridge	32.7	0.89	27.7	24.4	1.5	91.9(41.6)	83.4(58)
($\lambda = 5$)	(11.9)	(0.06)	(11.4)	(12)	(1.1)	1.1\5.9\1.1	7.2\5.9\3.5
EN	31.2	0.89	26.4	22.6	1.4	93.0(39.4)	83.9(59.6)
($\alpha = 0.4, \lambda = 0.01$)	(11.8)	(0.06)	(11.2)	(10.7)	(2.4)	0.7\5.1\1.2	7.9\5.2\3

Table 11.1: Indicators of model performance for internal validation, i.e. when glucose profiles are estimated from the same data subset "part 1" used for identify the models. In brackets is the complexity model parameter chosen by means of cross-validation. RMSE root mean squared error, R^2 Pearson coefficient of determination, MAD mean absolute difference, MARD mean absolute relative difference, ESOD energy of second-order differences, EGA (Clarke) error grid analysis, CEGA continuous error grid analysis.

In this section, Table 11.1 and Table 11.2 represent the results of the so-called "internal validation", namely when glucose profiles are estimated with the same data used to identify the models. In particular, Table 11.1 shows internal validation results for and data subset "part 1" and Table 11.2 data subset "part 2".

Results in terms of accuracy of estimated glucose profiles are presented through indicators widely discussed in Chapter 5. As expected, Table 11.1 and Table 11.2 indicate that, in the model identification stage, OLS outperforms the other models. Indeed,

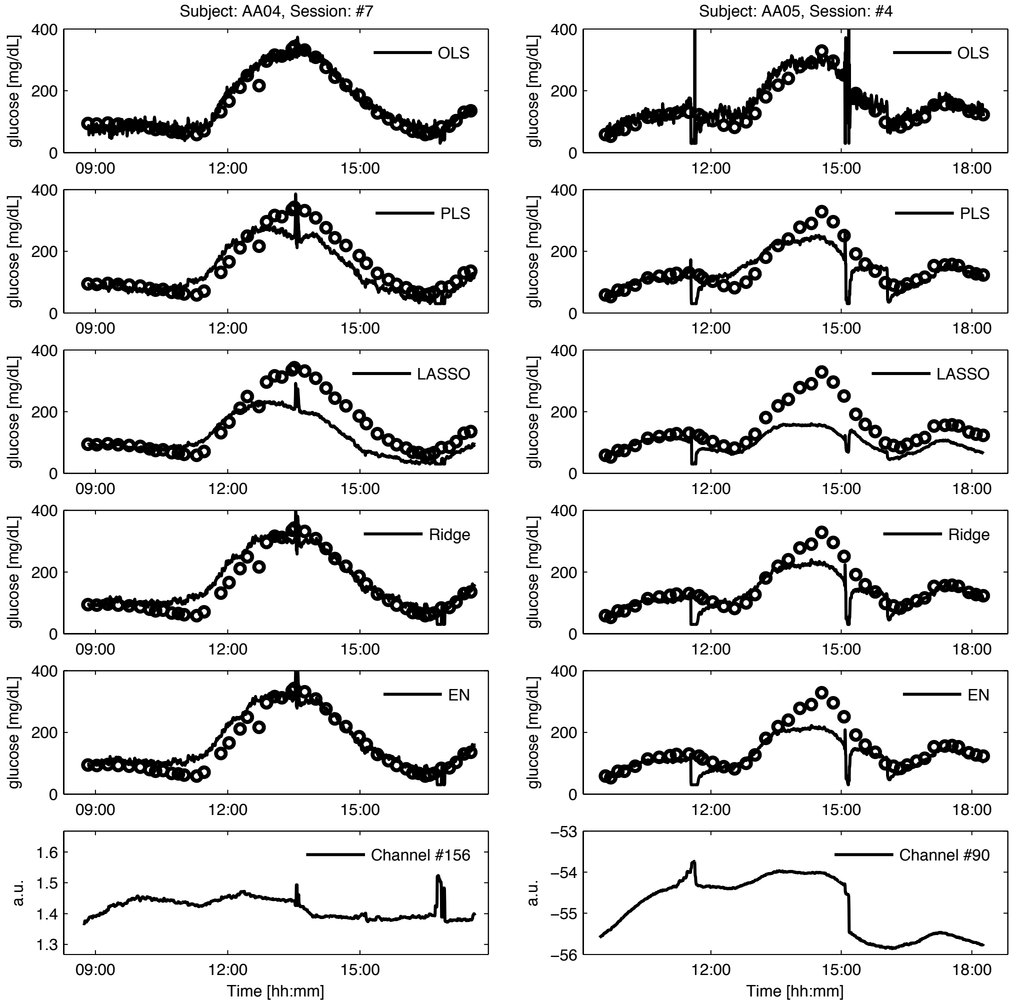


Figure 11.2: Representative recording sessions of Subjects AA04 (left) and AA05 (right). OLS, PLS, LASSO, Ridge and EN fit (continuous lines) *vs.* reference BGL (open bullets). Bottom panels display two representative channels (#156 an #90 for subject on the left and on the right respectively) entering the models, where occasional spikes and jumps are evident.

OLS identifies model parameters in such a way as to maximize the adherence to the identification data without any constraint on the complexity. As we will see in Section 11.3, this will results in a clear overfitting in the model test phase. Figure 11.2 shows a representative “internal validation” plot for data subsets “part 1” (left subplots) and data subsets “part 2” (right subplots). By visual inspection, it is possible to note how the (calibrated) glucose profiles fitted by the OLS model outperforms the other models. Moreover, Figure 11.2 (right) shows that despite OLS reach better accuracy, it is more sensitive than the other models to occasional jumps or spikes occurring on

	RMSE [mg/dL]	R^2	MAD [mg/dL]	MARD [%]	ESOD	EGA [%] $A+B(A)$ $C\backslash D\backslash E$	CEGA [%] $A_R+B_R(A_R)$ $C_R\backslash D_R\backslash E_R$
OLS	27.3	0.93	23.4	19.9	2.5	97.5(68.1)	84.1(58.3)
	(11)	(0.03)	(11.2)	(13.8)	(2.1)	0\2.5\0	4\8.2\3.7
PLS	44.7	0.85	38.7	30	1.4	94.6(47.7)	88.8(62.7)
(m=15)	(22.1)	(0.1)	(20.9)	(18.1)	(1.4)	0.9\4.4\0.1	5.3\3.5\2.4
LASSO	55.1	0.78	46.6	34.9	0.4	89.9(44.4)	93.6(69.8)
(j=16)	(26.6)	(0.19)	(24.7)	(21.8)	(0.1)	2.6\6.4\1.1	4.1\1.5\0.8
Ridge	42.4	0.86	36.3	28	1.1	95.1(51.3)	88.9(65.4)
($\lambda = 14$)	(21)	(0.11)	(11.4)	(19.7)	(0.8)	0\4.8\0.1	5.2\3.9\2
EN	45.3	0.85	38.8	29.8	0.8	93.9(48.9)	91.6(66.5)
($\alpha = 0.3, \lambda = 0.05$)	(24.4)	(0.14)	(22.8)	(19.3)	(0.4)	0.7\5.1\0.3	4.5\2.5\1.4

Table 11.2: Indicators of model performance for internal validation, i.e. when glucose profiles are estimated from the same data subset "part 2" used for identify the models. In brackets is the complexity model parameter chosen by means of cross-validation.

the Multisensor channels entering the model, as channel # 90 shown in the bottom panel. This characteristic will be more clear when glucose profiles will be estimated from Multisensor data not used during model identification. For sake of completeness, the full "internal validation" plots with all the (22+23 of the two data subsets) recording sessions are shown in Appendix A.

11.3 Model Test

This section presents the model test phase results, when the identified models in the previous section over data subsets "part 1" and "part 2" are tested over data subsets "part 2" and "part 1" respectively.

Indicators reported in Table 11.3 and Table 11.4 show that OLS model is the worst, confirming the occurrence of overfitting previously speculated. This point is further strengthened by visual inspection of the box-plots in Figure 11.4 and Figure 11.7. The OLS model results in indicators more scattered with respect to those of the other models the other models which limit their complexity. Moreover, as can be seen from the CEGA analysis of Figure 11.5 and Figure 11.6, the cloud of points (given by the couples of reference *vs.* estimated BGL) for OLS is the most scattered, with many points lying within the dangerous zones C,D and E.

Regularization based methods, i.e. LASSO, Ridge and EN, seem to outperform PLS. In particular, PLS shows that RMSE, R^2, MAD, MARD and ESOD are worse than for the other models controlling complexity. However, PLS shows EGA and CEGA only

	RMSE [mg/dL]	R^2	MAD [mg/dL]	MARD [%]	ESOD	EGA [%] $A+B(A)$ $C\backslash D\backslash E$	CEGA [%] $A_R+B_R(A_R)$ $C_R\backslash D_R\backslash E_R$
OLS	94	0.69	76.7	59	10.4	85.3(30)	82.6(53)
	(125.2)	(0.26)	(87.2)	(72.2)	(36.5)	8\4\2.7	5\7.9\4.5
PLS	61.2	0.65	51.7	40.1	2.7	90.6(51.3)	85.9(60.5)
(m=10)	(27.5)	(0.25)	(24.7)	(21.2)	(4.4)	0.4\9\0	4.4\6.1\3.6
LASSO	57.9	0.69	48.6	37.8	0.9	89.4(42.2)	89.2(62.1)
(j=15)	(27.1)	(0.25)	(23.7)	(20)	(1.1)	0.9\9.6\0.1	6.3\2.5\2
Ridge	52.3	0.71	44.1	35	2	91(58.7)	88(63)
($\lambda = 5$)	(22.8)	(0.21)	(19.2)	(17.7)	(2.7)	0.1\8.9\0	4.9\4.8\2.3
EN	51.8	0.71	43.9	34.1	2	92.3(59.9)	88.6(65)
($\alpha = 0.4, \lambda = 0.01$)	(24.3)	(0.22)	(20.5)	(17.2)	(2.4)	0.1\7.6\0	4.9\4.4\2.1

Table 11.3: Indicators of model performance when "part 1" of the data set is used for model identification and "part 2" for model test. In brackets is the complexity model parameter chosen by means of cross-validation. RMSE root mean squared error, R^2 Pearson coefficient of determination, MAD mean absolute difference, MARD mean absolute relative difference, ESOD energy of second-order differences, EGA (Clarke) error grid analysis, CEGA continuous error grid analysis.

slightly worse than the other models, indicating that although it can give good prediction of glucose trends it is too sensitive to noisy channels (Figure 11.3 (right)). This happens because the PLS model has all non-zero coefficients, resulting particularly sensitive to occasional jumps or spikes present in the Multisensor channels, as channel # 167 shown in the bottom panel of Figure 11.3. This is also confirmed from the higher ESOD values for PLS in Table 11.3 and Table 11.4 with respect to the other models.

Regularization methods provide, in general, better accuracy performance with respect to PLS. This point is confirmed when the models are tested in both the test data subsets (see Table 11.3 and Table 11.4). In particular, the LASSO model is the one estimating glucose profiles with the lowest ESOD. The reason is two-fold: first, the regularization performed by the ℓ_1 norm prevents the model coefficients from assuming large values thus predicting glucose profiles that are more flat than the other models (see for example Figure 11.8 (right)); second, channels more sensitive to noise that contain also glucose-related information are considered by PLS, and also by Ridge and EN exploiting the effect of the ℓ_2 norm, but are less probable to be selected by LASSO, thus yielding to smoother estimates (see also box-plots in Figure 11.4 and Figure 11.7). Indeed the ℓ_1 norm shrinks many coefficients to zero according to the value of the parameter j controlling complexity. This allows an easier interpretation of the results with a reduced number of original variables, representing the strongest effects, considered important for estimating glucose

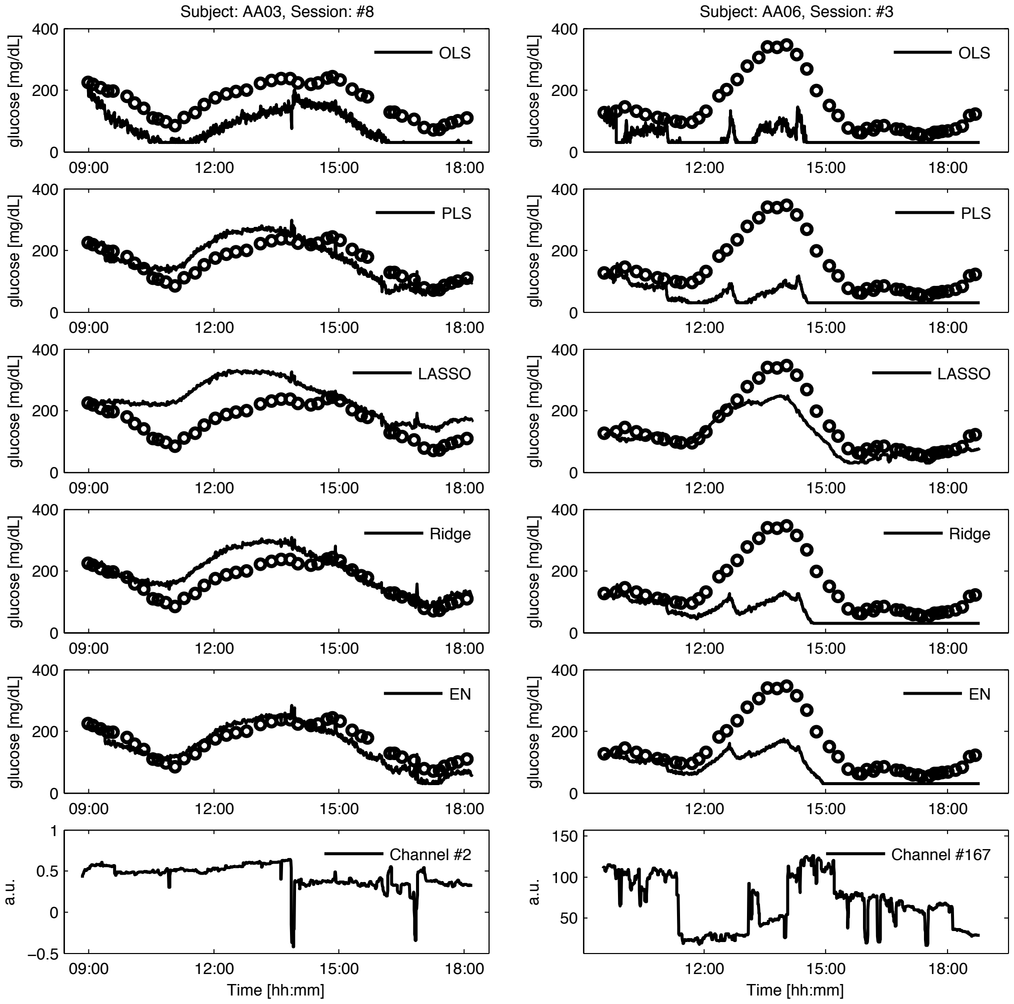
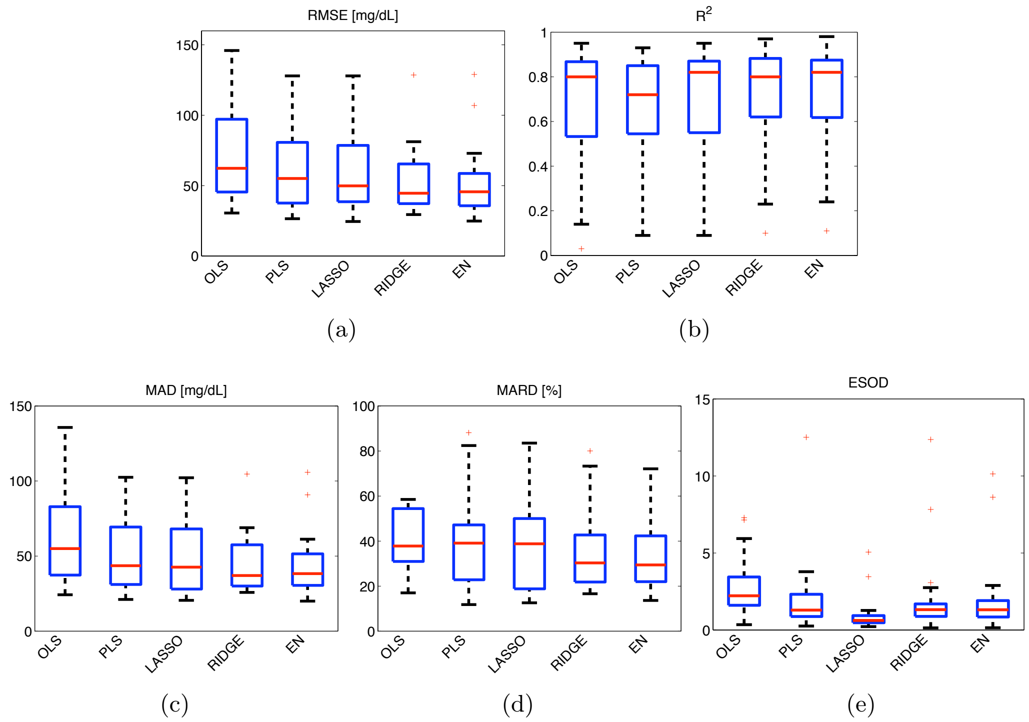


Figure 11.3: Representative recording sessions of Subjects AA03 (left) and AA06 (right). OLS, PLS, LASSO, Ridge and EN model test over independent test data subset (continuous lines) *vs.* reference BGL (open bullets). Bottom panels display two representative channels (#2 an #167 for subject on the left and on the right respectively) entering the models, where occasional spikes and jumps are evident.

profiles. This is a typical feature of the LASSO to act as a variable selection method.

Most of the time, a good agreement between glucose estimated profiles and reference glucose measures is achieved. However, unpredictable events might sometime lead to signals behaviour different from what is expected, yielding un-physiological glucose estimated levels by the model. In these cases, a lower limit of 30 mg/dL for estimated glucose levels is introduced [146]. For instance, Figure 11.3 (right) and Figure 11.8 (left) show two representative recording sessions where the estimated glucose profiles are



(a) (b)

(c) (d) (e)

Figure 11.4: Boxplots for 5 indicators in Table 11.3. RMSE (a), R^2 (b), MAD (c), MARD (d) and ESOD (e).

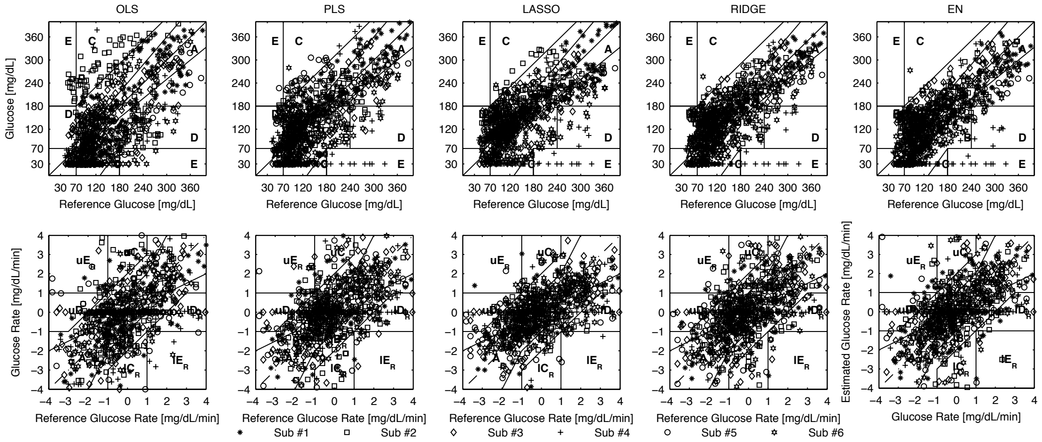


Figure 11.5: Clarke error grid (top) and Rate error grid (bottom) for the different models for test data subset "part 2".

set to the above limit given the presence of a jump affecting some of the Multisensor channels entering the model (see bottom panels of same figures, where the artifacts in

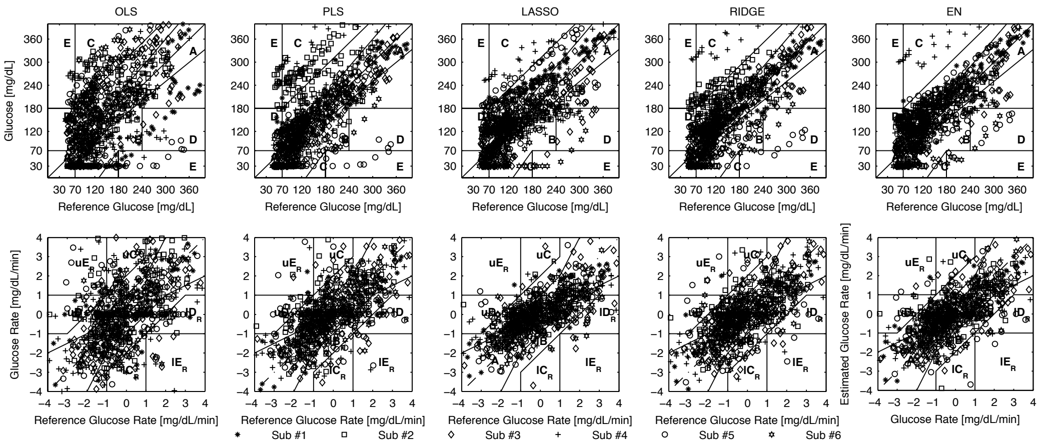


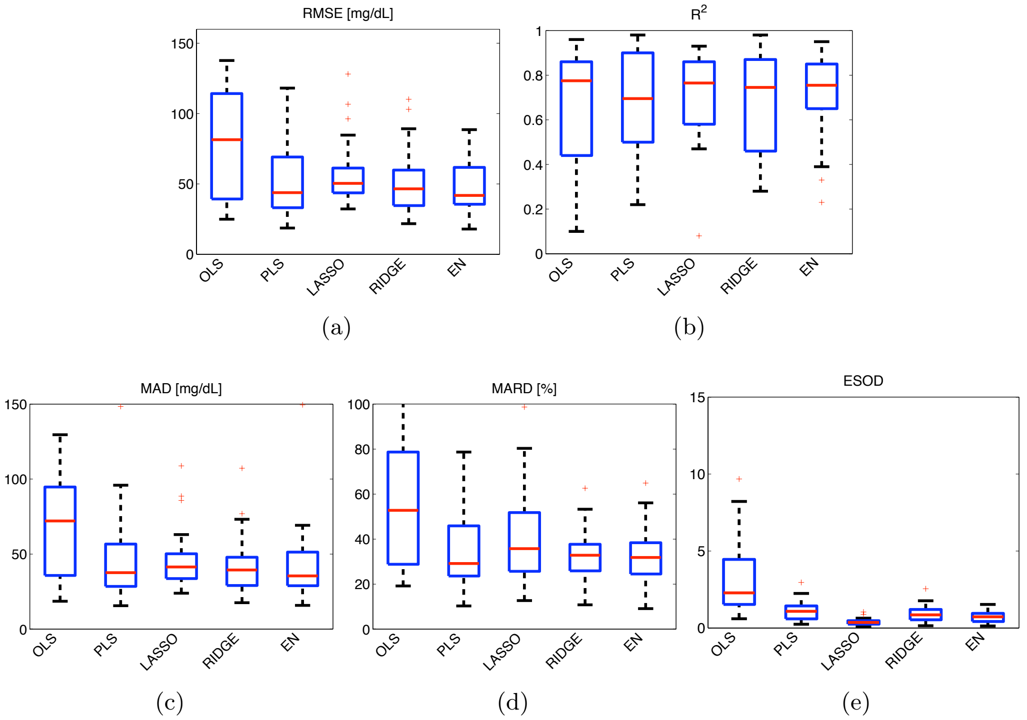
Figure 11.6: Clarke error grid (top) and Rate error grid (bottom) for the different models for test data subset “part 2”.

