UNIVERSITY OF PADOVA

DEPARTMENT OF INFORMATION ENGINEERING
Ph.D. Course on Information Engineering
Curriculum: Bioengineering
Series: XXXI

Calibration of continuous glucose monitoring sensors by time-varying models and Bayesian estimation

Course director:
Prof. Andrea Neviani
Advisor:
Prof. Giovanni Sparacino
Co-advisor:
Prof. Andrea Facchinetti

Ph.D. candidate:
Giada Acciaroli

September 2018

Acknowledgements

While completing this Ph.D. thesis I would like to thank a few persons that, in different ways, helped and guided me through these three years of studies.

Firstly, I thank my advisor, Prof. Giovanni Sparacino, for his precious and constant guidance and most of all for having believed in my talent and encouraged me to undertake a career in research.

I thank my co-advisor, Prof. Andrea Facchinetti, for his valuable help and support.

I thank all my colleagues at the University of Padova, with a special mention to my loved office mates at the GANZI lab.

I thank my family for their constant support and for being always by my side.

Finally, I thank Dexcom Inc. (San Diego, CA) for providing the data used in this Ph.D. thesis and for the valuable research collaboration carried out in these years.

Abstract

Minimally invasive continuous glucose monitoring (CGM) sensors are wearable medical devices that provide frequent (e.g., 1-5 min sampling rate) real-time measurements of glucose concentration for several consecutive days. This can be of great help in the daily management of diabetes. Most of the CGM systems commercially available today have a wire-based electrochemical sensor, usually placed in the subcutaneous tissue, which measures a "raw" electrical current signal via a glucose-oxidase electrochemical reaction. Observations of the raw electrical signal are frequently revealed by the sensor on a fine, uniformly spaced, time grid. These samples of electrical nature are in real-time converted to interstitial glucose (IG) concentration levels through a calibration process by fitting a few blood glucose (BG) concentration measurements, sparsely collected by the patient through fingerprick. Usually, for coping with such a process, CGM sensor manufacturers employ linear calibration models to approximate, albeit in limited time-intervals, the nonlinear relationship between electrical signal and glucose concentration. Thus, on the one hand, frequent calibrations (e.g., two per day) are required to guarantee a good sensor accuracy. On the other, each calibration requires patients to add uncomfortable extra actions to the many already needed in the routine of diabetes management.

The aim of this thesis is to develop new calibration algorithms for minimally invasive CGM sensors able to ensure good sensor accuracy with the minimum number of calibrations. In particular, we propose i) to replace the time-invariant gain and offset conventionally used by the linear calibration models with more sophisticated time-varying functions valid for multiple-day periods, with unknown model parameters for which an a priori statistical description is available from independent training sets; ii) to numerically estimate the calibration model parameters by means of a Bayesian estimation procedure that exploits the a priori information on model parameters in addition to some BG samples sparsely collected by the patient.

The thesis is organized in 6 chapters. In Chapter 1, after a background introduction on CGM sensor technologies, the calibration problem is illustrated. Then, some state-of-art calibration techniques are briefly discussed with their open problems, which result in the aims of the thesis illustrated at the end of the chapter.

In Chapter 2, the datasets used for the implementation of the calibration techniques are described, together with the performance metrics and the statistical analysis tools which will be employed to assess the quality of the results.

In Chapter 3, we illustrate a recently proposed calibration algorithm (Vettoretti et al., IEEE Trans Biomed Eng 2016), which represents the starting point of the study proposed in this thesis. In particular, we demonstrate that, thanks to the development of a time-varying day-specific Bayesian prior, the algorithm can become able to reduce the calibration frequency from two to one per day. However, the linear calibration model used by the algorithm has domain of validity limited to certain time intervals, not allowing to further reduce calibrations to less then one per day and calling for the development of a new calibration model valid for multiple-day periods like that developed in the remainder of this thesis.

In Chapter 4, a novel Bayesian calibration algorithm working in a multi-day framework (referred to as Bayesian multi-day, BMD, calibration algorithm) is presented. It is based on a multiple-day model of sensor time-variability with second order statistical priors on its unknown parameters. In each patient-sensor realization, the numerical values of the calibration model parameters are determined by a Bayesian estimation procedure exploiting the BG samples sparsely collected by the patient. In addition, the distortion introduced by the BG-to-IG kinetics is compensated during parameter identification via non-parametric deconvolution. The BMD calibration algorithm is applied to two datasets acquired with the "present-generation" Dexcom (Dexcom Inc., San Diego, CA) G4 Platinum (DG4P) CGM sensor and a "next-generation" Dexcom CGM sensor prototype (NGD). In the DG4P dataset, results show that, despite the reduction of calibration frequency (on average from 2 per day to 0.25 per day), the BMD calibration algorithm significantly improves sensor accuracy compared to the manufacturer calibration algorithm. In the NGD dataset, performance is even better than that of present generation, allowing to further reduce calibrations toward zero.

In Chapter 5, we analyze the potential margins for improvement of the

BMD calibration algorithm and propose a further extension of the method. In particular, to cope with the inter-sensor and inter-subject variability, we propose a multi-model approach and a Bayesian model selection framework (referred to as multi-model Bayesian framework, MMBF) in which the most likely calibration model is chosen among a finite set of candidates. A preliminary assessment of the MMBF is conducted on synthetic data generated by a well-established type 1 diabetes simulation model. Results show a statistically significant accuracy improvement compared to the use of a unique calibration model.

Finally, the major findings of the work carried out in this thesis, possible applications and margins for improvement are summarized in Chapter 6.

Sommario

I sensori minimamente invasivi per il monitoraggio in continua della glicemia, indicati con l'acronimo CGM (continuous glucose monitoring), sono dei dispositivi medici indossabili capaci di misurare la glicemia in tempo reale, ogni 1-5 minuti, per più giorni consecutivi. Questo tipo di misura fornisce un profilo di glicemia quasi continuo che risulta essere un'informazione molto utile per la gestione quotidiana della terapia del diabete. La maggior parte dei dispositivi CGM ad oggi disponibili nel mercato dispongono di un sensore di tipo elettrochimico, solitamente inserito nel tessuto sottocutaneo, che misura una corrente elettrica generata dalla reazione chimica di glucosio-ossidasi. Le misure di corrente elettrica sono fornite dal sensore con campionamento uniforme ad elevata frequenza temporale e vengono convertite in tempo reale in valori di glicemia interstiziale attraverso un processo di calibrazione. La procedura di calibrazione prevede l'acquisizione da parte del paziente di qualche misura di glicemia plasmatica di riferimento tramite dispositivi pungidito. Solitamente, le aziende produttrici di sensori CGM implementano un processo di calibrazione basato su un modello di tipo lineare che approssima, sebbene in intervalli di tempo di durata limitata, la più complessa relazione tra corrente elettrica e glicemia. Di conseguenza, si rendono necessarie frequenti calibrazioni (per esempio, due al giorno) per aggiornare i parametri del modello di calibrazione e garantire una buona accuratezza di misura. Tuttavia, ogni calibrazione prevede l'acquisizione da parte del paziente di misure di glicemia tramite dispositivi pungidito. Questo aumenta la già numerosa lista di azioni che i pazienti devono svolgere quotidianamente per gestire la loro terapia.

Lo scopo di questa tesi è quello di sviluppare un nuovo algoritmo di calibrazione per sensori CGM minimamente invasivi capace di garantire una buona accuratezza di misura con il minimo numero di calibrazioni. Nello specifico, si propone i) di sostituire il guadagno ed offset tempo-invarianti solitamente utilizzati nei modelli di calibrazione di tipo lineare con delle funzioni tempo-varianti, capaci di descrivere il comportamento del sensore per inter-

valli di tempo di più giorni, e per cui sia disponibile dell'informazione a priori riguardante i parametri incogniti; ii) di stimare il valore numerico dei parametri del modello di calibrazione con metodo Bayesiano, sfruttando l'informazione a priori sui parametri di calibrazione in aggiunta ad alcune misure di glicemia plasmatica di riferimento.

La tesi è organizzata in 6 capitoli. Nel Capitolo 1, dopo un'introduzione sulle tecnologie dei sensori CGM, viene illustrato il problema della calibrazione. In seguito, vengono discusse alcune tecniche di calibrazione che rappresentano lo stato dell'arte ed i loro problemi aperti, che risultano negli scopi della tesi descritti alla fine del capitolo.

Nel Capitolo 2 vengono descritti i dataset utilizzati per l'implementazione delle tecniche di calibrazione. Ioltre, vengono illustrate le metriche di accuratezza e le tecniche di analisi statistica utilizzate per analizzare la qualità dei risultati.

Nel Capitolo 3 viene illustrato un algoritmo di calibrazione recentemente proposto in letteratura (Vettoretti et al., IEEE, Trans Biomed Eng 2016). Questo algoritmo rappresenta il punto di partenza dello studio svolto in questa tesi. Più precisamente, viene dimostrato che, grazie all'utilizzo di un prior Bayesiano specifico per ogni giorno di utilizzo, l'algoritmo diventa efficace nel ridurre le calibrazioni da due a una al giorno senza perdita di accuratezza. Tuttavia, il modello lineare di calibrazione utilizzato dall'algoritmo ha dominio di validità limitato a brevi intervalli di tempo tra due calibrazioni successive, rendendo impossibile l'ulteriore riduzione delle calibrazioni a meno di una al giorno senza perdita di accuratezza. Questo determina la necessità di sviluppare un nuovo modello di calibrazione valido per intervalli di tempo più estesi, fino a più giorni consecutivi, come quello sviluppato nel resto di questa tesi.

Nel Calitolo 4 viene presentato un nuovo algoritmo di calibrazione di tipo Bayesiano (Bayesian multi-day, BMD). L'algoritmo si basa su un modello della tempo-varianza delle caratteristiche del sensore nei suoi giorni di utilizzo e sulla disponibilità di informazione statistica a priori sui suoi parametri incogniti. Per ogni coppia paziente-sensore, il valore numerico dei parametri del modello è determinato tramite stima Bayesiana sfruttando alcune misure plasmatiche di riferimento acquisite dal paziente con dispositivi pungidito. Inoltre, durante la stima dei parametri, la dinamica introdotta dalla cinetica plasma-interstizio viene compensata tramite deconvoluzione nonparametrica. L'algoritmo di calibrazione BMD viene applicato a due differenti set di dati acquisiti con il sensore commerciale Dexcom (Dexocm Inc., San Diego, CA) G4

Platinum (DG4P) e con un prototipo di sensore Dexcom di nuova generazione (NGD). Nei dati acquisiti con il sensore DG4P, i risultati dimostrano che, nonostante le calibrazioni vengano ridotte (in media da 2 al giorno a 0.25 al giorno), l' algoritmo BMD migliora significativamente l'accuratezza del sensore rispetto all'algoritmo di calibrazione utilizzato dall'azienda produttrice del sensore. Nei dati acquisiti con il sensore NGD, i risultati sono ancora migliori, permettendo di ridurre ulteriormente le calibrazioni fino a zero.

Nel Capitolo 5 vengono analizzati i potenziali margini di miglioramento dell'algoritmo di calibrazione BMD discusso nel capitolo precedente e viene proposta un'ulteriore estensione dello stesso. In particolare, per meglio gestire la variabilità tra sensori e tra soggetti, viene proposto un approccio di calibrazione multi-modello e un metodo Bayesiano di selezione del modello (Multi-model Bayesian framework, MMBF) in cui il modello di calibrazione più probabile a posteriori viene scelto tra un set di possibili candidati. Tale approccio multi-modello viene analizzato in via preliminare su un set di dati simulati generati da un simulatore del paziente diabetico di tipo 1 ben noto in letteratura. I risultati dimostrano che l'accuratezza del sensore migliora in modo significativo con MMBF rispetto ad utilizzare un unico modello di calibrazione.