	RMSE [mg/dL]	R^2	MAD [mg/dL]	MARD [%]	ESOD	EGA [%] $A+B(A)$ $C\backslash D\backslash E$	CEGA [%] $A_R+B_R(A_R)$ $C_R\backslash D_R\backslash E_R$
OLS	80.1	0.65	70.1	57.8	3	73.3(30.2)	79.9(52)
	(38.3)	(0.27)	(35.2)	(33.5)	(2.4)	16.6\6.8\3.3	6.8\8.3\5
PLS	59.8	0.67	51.3	43.4	1.1	84.7(45.3)	88.3(65)
(m=15)	(42.8)	(0.24)	(39.6)	(38.5)	(0.7)	9.2\3.6\2.5	5.8\3.9\2
LASSO	57.5	0.7	47.2	39.4	0.4	86.4(43)	92.5(68)
(j=16)	(25.1)	(0.2)	(21.8)	(20.1)	(0.2)	3.8\8\1.8	5.8\0.9\0.8
Ridge	56.2	0.69	47.1	40.2	0.9	87.3(45.3)	91.1(66.4)
($\lambda = 14$)	(37.5)	(0.22)	(34)	(30.6)	(0.5)	5.8\5.1\1.8	4.6\2.8\1.5
EN	52.6	0.71	44.2	37.5	0.7	88.6(42.6)	91.7(69.4)
($\alpha = 0.3, \lambda = 0.05$)	(31.6)	(0.2)	(27.3)	(23.8)	(0.3)	2.7\7.7\1	5\1.9\1.4

Table 11.4: Indicators of model performance when “part 1” of the data set is used for model identification and “part 2” for model test. In brackets is the complexity model parameter chosen by means of cross-validation.

the Multisensor channels are clearly visible). Interestingly, the LASSO model seems more robust than the other models to these jumps in the data, not requiring the onset of the lower limit cut off, and preserving glucose profile with elevated smoothness and reasonably accurate trend. This behavior can be attributed to the shrinking properties of the ℓ_1 norm. Finally, by looking at the last columns of Table 11.3 and Table 11.4, it is interesting to note how the LASSO model is able to estimate glucose profiles with a better trend accuracy than the other models.

The Ridge model is identified minimizing the RSS cost function subject to a bound on the ℓ_2 norm of the coefficients. This norm does not have the ability of inducing sparseness on the coefficients of the multivariate linear regression model, thus a parsimonious model is not identified and all the predictors are kept in the model. This might cause the estimated glucose profiles by the Ridge model to be sensitive to occasional spikes or jumps in the Multisensor channels, as happened for PLS. However, this influence seems lower than in the PLS model as indicated by the lower ESOD for Ridge and by looking at Figure 11.3 (right). It can be shown that Ridge is related to PLS, since PLS shrinks low variance directions inflating the high variance ones, while Ridge shrinks more the principal components of the predictor matrix $\boldsymbol{X}$ presenting low variance [116]. Estimated glucose profiles by the Ridge model show accuracy indicators slightly better than those of LASSO (see Table 11.3 and Table 11.4). This might indicate that channels discharged by the ℓ_1 norm because sensitive to occasional spikes or jumps actually contain useful glucose related information. Thus, it is reasonable that a combination of the ℓ_1 and ℓ_2 norms could identify a model sharing both properties of sparseness and grouping effect.



(a) (b)

(c) (d) (e)

Figure 11.7: Boxplots for 5 indicators in Table 11.3. RMSE (a), R^2 (b), MAD (c), MARD (d) and ESOD (e).

From Table 11.3 and Table 11.4, one can note that the EN model outperforms the others in terms of accuracy of estimated glucose profiles. In particular, EN is the model

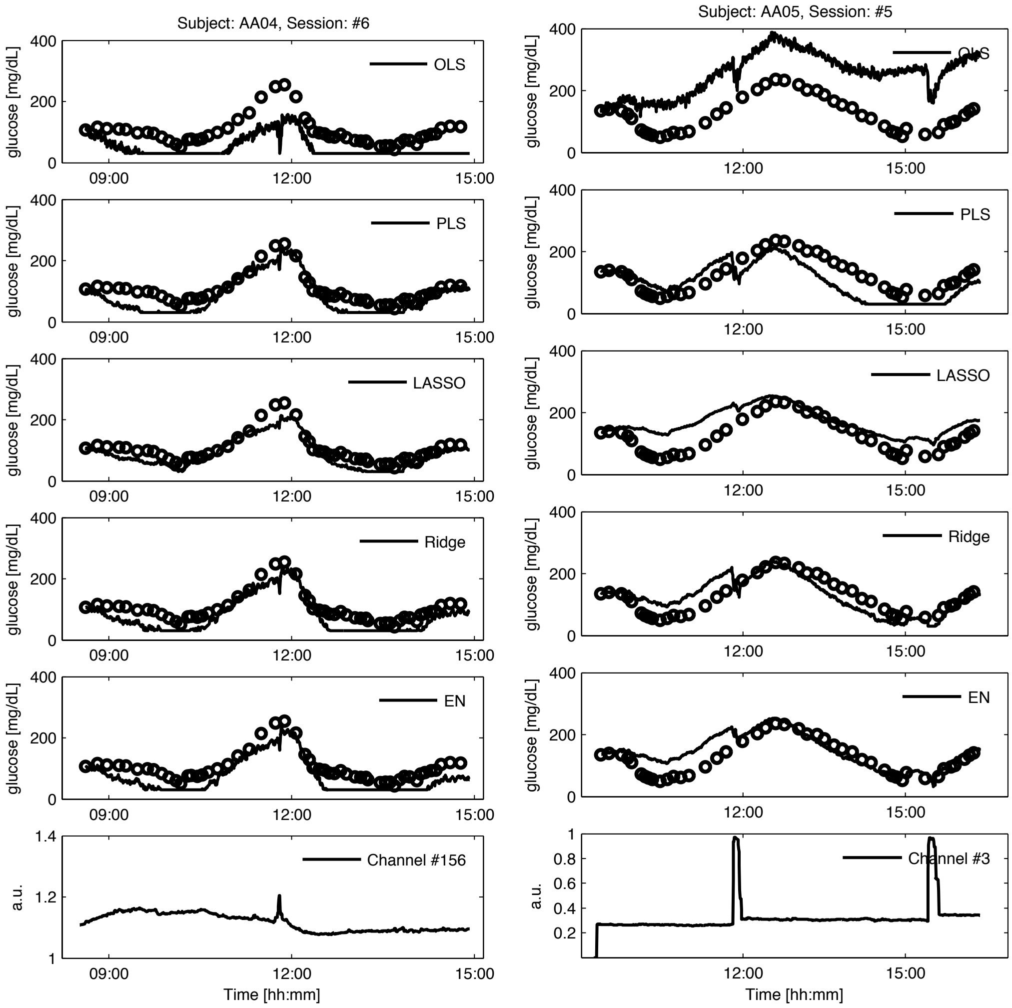


Figure 11.8: Representative recording sessions of Subjects AA04 (left) and AA05 (right). OLS, PLS, LASSO, Ridge and EN model test over independent test data subset (continuous lines) *vs.* reference BGL (open bullets). Bottom panels display two representative channels (#156 an #3 for subject on the left and on the right respectively) entering the models, where occasional spikes and jumps are evident.

presenting the best indicators and is only slightly worse than LASSO in accuracy for glucose trends (see CEGA results). Moreover, its clinical accuracy results on the Clarke Error Grid are substantially close to that of minimally invasive devices that present a percentage of points within the A+B zone spanning from 84.4 to 98.9 [118].

The good results obtained by the EN model are likely due to the combination of the ℓ_1 and ℓ_2 norms, giving to this model both the advantages of LASSO and Ridge. Indeed, a limitation of the LASSO is that if there is a group of correlated variables, then it tends to select only one variable from the group and does not care which one is selected, thus lacking in the ability of revealing grouping information. On the opposite, the ℓ_2 norm allows all coefficients to enter the model, resulting more sensitive to noisy channels. Thus, the ℓ_1 norm shrinks channel weights to zero (eliminating Multisensor channels not useful for predicting glucose) while the ℓ_2 norm encourages a grouping effect (automatically including whole groups into the model once one channel among them is selected). This combination results in indicators outperforming those of the other models (see Figure 11.4 and Figure 11.7) and in estimated glucose profiles with a good trade-off between sparseness of the model coefficients and robustness due to the grouping effect (see for example Figure 11.3 (left)). For sake of completeness, all the model test plots for all the 22+23 available recording sessions are shown in the Appendix B.

11.4 Concluding Remarks

This chapter showed the application of the identification techniques illustrated in Part II of the present thesis to a case study represented by the Solianis Multisensor data with the aim of estimating glucose profiles. We showed that the OLS model outperforms the others in “internal validation” conditions at the cost of overfitting. Indeed, OLS is the worst during model test because the bias of the methods controlling complexity in model identification leads to a better performance when glucose profiles are obtained from an independent test data set. PLS performed better than OLS, but slightly worse than regularization based methods. This is because PLS allows all the Multisensor channels to enter the model, also those affected by occasional jumps or shifts. The same behavior was shown by the Ridge model that allowed all the channels to enter the model. On the opposite, the LASSO model seemed particularly robust to this particular noise, because it shrunk many channel weights to zero [147]. Finally, we showed that EN is the best performing model, representing a good trade-off between Ridge and LASSO. EN is robust to occasional noise occurring in the Multisensor data, sharing the ℓ_1 norm properties with LASSO, but at the same time averages channels with correlated predictors allowing a more accurate estimation of the glucose profiles, exploiting the same ℓ_2 norm properties of Ridge.

12

Conclusions and Further Developments

12.1 Discussion of the Thesis Main Achievements

In diabetes management, tight monitoring of glycaemic levels is important for avoiding long and short term complications related to hypo- and hyper-glycaemia excursions. As reviewed in Chapters 1 and Chapter 2 of the present thesis, many sensors have been proposed for CGM. Most of them have a certain degree of invasiveness because they exploit needle based electrodes. On the other side, non-invasive devices are potentially more appealing, but their development is challenging for several reasons (see Chapter 4). In the last years, a new approach in the development of NI-CGM devices based on the embedding of sensors of different nature within the same device in order to obtain a better bio-physical characterization of the skin and underlying tissues gained particular attention. As seen in Chapter 4, this multisensor concept has been shown to be more robust in the daily-life use of these devices to possible environmental and physiological processes that can deteriorate accuracy of estimated glucose profiles [146, 103].

However, a model linking the measured multisensor data to glucose is needed, together with a set of techniques that can be used to identify the parameters of the multivariate linear regression model, as OLS, PLS, LASSO, Ridge and EN described in Part II (from Chapter 6 to 8), that are tested over the recently proposed Multisensor device by Solianis. The main aim of the thesis was to focus on the problem of identification of suitable regression models for modeling multisensor data with the aim of estimating glucose levels non-invasively (Chapter 11). Results indicate that: as expected, OLS results are superior only in "internal validation" (see Section 11.2), while overfitting clearly appears when models are tested on data previously unseen to the model; the PLS model estimates glucose profiles with reasonable good trends although this model is too sensitive to noisy channels, presenting an higher ESOD value with respect to the other models; the EN model outperforms, in general, the other models thanks to the combination of the ℓ_1 and ℓ_2 norms that allow it to share both the advantages of the LASSO, shrinking many model weights to zero being more robust to possible occasional jumps or spikes occurring on the Multisensor data, and of the Ridge model, averaging the contribution of correlated channels allowing a more robust estimation of glucose profiles.

With respect to the previous literature, this thesis demonstrated that while PLS is the current state-of-art for regression problems involving spectroscopy data (see [148, 149, 105] to mention just a few), EN can become very useful when dealing with regression problem with multisensor data. While retaining information from a group of variables (as PLS does), it also automatically selects those channels representing the strongest effects, giving more insights into the specific problem at hand.

Results obtained in the thesis also demonstrated that, while accuracy indexes defined in Section 5.3.2 are not yet comparable with those of current state of the art, enzyme-based, needle sensors [118], glucose trends estimated by the considered NI-CGM device plus a suitable model exhibit a good accuracy (see CEGA results in the last columns of Table 11.3 and Table 11.4). This result is important in the treatment of diabetes since the glucose trend can be a valid additional information to complement standard SMBG devices that measure glucose by fingerprick. Knowing the glucose trend in real time can greatly help the diabetic patient in preventing the occurrence of critical events, such as hypoglycaemia. To better illustrate this point, consider the example in Figure 12.1. Top panel shows a portion of data (open bullets are SMBG samples, continuous line is the glucose concentration estimated by the EN model in a representative subject (20090806_S4WP4_AA04 in Appendix -see Appendix for label's meaning-). Bottom panel shows the estimate of the glucose concentration time-derivative, computable, also in real time, through regularization algorithms (see [150] for details) starting from the glucose profile returned by the EN model. By using the static risk (SR) concept introduced in [151], the SMBG measures can be mapped into a symmetric risk space ranging from 0 (low risk) to 100 (high risk of hypo/hyperglycaemia, respectively). If only SMBG samples were available, at time 15:00 and subsequent values (labelled as A and B in the picture) similar SR values, equal to -16.8 and -18.4 respectively, would be estimated.

Following the ideas presented in [150], a reliable glucose trend estimation can be used to integrate SMBG information for calculating the dynamic risk (DR) in situations A and B. DR values in A and B are equal to -39 and -0.4, respectively, and allow the patient to interpret differently the situation of a glucose level near the hypoglycaemic threshold of 70 mg/dL with a negative (point A) rather than a positive (point B) trend: In situation A, an alert can be generated to solicit the patient to take sugar to mitigate, or even prevent, the hypoglycaemic event.

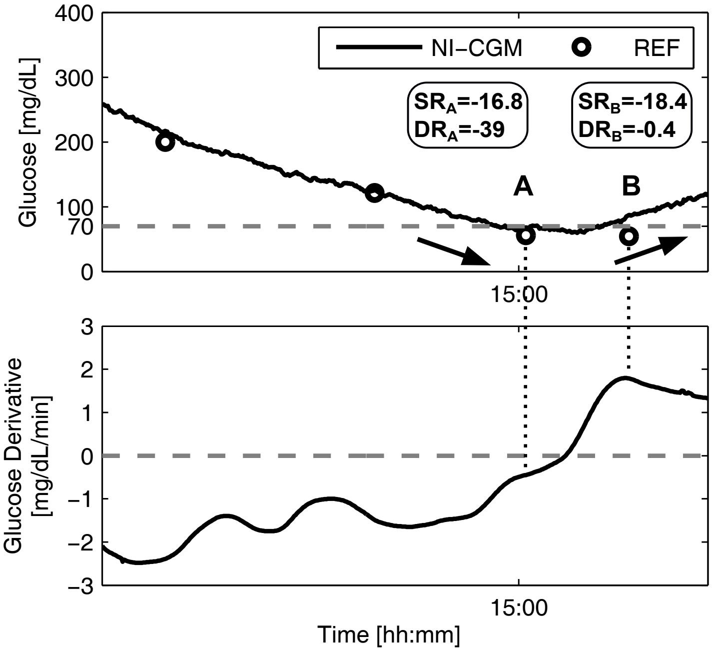


Figure 12.1: Application of dynamic risk concept exploiting NI-CGM data in diabetes management. Example of sparse SMBG values (A, B) (Top panel) complemented by NI-CGM trend information (Bottom panel).

Thus, the NI-CGM multisensor system (Solianis device plus the EN model) can not be considered yet a replacement of current needle-base glucose sensors. However, the accuracy in estimating glucose trends makes the system suitable to be used in the current diabetes therapy as a complement to standard SMBG devices. Promising results obtained with the EN model makes the system even more appealing given the incremental accuracy performance achieved.

12.2 Future Developments: Monte Carlo MC Methodology to Assess Robustness of Multisensor Models

As far as possible future developments of the present thesis is concerned, we briefly discuss a methodology for testing the robustness of the calibration parameter (see Section 10.2.3) against environmental and physiological processes that can occur during daily-life. The methodology is general and can be used also for multisensor devices for NI-CGM other than the Solianis Multisensor considered in this thesis.

12.2.1 Case Study: Effects of Sweat Events on Model Calibration

The parameter b in eq. (10.1) discussed in Section 10.2.3 is estimated by the calibration procedure of eq. (10.2) at the beginning of each experimental session and is not updated for the entire duration of an experiment, i.e. whilst the Multisensor device remains in contact with the skin. While this does not necessarily introduce issues in very controlled, i.e. hospital, conditions, in real life, uncontrollable events may occasionally disturb the Multisensor monitoring. In particular, a sweat event involves the creation of a conductive saline layer at the sensor-skin interface. As long as the sweat activity diminishes, the signal is expected to return to a level close to its initial value. However, as shown in Figure 12.2 (top), there still could be a large off-set in the signals measuring sweats (interdigitated electrode in the frequency range 1-200 KHz, from now on identified as channel #36, black line) that after the occurrence of sweat does not always return to its value before the sweat event, a condition already observed in the literature [152]. This off-set, together with changes in the hydratation levels of the skin and underlying tissues resulting from sweat, could also affect the DS electrodes measuring the main glucose related signals (see Figure 12.2 (top), channel #115, grey line) despite the fact that these electrodes are designed to sample the most microvascularized area (i.e. the upper and deep vascular plexus). If effects of sweat events impaired the calibration parameter calculated at the beginning of each experimental session, glucose levels after the occurrence of sweats would be estimated with less accuracy.

It is useful to assess potential benefits obtained by recalculating b in eq. (10.2) exploiting the first reference BGL samples collected after the occurrence of sweat events. To perform such a study, the first problem is to identify a sweat event using the Multisensor data that appear more sensitive to sweat. As shown in Figure 12.2, calculating the derivative (middle panel) of channel #36 (black line in top panel), measured by the interdigitated electrode with specific geometrical shape and at specific frequency for being sensitive to sweat, provides a rough but effective procedure for the on-line detection of sweat events by setting a proper threshold TH (shown in grey in middle panel). Here the threshold is chosen, in a pool of candidate values, as the one giving the better trade-off between missed and identified sweat events. After a sweat event is detected, a new calculation of the calibration parameter is performed according to eq. (10.2): the new b is calculated at the time instant t_i of the first available reference BGL after the detection of the sweat event.

The multivariate linear regression model used by the Multisensor is expected to properly combine the information contained in the Multisensor channels to compensate non-glucose related physiological processes such as sweat events. However, the compensation of sweat effects on the main glucose related signals that is expected to occur on the Multisensor channels # 36 (which contains information about the electrolyte balance changes on the skin surface) is principally performed by channels exploiting frequencies in the GHz range, that measure water balance variations in the tissue because sweating also results in changes in hydratation. Assuming that the model is not able to properly compensate these sweat related processes, a new calibration point would be needed for re-adjusting the glucose baseline every time a sweat event is occurring. This need results in the collection of a new reference BGL sample obtained by blood fingerprick, reducing, in a practical perspective, the usefulness of NI-CGM.

12.2.2 Assessment of Model Calibration Robustness by Monte Carlo Methodology

Generally speaking, a MC simulation is a stochastic technique widely used to explore the distribution of a target outcome when its direct calculation from available inputs is not feasible. More specifically, when performing a MC simulation, first a pool of N repeated (and randomly sampled) input vectors from their domain or distribution, usually with $N \geq 100$, is generated. Then, for each input vector, the computation of the outcome of the system under analysis is deterministically calculated (each of the N iteration is called simulation). Finally, the distribution of the target outcome is derived aggregating the result of each simulation. In our specific case, the domain over which the inputs are sampled corresponds to the set of time instants where reference BGL values for calibration are available, while the deterministic computation refers to the specific calibration procedure adopted or under test. The number of iterations considered is N=1000. At each iteration of the MC simulation, each glucose profile estimated by the multivariate model in the test data set undergoes the initial calibration (as explained in Section 10.2.3), which is fixed and does not change from simulation to simulation.

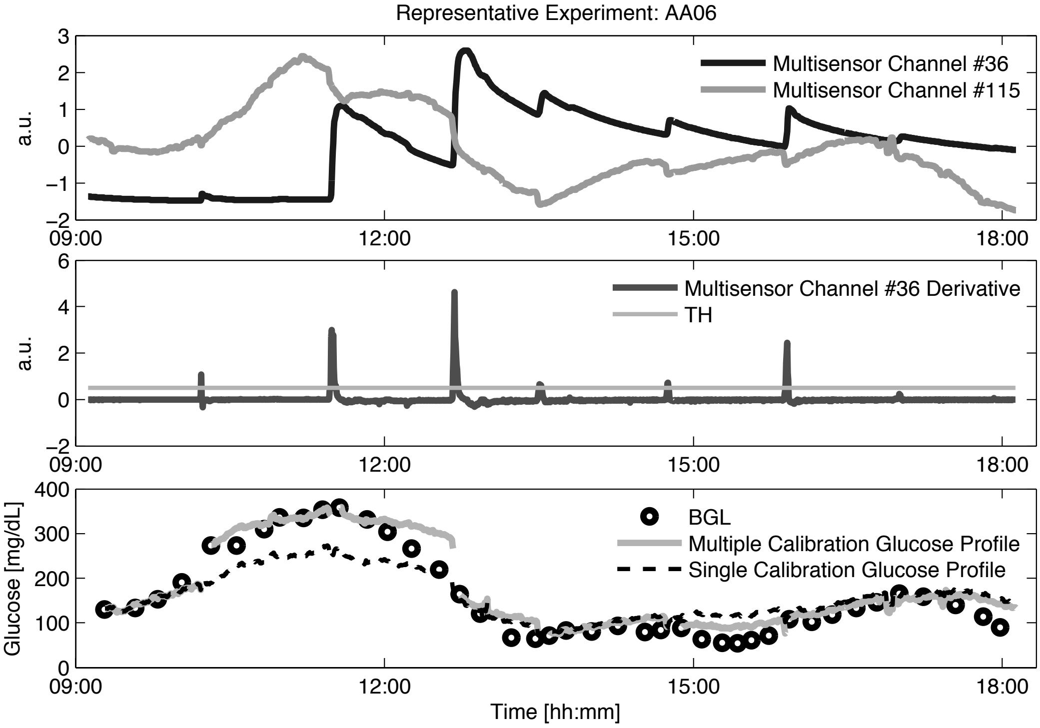


Figure 12.2: Representative experimental session recalibrated after sweat events. Top: Two of the 150 Multisensor channels recorded: channel sensitive to sweat events, i.e. channel #36, (black line) and channel particularly sensitive to glucose changes, i.e. channel #115 (grey line). Middle: derivative of the channel 36 signal (black line) with the chosen threshold TH (thin grey line). Bottom: Glucose profiles estimated by using single baseline calibration (black dashed line) and multiple calibrations (grey line). Reference BGL samples collected in parallel are also shown to allow qualitative visual assessment of accuracy (black circles).

Then, the calibration parameter b is recalculated, according to (10.2), one or several times over a grid of random time instants. Note that the number of recalculations of the parameter b performed at each simulation is fixed and depends on the number of events that characterizes the scenario under analysis. In the sweat events scenario, b will be recalculated N_s times in random time instants within the experimental session, where N_s is the average number of sweat events occurring in the test data experimental sessions. At the end of each MC iteration, accuracy of glucose profiles is measured through a subset of indicative indexes RMSE, MAD and MARD measuring point accuracy. Finally, after all N MC iterations are performed, the sample distribution of the above indexes is obtained, and compared with the result obtained with the specific calibration procedure under evaluation.

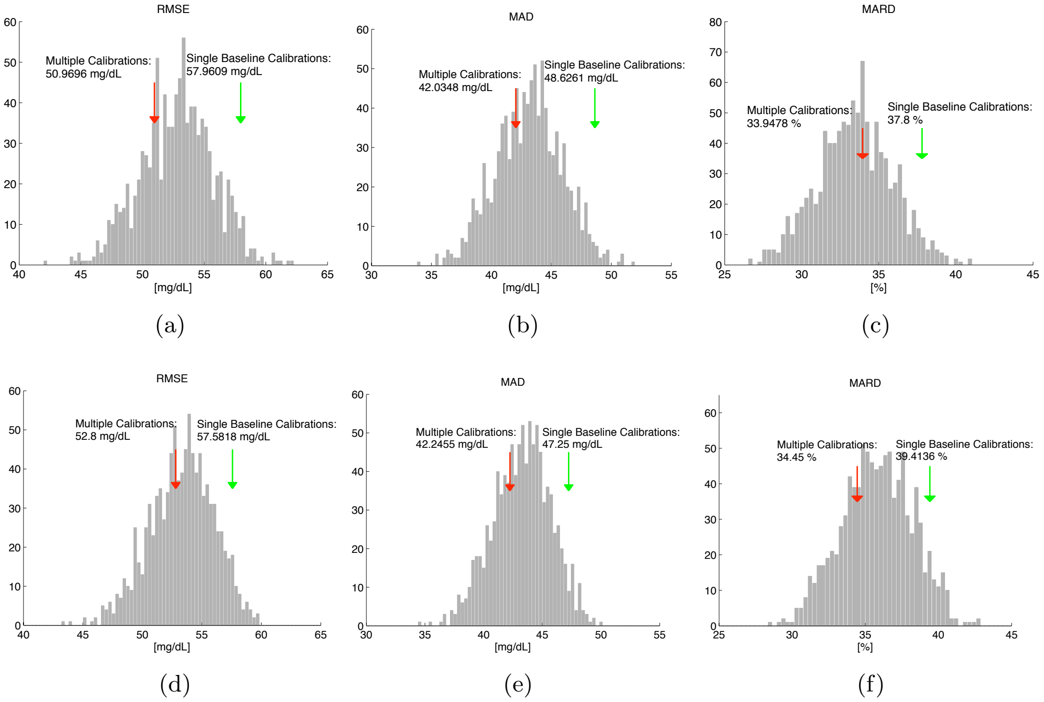
12.2.3 Robustness of Model Calibration to Sweat Events: Results

Table 12.1 shows average and standard deviation (in parentheses) of RMSE, MAD and MARD obtained for the standard working case, i.e. the calibration parameter b is calculated only once, as baseline value, at the beginning of the experiment (first line in Table 12.1), and for the multiple calibration strategy under assessment, i.e. b is updated using the first reference BGL available every time a sweat event is detected (second line in Table 12.1). These preliminary results are obtained with the LASSO model, given its earlier use for NI-CGM [147]. Both the test datasets are documented, i.e. test data subset "part 2" when data subset "part 1" is used for model identification (1 → 2) and test data subset "part 1" when data subset "part 2" is used for model identification (2 → 1). Statistical significance of the differences (computed according to the Students t-test) is also indicated by the p values. Though there is not a statistically significance difference for all the considered key indicators in both the test sets, the multiple calibration strategy for compensating sweat events seems to result in a reduction of the variability of the indicators. To assess if this improvement could be related to the higher number of reference BGL data points used rather than to a real benefit deriving from recalibrating exactly after sweat events, the MC simulation described in the previous subsection is performed.

	RMSE [mg/dL]		MAD [mg/dL]		MARD [%]	
	$1 \to 2$ p =0.07	$2 \to 1$ p =0.7	$1 \to 2$ p =0.06	$2 \to 1$ p =0.6	$1 \to 2$ p =0.09	$2 \to 1$ p =0.7
Single Baseline Calibration	57.9 (27.1)	57.5 (25.1)	48.6 (23.7)	47.2 (21.8)	37.8 (20)	39.4 (20.1)
Multiple Baseline Calibrations	50.9 (20.8)	52.8 (19.5)	42 (19.8)	42.2 (15.5)	33.9 (18.8)	34.4 (10.9)

Table 12.1: Key indicators results for the single and multiple glucose calibration. Average and standard deviation (in parenthesis) -over experimental sessions- of RMSE, MAD, MARD obtained when database "part 1" and database "part 2" are used for model identification and model test, respectively, (1 → 2), or *viceversa* (2 → 1). Single Baseline Calibration: parameter b in eq. (10.2) is calculated only at the beginning of the experimental session; Multiple Calibrations: b in eq. (10.2) is updated everytime a sweat event is detected. The p value indicates the statistical difference between the two calibration strategies according to the Student t-test.

For each of the 1000 MC simulations, the mean accuracy of the random multiple calibrated glucose profiles was evaluated by the same key indicators used above. Then, the distributions of the key indicators on the 1000 repetitions were compared with the mean values results in Table 12.1 and showed in Figure 12.3 for RMSE, MAD and MARD, respectively, only for one test data subsets (comparable results are obtained switching identification and test data sets see 2 → 1 in Table 12.1). In Figure 12.3, the distribution of mean values of the key indexes calculated on the 1000 MC simulations is depicted with grey bars, while mean value obtained recalculating the calibration parameter after each sweat event is showed with a red arrow. Interestingly, the peaks of the distributions for the three indicators are exactly comparable with the results obtained with the proposed recalibration strategy. In addition, from the bottom panel of Figure 12.3 we can note that a significant portion of the MC simulations produce a mean value lower than the one represented by the red arrow (39%, 31% and 27.6% for RMSE, MAD and MARD, respectively). Thus, the results of the MC simulation suggest that the improvements (with respect to the single baseline calibration scenario) in terms of accuracy noticed in Table 12.1 are due to the increased number of reference BGL points used for calibration rather than to performing recalibration exactly after a sweat event to compensate for changes in the baseline of the main glucose signals induced by the event.



(a) (b) (c)

(d) (e) (f)

Figure 12.3: Histogram of RMSE, MAD and MARD obtained in the Monte Carlo simulation when data subset "part 2" (top) and "part 1" (bottom) are used for model test respectively. Green arrows report the value (also presents in Table 12.1) of the key indicator considered for single baseline calibration, while the red arrow for the multiple baseline calibration.

The MC methodology showed that re-calculating the glucose baseline after the occurrence of sweat events is not necessary because the multisensor system (device plus model) is able to compensate for this particular detrimental effect. This is particularly useful in the therapy of diabetes and appealing for the ever-day use of the device because a patient do not need to collect a SMBG measure every time a sweat occurs.

12.2.4 Other Possible uses of the MC Simulation Strategy

As we saw in this section, the MC methodology can be a valid tool for assessing the robustness of model calibration by judging whether the improvement due to a proposed calibration scheme is really useful or rather due to the increased quantity of information considered (in previous case more reference BGL used for calibration). Within the same framework, other possible uses of the proposed MC methodology is to assess the validity of new strategies for calibration. For example, calibration scheduling are widely used also by minimally-invasive devices for improving accuracy of estimated glucose profiles by re-calculating the calibration parameters according to a temporal scheduling [153]. Calibration scheduling is also exploited by NI-CGM devices, such as for example by Harman-Boehm et al. [103].

12.3 Future Developments: Other Possible Fields of Investigations

Identification techniques considered in Part II minimize a cost function where the error term measuring the adherence to the data is given as the sum of the distances between the target (reference BGL) and model output. However, this cost function does not take into consideration that errors in glucose estimates do not always have the same clinical implications, as also depicted from the CGA and CEGA in Chapter 11. For example, in [154] a new glucose specific metric is introduced that modifies the MSE as defined in eq. (5.3) of Chapter 5 with a Clark error grid inspired penalty function, which penalizes overestimation in hypoglycemia and underestimation in hyperglycemia, i.e., the most harmful conditions on a clinical perspective. This new cost function is formally given by:

$$gMSE(y, \hat{y}) = MSE(y, \hat{y})Pen(y, \hat{y}) \tag{12.1}$$

where y and $\hat{y}$ represent the reference BGL data and the estimated glucose by the model respectively, while $MSE(\cdot, \cdot)$ is the euclidean distance and $Pen(\cdot, \cdot)$ is the Clarke inspired loss function. For instance, this new cost function, which is graphically depicted in Figure **??**, can replace the RSS used to identify, for example, the regularization based methods,

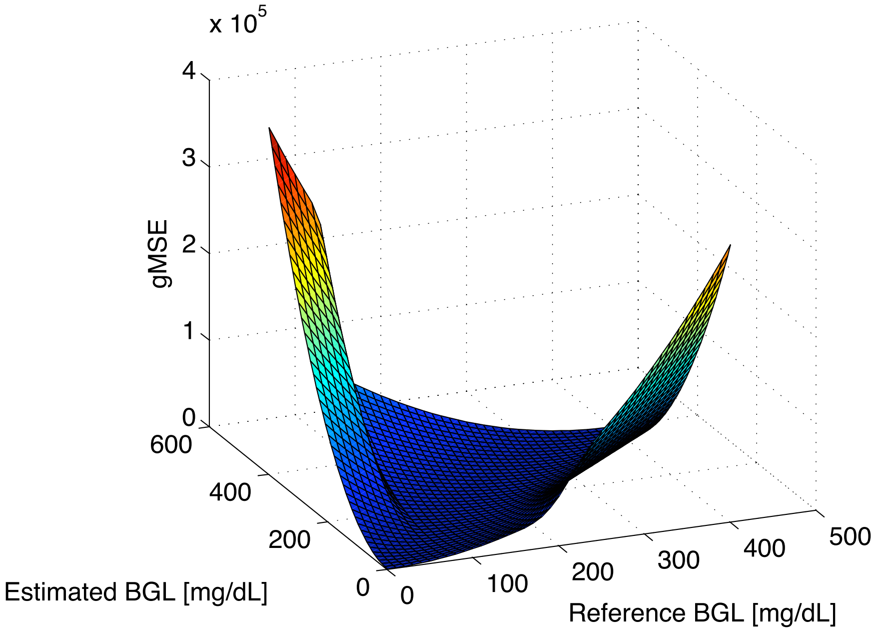


Figure 12.4: Clarke error grid inspired cost function gMSE.

i.e. LASSO, Ridge and EN.

Future investigations may also be focused on the application of the methodologies presented in this thesis to a wider data set possibly obtained in real-life situations, where environmental conditions are not controllable as those of in-clinic studies. This could be object of investigation for Biovotion AG (Zurich, Switzerland), the company that recently acquired IP and technology of the Multisensor data used in this thesis.

A

Full Model Identification Glucose Profiles

This appendix collects the full model identification plots when data subset “part 1” and “part 2” are used to identify the different models.

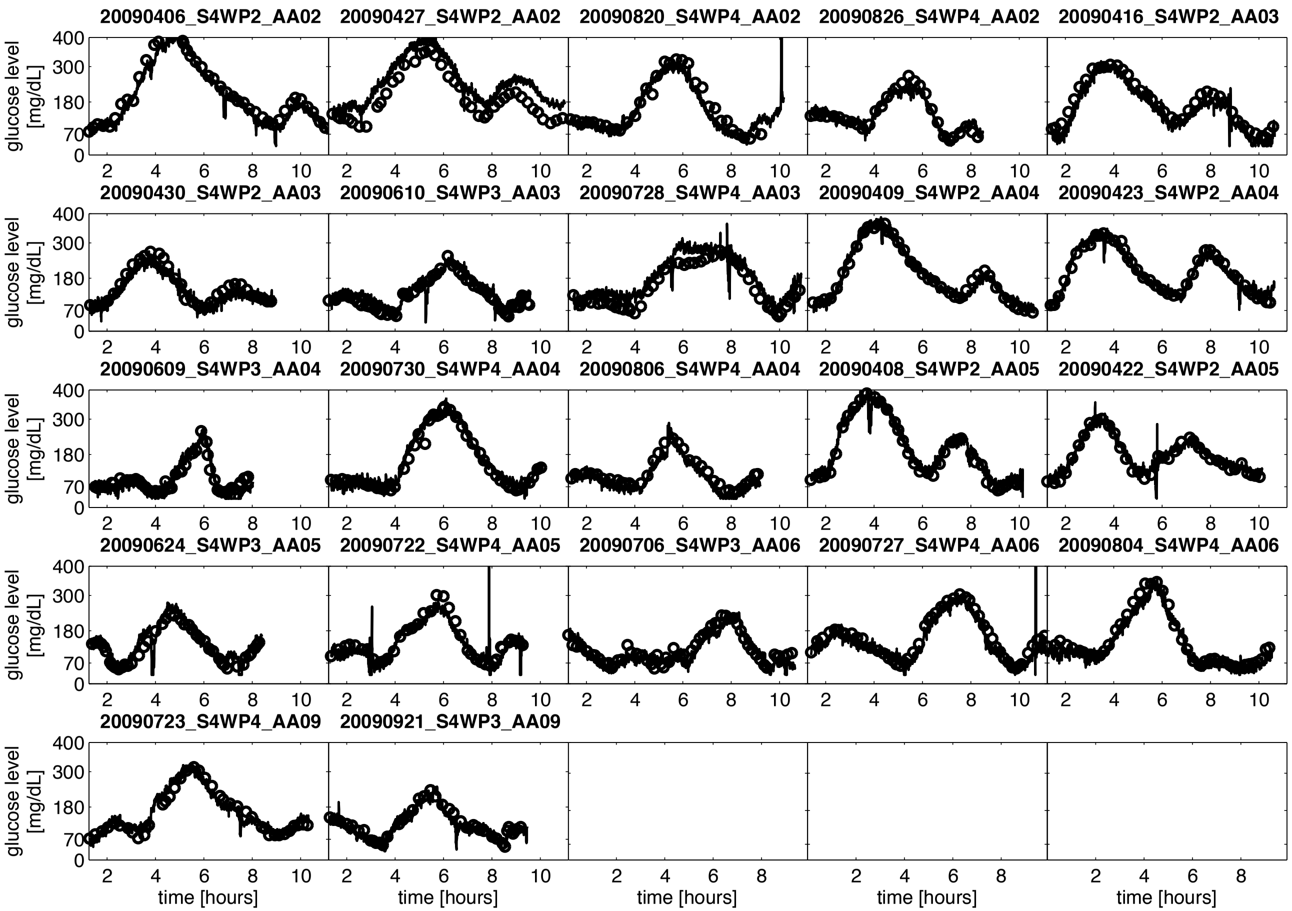


Figure A.1: Estimated glucose profiles by OLS (continuous black line) against reference BGL values (black circles) when the same multi-sensor data used for model identification, i.e. data subset "part 1", is considered ("internal validation"). The first part of the recording sessions' labels indicates the data acquisition day, the second part is an internal notation, and the third part states subject's id number.

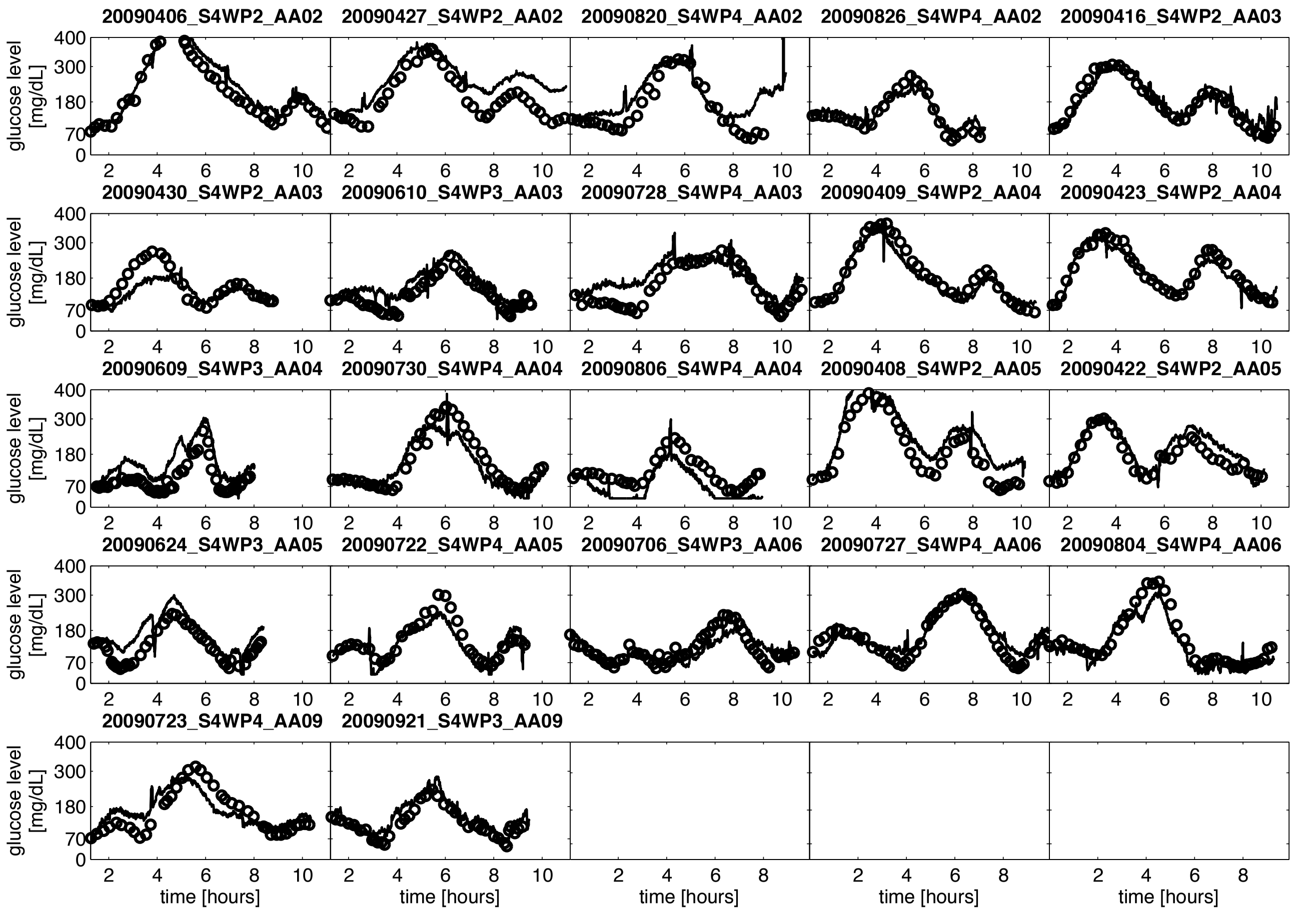


Figure A.2: Estimated glucose profiles by PLS (continuous black line) against reference BGL values (black circles) when the same multi-sensor data used for model identification, i.e. data subset "part 1", is considered ("internal validation"). The first part of the recording sessions' labels indicates the data acquisition day, the second part is an internal notation, and the third part states subject's id number.