Infine, nel Calitolo 6 vengono riassunti i principali risultati ottenuti in questa tesi, le possibili applicazioni, e i margini di miglioramento per gli sviluppi futuri.

Contents

Abbreviations

AD Absolute difference

AR Auto regressive

ARD Absolute relative difference

BMD Bayesian multi-day

BG Blood glucose

CEG Clarke error grid

CGM Continuous glucose monitoring

DG4P Dexcom G4 Platinum

FDA Food and Drug Administration

IG Interstitial glucose

IQR Inter quartile range

ISO International Organization for Standardization

MAD Mean absolute difference

MAP Maximum a posteriori

MARD Mean absolute relative difference

MRD Mean relative difference

MMBF Multi-model Bayesian framework

NGD Next generation Dexcom

PAGE Percentage of accurate glucose estimates

RMSE Root mean square error

RD Relative difference

SD Standard deviation

SMBG Self-monitoring of blood glucose

T1D Type 1 diabetes

T1DS Type 1 diabetes simulator

T2D Type 2 diabetes

WHO World Health Organization

YSI Yellow Springs Instrument

Chapter 1

Continuous glucose monitoring (CGM) sensors and the problem of their calibration: state-of-art and open issues[1]

1.1 The diabetes disease and its management

Diabetes is a chronic disorder that occurs either when the pancreas is no longer able to produce insulin (type 1 diabetes, T1D), or if body tissues and organs cannot effectively utilize circulating insulin (type 2 diabetes, T2D) [3]. In people with diabetes, the deficiencies in glucose control lead to blood glucose (BG) values exceeding the safe range of 70-180 mg/dl. While hyperglycemia (i.e., BG > 180 mg/dl) can result in long-term complications, e.g., retinopathy, nephropathy, cardiovascular disease, hypoglycemia (i.e., BG < 70 mg/dl) can produce short-term adverse conditions that can cause coma, or even death [4] [5] [6].

Nowadays diabetes afflicts more than 350 millions of people worldwide. T2D accounts for about 90% of all cases [7] and its onset is largely correlated with excessive body weight, physical inactivity, and unhealthy diet. The World Health Organization (WHO) predicts T2D prevalence to significantly grow in the next years, due to aging populations and sedentary lifestyles [3] [8]. Consequently, the total cases of diabetes is expected to exceed 500 million by 2030, becoming one of the most challenging socio-health emergencies of the third

[1] Part of this chapter is taken from the published review papers [1] and [2].

millennium. Although it is not possible with current knowledge to definitely cure diabetes, a constant and appropriate management of the disease can control and prevent many complications [3] [9] [10]. While T2D management mainly consists of healthy diet, physical exercise, and drug administration, T1D therapy requires daily insulin administration. The most challenging issue is related to insulin administration, being the correct amount of injected insulin determined on the basis of BG concentration levels, which needs to be monitored by suitable technologies.

At-home BG monitoring became available only in the 1970s, when the first self-monitoring BG (SMBG) meters were commercialized [11]. The early SMBG portable devices were based on optical detection of a color change on glucose oxidase-based strips, while the most recent systems rely on electrochemical-based sensing techniques [12] becoming, by the mid-1980s, the landmark in diabetes management [11]. Since the advent of SMBG devices for home BG testing, the standard therapy for diabetes management consisted of 3–4 SMBG measurements per day, individually acquired by fingerprick tests [3]. Although the introduction of self-BG monitoring in everyday life resulted in a general improvement of metabolic control [13] [14] [15], sparse BG measures cannot provide a complete description of glucose dynamics during the day. For instance, hypoglycemic or hyperglycemic events occurring between two BG acquisitions cannot be detected. An example of SMBG time series (data derive from [16]) is depicted in Figure 1.1 (triangles), where it is apparent that some critical episodes, e.g., a hypoglycemia event and several hyperglycemic conditions, cannot be revealed due to insufficient sampling frequency. On the one hand, more frequent SMBG measurements would be required to optimize glucose control but, on the other hand, this would increase patient's discomfort and increment the number of actions needed daily to manage the disease.

In recent years, technological innovations have been introduced for the treatment of diabetes. In particular, the development of continuous glucose monitoring (CGM) sensors [17] [1] have revolutionized diabetes management. CGM systems are wearable devices able to measure subcutaneous glucose concentration almost continuously, e.g., every 1–5 minutes [18], for several consecutive days, greatly increasing the information on glucose dynamics compared to standard SMBG-based monitoring, with consequent improvement of glycemic control, quality of life, and reduction of diabetes-related complications [19] [20] [21] [22] [23] [24] [25] [26]. See, for instance, in Figure 1.1 (continuous line) the hypo- and hyperglycemic episodes detected by the Dex-

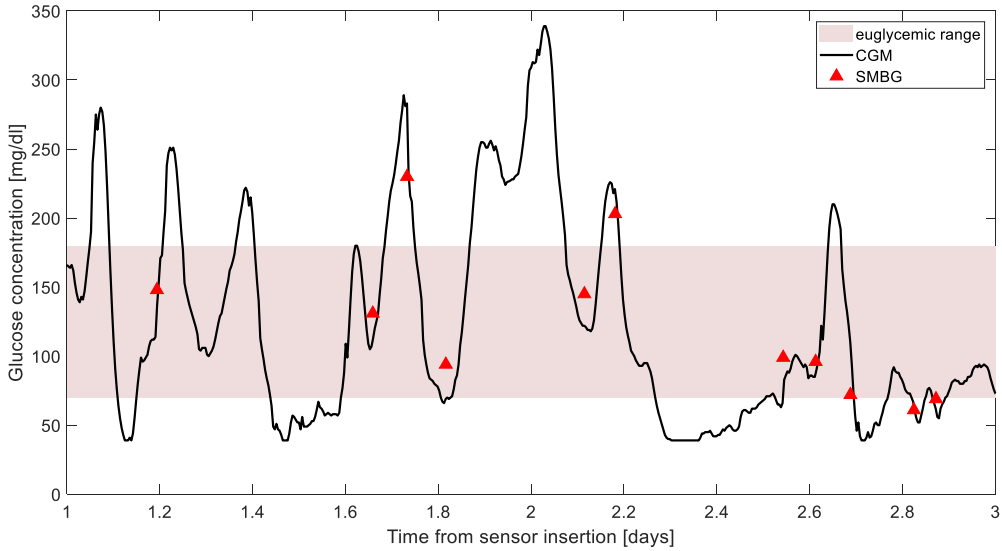

Figure 1.1: Representative two-days of blood glucose (BG) monitoring obtained with self-monitoring BG (SMBG), triangles, and with continuous glucose monitoring (CGM), continuous line. The pink band represents the euglycemic range.

com G4 Platinum (Dexcom Inc., San Diego, CA) CGM sensor (data previously published in [16]) and otherwise undetectable with the standard SMBG measurements. This high frequency information on glucose dynamics provided by CGM sensors is a key component for several open loop applications, such, e.g., sensor-augmented pump therapies [27] and decision support systems [28], [29], and closed loop applications like the artificial pancreas (see, e.g., [30], [31], [32], [33], [34], [35], [36], or [37] for an extensive review).

The following section gives a more detailed description of the CGM sensor technologies, from the early age devices to the products currently available on the market.

1.2 CGM sensor technologies

In the last years, various glucose-sensing mechanisms for non-invasive, or at least minimally invasive, CGM have been proposed [38] [12] [39] [40] [41] [42] [43], trying to match all fundamental requirements for an extended in vivo use, e.g., sensitivity, specificity, linearity within biological relevant range, biocompatibility, and lifetime [44]. Among all the proposed techniques, i.e. electrochemical, optical, and piezoelectric, the one that is today exploited by most of the commercialized CGM systems is the glucose-oxidase electrochemical prin-

Figure 1.2: Example of patient wearing a CGM sensor in the abdomen. (Source: https://www.medicalexpo.com)

ciple [12]. The devices based on this principle employ a minimally-invasive wire-based sensor, usually inserted through a thin needle in the subcutaneous tissue of the abdomen or the arm (see Figure 1.2), which measures an electrical current signal generated by the glucose-oxidase reaction. This signal is proportional to the glucose concentration in the interstitial fluid. The process that converts the electrical current signal into a glucose concentration estimate is referred to as a calibration procedure.

1.2.1 The early age of glucose-oxidase CGM sensors

The first prototypes of CGM systems based on the glucose-oxidase electrochemical principle were proposed in the late 90's [45], but biocompatibility problems often led to commercially unsuccessful products. The first generation of commercial CGM systems became available for personal use only a few years later, starting from 2005. The first three successfully commercialized products were the Medtronic Guardian (Medtronic Minimed, Northridge, CA), the Dexcom Seven Plus (Dexcom Inc., San Diego, CA) and the Abbott Navigator (Abbott Diabetes Care, Alameda, CA).

Accuracy of these systems is commonly computed by comparing the CGM

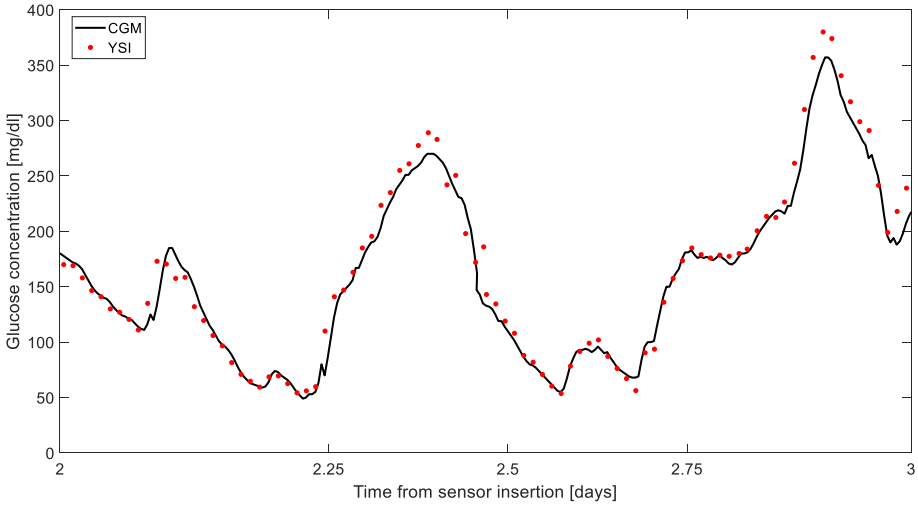

Figure 1.3: Example of Yellow Spring Instrument (YSI) measurements (red points) versus CGM data (continuous line) collected during an in-clinic session of about 1 day. Accuracy of the CGM sensor can be quantified by using the YSI measurements as ground truth, e.g., by computing the mean absolute relative difference (MARD) as the average of the ratio between the YSI-CGM absolute difference over the YSI.

sensor output with a reference BG measurement acquired with highly accurate laboratory instruments, i.e., Yellow Spring Instrument Inc. (YSI) instruments. An example of dataset provided to assess accuracy of a CGM sensor is reported in Figure 1.3. Of note is that high frequency YSI references (one every 15 minutes in the picture) can be collected only in a hospital setting. Along the different metrics that have been proposed in the literature to quantify accuracy of CGM sensors, e.g., mean absolute difference (MAD), mean absolute relative difference (MARD), and Clarke error grid analysis [46], MARD is the most popular accuracy index used in recent studies.

From a number of datasets like that of Figure 1.3 it is trivial to estimate the MARD as the average of the ratio between the YSI-CGM absolute difference over the YSI. For instance, the MARD of the first commercialized sensors was 15.8% for the Medtronic Guardian [47], 16.7% for the Dexcom Seven Plus [48], and 12.8% for the Abbott Navigator [49], for a mean lifetime of 4-5 days.