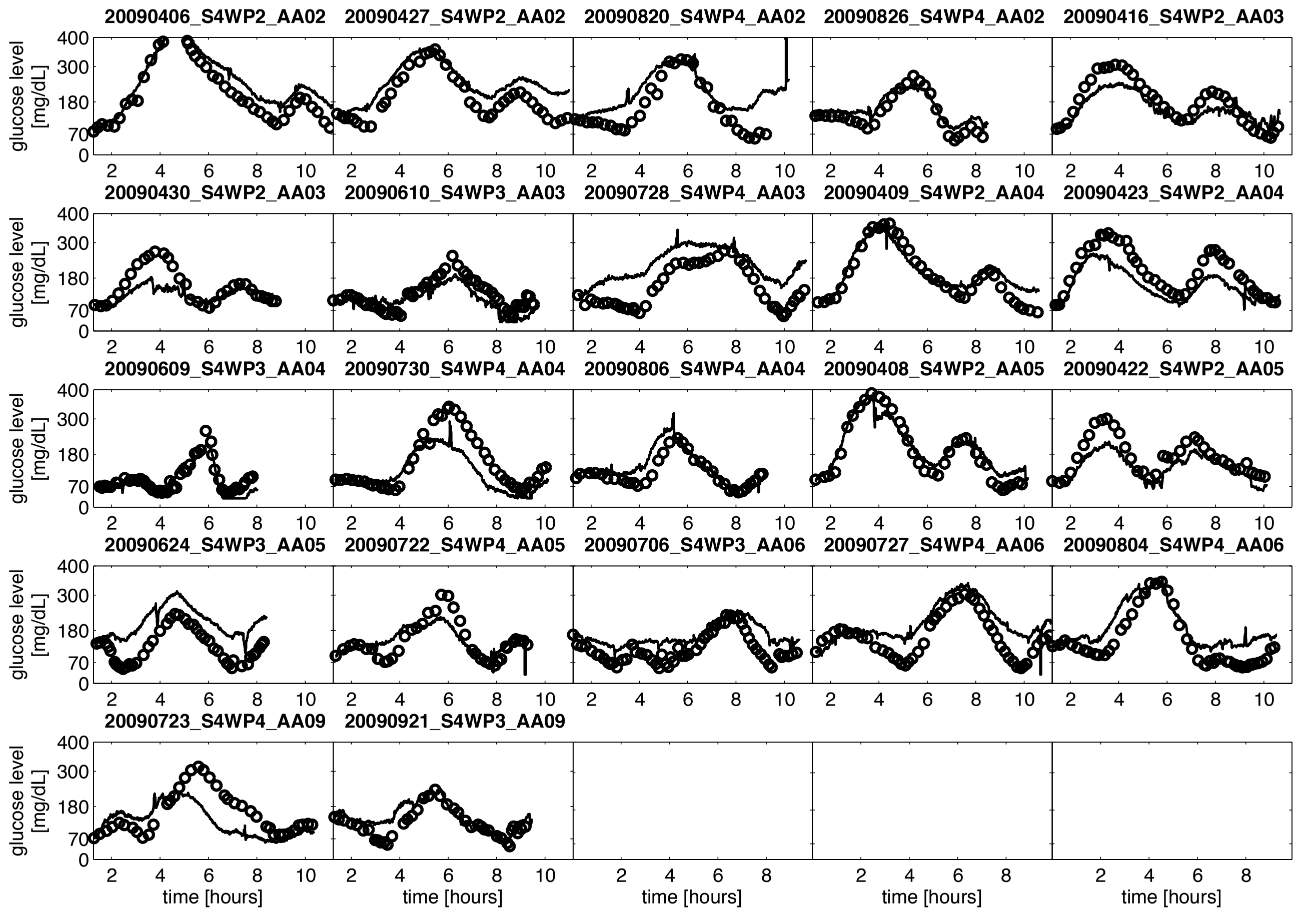


Figure A.3: Estimated glucose profiles by LASSO (continuous black line) against reference BGL values (black circles) when the same multi-sensor data used for model identification, i.e. data subset “part 1”, is considered (“internal validation”). The first part of the recording sessions’ labels indicates the data acquisition day, the second part is an internal notation, and the third part states subject’s id number.

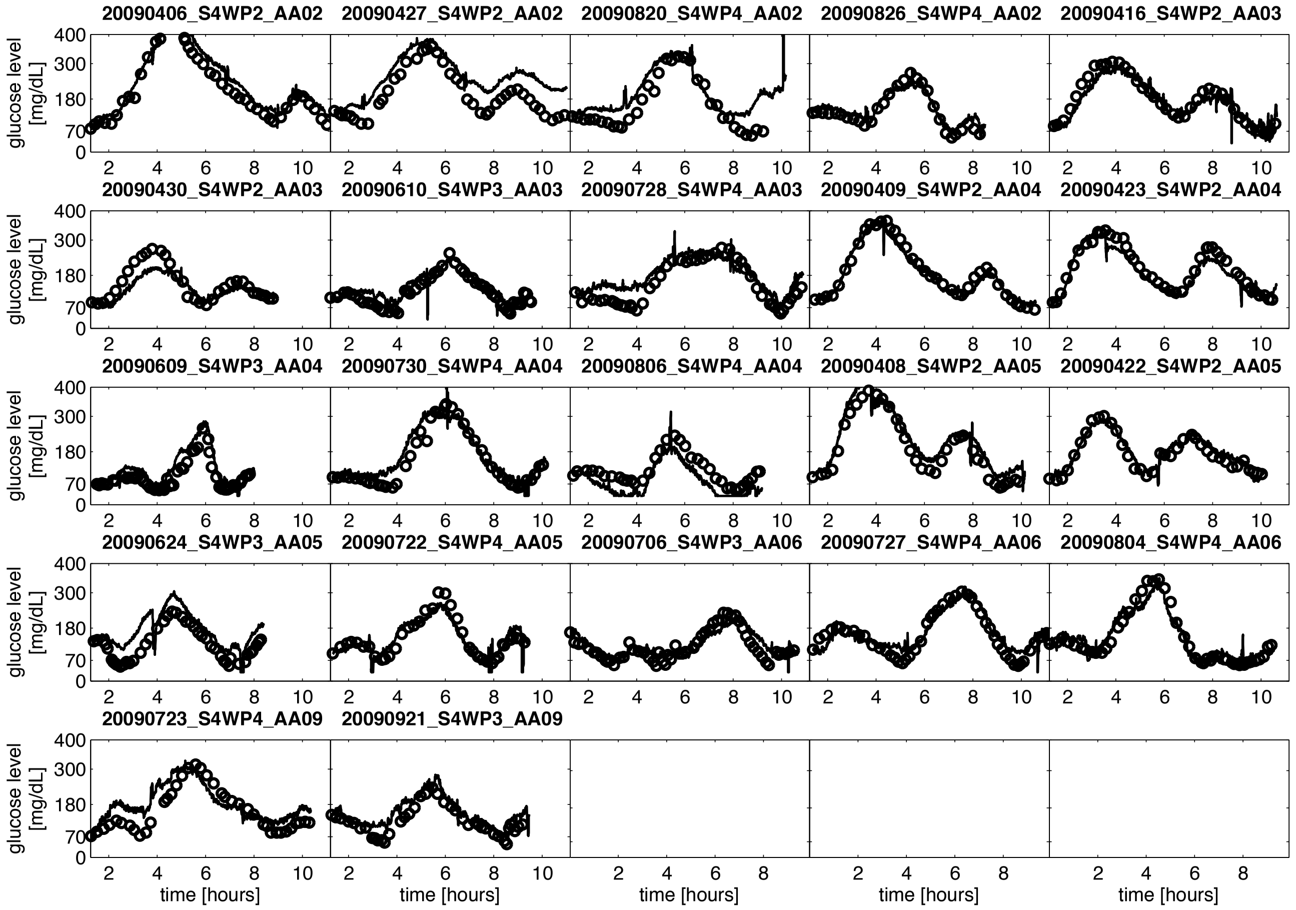


Figure A.4: Estimated glucose profiles by Ridge (continuous black line) against reference BGL values (black circles) when the same multi-sensor data used for model identification, i.e. data subset "part 1", is considered ("internal validation"). The first part of the recording sessions' labels indicates the data acquisition day, the second part is an internal notation, and the third part states subject's id number.

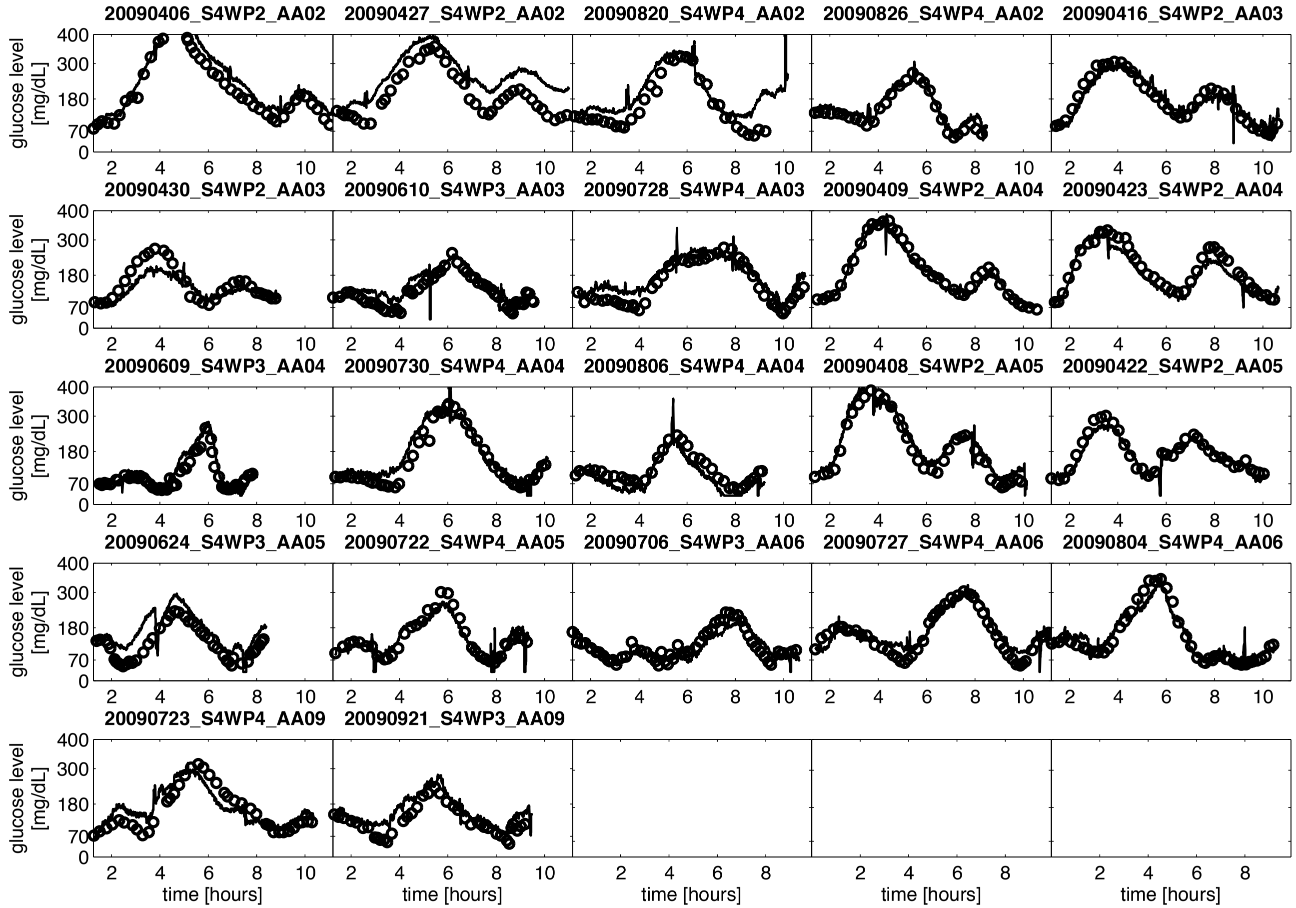


Figure A.5: Estimated glucose profiles by EN (continuous black line) against reference BGL values (black circles) when the same multi-sensor data used for model identification, i.e. data subset “part 1”, is considered (“internal validation”). The first part of the recording sessions’ labels indicates the data acquisition day, the second part is an internal notation, and the third part states subject’s id number.

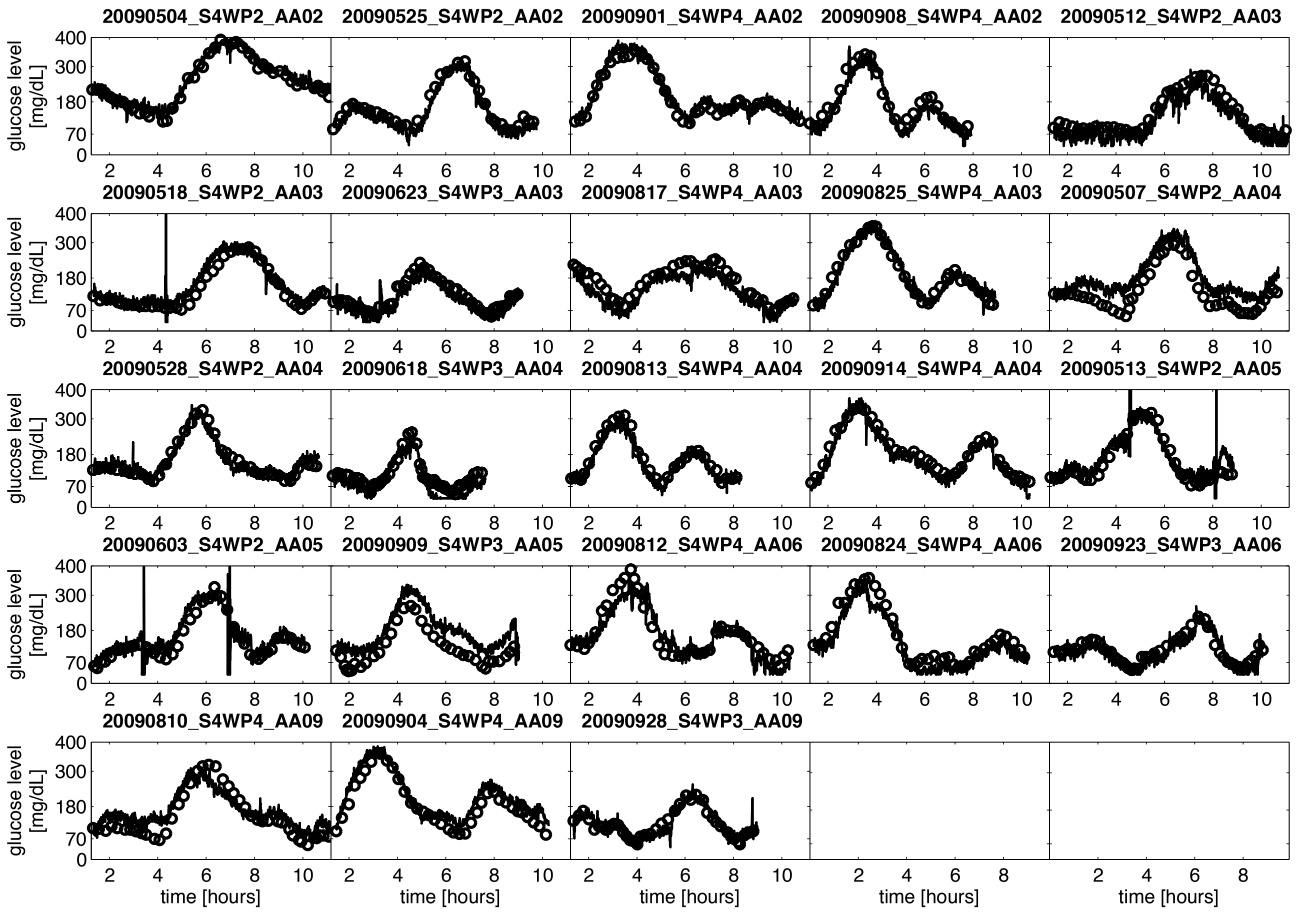


Figure A.6: Estimated glucose profiles by OLS (continuous black line) against reference BGL values (black circles) when the same multi-sensor data used for model identification, i.e. data subset "part 2", is considered ("internal validation"). The first part of the recording sessions' labels indicates the data acquisition day, the second part is an internal notation, and the third part states subject's id number.

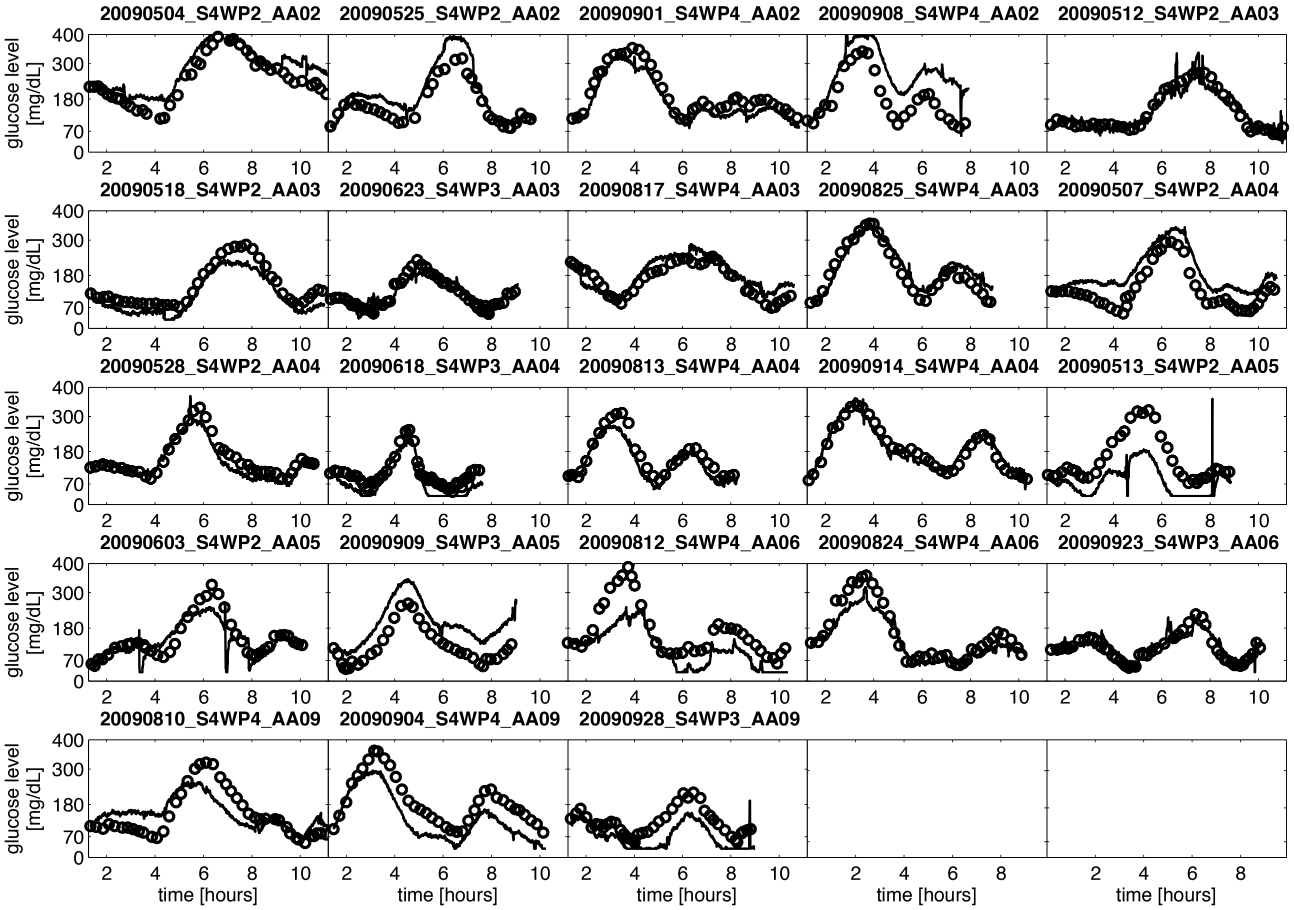


Figure A.7: Estimated glucose profiles by PLS (continuous black line) against reference BGL values (black circles) when the same multi-sensor data used for model identification, i.e. data subset “part 2”, is considered (“internal validation”). The first part of the recording sessions’ labels indicates the data acquisition day, the second part is an internal notation, and the third part states subject’s id number.

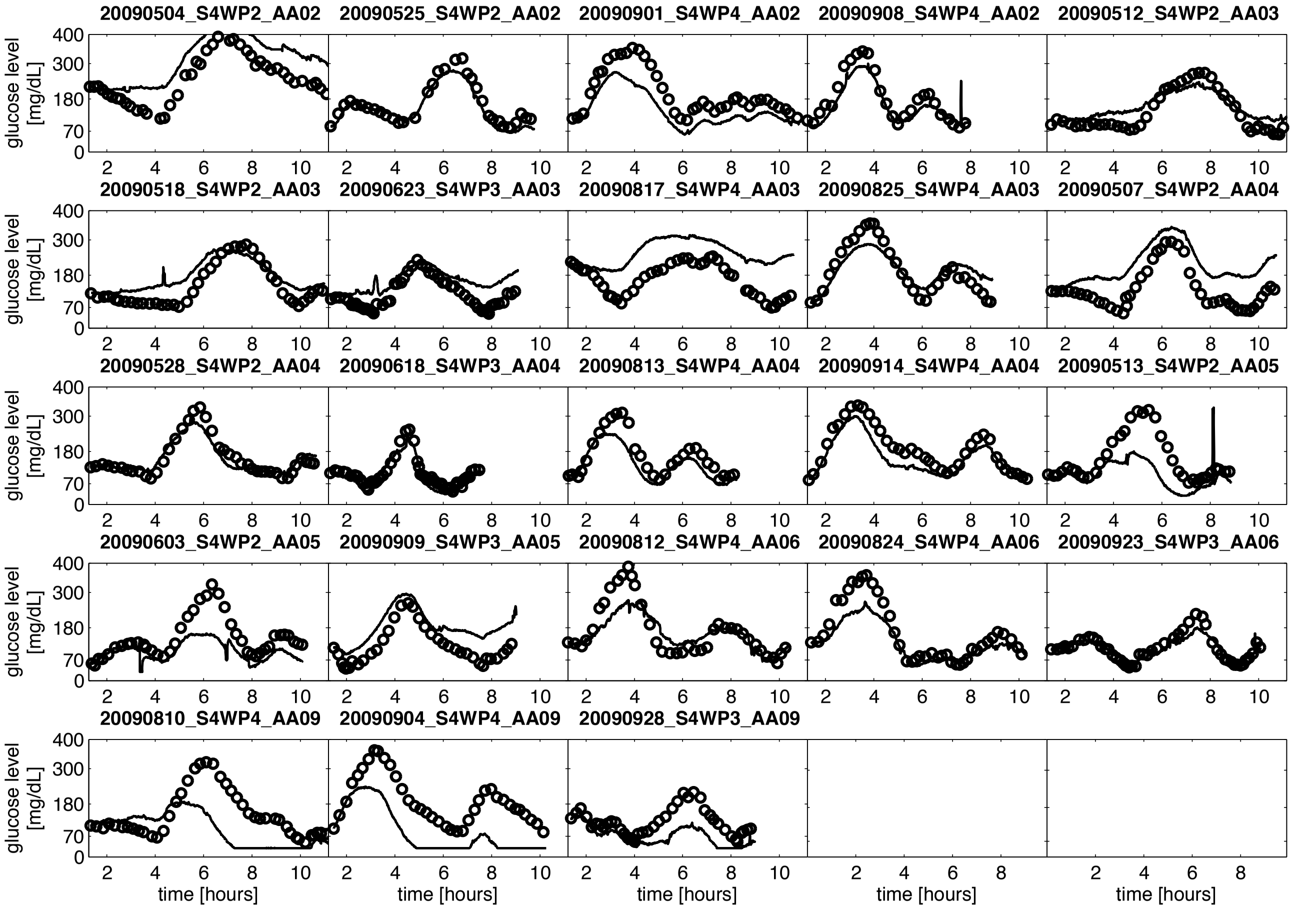


Figure A.8: Estimated glucose profiles by LASSO (continuous black line) against reference BGL values (black circles) when the same multi-sensor data used for model identification, i.e. data subset “part 2”, is considered (“internal validation”). The first part of the recording sessions’ labels indicates the data acquisition day, the second part is an internal notation, and the third part states subject’s id number.

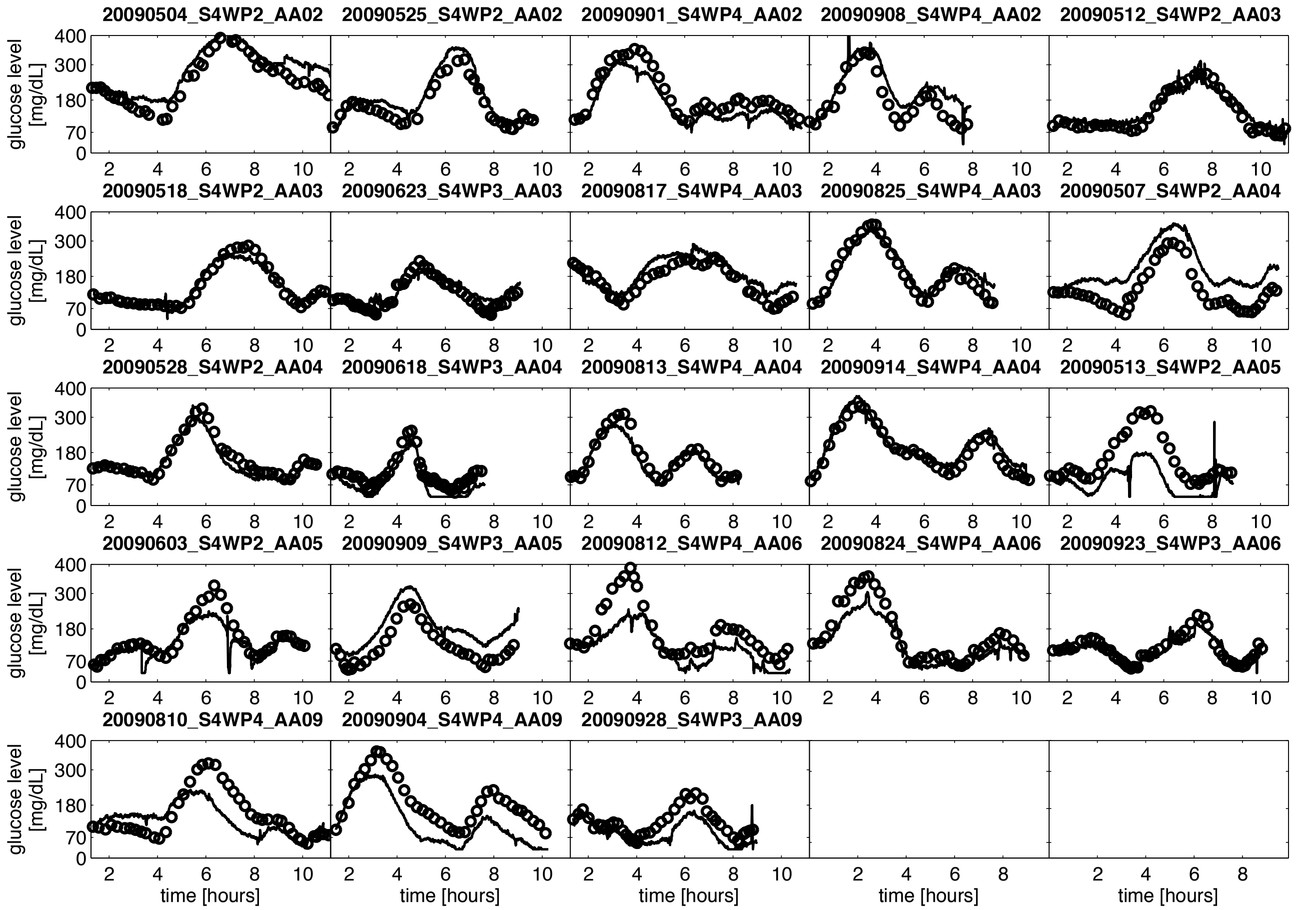


Figure A.9: Estimated glucose profiles by Ridge against reference BGL values (black circles) when the same multi-sensor data used for model identification, i.e. data subset "part 2", is considered ("internal validation"). The first part of the recording sessions' labels indicates the data acquisition day, the second part is an internal notation, and the third part states subject's id number.

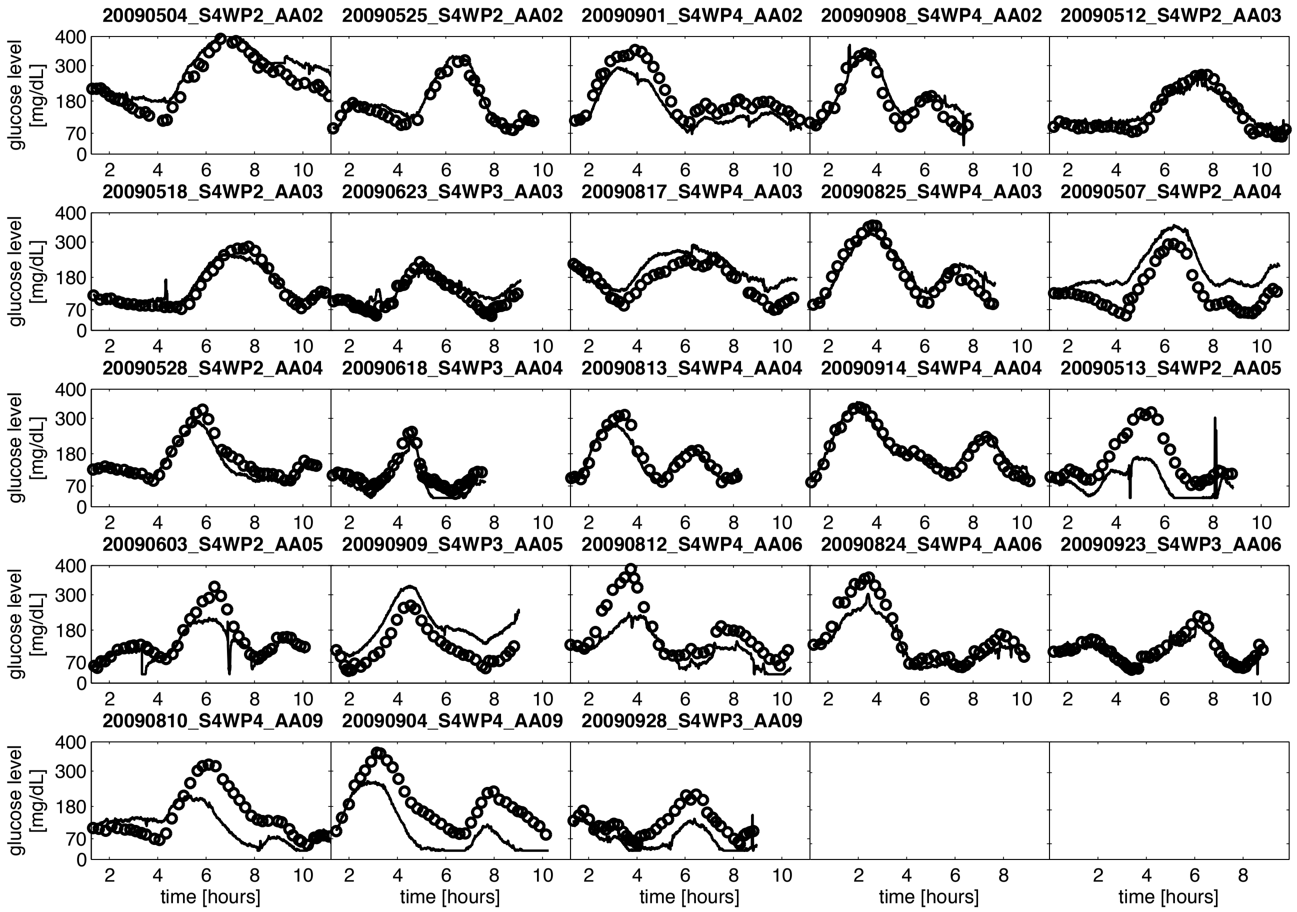


Figure A.10: Estimated glucose profiles by EN (continuous black line) against reference BGL values (black circles) when the same multi-sensor data used for model identification, i.e. data subset “part 2”, is considered (“internal validation”). The first part of the recording sessions’ labels indicates the data acquisition day, the second part is an internal notation, and the third part states subject’s id number.

B

Full Model Test Glucose Profiles

This section collects the full model test plots when data subset “part 2” and “part 1” are used to test the different models.

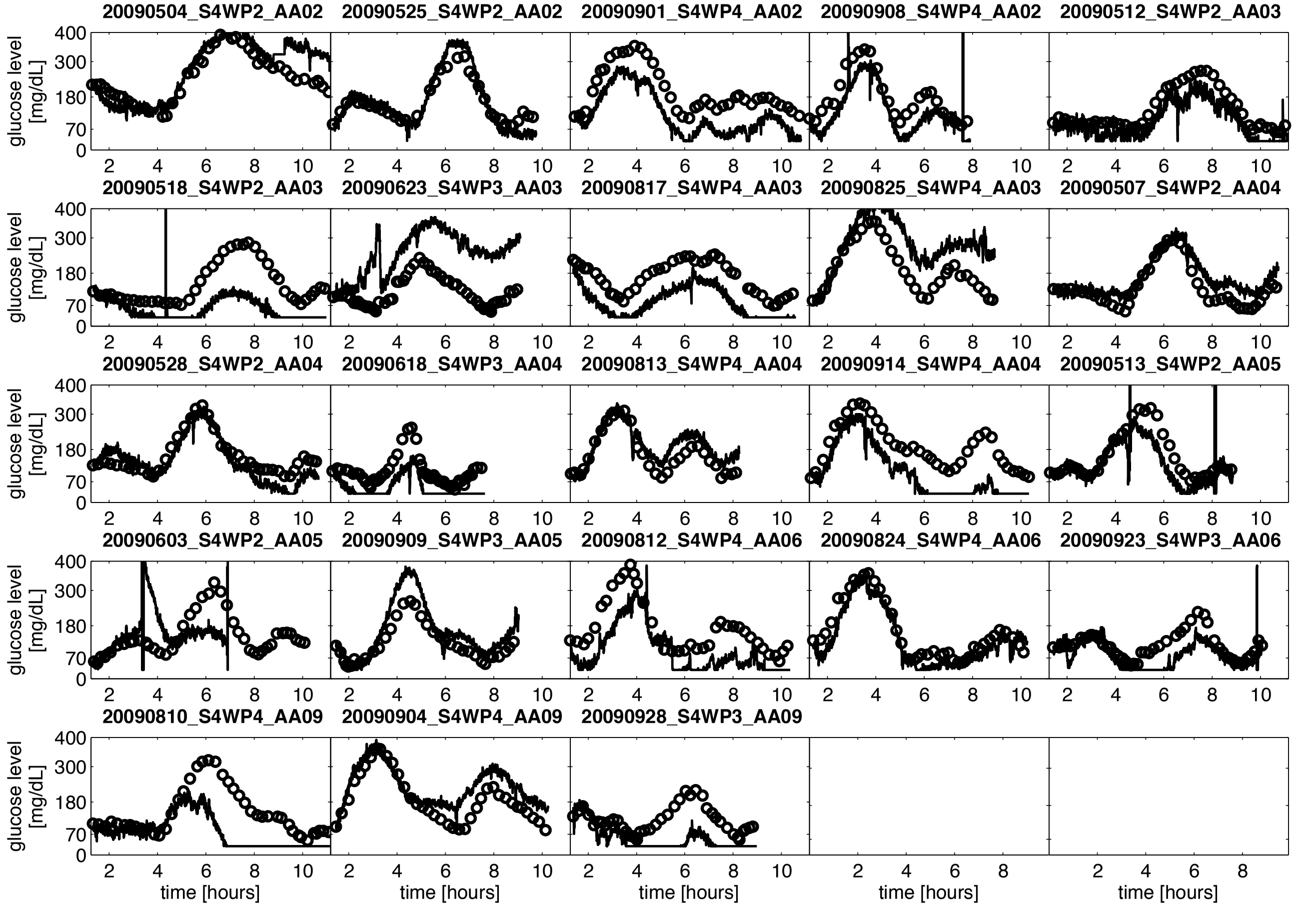


Figure B.1: Estimated glucose profiles by OLS (continuous black line) (continuous black line) against reference BGL values (black circles) when the same multi-sensor data used for model test, i.e. data subset "part 2", is considered ("external validation"). The first part of the recording sessions' labels indicates the data acquisition day, the second part is an internal notation, and the third part states subject's id number.

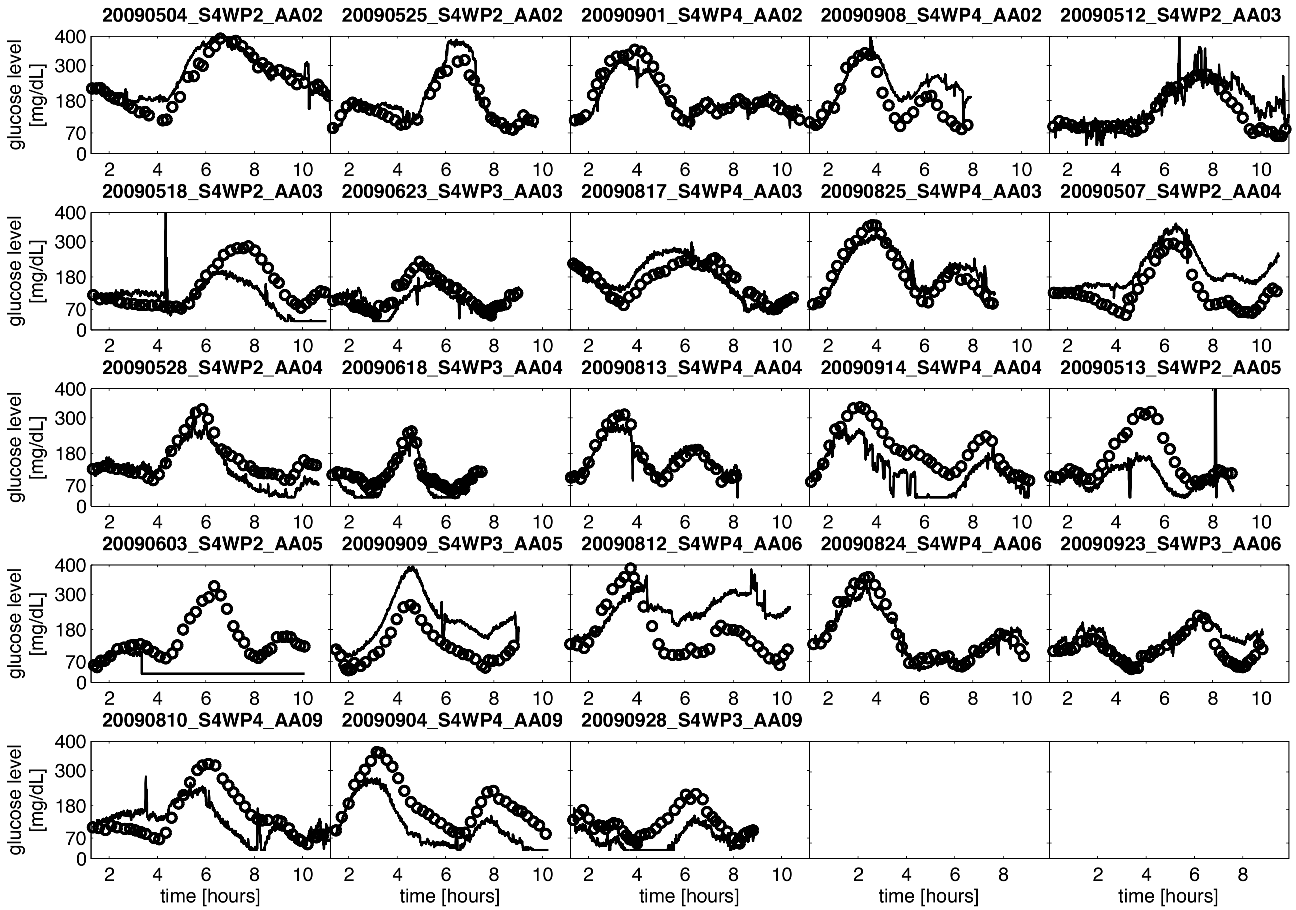


Figure B.2: Estimated glucose profiles by PLS (continuous black line) against reference BGL values (black circles) when the same multi-sensor data used for model identification, i.e. data subset "part 2", is considered ("external validation"). The first part of the recording sessions' labels indicates the data acquisition day, the second part is an internal notation, and the third part states subject's id number.