As for any other medical devices, accuracy represents one of the most important requirements, since critical (and possibly harmful for the patient) therapeutic decisions and actions, like insulin dosing, are based on BG readings. At the end of the past decade, accuracy values of commercial grade CGM sensors were still significantly worse than those of SMBG systems (that have a

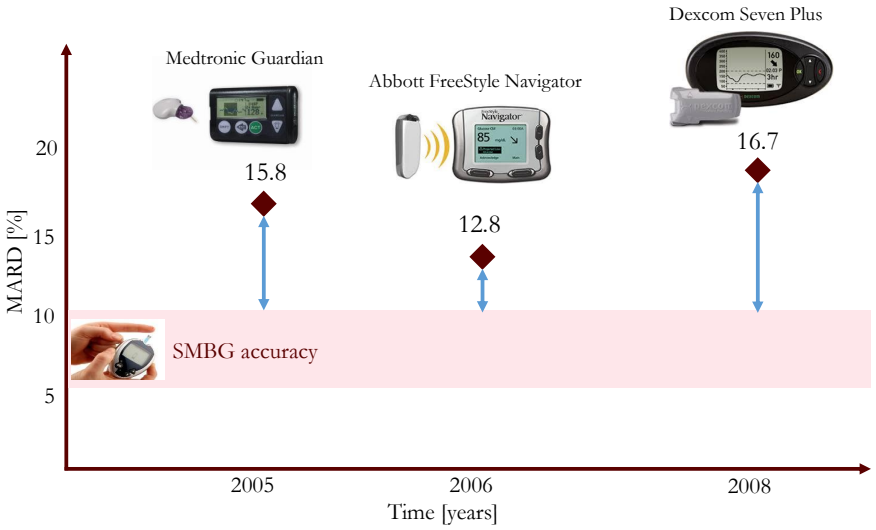

Figure 1.4: Early-age CGM devices commercialized between 2005 and 2008 and schematic representation of their accuracy, measured as MARD. From left to right: Medtronic Guardian (Medtronic Minimed, Northridge, CA), Abbott FreeStyle Navigator (Abbott Diabetes Care, Alameda, CA), and Dexcom Seven Plus (Dexcom Inc., San Diego, CA). Accuracy range of standard SMBG devices (5-10% MARD) is also depicted in pink for comparison.

MARD between 5 and 10%), as represented in Figure 1.4, making therefore CGM sensors still unsafe for taking therapeutic decisions. Thus, in the past 10 years, CGM companies concentrated a big effort in developing improved, more accurate, sensor systems.

1.2.2 State-of-art glucose-oxidase CGM sensors

State-of-art CGM devices are depicted in a timeline in Figure 1.5, where their accuracy is also schematized in terms of MARD, compared to the accuracy of standard SMBG devices.

In 2011, Medtronic launched the new Enlite sensor, which received the CE approval in 2011 and the Food and Drug Administration (FDA) approval in 2013, providing users with a more accurate and comfortable sensor. Indeed, with respect to the previous first generation device, the Enlite sensor shows significant clinical and human factors improvements. The electrode layout was redesigned to face inflammatory response effects, the size of the sensor decreased, the wear time was extended from 3 to 6 days, and the inserter device became more user-friendly. All these changes improved usability and accuracy, reaching 13.6% MARD [50]. In 2016, the same company proposed a

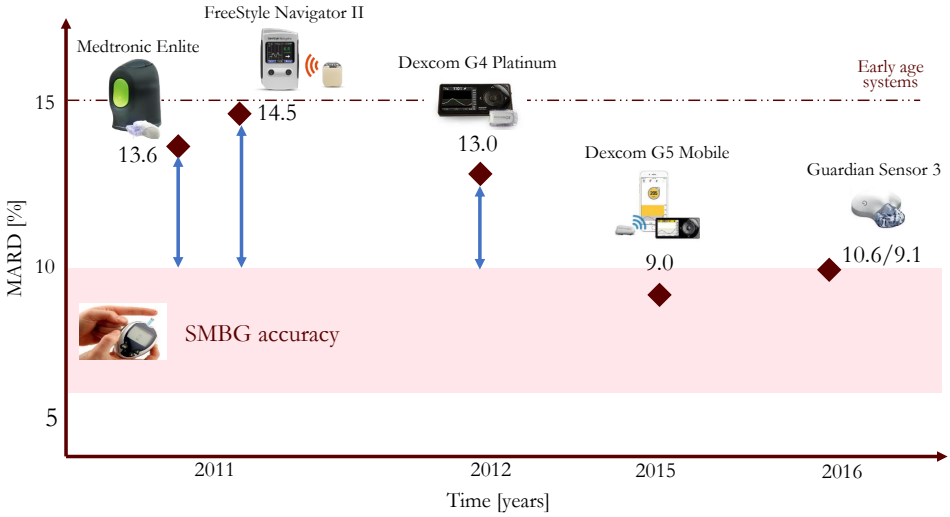

Figure 1.5: State of art glucose sensor devices commercialized between 2011 and 2016 and schematic representation of their accuracy, measured as MARD. From left to right: Medtronic Enlite (Medtronic Minimed, Northridge, CA), Abbott FreeStyle Navigator II (Abbott Diabetes Care, Alameda, CA), Dexcom G4 Platinum, Dexcom G5 Mobile (Dexcom Inc., San Diego, CA), and Guardian Sensor 3 (Medtronic Minimed, Northridge, CA). Accuracy range of standard SMBG devices (5-10% MARD) is also depicted in pink for comparison.

new and more accurate sensor, the Guardian Sensor 3 [51], which has been recently FDA approved to be used in the Minimed 670G hybrid closed loop system [52].

Dexcom launched the new G4 Platinum CGM sensor in 2012, after receiving CE Mark in June and FDA approval in October of the same year. The new Dexcom G4 Platinum system, which lasts for seven days, employs a completely redesigned smaller transmitter and several sensor improvements that allowed enhancing accuracy to 13% MARD [16]. In 2014, algorithmic changes in the Dexcom G4 Platinum sensor further enhanced accuracy decreasing MARD to 9% [53]. In addition to accuracy improvements, Dexcom pushed to render their new products more effective and user-friendly by embedding the G4 Platinum sensor with the Share technology (2015) [54]. The Share technology allows secure wireless connection via Bluetooth Low Energy between a patient's receiver and an app on the patient's smartphone and up to five designated recipients. In the same direction, in 2015, Dexcom launched the G5 Mobile CGM system that allows direct wireless communication to a smartphone without the need to have a dedicated receiver [54].

Abbott launched the FreeStyle Navigator II CGM system in 2011 in some

European countries. The new sensor, which lasts for five days, shows a completely redesigned receiver, a smaller transmitter, a slightly smaller and redesigned sensor that allowed reducing warm-up time from 10 to 1 h [55].

1.2.3 Technological trends and challenges for the next generation CGM sensors

From an hardware point of view, the size and lifetime of a CGM system represent critical factors in device development. In general, among a series of equally performing devices, the smallest and more long-lasting product often has the greatest appeal and success. It is reasonable to extend this general concept to CGM devices, expecting that next generation CGM systems will be small, easy to wear, and possibly guarantee longer duration.

Many companies are currently working to match these fundamental requirements by employing different strategies. In August 2015, Dexcom announced the agreement with the Life Science team at Google (Google Inc., Mountain View, CA) to develop jointly a series of next generation CGM products designed to be smaller, cheaper, and lasting for 10–14 consecutive days [54]. On the other hand, in 2014 Abbott already launched in Europe the new FreeStyle Libre flash glucose monitoring system (see Figure 1.6), which can be worn for 14 consecutive days showing glucose levels when scanned by the user with an accuracy of 11.4% MARD [56] (accuracy reported in the scheme of Figure 1.5). Also, in 2016 Senseonics Inc. (Germantown, MD) launched the Eversense sensor, the first fully implantable CGM system based on optical sensor technology that lasts for 90 consecutive days and has a MARD of 11.4% [57] (see Figure 1.6).

Beyond the requirements just discussed in terms of sensor lifetime, size, and user acceptability, which can be achieved mostly by hardware adjustments, next generation CGM systems need to address several other requirements through software updates. Currently available CGM systems already incorporate smart features, such as alerts and alarms generation in case of hypoglycemic or hyperglycemic events, incorporation of glucose trend information, detection of sensor faults and artifacts [58], [38]. Next generation systems certainly need to incorporate these features and to deal with data management and integration into external devices, e.g., smartphone apps or dedicated cloud platforms.

Another crucial aspect that needs to be considered to improve the ease of

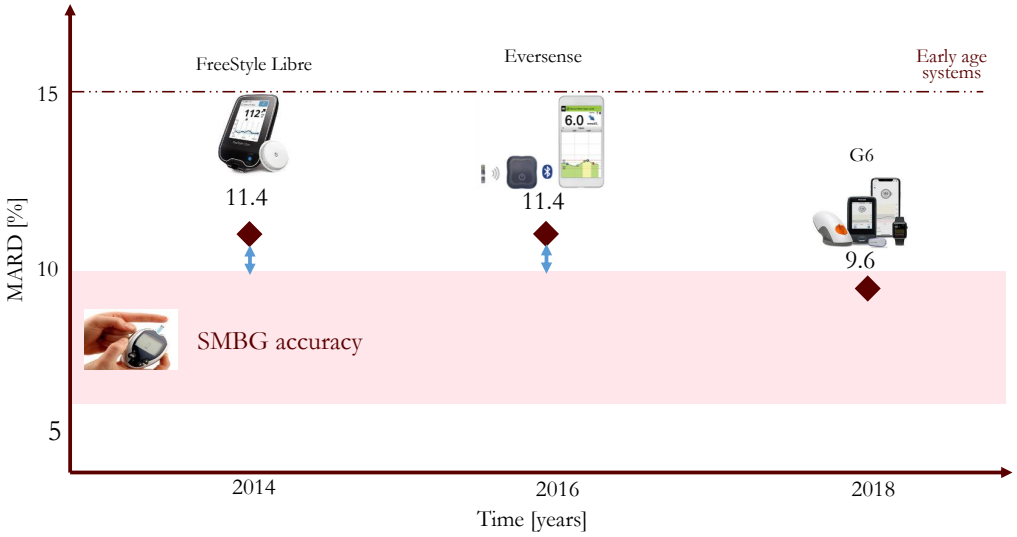

Figure 1.6: Next generation glucose sensor devices and schematic representation of their accuracy, measured as MARD. From left to right: Abbott FreeStyle Libre (Abbott Diabetes Care, Alameda, CA), Eversense (Senseonics Inc., Germantown, MD), and Dexcom G6 (Dexcom Inc., San Diego, CA). Accuracy range of standard SMBG devices (5-10% MARD) is also depicted in pink for comparison.

use of next generation sensors is calibration requirement. Most of the commercially available glucose sensors need to be calibrated to convert the raw measurements (e.g., the electrical current) to glucose values, usually twice per day, requiring the collection of one or more SMBG measurements to use as ground truth [58]. In order to facilitate the use of CGM technology, next generation systems need to be less calibration dependent. On the other hand, sensor accuracy needs to be maintained despite the reduction of calibration requirements. These apparently conflicting demands can be managed by developing ad hoc calibration techniques.

The very last FDA approved CGM sensor that matches many of the aforementioned requirements of next generation systems, showing a 9.6% MARD with no calibrations, is the Dexcom G6 (Dexcom Inc., San Diego, CA) [59] (see Figure 1.6), which starts now, at the time of writing, to be commercialized.[2]

[2] All the devices reported in the timeline of Figure 1.5 and Figure 1.6 between 2015 and 2018 were not yet commercialized at the beginning of this Ph.D. research program.

1.3 Calibration of minimally invasive glucose-oxidase CGM sensors

Minimally invasive CGM sensors measure a signal that reflects glucose concentration only indirectly. Indeed, the wired-based sensor placed through a needle in the subcutaneous tissue measures a current signal derived from the glucose-oxidase electrochemical reaction [12], [60]. The calibration process consists in the estimation of a mathematical model that converts the current signal (given in fractions of ampere) into meaningful glucose concentration values (in mg/dl). The following subsections illustrate the calibration problem and the critical aspects related to calibration.

1.3.1 Calibration problem statement

Letting $u(t)$ be the glucose concentration profile, $y(t)$ the electrical current profile, and $f(\cdot)$ the function of parameters $\mathbf{P} = [p_1, p_2, ..., p_n]$ that relates $u(t)$ and $y(t)$, the calibration process can be schematized in two steps (schematized in panel (a) and (b) of Figure 1.7). The first step consists in the identification of the calibration model parameters. In formal terms, the current signal $y(t)$ collected by the sensor and corrupted by measurement error $w(t)$, and the BG measurements (samples of $u(t)$) acquired by the patient at correspondent time instants, are described by the model (see Figure 1.7, panel (a)):

$$y(t) = f(\mathbf{P}, u(t)) + w(t) \tag{1.1}$$

from which a numerical value $\hat{\mathbf{P}}$ of the calibration parameter vector can be provided using, for instance, parametric estimation techniques. This step can be repeated each time a new BG reference is available, with consequent updates of the calibration parameters $\hat{\mathbf{P}}$ (e.g., every 12–24 h by acquiring SMBG samples). The second step leads to the estimation of the glucose concentration profile (see Figure 1.7, panel (b)). Formally, from the vector of estimated parameters $\hat{\mathbf{P}}$ and the measured current profile $y(t)$, the calibrated glycemic profile $\hat{u}(t)$ is obtained in real-time by inverting the calibration function $f(\cdot)$:

$$\hat{u}(t) = f^{-1}(\hat{\mathbf{P}}, y(t)) \tag{1.2}$$

The choice of the calibration function $f(\cdot)$ is critical. It has to be invertible and it has to precisely describe the relationship between the electrical cur-

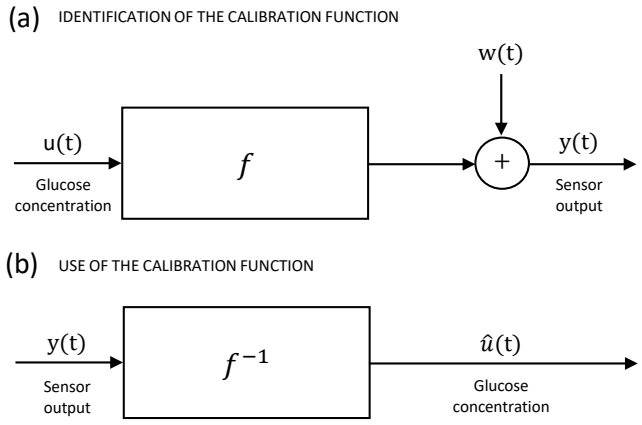

Figure 1.7: Schematic representation of the two steps characterizing the calibration process. Panel (a): first step, identification of the parameters of the calibration function f from the glucose concentration samples, $u(t)$ and the sensor measurements, $y(t)$, corrupted by measurement noise, $w(t)$. Panel (b): second step, estimation of the glucose concentration profile, $\hat{u}(t)$, from the sensor measurements, $y(t)$, by inverting the calibration function.

rent signal and glucose concentration, which can be, in the most general case, non-linear and time-variant (in this case, time t would be, explicitly, an input of $f(\cdot)$). Moreover, the choice of using either the electrical current or the BG measurements as independent variable in the calibration model may affect the calibration performance [61].

The most common and simplest calibration model adopted by manufacturers of CGM systems is a first-order time-independent linear function [62], [63], [64], [65], with parameters $\mathbf{P} = [s, b]$, where s and b are referred to as sensor sensitivity and baseline (or offset), respectively. In this case, the model of the measurements reported in Equation 1.1 in a general form turns into:

$$y(t) = f(\mathbf{P}, u(t)) + w(t) = s \cdot u(t) + b + w(t) \tag{1.3}$$

The numerical determination of the estimates $\hat{s}$ and $\hat{b}$ is, thus, required. For such a scope, if for instance two BG references $u(t_1)$ and $u(t_2)$ are available at times t_1 and t_2, knowing the electrical current values given by the sensor at the same time instants, $y(t_1)$ and $y(t_2)$, the so-called two-point calibration can be performed [66], which allows the estimation of sensitivity, $\hat{s}$, and baseline, $\hat{b}$,

from the two measured pairs as:

$$\begin{cases} \hat{s} = \frac{y(t_2)-y(t_1)}{u(t_2)-u(t_1)} \\ \hat{b} = y(t_2) - \frac{y(t_2)-y(t_1)}{u(t_2)-u(t_1)} \cdot u(t_2) \end{cases} \tag{1.4}$$

In general, when multiple pairs of electrical current and BG samples are available at times t_i ($i = 1, 2, ..., N$), a linear regression is used to fit the sensitivity and baseline to the data. In particular, including the measurement noise $w(t_i)$, the model of the measurements becomes:

$$y(t_i) = s \cdot u(t_i) + b + w(t_i) \tag{1.5}$$

and the numerical determination of model parameters is done by minimizing the residual sum of squares:

$$[\hat{s}, \hat{b}] = \arg\min_{s,b} \sum_{i=1}^{N} (y(t_i) - s \cdot u(t_i) - b)^2 \tag{1.6}$$

Finally, the calibrated glucose profile $\hat{u}(t)$ is obtained from the measured current signal $y(t)$ and the estimated calibration parameters $\hat{s}$ and $\hat{b}$ by inverting the calibration function:

$$\hat{u}(t) = \frac{y(t) - \hat{b}}{\hat{s}} \tag{1.7}$$

The quality of the estimate of the calibration parameters is expected to increase with N, i.e., the more electrical current-BG pairs that are available, the more accurately the calibration parameters are estimated. On the other hand, increasing N is difficult to satisfy, for practical reasons, e.g., for the discomfort related to the acquisition of SMBG samples, and because CGM manufacturers push to minimize the calibration points to facilitate the ease-of-use of their devices. Moreover, in the presence of measurement uncertainty and/or when only a few data points are available, the standard two-point calibration of Equation 1.4 could be simplified to a one-point calibration by considering a zero baseline. This simplification may improve the calibration performance by reducing the effect of the noise [66], [67].

Although the use of such linear calibration techniques is appealing for its simplicity and ease of implementation, it introduces several critical aspects that, together with uncertainty on the measured sensor output and BG ref-

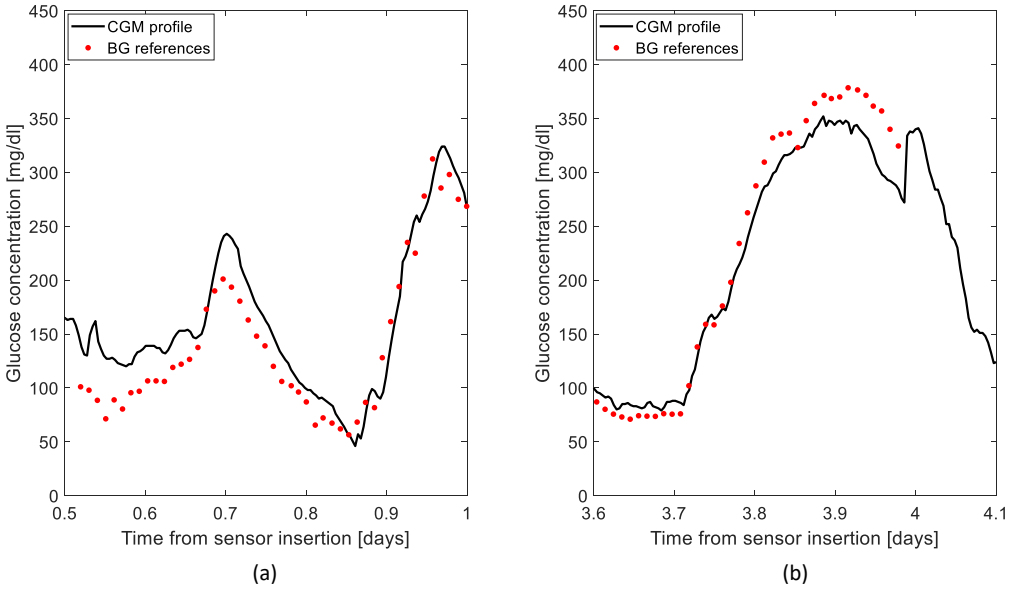

Figure 1.8: Examples in which the CGM sensor output (continuous line) (a) overestimates and (b) underestimates the reference blood glucose (points).

erences, are often a cause of CGM sensor inaccuracy. Two examples of sensor inaccuracy are illustrated in Figure 1.8, where two representative CGM profiles (continuous lines) acquired by the Dexcom G4 Platinum CGM sensor are compared with reference BG concentrations measured in parallel by gold-standard laboratory instruments (points). Two major causes of these deviations are discussed in the following section.

1.3.2 Critical aspects affecting calibration

A first aspect explaining the discrepancies evidenced in Figure 1.8 is the distortion introduced by the BG-to-IG kinetics. Indeed, the needle sensor is inserted in the subcutaneous tissue and measures a current signal that is proportional to the glucose concentration in the interstitial fluid. This is due to the fact that, in order to reduce invasiveness of CGM devices, sensors are placed in the subcutis and measure the glucose-related current signal from the interstitial fluid rather than directly from the blood. Thus, the two measurements available during the calibration process, i.e., the electrical current signal measured by the sensor and the BG references acquired through fingerprick devices, belong to different physiological sites. A widely-established description of the relationship between BG and IG is based on a two-compartment model (Figure 1.9 panel a) [68], [69]. According to this representation, and noting that, in steady-

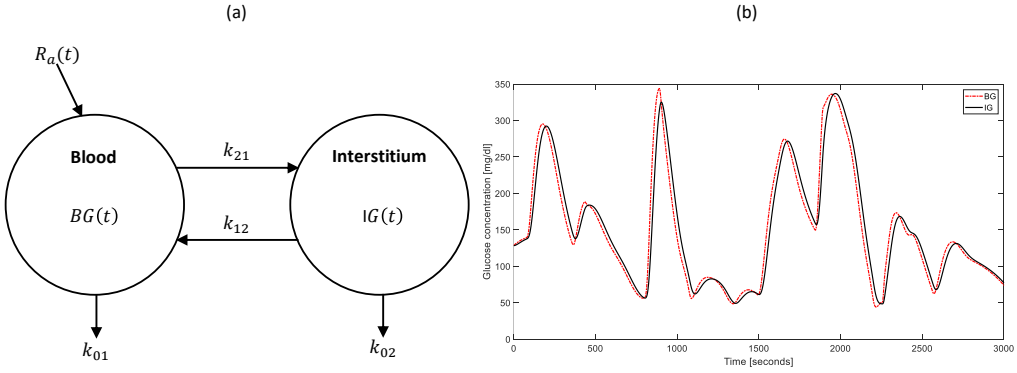

Figure 1.9: (a) Two-compartment model describing the blood glucose to interstitial glucose (BG-to-IG) kinetics. R_a is the rate of appearance; k_{01}, k_{02}, k_{12}, k_{21} are rate constants. The time constant of the BG-to-IG system is $\tau = \frac{1}{k_{02}+k_{12}}$. (b) Representative blood glucose (dashed line) and interstitial glucose (continuous line) concentration profiles simulated assuming $\tau = 11$ min.

state, BG and IG have equal values, the IG profile can be described by the convolution of the BG profile with a single exponential $\frac{1}{\tau}e^{-\frac{t}{\tau}}$, a decay function having unitary area under the curve and time constant τ. The time constant τ is related to the parameters of the two-compartment model by the equation $\tau = \frac{1}{k_{02}+k_{12}}$ [68], [69]. Given the low-pass filtering nature of the system described in Figure 1.9 panel a, the IG signal is a smoothed and delayed version of the BG concentration [70]. An example is reported in Figure 1.9 panel b, where the IG profile (obtained by convolving a given, simulated, BG profile with a single exponential with $\tau = 11$ minutes) shows both amplitude attenuation and phase delay compared to the BG profile. Notably, τ shows inter- and intra-subject variability and its numerical identification requires suitable collection of both BG and IG samples. Published values of the time constant τ range from 6 to 15 min [69]. In practice, the BG-to-IG time constant τ is treated as a user parameter, but its role in the calibration process needs to be carefully considered [71], [72].