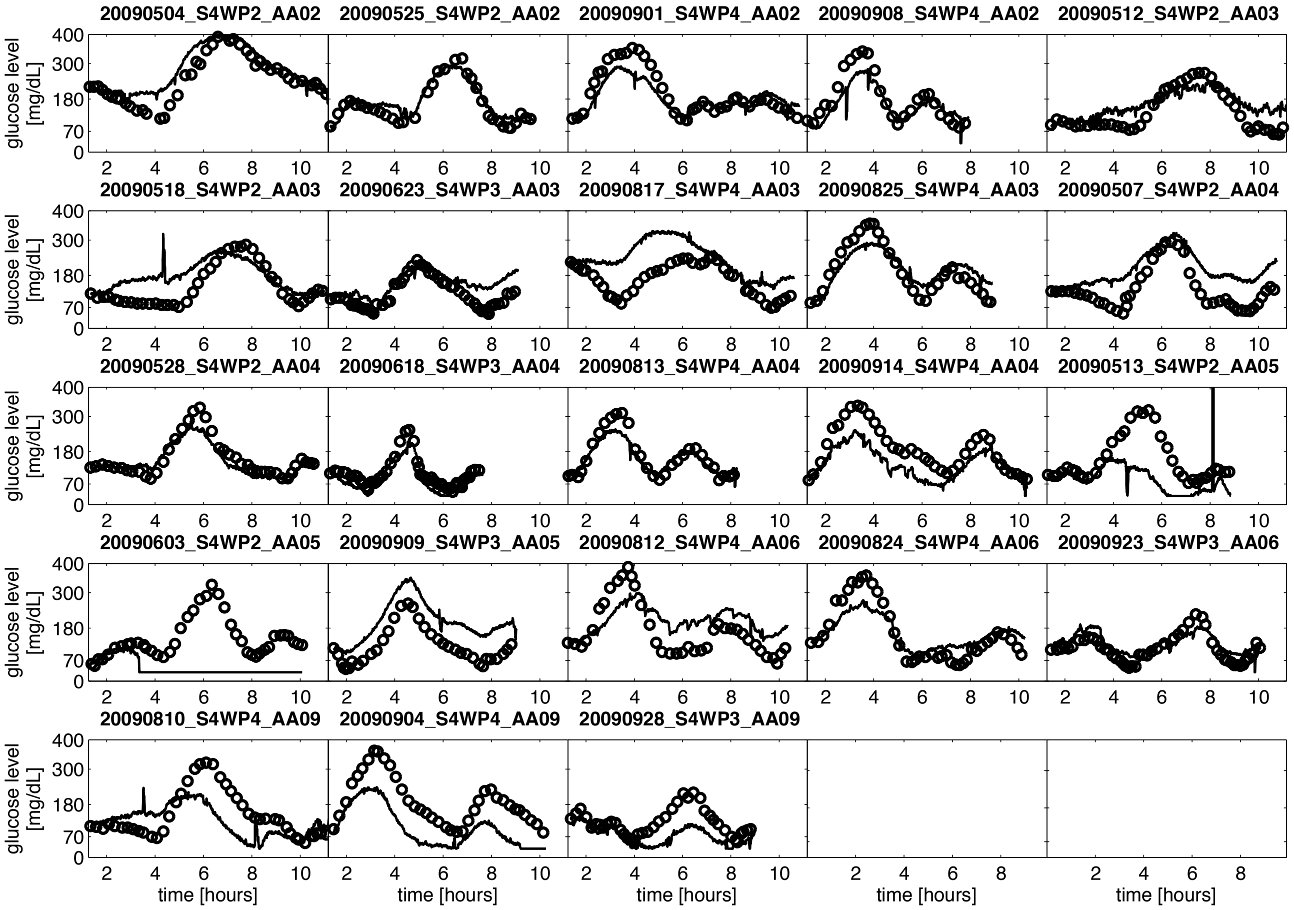


Figure B.3: Estimated glucose profiles by LASSO (continuous black line) against reference BGL values (black circles) when the same multi-sensor data used for model identification, i.e. data subset “part 2”, is considered (“external validation”). The first part of the recording sessions’ labels indicates the data acquisition day, the second part is an internal notation, and the third part states subject’s id number.

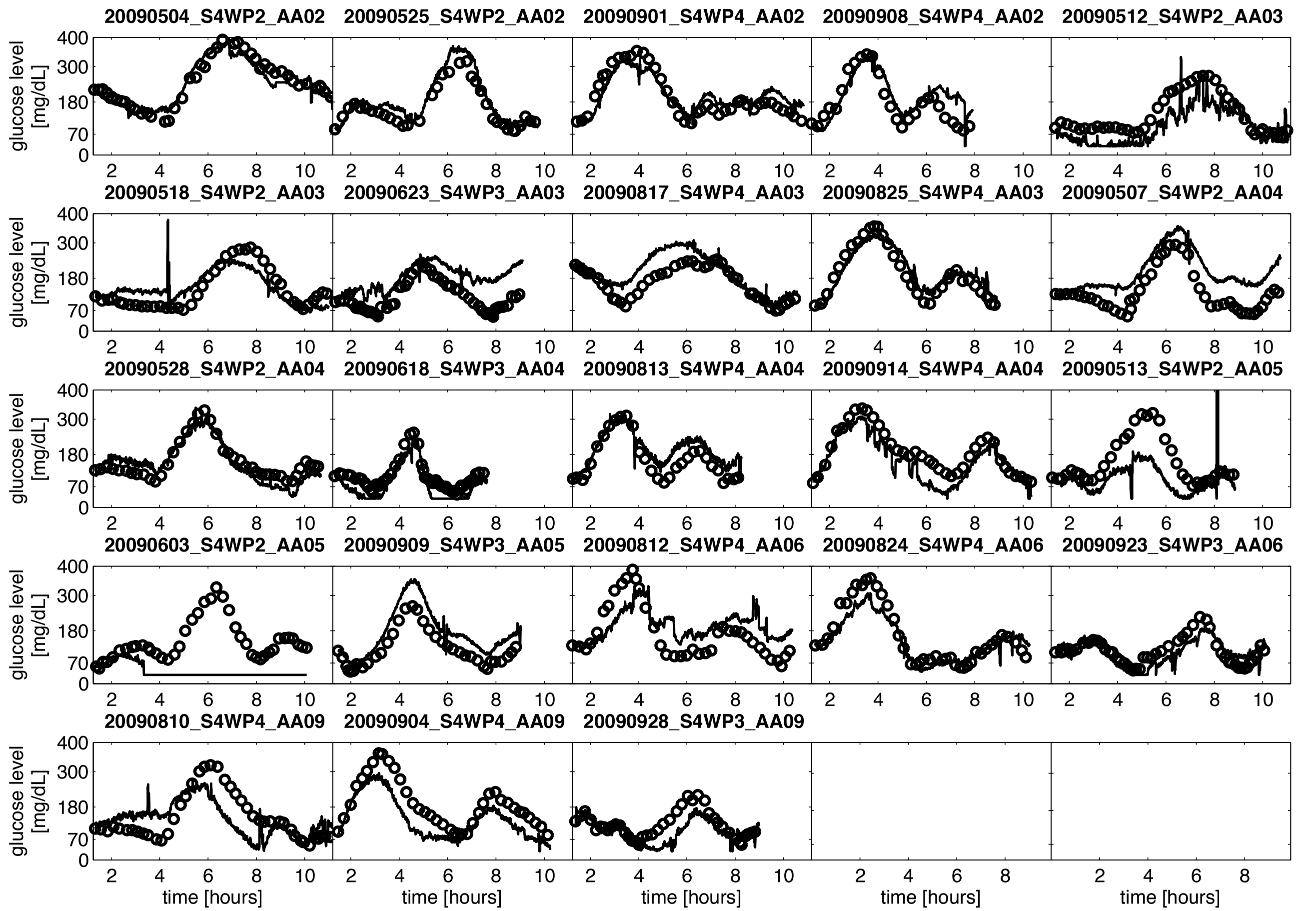


Figure B.4: Estimated glucose profiles by Ridge (continuous black line) against reference BGL values (black circles) when the same multi-sensor data used for model identification, i.e. data subset "part 2", is considered ("external validation"). The first part of the recording sessions' labels indicates the data acquisition day, the second part is an internal notation, and the third part states subject's id number.

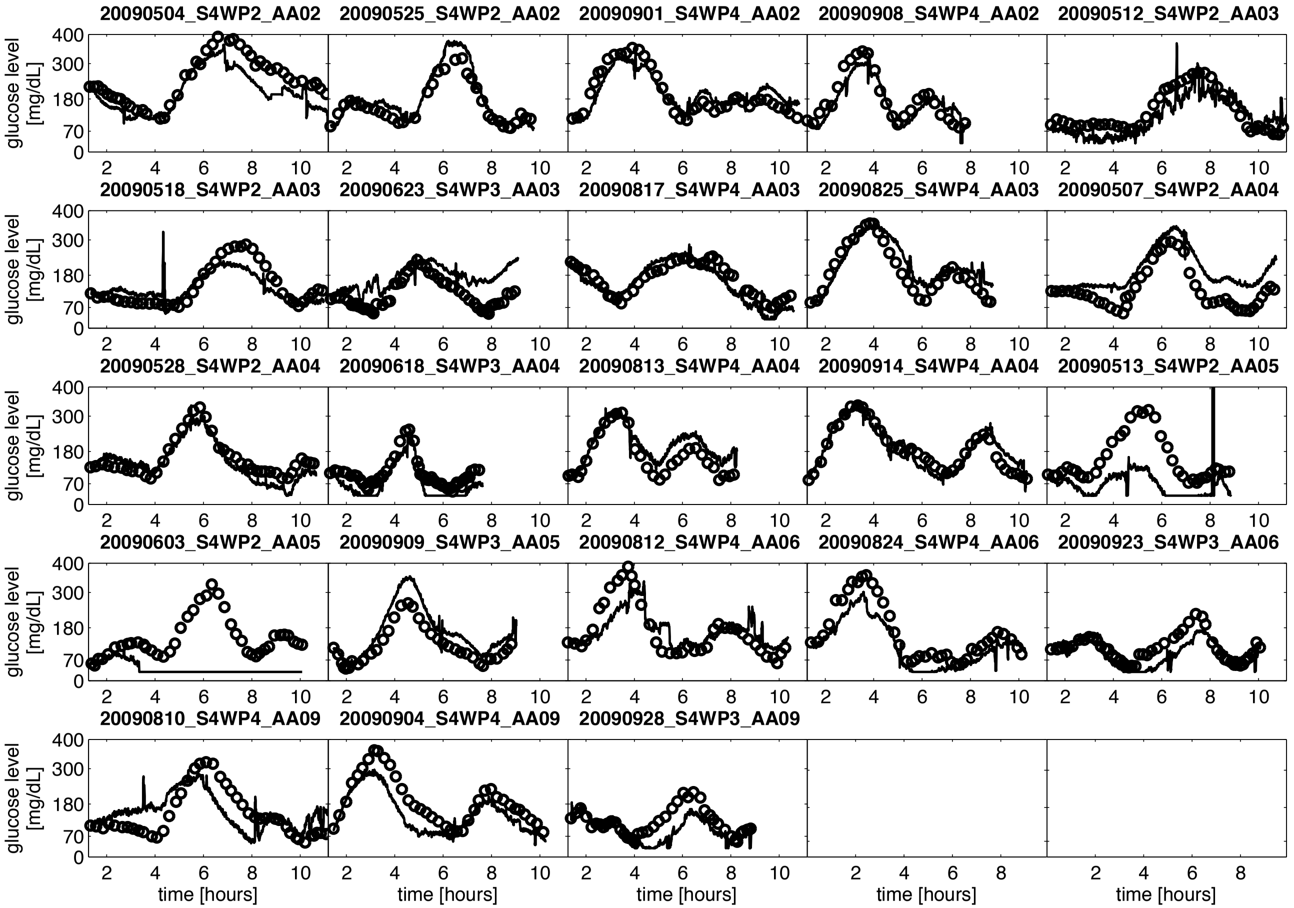


Figure B.5: Estimated glucose profiles by EN (continuous black line) against reference BGL values (black circles) when the same multi-sensor data used for model identification, i.e. data subset “part 2”, is considered (“external validation”). The first part of the recording sessions’ labels indicates the data acquisition day, the second part is an internal notation, and the third part states subject’s id number.

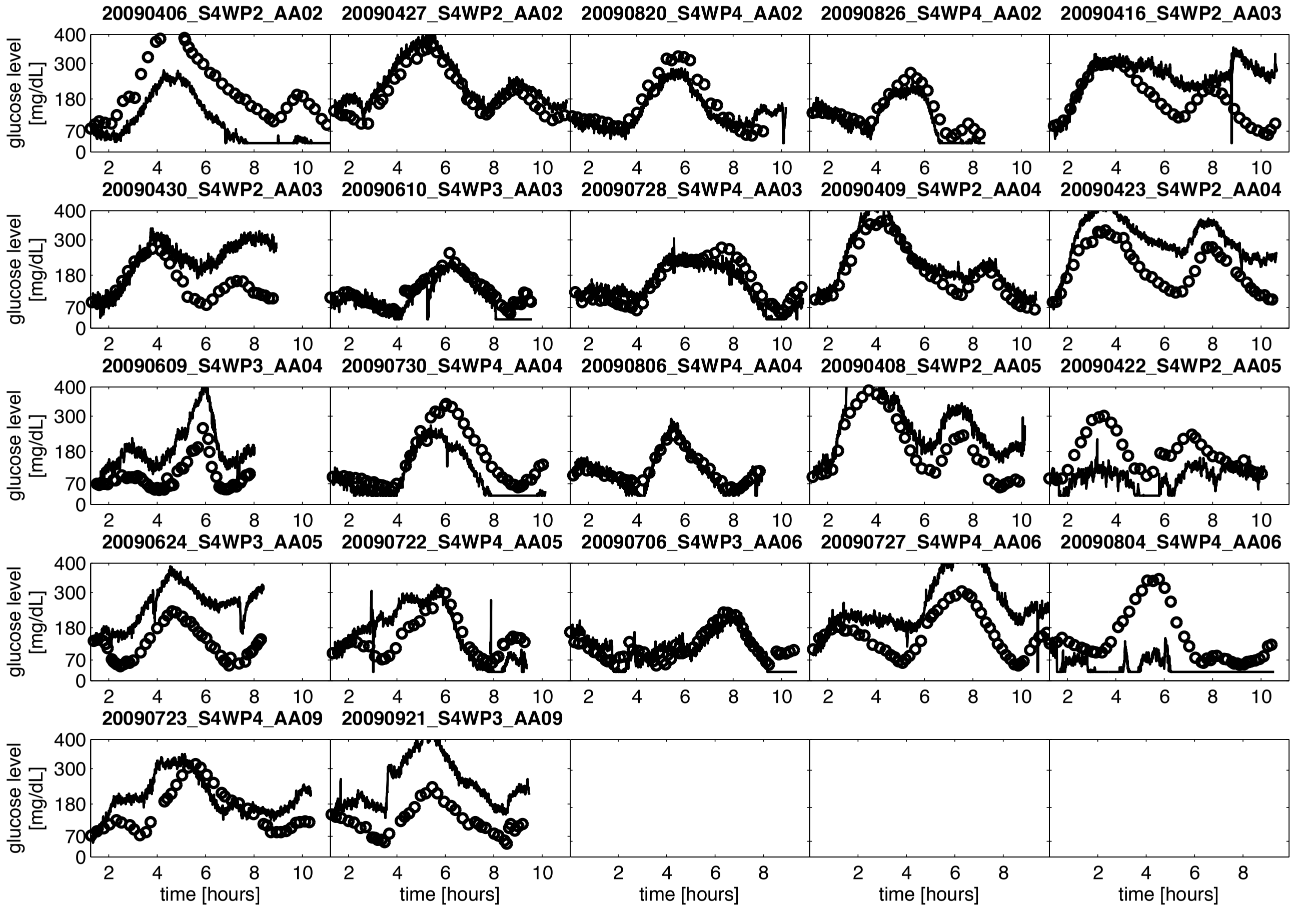


Figure B.6: Estimated glucose profiles by OLS (continuous black line) against reference BGL values (black circles) when the same multi-sensor data used for model identification, i.e. data subset “part 1”, is considered (“external validation”). The first part of the recording sessions’ labels indicates the data acquisition day, the second part is an internal notation, and the third part states subject’s id number.

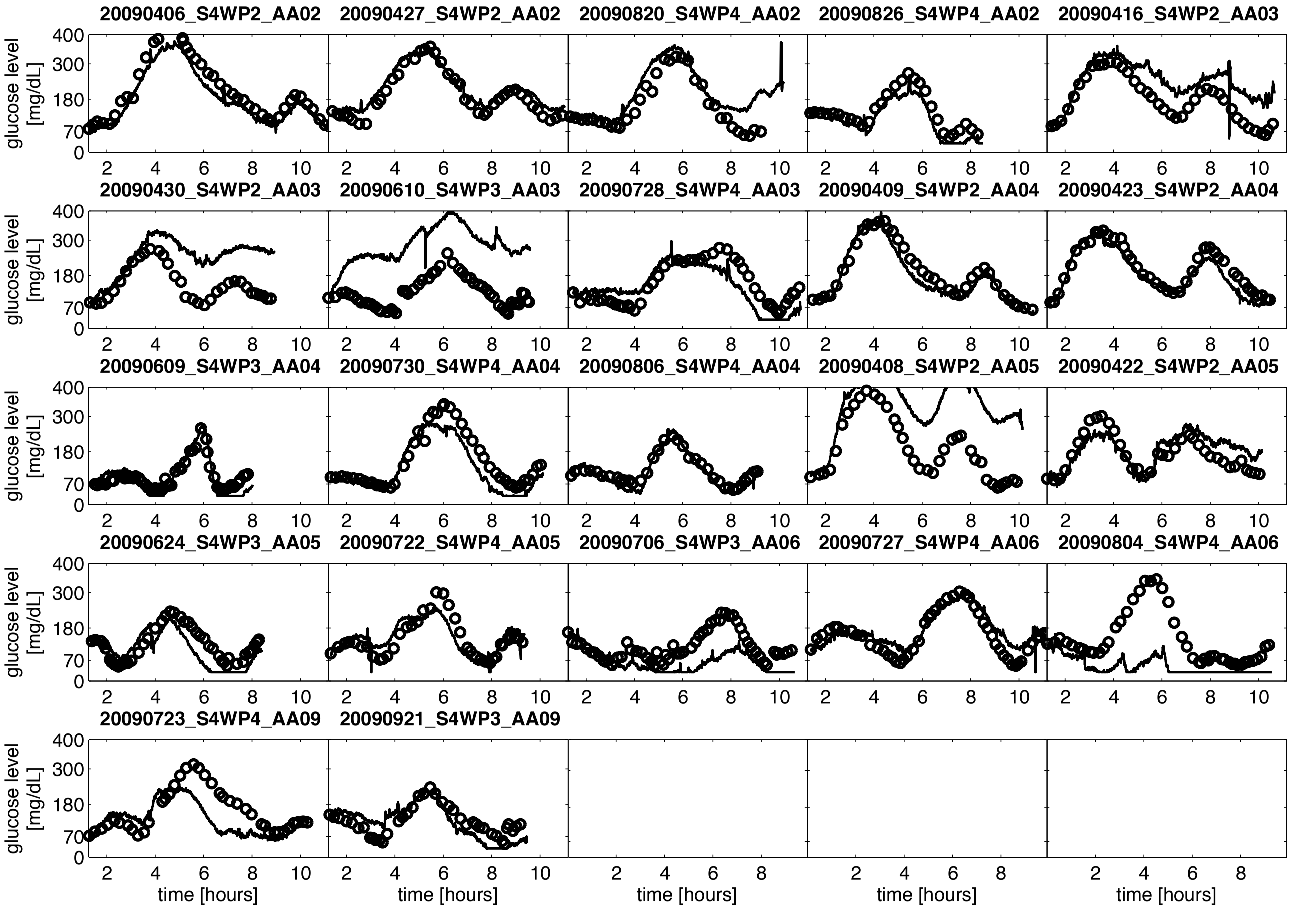


Figure B.7: Estimated glucose profiles by PLS (continuous black line) against reference BGL values (black circles) when the same multi-sensor data used for model identification, i.e. data subset “part 1”, is considered (“external validation”). The first part of the recording sessions’ labels indicates the data acquisition day, the second part is an internal notation, and the third part states subject’s id number.

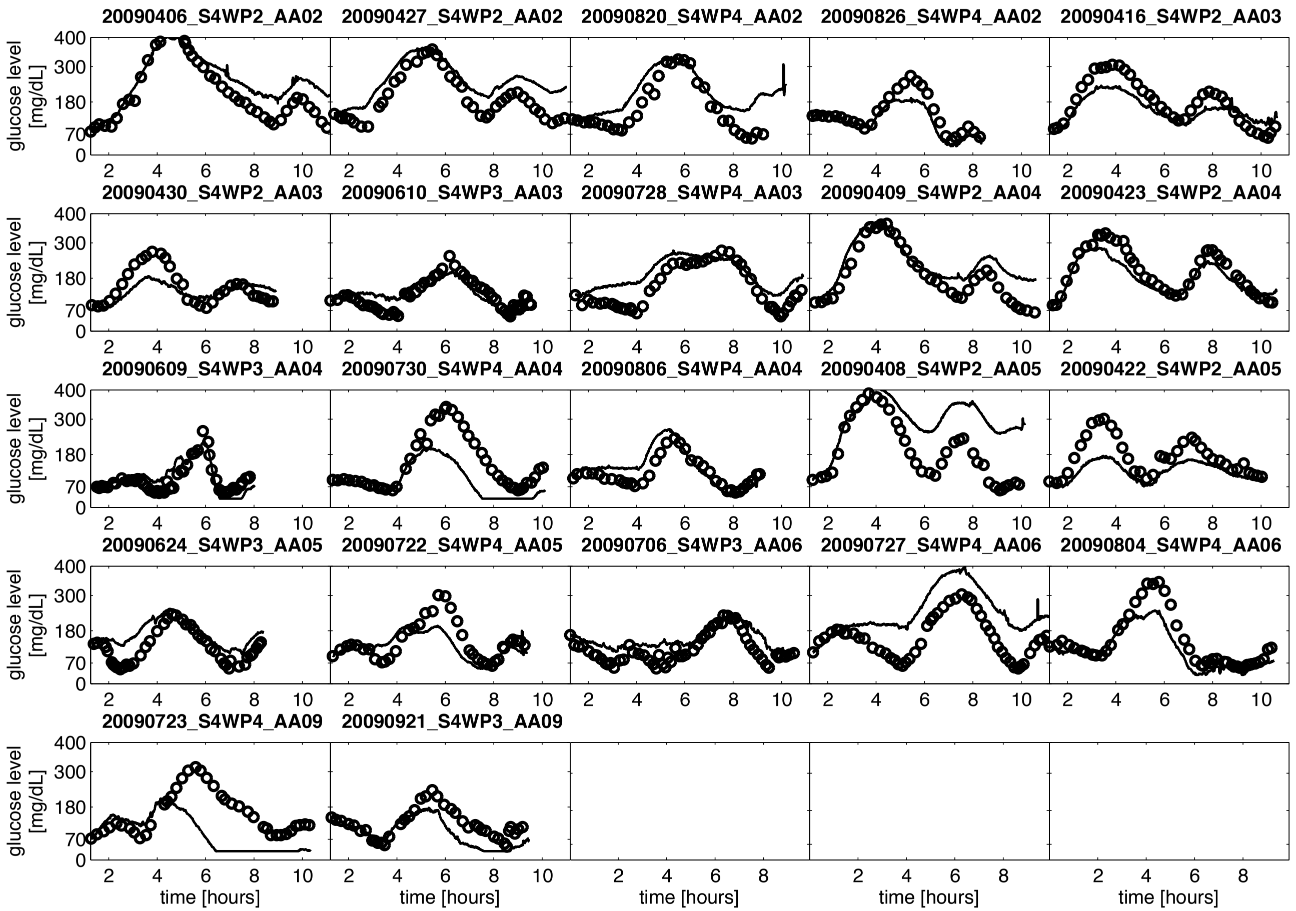


Figure B.8: Estimated glucose profiles by LASSO (continuous black line) against reference BGL values (black circles) when the same multi-sensor data used for model identification, i.e. data subset "part 1", is considered ("external validation"). The first part of the recording sessions' labels indicates the data acquisition day, the second part is an internal notation, and the third part states subject's id number.

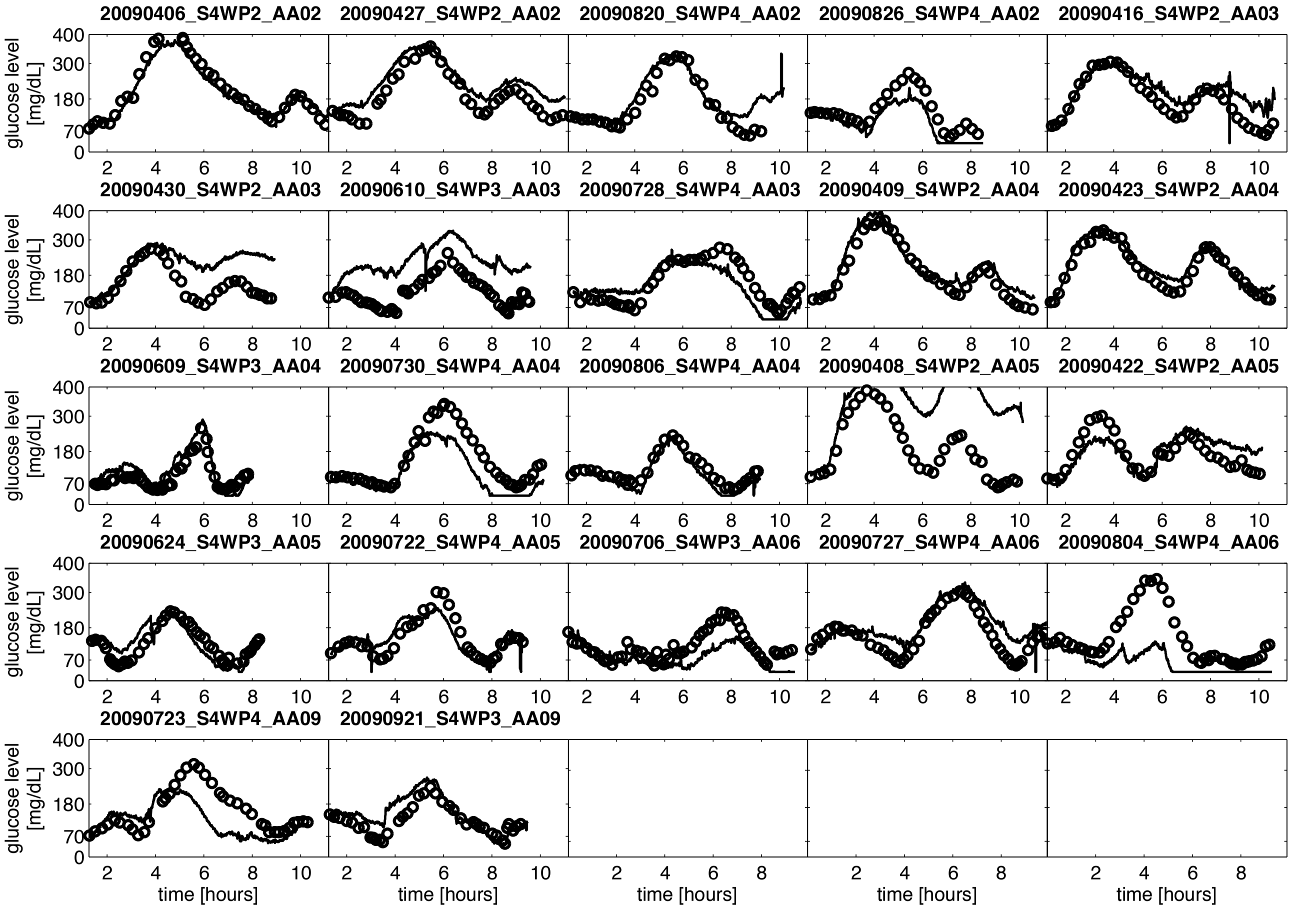


Figure B.9: Estimated glucose profiles by Ridge (continuous black line) against reference BGL values (black circles) when the same multi-sensor data used for model identification, i.e. data subset “part 1”, is considered (“external validation”). The first part of the recording sessions’ labels indicates the data acquisition day, the second part is an internal notation, and the third part states subject’s id number.

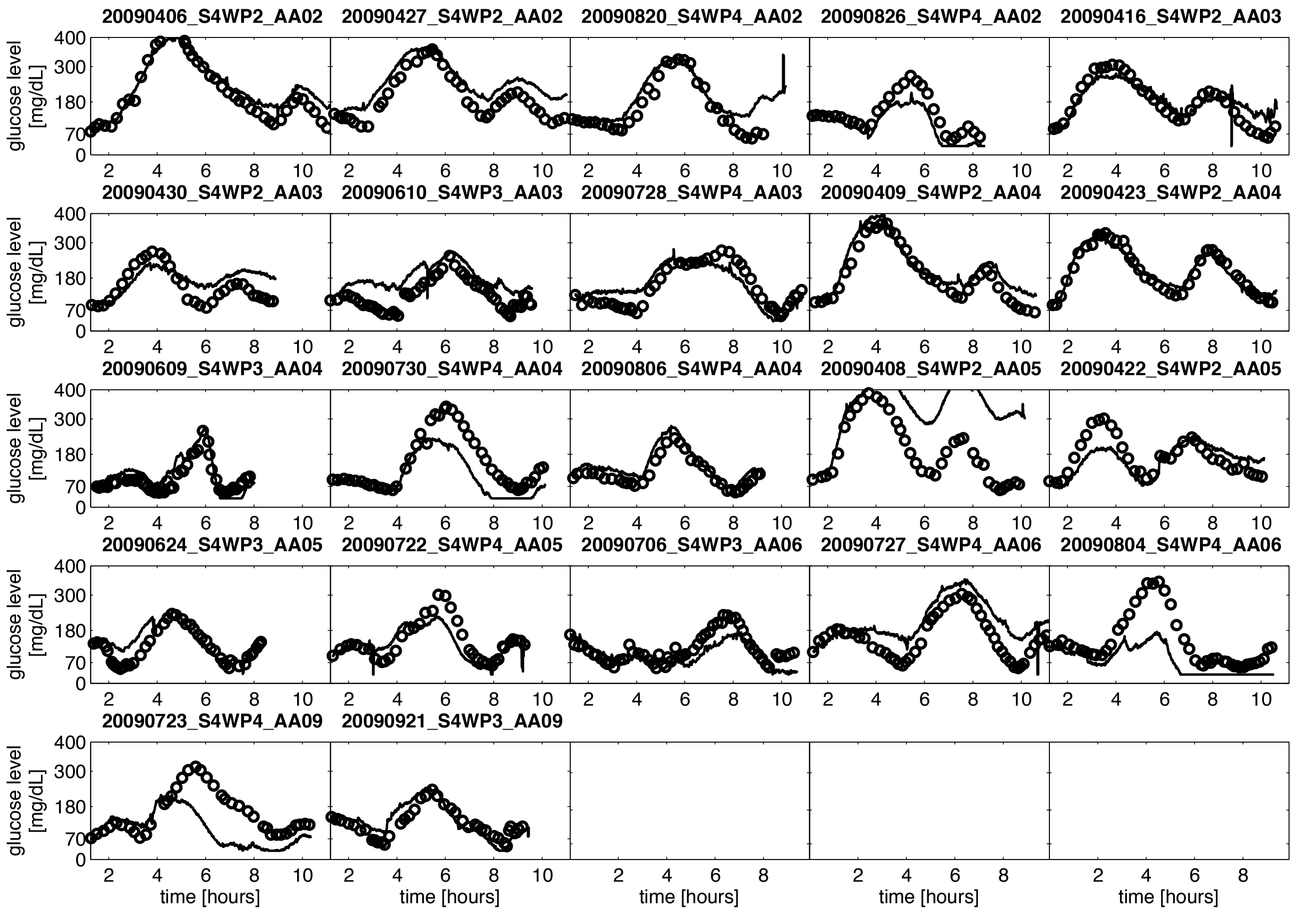


Figure B.10: Estimated glucose profiles by EN (continuous black line) against reference BGL values (black circles) when the same multi-sensor data used for model identification, i.e. data subset "part 1", is considered ("external validation"). The first part of the recording sessions' labels indicates the data acquisition day, the second part is an internal notation, and the third part states subject's id number.

Bibliography

[1] World Health Organization. `http://www.who.int/mediacentre/factsheets/fs312/en/`. Accessed 31 January 2013.

[2] IDF Diabetes Atlas; 5th ed. `http://www.idf.org/diabetesatlas/`. Accessed 31 January 2013.

[3] P. Zimmet, K. Alberti, J. Shaw, et al. Global and societal implications of the diabetes epidemic. *Nature*, 414(6865):782–787, 2001.

[4] A R Saltiel and R Kahn. Insulin signalling and the regulation of glucose and lipid metabolism. *Nature*, 414:799–806, 2001.

[5] I.M. Stratton, A.I. Adler, H.A.W. Neil, D.R. Matthews, S.E. Manley, C.A. Cull, D. Hadden, R.C. Turner, and R.R. Holman. Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes: prospective observational study. *British Medical Journal*, 321(7258):405, 2000.

[6] S.N. Davis and G. Lastra-Gonzalez. Diabetes and Low Blood Sugar (Hypoglycemia). *Journal of Clinical Endocrinology & Metabolism*, 93(8), 2008.

[7] S.R. Gambert and S. Pinkstaff. Emerging epidemic: diabetes in older adults: demography, economic impact, and pathophysiology. *Diabetes Spectrum*, 19(4):221, 2006.

[8] L. Heinemann and D. Boecker. Lancing: Quo vadis? *Journal of Diabetes Science and Technology*, 5(4):966, 2011.

[9] P. Magni and R. Bellazzi. A stochastic model to assess the variability of blood glucose time series in diabetic patients self-monitoring. *Biomedical Engineering, IEEE Transactions on*, 53(6):977–985, 2006.

[10] G.V. McGarraugh, W.L. Clarke, and B.P. Kovatchev. Comparison of the clinical information provided by the freestyle navigator continuous interstitial glucose monitor versus traditional blood glucose readings. *Diabetes Technology & Therapeutics*, 12(5):365–371, 2010.

[11] V. Srinivasan, V.K. Pamula, M.G. Pollack, and R.B. Fair. Clinical diagnostics on human whole blood, plasma, serum, urine, saliva, sweat, and tears on a digital microfluidic platform. In *Proc. μTAS*, pages 1287–1290, 2003.

[12] B H Ginsberg. The current environment of CGM technologies. *Journal of Diabetes Science and Technology*, 1(1):117–121, 2007.

[13] D. Rodbard. New and improved methods to characterize glycemic variability using continuous glucose monitoring. *Diabetes Technology & Therapeutics*, 11(9):551–565, 2009.

[14] B.P. Kovatchev, W.L. Clarke, M. Breton, K. Brayman, and A. McCall. Quantifying temporal glucose variability in diabetes via continuous glucose monitoring: mathematical methods and clinical application. *Diabetes Technology & Therapeutics*, 7(6):849–862, 2005.

[15] D. Deiss, J. Bolinder, J.P. Riveline, T. Battelino, E. Bosi, N. Tubiana-Rufi, D. Kerr, and M. Phillip. Improved glycemic control in poorly controlled patients with type 1 diabetes using real-time continuous glucose monitoring. *Diabetes Care*, 29(12):2730–2732, 2006.

[16] T. Battelino, M. Phillip, N. Bratina, R. Nimri, P. Oskarsson, and J. Bolinder. Effect of continuous glucose monitoring on hypoglycemia in type 1 diabetes. *Diabetes Care*, 34(4):795–800, 2011.

[17] G Sparacino, A Facchinetti, and C Cobelli. “Smart” continuous glucose monitoring sensors: On-line signal processing issues. *Sensors*, 10(6):6751–6772, 2010.

[18] G. Sparacino, M. Zanon, A. Facchinetti, C. Zecchin, A. Maran, and C. Cobelli. Italian contributions to the development of continuous glucose monitoring sensors for diabetes management. *Sensors*, 12(10):13753–13780, 2012.

[19] B.W. Bequette. Continuous glucose monitoring: real-time algorithms for calibration, filtering, and alarms. *Journal of Diabetes Science and Technology*, 4(2):404, 2010.

[20] Andrea Facchinetti, Giovanni Sparacino, Stefania Guerra, Yoeri M. Luijf, J.Hans DeVries, Julia K. Mader, Martin Ellmerer, Carsten Benesch, Lutz Heinemann,

Daniela Bruttomesso, Angelo Avogaro, Claudio Cobelli, and on behalf of the AP@home Consortium. Real-time improvement of continuous glucose-monitoring accuracy: The smart sensor concept. *Diabetes Care*, 2012.

[21] A. Facchinetti, G. Sparacino, and C. Cobelli. Online denoising method to handle intraindividual variability of signal-to-noise ratio in continuous glucose monitoring. *Biomedical Engineering, IEEE Transactions on*, 58(9):2664–2671, 2011.

[22] J.G. Chase, C.E. Hann, M. Jackson, J. Lin, T. Lotz, X.W. Wong, and G.M. Shaw. Integral-based filtering of continuous glucose sensor measurements for glycaemic control in critical care. *Computer Methods and Programs in Biomedicine*, 82(3):238–247, 2006.

[23] S. Guerra, A. Facchinetti, G. Sparacino, G. De Nicolao, and C. Cobelli. Enhancing the accuracy of subcutaneous glucose sensors: a real-time deconvolution-based approach. *Biomedical Engineering, IEEE Transactions on*, 59(6):1658–1669, 2012.

[24] C. King, S.M. Anderson, M. Breton, W.L. Clarke, and B.P. Kovatchev. Modeling of calibration effectiveness and blood-to-interstitial glucose dynamics as potential confounders of the accuracy of continuous glucose sensors during hyperinsulinemic clamp. *Journal of Diabetes Science and Technology (Online)*, 1(3):317, 2007.

[25] D.T. Ther. Evaluation of factors affecting cgms calibration. *Diabetes Technology & Therapeutics*, 8(3):318–325, 2006.

[26] 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. *Biomedical Engineering, IEEE Transactions on*, 59(6):1550–1560, 2012.

[27] M. Eren-Oruklu, A. Cinar, L. Quinn, and D. Smith. Estimation of future glucose concentrations with subject-specific recursive linear models. *Diabetes Technology & Therapeutics*, 11(4):243–253, 2009.

[28] D.A. Finan, F.J. Doyle III, C.C. Palerm, W.C. Bevier, H.C. Zisser, L. Jovanovič, and D.E. Seborg. Experimental evaluation of a recursive model identification technique for type 1 diabetes. *Journal of Diabetes Science and Technology (Online)*, 3(5):1192, 2009.

[29] V. Naumova, SV Pereverzyev, and S. Sivananthan. A meta-learning approach to the regularized learningcase study: Blood glucose prediction. *Neural Networks*, 33:181–193, 2012.

[30] D. Bruttomesso, A. Farret, S. Costa, M.C. Marescotti, M. Vettore, A. Avogaro, A. Tiengo, C. Dalla Man, J. Place, A. Facchinetti, S. Guerra, L. Magni, G. De Nicolao, C. Cobelli, E. Renard, and A. Maran. Closed-loop artificial pancreas using subcutaneous glucose sensing and insulin delivery and a model predictive control algorithm: Preliminary studies in Padova and Montpellier. *Journal of Diabetes Science and Technology*, 3(5):1014–1021, 2009.

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

[32] R. Hovorka, J.M. Allen, D. Elleri, L.J. Chassin, J. Harris, D. Xing, C. Kollman, T. Hovorka, A.M.F. Larsen, M. Nodale, et al. Manual closed-loop insulin delivery in children and adolescents with type 1 diabetes: a phase 2 randomised crossover trial. *The Lancet*, 375(9716):743–751, 2010.

[33] S. Garg and IB Hirsch. Self-monitoring of blood glucose. *International Journal of Clinical Practice*, 64(166):1–10, 2010.

[34] C.M. Girardin, C. Huot, M. Gonthier, and E. Delvin. Continuous glucose monitoring: A review of biochemical perspectives and clinical use in type 1 diabetes. *Clinical Biochemistry*, 42(3):136–142, 2009.

[35] G. McGarraugh. The chemistry of commercial continuous glucose monitors. *Diabetes Technology & Therapeutics*, 11(S1):17–24, 2009.

[36] F. Ricci, D. Moscone, and G. Palleschi. Ex vivo continuous glucose monitoring with microdialysis technique: The example of glucoday. *Sensors Journal, IEEE*, 8(1):63–70, 2008.

[37] A. Tura, A. Maran, and G. Pacini. Non-invasive glucose monitoring: assessment of technologies and devices according to quantitative criteria. *Diabetes Research and Clinical Practice*, 77(1):16–40, 2007.

[38] S.K. Vashist. Non-invasive glucose monitoring technology in diabetes management: A review. *Analytica Chimica Acta*, 2012.

[39] E Renard. Implantable closed-loop glucose-sensing and insulin delivery pump therapy. *Current Opinion in Pharmacology*, 2(6):708–716, 2002.

[40] G. Bochicchio, J. Joseph, M. Magee, A. Gulino, M. Higgins, E. Lifesciences, T. Peyser, P. Simpson, J. Leach, and A. Kamath. Multicenter evaluation of a first

generation automated blood glucose monitor in the or/icu. *Critical Care Medicine*, 39(12):55, 2011.

[41] J.C. Pickup, F. Hussain, N.D. Evans, O.J. Rolinski, and D.J.S. Birch. Fluorescence-based glucose sensors. *Biosensors and Bioelectronics*, 20(12):2555–2565, 2005.

[42] T. Peyser, H. Zisser, U. Khan, L. Jovanovič, W. Bevier, M. Romey, J. Suri, P. Strasma, S. Tiaden, and S. Gamsey. Use of a novel fluorescent glucose sensor in volunteer subjects with type 1 diabetes mellitus. *Journal of Diabetes Science and Technology*, 5(3):687, 2011.

[43] B Beier, K Musick, A Matsumoto, A Panitch, E Naumann, and P Irazoqui. Toward a continuous intravascular glucose monitoring system. *Sensors*, 11(1):409–424, 2011.

[44] A. Facchinetti, G. Sparacino, C. Cobelli, et al. Reconstruction of glucose in plasma from interstitial fluid continuous glucose monitoring data: role of sensor calibration. *Journal of Diabetes Science and Technology*, 1(5):617, 2007.

[45] Abbott Diabetes Care. Freestyle Navigator. Available online:. `www.freestylenavigator.com`. Accessed 31 January 2013.

[46] R L Weistein, S L Schwartz, R L Brazg, J R Bugler, T A Peyser, and G V McGarraugh. Accuracy of the 5-day freestyle navigator continuous glucose monitoring system. *Diabetes Care*, 30(5):1125–1130, 2007.

[47] D M Wilson, R W Beck, W V Tamborlane, M J Dontchev, C Kollman, P Chase, L A Fox, K J Ruedy, E Tsalikian, and S A Weinzimer. The accuracy of the freestyle navigator continuous glucose monitoring system in children with type 1 diabetes. *Diabetes Care*, 30(1):59–64, 2007.

[48] IDF Diabetes Atlas; 5th ed. `http://www.dexcom.com/seven-plus`. Accessed 31 January 2013.

[49] S Garg, H Zisser, S Schwartz, T Bailey, R Kaplan, S Ellis, and L Jovanvic. Improvement in glycemic excursion with a trancutaneous, real-time continuous glucose sensor. *Diabetes Care*, 29(1):44–50, 2006.

[50] Medtronic Diabetes. Guardian CGM System. Available online:. `http://www.medtronicdiabetes.com/products/guardiancgm`. Accessed 31 January 2013.

[51] J Mastrototaro and S Lee. The integrated minimed paradigm real-time insulin pump and glucose monitoring system: Implications for improved patient outcomes. *Diabetes Technology & Therapeutics*, 11(1):S37–S44, 2009.

[52] A Maran. Continuous subcutaneous glucose monitoring in diabetic patients. *Diabetes Care*, 25(2):347–352, 2002.

[53] T Kubiak, B Woerle, B Kuhr, I Nied, G Glaesner, N Hermanns, B Kulzer, and T Haak. Microdialysis-based 48-hour continuous glucose monitoring with glucoday: clinical performance and patient's acceptance. *Diabetes Technology & Therapeutics*, 8(5):570–575, 2006.

[54] F. Valgimigli, F. Lucarelli, C. Scuffi, S. Morandi, and I. Sposato. Evaluating the clinical accuracy of glucomen® day: a novel microdialysis-based continuous glucose monitor. *Journal of Diabetes Science and Technology*, 4(5):1182, 2010.

[55] F. Ricci, F. Caprio, A. Poscia, F. Valgimigli, D. Messeri, E. Lepori, G. Dall'Oglio, G. Palleschi, and D. Moscone. Toward continuous glucose monitoring with planar modified biosensors and microdialysis: Study of temperature, oxygen dependence and in vivo experiment. *Biosensors and Bioelectronics*, 22(9):2032–2039, 2007.

[56] F. Lucarelli, F. Ricci, F. Caprio, F. Valgimigli, C. Scuffi, D. Moscone, and G. Palleschi. Glucomen day continuous glucose monitoring system: A screening for enzymatic and electrochemical interferents. *Journal of Diabetes Science and Technology*, 6(5):1172, 2012.