A second critical aspect behind the differences pointed out in Figure 1.8 is related to the time variability of sensor sensitivity. The raw electrical current signals acquired by CGM sensors often exhibit a non-physiological drift, especially in the first day after sensor insertion. An example of non-physiological drift observed in a raw CGM signal acquired by the Dexcom G4 Platinum CGM sensor is depicted in Figure 1.10, where the continuous line represents the electrical current signal (in units not specified by the manufacturer) and the dashed line shows the drift. This phenomenon is related to a variation of

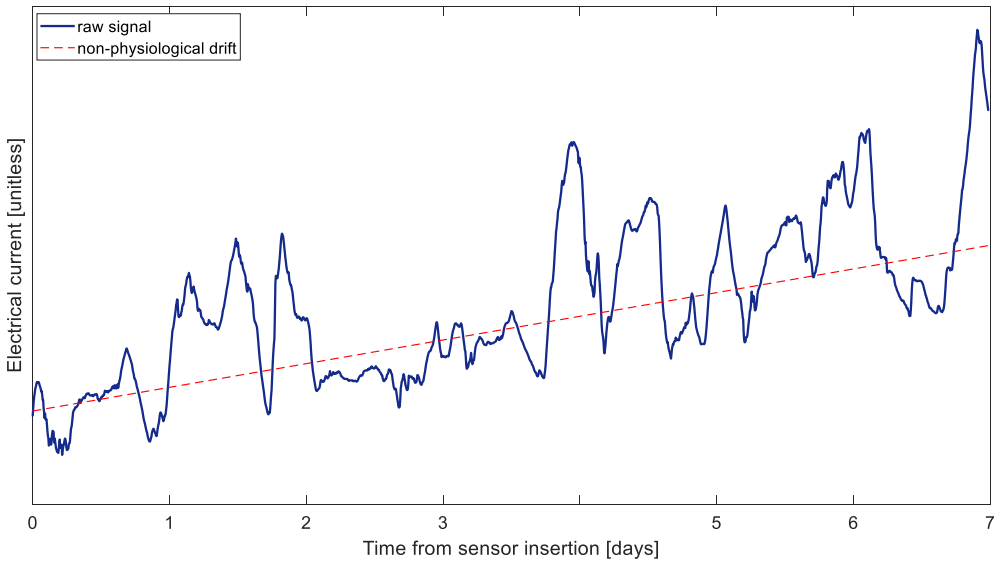

Figure 1.10: Representative raw CGM sensor signal (continuous line, units not specified by the manufacturer) that exhibits a non-physiological drift (dashed line) due to the time variability of sensor sensitivity.

sensor sensitivity after its insertion in the body, when the sensor membrane enters in contact with the biological environment and undergoes the immune system reaction [73], [74]. The calibration model has to properly compensate for such time variability, which is often non-linear.

Several techniques have been proposed to deal with these issues affecting CGM sensor calibration. The next section presents and discusses the most recent algorithms proposed in the literature.

1.4 State-of-art calibration algorithms

One of the major limitations of the calibration linear regression techniques presented in Section 1.3.1 is that they all neglect the time lag between the BG and the raw sensor signal, which can lead to a suboptimal estimation of the parameters of the calibration function. Therefore, most of the calibration algorithms developed by the scientific community included more or less sophisticated approaches to overcome this limitation and take BG-to-IG dynamics into account.

The first simple approach is to require calibration of the sensor when glucose is relatively stable. This approach can be applied to any calibration algorithm. The rationale of this heuristic is that, in such a condition, BG and IG concentrations should be at equilibrium and, thus, the estimation of the linear

regression parameters should not be influenced by not considering the BG-to-IG dynamics [75]. Following this rationale, Aussedat et al. [76] developed an automated algorithm that requests sensor calibration only when a window of stable signal is detected, i.e., when the sensor signal has not changed by more than 1% over a four-minute window, and when the raw current value for the second calibration point differs from the first by $\geq$ 2 nA. The study proved that performing calibrations during periods of relative glucose stability minimizes difference between BG references and raw sensor measurements due to the BG-to-IG kinetics.

More sophisticated model-based approaches to account for the BG-to-IG dynamics have been developed relying on Kalman filter theory. In particular, Knobbe et al. [77] proposed a five-state extended Kalman filter, which estimates subcutaneous glucose levels, BG levels, time lag between the sensor measured subcutaneous glucose and BG, time-rate-of-change of the BG level, and the subcutaneous glucose sensor scale factor [78]. In this study, BG levels are reconstructed in continuous time from CGM measurements, employing a state-space Bayesian framework with a priori knowledge of unknown variables. A direct application of a Kalman filter to improve CGM sensor accuracy was proposed by Kuure-Kinsey and colleagues [79], employing a dual-rate Kalman filter and exploiting sparse SMBG measurements to estimate the sensor sensitivity in real-time. Although designed for real-time glucose and its rate of change estimation, the algorithm does not account for BG-to-IG kinetics. A further development, with direct application to the calibration problem and incorporation of the BG-to-IG dynamic model, was given by Facchinetti et al. [80]. The authors proposed an extended Kalman filter method that works in cascade to the standard device calibration to enhance sensor accuracy. By taking into account BG-to-IG kinetics, using a model to describe the variability of sensor sensitivity, and exploiting four BG reference samples per day, the method significantly improves CGM accuracy when applied to synthetic data. However, its real-time implementation is not straightforward, requiring the knowledge of the variances of both state and measurement error processes, as well as an initial burn-in interval.

Another approach for real-time glucose estimation based on auto regressive (AR) models was proposed by Leal et al. [81]. The study used AR models to estimate BG from raw CGM measurements. Data acquired from 18 T1D patients were used to train a population AR model, which was then incorporated into a calibration algorithm for real-time BG estimation. The raw sensor sig-

nal, used as the independent variable, and the BG concentration, considered as a dependent variable, were both normalized based on the maximum range of the available signals. The best overall estimated model, with a third-order Box-Jenkins structure and fixed parameters, enhanced CGM performance, especially in hypoglycemia detection. Significant improvement in hypoglycemia detection was also obtained by the same authors in another study performed on 21 patients with T1D where a new linear regression algorithm with enhanced offset estimation was proposed [82].

Barceló-Rico and colleagues proposed an alternative calibration algorithm based on a dynamic global model of the relationship between BG and interstitial CGM signal [83]. The algorithm integrates several local dynamic models, each one representing a different metabolic condition and/or sensor-subject interaction. The local models are then weighted and added to compose the global calibration model. Inputs of the model are the signal measured by the sensor and other signals containing information relevant to glucose dynamics, which are normalized in magnitude using population parameters. The algorithm showed improvements in CGM sensor accuracy, although it was tested on only eight healthy subjects and a more extensive assessment on the diabetic population would be needed to confirm the findings. A further development of the algorithm was proposed in [84], where an adaptive scheme is used to estimate patient's normalization parameters in real-time instead of using simple population parameters. Results on 30 virtual patients showed that the adaptation of normalization parameters further improved the performance of the algorithm, since they were able to compensate for sensor sensitivity variations.

Most of the algorithms proposed for improving CGM performance employ sophisticated models and signal processing features that, although still allowing the implementation on wearable devices/smartphones, increase the computational complexity and processing delay compared to the simple linear regression techniques. With the aim of reducing the delay due to signal processing, Mahmoudi et al. proposed a multistep calibration algorithm based on rate-limiting filtering, selective smoothing, and robust regression [85]. The rate-limiting filter limits the rate-of-change if a physiological threshold is exceeded; the selective smoothing is applied if the signal is noisy, i.e., if the number of zero crossings of the signal first-order differences exceeds a predefined threshold; the robust regression then converts the raw measured current to BG levels using reference SMBG measurements (for a maximum of four references per day). The application of the filtering step to only the noisy parts of the sig-

nal lowered the delay introduced by the signal processing of the CGM profile.

Another approach that has the low computational complexity as a major strength was proposed by Kirchsteiger and colleagues employing linear matrix inequalities techniques, resulting in convex optimization problems of low complexity [86], [87]. The authors proposed two different parametric descriptions of the relationship between IG and BG and a constructive algorithm to adaptively estimate the unknown parameters. The algorithm explicitly considers the measurement uncertainty of the device used to collect the calibration measurements, which was firstly pointed out by Choleau and colleagues [66]. Moreover, the algorithm embeds an automatic feature to detect fingerprtick measurements, which are not suitable to be used for calibration.

The uncertainty in the reference SMBG samples used for calibration is a key issue in the development of robust calibration algorithms. The real-time deconvolution-based approach proposed by Guerra et al. [88] demonstrated its robustness against both temporal misplacement of the SMBG references and uncertainty in the BG-to-IG kinetics model. The authors proposed a real-time signal-enhancement module to be applied to the CGM sensor output to improve the accuracy of the device. The algorithm compensates the distortion due to the BG-to-IG dynamic by means of regularized deconvolution [89] and relies on a linear regression model that is updated each time a pair of SMBG references is collected. Significant accuracy improvements were observed both on simulated and real datasets. The deconvolution-based approach of [88] was further developed in [90], where it was directly applied to the raw measured signal rather than in cascade to the CGM sensor output. The algorithm fits the raw current signal against BG references (collected twice a day) using a time-varying linear calibration function whose parameters are identified in the Bayesian framework using a priori knowledge on their statistical distribution. The BG-to-IG kinetics is compensated, as in [88], via nonparametric deconvolution. Results showed significant accuracy improvements compared to the manufacturer calibration.

Current CGM products are available for continuous use and are replaced after several days. However, none of the methods discussed so far have embedded any features able to capture this essential cyclic nature by exploiting, e.g., the data from prior weeks to better calibrate new CGM data. A first attempt on this direction was made by Lee and colleagues in [91], where a run-to-run strategy that personalizes sensor calibration parameters using data from previous weeks' use was proposed. Before each weekly new sensor insertion,

the algorithm minimizes a cost function that penalizes differences between fingerprtick reference values and CGM output of previous weeks. Repeated iterations of the run-to-run procedure demonstrated improved performance on synthetic data (summed square error reduced by 20% after two weeks, and up to 50% after six weeks). On the same line, another calibration algorithm, employing a time-varying linear calibration function as in [90], was augmented with a weekly updating feature for parameter optimization [92]. The algorithm estimates the calibration parameters through the recursive least squares to fit SMBG measurements taken approximately every 12 hours. Then, personalized calibration parameters are optimized after the first week of use using past data, employing a forgetting factor to give more weight to the most recent data.

The literature calibration algorithms discussed so far showed, in general, several performance improvements compared to the simple linear regression methods described in Section 4.1.3 and implemented in the first commercialized CGM sensors. However, none of them explicitly aimed at enhancing sensor accuracy while reducing, at the same time, the frequency of calibrations, i.e., the number of SMBG fingerprick measurements needed as input to the algorithm, which are an obvious reason of discomfort for the patients.

To pursue this objective, the use of the Bayesian estimation in the calibration process, as proposed in [90], appears the most promising technique. Indeed, by setting the calibration problem in the Bayesian framework, the information brought by additional BG references could be substituted by the a priori knowledge on calibration parameters derived from ad hoc training sets, allowing, in principle, the reduction of calibration frequency without scarifying sensor accuracy. Thus, the calibration algorithm of [90] and, in particular, the Bayesian strategy employed in the estimation of the calibration parameters, represents the starting point of the study described in this thesis. A more detailed description of the method will be given in Chapter 3.

1.5 Aim of the thesis and outline

The primary aim of this thesis is to develop a calibration algorithm able to improve CGM sensor accuracy while, at the same time, reducing the frequency of calibrations by using a time-varying calibration function and the Bayesian estimation.

In particular, in Chapter 3 we will demonstrate that a recently proposed state-of-art calibration algorithm, when fed with day-specific Bayesian pri-

ors, enables the reduction of calibration frequency from two to one per day. Then, in Chapter 4, we will propose to replace the time-invariant sensitivity and baseline conventionally used by linear calibration functions with more sophisticated time-varying functions, valid for multiple-day periods, with unknown parameters for which an a priori statistical description is available. We will refer to this algorithm as Bayesian multi-day (BMD) calibration algorithm. As a further development of the BMD calibration algorithm, in Chapter 5 we will propose a multi-model Bayesian framework (MMBF) that, in order to cope with inter-sensor and inter-subject variability, enables the choice of the calibration model among a finite specified set of candidates. Finally, in Chapter 6 we will summarize the major findings of the work carried out in this thesis, the possible applications and margins for improvement.

Chapter 2

Available datasets and criteria for the assessment of the calibration methods

In this chapter, we describe the two datasets - kindly provided by Dexcom Inc. (San Diego, CA) - used to test the calibration algorithms discussed in the thesis and the relative performance metrics employed to assess the accuracy of the calibrated profiles. The first dataset was acquired with a commercial CGM sensor, while the second dataset was acquired with a new CGM sensor prototype. We will define specific metrics for each of the two datasets. This choice was made to ensure agreement of the metrics employed with other literature studies on the same datasets.

2.1 The Dexcom G4 Platinum (DG4P) dataset

2.1.1 Dataset description

The first dataset available and used in the thesis was collected during a multicenter pivotal study involving 72 subjects with diabetes (60 subjects with T1D, 12 subject with T2D) [16]. Subjects wore the DG4P CGM sensor for a 7-day period. In total, a pool of 108 datasets was collected, since 36 subjects wore two sensors during the study. Each of the 108 datasets includes:

- the raw, unprocessed, electrical current signal measured by the sensor with 5-min sampling period;

- the CGM profile as originally calibrated by the manufacturer, i.e., the output in mg/dl displayed to the user, with 5-min sampling period;
- the SMBG measurements acquired during the study and used by the manufacturer for calibrating the sensor (about 2 SMBG samples per day);
- highly accurate YSI laboratory measurements acquired during three in-clinic sessions of 12 h duration in days 1, 4, and 7 after sensor insertion, with 15-min sampling period, during which carbohydrate consumption, insulin dosing, and meal timing were manipulated to obtain a wide range of glucose values (from <60 mg/dl up to 400 mg/dl).

In Figure 2.1, a representative dataset is reported. Top panel shows the raw electrical current signal (blue line, in units not specified by the manufacturer) and the SMBG measurements (orange triangles, in mg/dl) used for calibration. To note that, the same SMBG and raw current data will be used as input to the calibration algorithms proposed in the thesis, producing a calibrated CGM output comparable to that originally obtained by the manufacturer. Bottom panel reports the CGM profile (black line, in mg/dl) as originally calibrated by the manufacturer and the YSI measurements (red points, in mg/dl) acquired during the three in-clinic sessions and used for assessing the accuracy of the calibrate profiles.

2.1.2 Performance analysis

Performance of the DG4P dataset is assessed by comparing the CGM calibrated profiles under test and the YSI laboratory references. Firstly, the two measurements are matched in time. Then, accuracy is quantified by computing three commonly used metrics.

2.1.2.1 Estimated vs. reference glucose matching

The sampling period of the calibrated CGM profiles corresponds to the sampling frequency of the unprocessed electrical current signal, i.e., 5 minutes. The sampling period of the YSI laboratory references is instead 15 minutes, and the two time-grids do not necessarily contain the same time instants. Thus, in order to compute any accuracy metric, the two measurements need to be matched in time. In performing the time-matching, we consider that the calibrated CGM and the YSI measurements reflects samples of the interstitial and

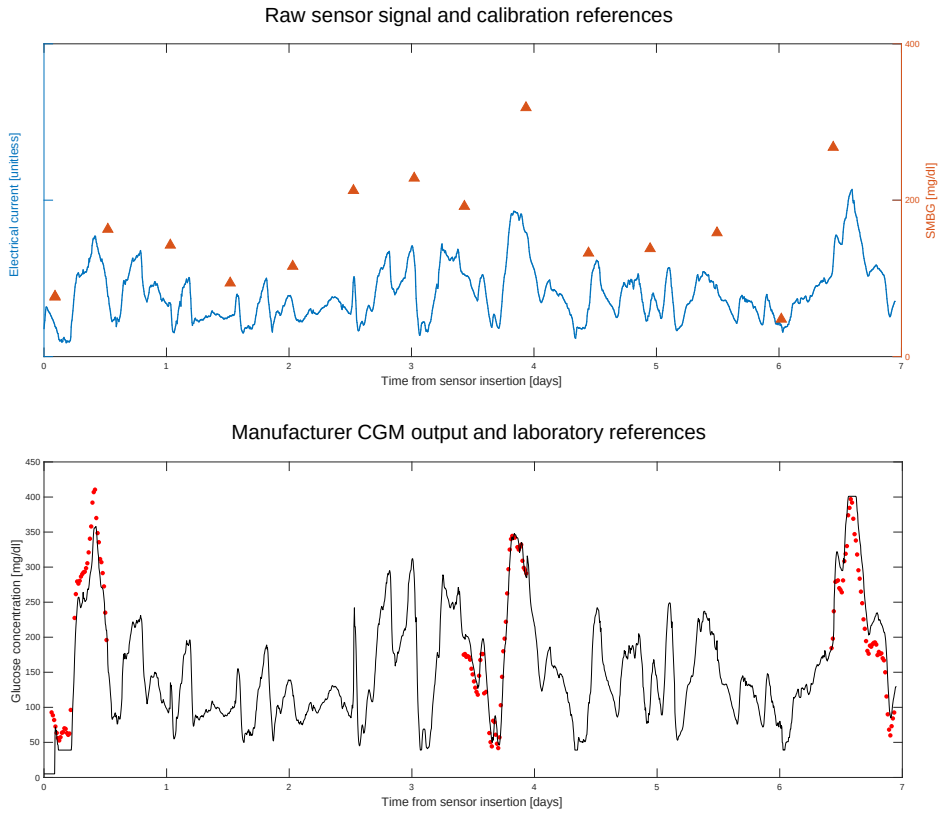

Figure 2.1: Representative Dexcom G4 Platinum (DG4P) dataset. Top panel: raw, unprocessed, current signal (blue line) and calibration SMBG references (orange triangles). Bottom panel: CGM profile as originally calibrated by the manufacturer (black line) and laboratory references for accuracy assessment (red points).

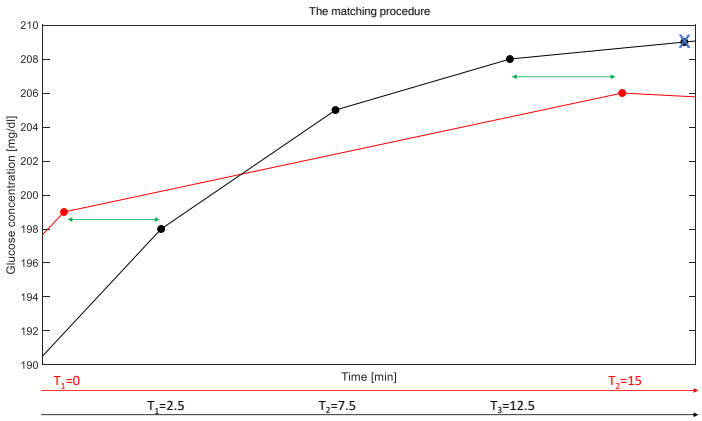

Figure 2.2: Illustrative description of the matching procedure. Each reference measurement (red points) is matched with the closest calibrated glucose estimation (black points) in the following 5-minutes window. For instance, the reference acquired at time $T_1 = 0$ is matched with the glucose estimation sample at time 2.5 min. If no CGM sample is available in the following 5-minutes window, due to, e.g., a missing sample (blue cross), the matching is performed with the closest glucose estimation sample in the 5-minutes window preceding the reference. For instance, in the example depicted here, the reference at time $T_2 = 15$ min is matched with the CGM sample at time 12.5 min.

blood glucose, respectively. As discussed in Section 1.3.2, the IG profile is often a delayed version of the BG profile. Thus, the YSI measurement acquired at time $t = 0$ is matched with the correspondent closest CGM sample in the time window between $t = 0$ and $t = 5$ min. If no CGM sample is available in the specified time window due, e.g., to missing samples in the raw signal acquired by the sensor, the matching is performed with the closest CGM sample in the tine time window between $t = -5$ min and $t = 0$. If no CGM sample is available neither in this second time interval, the YSI measurement is discarded and not used for accuracy assessment.

An example of matching procedure is reported in Figure 2.2, where the YSI reference acquired at time instant $T_1 = 0$ is matched with the closest CGM sample in the following 5 min, i.e., the sample at 2.5 min, while the YSI reference at time instant $T_2 = 15$ min is matched with a previous CGM sample, i.e., the one at 12.5 min, due to a missing sample in the 5-min window following the reference time T_2.

The matching procedure produces a set of YSI-CGM data pairs that are used to compute the accuracy metrics described in the following section.

2.1.2.2 Performance metrics

Performance of the DG4P dataset is assessed by computing three common metrics (also considered in other studies analyzing the same dataset [16], [90]):

- Mean absolute relative difference (MARD) between the calibrated CGM profile and the YSI measurements:

$$MARD = \frac{1}{n}\sum_{k=1}^{n}(100 \cdot \frac{|u_k - \hat{u}_k|}{u_k}) \tag{2.1}$$

  where n is the number of YSI measurements, u_k are the YSI measurements, and $\hat{u}_k$ are the values of the calibrated glucose profile at correspondent matched times.

- Percentage of accurate glucose estimates (PAGE), i.e., the percentage of estimated glucose values $\hat{u}_k$ falling within either 20 mg/dl from the relative reference u_k if u_k is lower than 80 mg/dl or within 20% of the relative reference u_k if u_k is above 80 mg/dl.

- Percentage of $(u_k, \hat{u}_k)$ pairs lying in the "A" zone of the Clarke error grid (CEG-A). The CEG is an error grid divided into five zones indicating the accuracy of BG estimates generated by meters as compared to a reference value. In particular, zone "A" indicates accurate glucose results that do not lead to subsequent wrong or dangerous treatments. An example of CEG plot is reported in Figure 2.3.

These three metrics are used to assess the accuracy of both the CGM profiles as originally calibrated by the manufacturer and the CGM profiles as calibrated by the algorithms analyzed and proposed in this thesis.

For each CGM profile under test, a subject-level analysis is performed by computing the performance metrics for each dataset and then taking the population statistics. In particular, the three metrics described above can be computed by considering, for each dataset, the YSI-CGM pairs of the entire monitoring session, or by considering the YSI-CGM pairs of days 1, 4, and 7 separately (this last option to evaluate performance in each specific day). The population performance indexes are obtained by computing the mean and standard deviation of the metrics obtained in each dataset for normally distributed metrics and the median and interquartile range of the metrics obtained in each dataset for non-normally distributed metrics. Normality is assessed for each metric by Lilliefors test.