[57] C. Kapitza, V. Lodwig, K. Obermaier, K.J.C. Wientjes, K. Hoogenberg, K. Jungheim, and L. Heinemann. Continuous glucose monitoring: reliable measurements for up to 4 days with the scgm1 system. *Diabetes Technology & Therapeutics*, 5(4):609–614, 2003.

[58] D. Cooke, SJ Hurel, A. Casbard, L. Steed, S. Walker, S. Meredith, AJ Nunn, A. Manca, M. Sculpher, M. Barnard, et al. Randomized controlled trial to assess the impact of continuous glucose monitoring on hba1c in insulin-treated diabetes (mitre study). *Diabetic Medicine*, 26(5):540–547, 2009.

[59] K Pitzer, S Desai, T Dunn, S Edelman, Y Jayalakshimi, J Kennedy, J A Tamada, and R O Potts. Detection oh hypoglycemia with the glucowatch biographer. *Diabetes Care*, 24(5):881–885, 2001.

[60] A Tura. Advances in the development of devices for noninvasive glycemia monitoring: who will win the race? *Nutritional Therapy & Metabolism*, 28(1):33–39, 2010.

[61] S Y Rhee, S Chon, G Koh, J R Paeng, S Oh, J Woo, S W Kim, J Kim, and Y S Kim. Clinical experience of an iontophoresis based glucose measuring system. *J Korean Med Sci*, 22(5):70–73, 2007.

[62] C T S Ching, T P Sun, S H Huang, H L Shieh, and C Y Chen. A mediated glucose biosensor incorporated with reverse iontophoresis function for noninvasive glucose monitoring. *Annals of Biomedical Engineering*, 38(4):1548–1555, 2010.

[63] Echo Therapeutics. Available online:. `www.echotx.com/symphony-tcgm-system.html`. Accessed 31 January 2013.

[64] B M Becker, S Helfrich, E Backer, K Lovgren, A Minugh, and J Machan. Ultrasound with topical anesthetic rapidly decreases pain of intravenous cannulation. *Acad Emerg Med*, 12(4):289–285, 2005.

[65] H Chuang, M Trieu, J Hurley, E J Taylor, M R England, and S A Nasraway. Pilot studies of transdermal continuous glucose measurement in outpatient diabetic and in patients during and after cardiac surgery. *Journal of Diabetes Science and Technology*, 2(4):595–602, 2008.

[66] C E Ferrante do Amaral and B Wolf. Current development in non-invasive glucose monitoring. *Medical Engineering & Physics*, 30(5):541–549, 2008.

[67] P. Zakharov, F. Dewarrat, A. Caduff, and MS Talary. The effect of blood content on the optical and dielectric skin properties. *Physiological Measurement*, 32(1):131–151, 2011.

[68] C S Chen, K K Wang, M Y Jan, W C Hsu, S P Li, Y Y Wang-Lin, and J G Bau. Noninvasive blood glucose monitoring using the optical signal of pulsatile microcirculation: a pilot study in subjects with diabetes. *Journal of Diabetes and Its Complications*, 22(6):371–376, 2008.

[69] O Amir, D Weinstein, S Zilberman, M Less, D Perl-Treves, H Primack, A Weinstein, E Gabis, B Fikhte, and A Karasik. Continuous non invasive glucose monitoring technology based on “occlusion spectroscopy”. *Journal of Diabetes Science and Technology*, 1(4):463–469, 2007.

[70] R. A. Gabbay and S. Sivarajah. Optical coherence tomography-based continuous noninvasive glucose monitoring in patients with diabetes. *Diabetes Technology & Therapeutics*, 10(3):188–193, 2008.

[71] N S Oliver, C Toumazou, E G Cass, and G Johnston. Glucose sensors: a review of current and emerging technology. *Diabetic Medicine*, 26(3):197–210, 2009.

[72] D.D. Cunningham and J.A. Stenken. *In vivo glucose sensing*, volume 174. Wiley, 2009.

[73] G Yosipovitch, E Hodak, P Vardi, I Shraga, M Karp, E Sprecher, and M David. The prevalence of cutaneous manifestations in IDDM patients and their association with diabetes risk factors and microvascular complications. *Diabetes Care*, 21(4):506–509, 1998.

[74] Y. Yamakoshi, M. Ogawa, T. Yamakoshi, M. Satoh, M. Nogawa, S. Tanaka, T. Tamura, P. Rolfe, and K. Yamakoshi. A new non-invasive method for measuring blood glucose using instantaneous differential near infrared spectrophotometry. In *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pages 2964–2967. IEEE, 2007.

[75] Noninvasive Glucose InLight Solutions Bringing Light to Life. `http://www.inlightsolutions.com/prod-glu.html`. Accessed 31 January 2013.

[76] P. Trombetta and V. Londoni. Diode laser device for the non-invasive measurement of glycaemia, January 27 2011. US Patent App. 13/014,998.

[77] H.L. Berman, J.N. Roe, and R.N. Blair. Glucose measurement utilizing non-invasive assessment methods, February 18 2003. US Patent 6,522,903.

[78] D A Stuart, J M Yuen, N Shah, O Lyandres, C R Yonzon, M R Glucksberg, J T Walsh, and R P Van Duyne. In vivo glucose measurement by surface-enhanced raman spectroscopy. *Analytical Chemistry*, 78(20):7211–7215, 2006.

[79] Annika M K Enejder, Thomas G. Scecina, Martin Hunter, Wei-Chuan Shih, Michael S. Feld, Jeankun Oh, Slobodan Sasic, and Gary L. Horowitz. Raman spectroscopy for noninvasive glucose measurements. *Journal of Biomedical Optics*, 10(3):031114–1–031114–9, 2005.

[80] J. Lipson, J. Bernhardt, U. Block, W.R. Freeman, R. Hofmeister, M. Hristakeva, T. Lenosky, R. McNamara, D. Petrasek, D. Veltkamp, et al. Non-invasive technologies for glucose monitoring: Requirements for calibration in noninvasive glucose

monitoring by raman spectroscopy. *Journal of Diabetes Science and Technology*, 3(2):233, 2009.

[81] O. Cohen, I. Fine, E. Monashkin, and A. Karasik. Glucose correlation with light scattering patterns-a novel method for non-invasive glucose measurements. *Diabetes Technology & Therapeutics*, 5(1):11–17, 2003.

[82] Diabetes & Blood Glucose Orsense. `http://www.orsense.com/application.php?ID=6`. Accessed 31 January 2013.

[83] K.V. Larin, T.V. Ashitkov, I.V. Larina, I.Y. Petrova, M.S. Eledrisi, M. Motamedi, and R.O. Esenaliev. Optical coherence tomography and noninvasive blood glucose monitoring: a review. In *Saratov Fall Meeting 2003: Optical Technologies in Biophysics and Medicine V*, pages 285–290. International Society for Optics and Photonics, 2004.

[84] R. Badugu, J.R. Lakowicz, and C.D. Geddes. A glucose-sensing contact lens: from bench top to patient. *Current Opinion in Biotechnology*, 16(1):100 – 107, 2005.

[85] H. Shibata, Y.J. Heo, T. Okitsu, Y. Matsunaga, T. Kawanishi, and S. Takeuchi. Injectable hydrogel microbeads for fluorescence-based in vivo continuous glucose monitoring. *PNAS*, 1(5):1–5, 2010.

[86] E.A. Moschou, B.V. Sharma, S.K. Deo, and S. Daunert. Fluorescence glucose detection: advances toward the ideal in vivo biosensor. *Journal of Fluorescence*, 14(5):535–547, 2004.

[87] J. Sandby-Møller, T. Poulsen, and H.C. Wulf. Influence of epidermal thickness, pigmentation and redness on skin autofluorescence. *Photochemistry and Photobiology*, 77(6):616–620, 2003.

[88] Bilal H. Malik and Gerard L. Coté. Real-time, closed-loop dual-wavelength optical polarimetry for glucose monitoring. *Journal of Biomedical Optics*, 15(1):017002/1–017002/6, 2010.

[89] C.D. Malchoff, K. Shoukri, J.I. Landau, and J.M. Buchert. A novel noninvasive blood glucose monitor. *Diabetes Care*, 25(12):2268–2275, 2002.

[90] O.S. Khalil. Near-infrared thermo-optical response of the localized reflectance of diabetic and non-diabetic human skin. *Handbook of Optical Sensing of Glucose in Biological Fluids and Tissues*, page 181, 2008.

[91] R Weiss, Y Yegorchikov, A Shusterman, and I Raz. Non invasive continuous glucose monitoring using photoacoustic technology-results from the first 62 subjects. *Journal of Diabetes Science and Technology*, 9(1):68–74, 2007.

[92] A. Tura, S. Sbrignadello, D. Cianciavicchia, G. Pacini, and P. Ravazzani. A low frequency electromagnetic sensor for indirect measurement of glucose concentration: In vitro experiments in different conductive solutions. *Sensors*, 10(6):5346–5358, 2010.

[93] Y. Hayashi, L. Livshits, A. Caduff, and Y. Feldman. Dielectric spectroscopy study of specific glucose influence on human erythrocyte membranes. *Journal of Physics D: Applied Physics*, 36(4):369, 2003.

[94] P. Åberg. Skin cancer as seen by electrical impedance. *Stockholm, Sweden: Karolinska Instituttet*, 2004.

[95] A Caduff, M Talary, and P Zakharov. Cutaneous blood perfusion as a perturbing factor for non invasive glucose monitoring. *Diabetes Technology & Therapeutics*, 12(1):1–9, 2010.

[96] L. Livshits, A. Caduff, MS Talary, and Y. Feldman. Dielectric response of biconcave erythrocyte membranes to d-and l-glucose. *Journal of Physics D: Applied Physics*, 40(1):15, 2006.

[97] L. Livshits, A. Caduff, M.S. Talary, H.U. Lutz, Y. Hayashi, A. Puzenko, A. Shendrik, and Y. Feldman. The role of glut1 in the sugar-induced dielectric response of human erythrocytes. *The Journal of Physical Chemistry B*, 113(7):2212–2220, 2009.

[98] A. Tura, S. Sbrignadello, S. Barison, S. Conti, and G. Pacini. Impedance spectroscopy of solutions at physiological glucose concentrations. *Biophysical Chemistry*, 129(2):235–241, 2007.

[99] G. Gelao, R. Marani, V. Carriero, and A.G. Perri. Design of a dielectric spectroscopy sensor for continuous and non-invasive blood glucose monitoring. *International Journal of Advances in Engineering & Technology*, 3, 2012.

[100] K.V. Larin, M.S. Eledrisi, M. Motamedi, and R.O. Esenaliev. Noninvasive blood glucose monitoring with optical coherence tomography a pilot study in human subjects. *Diabetes Care*, 25(12):2263–2267, 2002.

[101] M.A. Arnold and G.W. Small. Noninvasive glucose sensing. *Analytical Chemistry*, 77(17):5429–5439, 2005.

[102] A Caduff, M Talary, M Mueller, F Dewarrat, J Klisic, M Donath, L Heinemann, and W A Stahel. Non-invasive glucose monitoring in patients with type 1 diabetes: a multisensor system combining sensors for dielectric and optical characterization of skin. *Biosensors and Bioelectronics*, 24(9):2778–2784, 2009.

[103] I. Harman-Boehm, A. Gal, A.M. Raykhman, E. Naidis, and Y. Mayzel. Noninvasive glucose monitoring: increasing accuracy by combination of multi-technology and multi-sensors. *Journal of Diabetes Science and Technology*, 4(3):583, 2010.

[104] Integrity Applications Ltd. GlucoTrack. Available online:. `http://www.integrity-app.com/description.html`. Accessed 31 January 2013.

[105] C.F. Amaral, M. Brischwein, and B. Wolf. Multiparameter techniques for non-invasive measurement of blood glucose. *Sensors and Actuators B: Chemical*, 140(1):12–16, 2009.

[106] A Caduff, E Hirt, Yu Feldman, Z Ali, and L Heinemann. First human experiments with a novel non-invasive, non-optical continuous glucose monitoring system. *Biosensors and Bioelectronics*, 19(3):209–217, 2003.

[107] A Caduff, F Dewarrat, M Talary, G Stalder, L Heinemann, and Yu Feldman. Non-invasive glucose monitoring in patients with diabetes: a novel system based on impedance spectroscopy. *Biosensors and Bioelectronics*, 22(5):598–604, 2006.

[108] T Forst, A Caduff, M Talary, M Weder, M Braendle, P Kann, F Flacke, C Friedrich, and A Pfuetzner. Impact of environmental temperature on skin thickness and microvascular blood flow in subjects with and without diabetes. *Diabetes Technology & Therapeutics*, 8(1):94–101, 2006.

[109] P. Zakharov, MS Talary, and A. Caduff. A wearable diffuse reflectance sensor for continuous monitoring of cutaneous blood content. *Physics in Medicine and Biology*, 54(17):5301, 2009.

[110] MD Dyar, ML Carmosino, EA Speicher, MV Ozanne, SM Clegg, and RC Wiens. Comparison of partial least squares and lasso regression techniques as applied to laser-induced breakdown spectroscopy of geological samples. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 2012.

[111] V. Pomareda, D. Calvo, A. Pardo, and S. Marco. Hard modeling multivariate curve resolution using lasso: application to ion mobility spectra. *Chemometrics and Intelligent Laboratory Systems*, 104(2):318–332, 2010.

[112] R.N. Bergman, Y.Z. Ider, C.R. Bowden, and C. Cobelli. Quantitative estimation of insulin sensitivity. *American Journal of Physiology-Endocrinology And Metabolism*, 236(6):E667, 1979.

[113] C. Cobelli, G.M. Toffolo, C. Dalla Man, M. Campioni, P. Denti, A. Caumo, P. Butler, and R. Rizza. Assessment of β-cell function in humans, simultaneously with insulin sensitivity and hepatic extraction, from intravenous and oral glucose tests. *American Journal of Physiology-Endocrinology And Metabolism*, 293(1):E1–E15, 2007.

[114] B.W. Bequette. A critical assessment of algorithms and challenges in the development of a closed-loop artificial pancreas. *Diabetes Technology & Therapeutics*, 7(1):28–47, 2005.

[115] S. Roweis and Z. Ghahramani. A unifying review of linear gaussian models. *Neural Computation*, 11(2):305–345, 1999.

[116] T. Hastie, R. Tibshirani, and J.H. Friedman. *The elements of statistical learning: data mining, inference, and prediction.* Springer Verlag, 2nd edition, 2009.

[117] R.E. Bellman. *Adaptive control processes: a guided tour.* Princeton University Press, 1961.

[118] B. Kovatchev, S. Anderson, L. Heinemann, and W. Clarke. Comparison of the numerical and clinical accuracy of four continuous glucose monitors. *Diabetes Care*, 31(6):1160–1164, 2008.

[119] AN Tikhonov and VY Arsenin. Solutions of ill-posed problems (Washington, DC: Winston–Wiley). 1977.

[120] W.L. Clarke, D. Cox, L.A. Gonder-Frederick, W. Carter, and S.L. Pohl. Evaluating clinical accuracy of systems for self-monitoring of blood glucose. *Diabetes Care*, 10(5):622–628, 1987.

[121] 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.

[122] S. Sivananthan, V. Naumova, C.D. Man, A. Facchinetti, E. Renard, C. Cobelli, and S.V. Pereverzyev. Assessment of blood glucose predictors: the prediction-error grid analysis. *Diabetes Technology & Therapeutics*, 13(8):787–796, 2011.

[123] S. de Jong. SIMPLS: an alternative approch to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, 18(3):pp. 251–263, 1993.

[124] C.M. Bishop et al. *Pattern recognition and machine learning*, volume 4. Springer New York, 2006.

[125] M. Schmidt, G. Fung, and R. Rosales. Optimization methods for l1-regularization. *University of British Columbia, Technical Report TR-2009-19*, 2009.

[126] W.J. Fu. Penalized regressions: the bridge versus the lasso. *Journal of Computational and Graphical Statistics*, 7(3):397–416, 1998.

[127] S.K. Shevade and S.S. Keerthi. A simple and efficient algorithm for gene selection using sparse logistic regression. *Bioinformatics*, 19(17):2246, 2003.

[128] S. Perkins, K. Lacker, and J. Theiler. Grafting: Fast, incremental feature selection by gradient descent in function space. *The Journal of Machine Learning Research*, 3:1333–1356, 2003.

[129] M.Y. Park and T. Hastie. L_1-regularization path algorithm for generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(4):659–677, 2007.

[130] S. Rosset. Following curved regularized optimization solution paths. In Lawrence K. Saul, Yair Weiss, and Léon Bottou, editors, *Advances in Neural Information Processing Systems 17*, pages 1153–1160. 2005.

[131] A. Galen and G. Jianfeng. Scalable training of L_1-regularized log-linear models. In *Proceedings of the 24th International Conference on Machine learning*, pages 33–40, 2007.

[132] S.I. Lee, H. Lee, P. Abbeel, and A.Y. Ng. Efficient L_1-regularized logistic regression. In *Proceedings of the National Conference on Artificial Intelligence*, volume 21, page 401, 2006.

[133] M.A.T. Figueiredo. Adaptive sparseness for supervised learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9):1050–1159, 2003.

[134] B. Krishnapuram, L. Carin, M.A.T. Figueiredo, and A.J. Hartemink. Sparse multinomial logistic regression: Fast algorithms and generalization bounds. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(6):957–968, 2005.

[135] Y.J. Lee and O.L. Mangasarian. SSVM: A smooth support vector machine for classification. *Computational Optimization and Applications*, 20(1):5–22, 2001.

[136] S.J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky. An Interior-Point Method for Large-Scale L_1-Regularized Least Squares. *Selected Topics in Signal Processing, IEEE Journal of*, 1(4):606–617, 2007.

[137] S.P. Boyd and L. Vandenberghe. *Convex optimization.* Cambridge Univ Pr, 2004.

[138] E.M. Gafni and D.P. Bertsekas. Two-metric projection methods for constrained optimization. *SIAM Journal on Control and Optimization*, 22(6):936–964, 1984.

[139] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *The Annals of Statistics*, 32(2):pp. 407–451, 2004.

[140] H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320, 2005.

[141] J. Friedman, T. Hastie, H. Höfling, and R. Tibshirani. Pathwise coordinate optimization. *The Annals of Applied Statistics*, 1(2):302–332, 2007.

[142] J. Friedman, T. Hastie, and R. Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1, 2010.

[143] A.J. Van Der Kooij et al. *Prediction accuracy and stability of regression with optimal scaling transformations.* Child & Family Studies and Data Theory (AGP-D), Department of Education and Child Studies, Faculty of Social and Behavioural Sciences, Leiden University, 2007.

[144] D.L. Donoho and I.M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90(432):1200–1224, 1995.

[145] T.A. Stamey, JN Kabalin, JE McNeal, IM Johnstone, F. Freiha, EA Redwine, N. Yang, et al. Prostate specific antigen in the diagnosis and treatment of adenocarcinoma of the prostate. ii. radical prostatectomy treated patients. *The Journal of Urology*, 141(5):1076, 1989.

[146] A. Caduff, M. Mueller, A. Megej, F. Dewarrat, R.E. Suri, J. Klisic, M. Donath, P. Zakharov, D. Schaub, W.A. Stahel, et al. Characteristics of a multisensor

system for non invasive glucose monitoring with external validation and prospective evaluation. *Biosensors and Bioelectronics*, 26(9):3794–3800, 2011.

[147] M. Zanon, G. Sparacino, A. Facchinetti, M. Riz, M.S. Talary, R.E. Suri, A. Caduff, and C. Cobelli. Non-invasive continuous glucose monitoring: improved accuracy of point and trend estimates of the multisensor system. *Medical and Biological Engineering and Computing*, 50(10):1047–1057, 2012.

[148] A.M.K. Enejder, T.G. Scecina, J. Oh, M. Hunter, W.C. Shih, S. Sasic, G.L. Horowitz, and M.S. Feld. Raman spectroscopy for noninvasive glucose measurements. *Journal of Biomedical Optics*, 10(3):031114–031114, 2005.

[149] M.A. Arnold, L. Liu, J.T. Olesberg, et al. Selectivity assessment of noninvasive glucose measurements based on analysis of multivariate calibration vectors. *Journal of Diabetes Science and Technology*, 1(4):454–462, 2007.

[150] S. Guerra, G. Sparacino, A. Facchinetti, M. Schiavon, C.D. Man, and C. Cobelli. A dynamic risk measure from continuous glucose monitoring data. *Diabetes Technology & Therapeutics*, 13(8):843–852, 2011.

[151] 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(11):1655–1658, 1997.

[152] C. Tronstad, G.K. Johnsen, S. Grimnes, and Ø.G. Martinsen. A study on electrode gels for skin conductance measurements. *Physiological Measurement*, 31(10):1395, 2010.

[153] T. Zueger, P. Diem, S. Mougiakakou, and C. Stettler. Influence of time point of calibration on accuracy of continuous glucose monitoring in individuals with type 1 diabetes. *Diabetes Technology & Therapeutics*, 14(7):583–588, 2012.

[154] S. Del Favero, A. Facchinetti, and C. Cobelli. A glucose-specific metric to assess predictors and identify models. *Biomedical Engineering, IEEE Transactions on*, 59(5):1281–1290, 2012.

Acknowledgements

I would like to thank all the people for the support, numerous discussions and inspirational comments during this PhD program, in particular my advisor Prof. Giovanni Sparacino and the Solianis Monitoring AG team.

I also would like to thank colleagues and office mates for useful discussions and comments. Amongst many others, I would like to thanks especially my Friends, whose names will not be mentioned here because they already have a spot in my heart.

Last but not least, I would like to thank my family for unconditional support and endless patience with me.

The work in this thesis has been supported by Solianis Monitoring AG and the University of Padova.