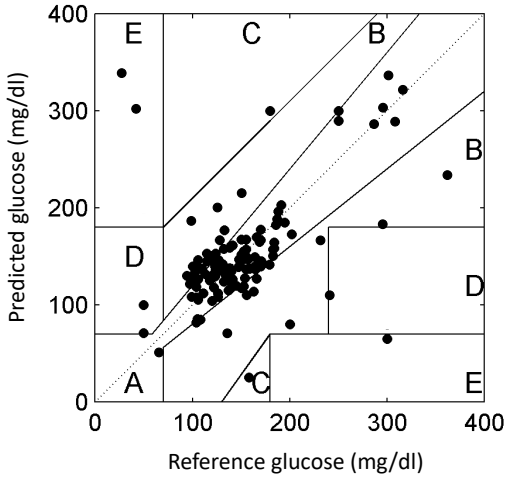

Figure 2.3: Clarke error grid (CEG) used for accessing the accuracy of a glucose estimation, e.g., the glucose estimation given by a CGM sensor, vs. a reference measurement (black points). It is subdivided in 5 zones associated to different grades of clinical danger (from zone "A", the safest zone, to zone "E", associated to the higher risk of taking dangerous therapeutic actions).

2.2 The next generation Dexcom (NGD) CGM sensor prototype dataset

2.2.1 Dataset description

The second available dataset derive from a clinical trial involving 60 subjects affected by diabetes (54 with T1D, 6 with T2D) which were monitored for a 10-day period with a NGD CGM sensor prototype.

For each subject, the following data were collected:

- the raw, unprocessed, electrical current signal measured by the sensor with 5-min sampling period;
- SMBG measurements acquired by the subjects three to four times per day;
- highly accurate YSI laboratory measurements acquired during one in-clinic session of about 6 h duration in day 1, 4, 5, 6, or 10 after sensor insertion, with 15-min sampling period, during which subjects were re-

quired to follow their normal diabetes treatment (with no glucose manipulation to obtain a wider glycemic range).

In 5 of the 60 subjects enrolled in the study, the laboratory YSI measurements were not available. Thus, in the analysis conducted in this thesis, we only considered the 55 subjects for which the YSI laboratory references were available to perform an accuracy analysis.

To note that, differently from the DG4P dataset described in Section 2.1, the CGM output as originally calibrated by the manufacturer is not available in the NGD dataset (data were acquired with a sensor prototype which was not yet commercialized).

Figure 2.4 reports a representative NGD dataset. Top panel shows the raw electrical current signal measured by the sensor (blue line, in units not specified by the manufacturer). Bottom panel reports the YSI laboratory measurements acquired during the 6 h in-clinic session (red points, in mg/dl) and the SMBG measurements acquired by the user (orange triangles, in mg/dl). A subset of these SMBG references will be processed by our BMD calibration algorithm (described in Chapter 4) together with the raw electrical current signal to obtain a calibrated CGM profile, whose accuracy will be assessed against the YSI laboratory measurements.

2.2.2 Performance analysis

The matching procedure preliminary to the computation of the NGD performance metrics is the same used for the DG4P dataset described in Section 2.1.2.1. Then, on the CGM-YSI data pairs obtained, the following performance metrics are computed:

- Mean absolute difference (MAD) between the calibrated CGM ($\hat{u}_k$) and the YSI references (u_k):

$$MAD = \frac{1}{n}\sum_{k=1}^{n}(|u_k - \hat{u}_k|) \tag{2.2}$$

  where n is the number of YSI-CGM matched pairs.

- MARD (as defined in Section 2.1.2.2, Equation 2.1).

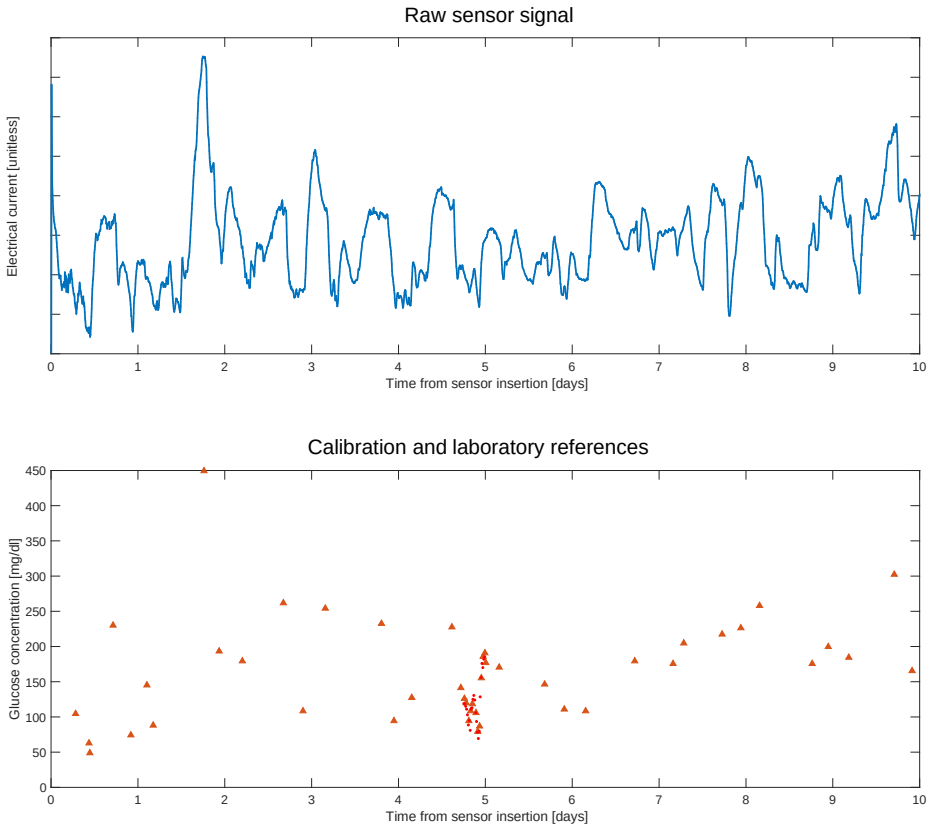

Figure 2.4: Representative next generation Dexcom (NGD) CGM sensor prototype dataset. Top panel: raw, unprocessed, current signal (blue line). Bottom panel: SMBG measurements (orange triangles) and laboratory references for accuracy assessment (red points).

- Mean relative difference (MRD):

$$MRD = \frac{1}{n}\sum_{k=1}^{n}(100 \cdot \frac{u_k - \hat{u}_k}{u_k}). \tag{2.3}$$

- Percentage of CGM-YSI data matching the SMBG standard International Organization for Standardization (ISO) 15197:2013 [93], that is, the percentage of data falling within either 15 mg/dl from the YSI measurement if the YSI measurement is lower than 100 mg/dl or within 15% of YSI if YSI is above 100 mg/dl (15/15%). We also consider other two ranges, 20/20% (which is the analogous of PAGE metric described in Section 2.1.2.2) and 30/30%.

- CEG-A (as defined in Section 2.1.2.2).

The metrics described above are used to assess the accuracy of the CGM profiles calibrated with the BMD algorithm described in Chapter 4. A subject-level analysis is performed by computing the metrics described above for each dataset. The population performance indexes are then obtained by computing the mean and standard deviation, or median and interquartile range, of the metrics obtained in each dataset for normally or non-normally distributed metrics, respectively (normality assessed by Lilliefors test).

Additional aggregated performance analysis

Differently from the DG4P dataset, in the NGD dataset only one YSI session per subject is available (in day 1, 4, 5, 6, or 10). Thus, with a subject-level analysis we can only compute population performance indexes for the entire monitoring period, and not for each specific day separately. Thus, to assess the accuracy also at different days of the monitoring session, we complement the performance analysis with an overall, aggregated, performance assessment. The aggregated assessment is performed by analyzing the CGM-YSI matched pairs from all subjects together. In particular, from the aggregated pool of YSI-CGM pairs (u_k,$\hat{u}_k$), the following metrics are computed:

- Absolute difference (AD):

$$AD = |u_k - \hat{u}_k| \tag{2.4}$$

computed for each single YSI-CGM data pair k. Then, the mean and standard deviation or median and interquartile range of all pairs AD is computed (for normally or non-normally distributed metrics, respectively).

- Absolute relative difference (ARD):

$$ARD = 100 \cdot \frac{|u_k - \hat{u}_k|}{u_k} \tag{2.5}$$

computed for each single YSI-CGM data pair k. Then, the mean and standard deviation or median and interquartile range of all pairs ARD is computed (for normally or non-normally distributed metrics, respectively).

- Relative difference (RD):

$$RD = 100 \cdot \frac{u_k - \hat{u}_k}{u_k} \tag{2.6}$$

computed for each single YSI-CGM data pair k. Then, the mean and standard deviation or median and interquartile range of all pairs RD is computed (for normally or non-normally distributed metrics, respectively).

- ISO metric 15/15% and the additional ranges 20/20% and 30/30% (as defined above).

- CEG-A (as defined in Section 2.1.2.2).

These metrics are computed by considering all YSI-CGM data pairs, obtaining overall aggregated performance indexed, and by grouping pairs from different days, to obtain aggregated performance indexes specific for different days of the monitoring sessions. In particular, we computed aggregated performance for day 1, day 10, and days 4 to 6, which represent the beginning, end, and middle phase of the monitoring session, respectively.

2.3 Statistical analysis

In both the DG4P dataset and the NGD dataset, different calibrated profiles obtained from the same unprocessed current signal by using different calibration algorithms or the same calibration algorithm with different calibration frequency, are computed and compared. In particular, for the DG4P dataset, the

accuracy of the calibrated profiles obtained with the algorithms analyzed and proposed in this thesis with different calibration frequencies is always compared to the accuracy of the CGM profiles as originally calibrated by the manufacturer (Chapter 3 and Chapter 4). For the NGD dataset, performance metrics obtained with the BMD algorithm described in Chapter 4 at different calibration frequencies are compared between each other. In general, the statistical significance of the differences in performance metrics between two calibration algorithms (e.g, our BMD algorithm vs. the manufacturer calibration) or between the same calibration algorithm at different calibration frequencies, is determined by:

- a Wilcoxon signed-rank test for non-normally distributed data;
- a t-test for normally distributed data (normality assessed by Lilliefors test).

In particular, we test the null hypothesis that the median/mean (in case of non-normally/normally distributed data) difference between the paired values of the two groups of metrics is zero, with a significance level of 0.005.

Chapter 3

The starting point: the calibration algorithm of Vettoretti et al. 2016. Comprehensive assessment and margins of improvement[1]

In 2016 Vettoretti et al. proposed a calibration algorithm [90] that employs the Bayesian methodology for parameters estimation. When used with the same calibration frequency as the manufacturer, i.e., two calibrations per day, this method showed improved performance compared to the manufacturer calibration. By using the DG4P dataset described in Section 2.1 in this chapter we assess whether the use of the Bayesian algorithm of [90] can enable the reduction of the number of daily in vivo calibrations required to maintain a good sensor accuracy. We will also demonstrate that there are some margins of improvement of the methodology for further reducing calibration frequency.

3.1 The calibration algorithm of Vettoretti et al. 2016

The Bayesian calibration method implemented in this chapter is fully described in [90]. Here, we summarize the main steps of the method.

[1] Methods and results described in this chapter have been published in [94].

3.1.1 The calibration method

The algorithm processes in real time the raw sensor current signal and the calibration SMBG references, acquired at time instants $t_i, i = 1, 2, ..., N$. At a given time t of the i-th calibration interval, the following model is used to relate the raw interstitial current signal, denoted by $y_I(t)$, to the correspondent IG, denoted by $u_I(t)$ and expressed in mg/dl:

$$y_I(t) = a \cdot u_I(t) + b + c \cdot \Delta t + w(t) \tag{3.1}$$

where a, b, and c are the calibration parameters, Δt (min) is the time from sensor insertion, and $w(t)$ is the noise affecting the measurement. It is important to note that the model of Equation 3.1 is defined to locally describe the relationship between sensor current and IG in the time interval between two consecutive calibrations. In particular, each time a new SMBG reference is available, a new set of calibration parameters are estimated and then used to calibrate the current signal until the following reference is provided. The entire procedure (schematically represented in Figure 3.1) is summarized in three steps:

1. Nonparametric deconvolution of the sensor current signal $y_I(t)$, to compensate for the BG-to-IG kinetics (see, e.g., [69], [95]). Notably, the deconvoluted signal $y_B(t)$ corresponds to a hypothetical electrical current signal measured in plasma.

2. Estimation of calibration parameters $\hat{a}$, $\hat{b}$, $\hat{c}$ in a Bayesian linear minimum variance framework using a day-specific prior information [96] (see Section 3.1.2), by fitting $y_B(t)$ against the last two SMBG references available.

3. Conversion of the sensor current signal $y_I(t)$ into the IG signal $u_I(t)$:

$$u_I(t) = \frac{y_I(t) - \hat{b} - \hat{c} \cdot \Delta t}{\hat{a}} \tag{3.2}$$

Remark

Equation 3.1 corresponds to a specific definition of the generic function f introduced in Chapter 1. Also, the schematic representation of Figure 3.1 is a more specific definition of the general scheme of Figure 1.7, where the additional block relative to the BG-to-IG kinetics is explicitly represented.

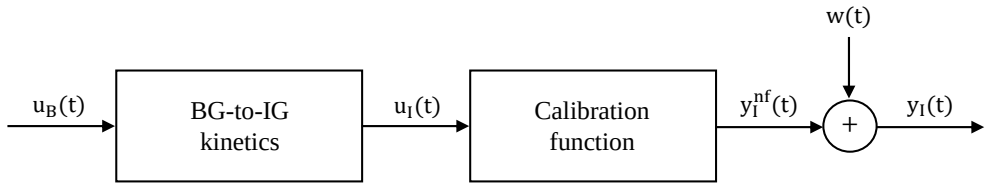

Figure 3.1: Schematic representation of the dynamic system relating the blood glucose $u_B(t)$, whose samples are given by SMBG measurements, to the interstitial current $y_I(t)$, measured by the sensor. The intermediate variables $u_I(t)$ and $y_I^{nf}(t)$ represents, respectively, interstitial glucose and noise-free interstitial current. The variable $w(t)$ represents the measurement noise.

3.1.2 The prior information

The estimation of the calibration parameters a, b, and c (step 2 of the procedure described in Section 4.1) is performed in a Bayesian framework that requires a priori information. Expectations on a, b, and c are obtained by identifying, on an independent training set (defined in Section 4.2), the following nonlinear model:

$$y_I(t) = a \cdot (u_B(t) * \frac{1}{\tau} e^{-\frac{t}{\tau}}) + b + c \cdot \Delta t \tag{3.3}$$

where $*$ is the convolution operator [97], $u_B(t)$ is the BG concentration at time t, measured in mg/dl, and τ (min) is the equilibration time characterizing the BG-to-IG kinetics [69], whereas all other variables were defined in Section 4.1.

In practice, the unknown parameters a, b, c, and τ of Equation 3.3 can be estimated by nonlinear least squares, provided that $u_B(t)$ is a smoothed profile obtained from highly accurate BG measurements, i.e., the YSI measurements acquired during the three in-clinic sessions in days 1, 4, and 7. Parameter τ is considered constant during all the monitoring session, whereas for a, b, and c, different estimates are obtained on days 1, 4, and 7. From the point parameter estimates, sample means and covariance matrices are then calculated and plugged in the Bayesian linear minimum variance estimation procedure (step 2 of the procedure described in Section 4.1). In addition, since no YSI data are available for identifying the model of Equation 3.3 in days 2, 3, 5, and 6, a sort of "global" prior is also built by considering the distributions of a, b, and c on days 1, 4, and 7 all together.

3.2 Assessment of the method with two vs. one calibrations per day

The calibration algorithm of Vettoretti et al. [90] showed improved performance compared to the manufacturer calibration when implemented with the same calibration frequency, i.e., two calibrations per day. Here, we assess whether the method can maintain the same performance when implemented with only one calibration per day instead of two.

3.2.1 Dataset selection

To investigate a two vs. one per day calibration approach, we selected from the DG4P dataset a subset of recordings which present adequate features in terms of SMBG scheduling. In particular, datasets presenting more, or less, than two calibrations per day were excluded from the analysis. Moreover, the calibration model of Equation 3.1 approximate the time dependence of the relationship between the sensor current and IG [73], [74] with a linear function inside each calibration window. This assumption can become critical for calibration intervals that are much wider than 24 hours. For this reason, the datasets having an irregular calibration schedule, or with inter-calibration intervals much wider than 12 h, were also excluded from the analysis. Therefore, the final database used for this assessment consists of 57 recordings (out of the 108 originally available in the DG4P dataset).

3.2.2 Implementation

The Bayesian calibration algorithm as described in Section 4.1 is designed to be used in real-time. It uploads the calibration model with a new set of parameters ($\hat{a}, \hat{b}, \hat{c}$ in Equation 3.2) each time a new SMBG measurement is available by exploiting the last two SMBG samples (i.e., the newly available and the previous most recent one). The so-determined calibration parameters are used throughout the following time interval; then, when a new SMBG sample is available, a new calibration is performed. To compare the two vs. one per day calibration, the original SMBG calibration vector needs to be downsampled, passing from 12 to 24 h inter-calibration intervals (details of the downsampling procedure are reported in Section 3.2.2.1).

Given the limited number of subjects, to create large training and test sets,

where to, respectively, calculate priors and assess calibration accuracy, we used a leave-one-out cross-validation approach. Iteratively, one subject was used to test and evaluate the method, whereas all other subjects formed the training set, which is needed to derive the prior information on calibration parameters (for those subjects wearing two sensors, when the first sensor was under test, the second was excluded from the training, and vice versa). To note that, differently from Vettoretti et al. [90], in which different priors are defined for day 1, 4, and 7, here we use a more specific prior for the first day of monitoring, in order to cope with the sensor time-variability that characterizes day 1 [73], [74]. In particular, instead of using a unique prior for day 1, we use a specific prior for the first and second 12 h of monitoring [96].

3.2.2.1 Downsampling of the calibration points

The SMBG samples downsampling is trivial in terms of practical implementation, but, given the sampling schedule of the dataset, the accuracy assessment by using only one SMBG sample per day is not straightforward. In fact, to assess the accuracy on a 24 h inter-calibration window, YSI reference measurements are required for the entire 24 h period between two consecutive calibrations. However, the available dataset was designed to assess accuracy of the DG4P sensor calibrated once every 12 h, that is, YSI values have been collected only for 12 h between two consecutive calibrations (see Section 2.1).

Note that, passing from two to one calibration per day, that is, having inter-calibration windows of 24 h, the 12 h YSI sections cannot match the entire 24 h window, but they only fill either the first or the second 12 h, depending on how the SMBG values are downsampled from the original SMBG vector. To overcome this problem and to be able to assess accuracy on the entire 24 h calibration window, the SMBG vector is downsampled in two different ways, so that YSI sections will alternatively match the first and the second 12 h. This procedure allows accuracy assessment in the whole 24 h inter-calibration period. In particular, a first SMBG schedule, consisting of selecting the SMBG value collected at the beginning of the YSI section and the one acquired 24 h earlier, allows to assess accuracy in the first half (0–12 h) of the 24 h inter-calibration window. Then, a second SMBG schedule, consisting of selecting the two SMBG samples acquired respectively 12 and 24 h before the beginning of the YSI section, allows to assess accuracy in the second half (12–24 h) of the window. The global accuracy on the 24 h inter-calibration window is then the

mean of the two values.

The creation of two different SMBG schedules to evaluate accuracy of the Bayesian calibration algorithm by using only one SMBG in the entire 24 h inter-calibration window is always possible on days 4 and 7. In these days, accuracy of the Bayesian calibration algorithm is assessed for all 57 datasets both in the first and in the second half of the 24 h inter calibration intervals.

Conversely, on day 1, a similar approach can be applied only to 14 (out of 57) datasets, in which YSI references are collected in the second 12 h of the day and for which there is a sufficient number of SMBG samples to apply both SMBG schedules mentioned earlier. For the remaining 43 (out of 57) datasets, the 12 h YSI window is collected at the beginning of the monitoring period; thus, it is not possible to apply both schedules (SMBG samples and sensor current acquired before the beginning of the monitoring session would be needed), but only to the first one, and hence the accuracy can be evaluated only in the first 12 h of the 24 h inter-calibration window.

In conclusion, for 14 out of 57 datasets, accuracy is assessed on day 1 both in the first and in the second 12 h; whereas for the remaining 43 out of 57 datasets, accuracy is assessed only in the first 12 h.

3.2.3 Results

Table 3.1 shows the population median value (metrics are all non-normally distributed) of each performance metric calculated from CGM profiles given by the original manufacturer calibration, performed every 12 h (third column), and obtained with the Bayesian calibration algorithm, applied every 12 and 24 h, in fourth and fifth columns, respectively. The overall accuracy, calculated on days 1, 4, and 7, and the accuracy calculated separately for each of the three days are reported.

Comparing the original calibration and the Bayesian calibration algorithm fed with two SMBG calibration samples per day (i.e., using the same number of SMBG samples of the original manufacturer calibration), the enhancement brought by the Bayesian approach is evident (confirming [90] and [96]) and viewable, for a representative subject, in first and second panels of Figure 3.2. In particular, first panel shows the CGM output given by the manufacturer versus YSI references; whereas second panel depicts the CGM output obtained using the Bayesian calibration algorithm with two calibrations per day versus YSI, evidencing a better accuracy with respect to the CGM profile of the first

Table 3.1: Performance metrics (median values).

Metric	Day	Original calibration (12 h period)	Bayesian calibration (12 h period)	Bayesian calibration (24 h period)	P (original vs. Bayesian 12 h)	P (original vs. Bayesian 24 h)
MARD	1	15.98	12.65	13.29	0.0063	0.0206
	4	9.07	9.06	9.59	0.1832	0.4624
	7	8.81	8.77	10.13	0.6421	0.7356
	1,4,7	13.05	11.67	11.81	0.0248	0.0411
PAGE	1	76.92	85.11	80.95	0.0026	0.0112
	4	93.33	95.83	94.87	0.0023	0.0300
	7	93.75	95.83	93.74	0.4684	0.6800
	1,4,7	78.01	88.65	87.23	0.0097	0.0238
CEG-A	1	81.25	89.13	85.11	0.0181	0.0435
	4	91.67	93.62	92.86	0.1485	0.0772
	7	93.62	91.49	91.62	0.9439	0.8195
	1,4,7	77.78	87.22	84.75	0.0792	0.1148

panel.

Comparing the Bayesian calibration algorithm using one SMBG calibration sample per day (fifth column of Table 3.1) against the Bayesian calibration algorithm using two SMBG calibration samples per day (fourth column of Table 1), a modest worsening of performance is observed. This result is expected, because the Bayesian calibration algorithm (which, we remind, is the same in both scenarios) is now fed with only half of the original SMBG calibration samples. However, looking at the global performances over the 57 datasets, the worsening is small and, in most cases, negligible, and not statistically significant for any of the considered metrics (P values not shown). In particular, the overall MARD is almost the same, that is, 11.7% and 11.8% for the Bayesian calibration algorithm using two and one SMBG samples per day, respectively.

Moreover, comparing the Bayesian calibration method with a 24 h inter-calibration interval against the original manufacturer calibration, the same statistically significant improvement in accuracy (MARD on day 1, PAGE on days 1 and 4, and CEG-A on day 1) achieved by the Bayesian calibration algorithm using two SMBG samples per day is still achieved by the Bayesian calibration algorithm using one SMBG per day. Overall, a statistically significant improvement is observed for the MARD value, passing from 13.1% to 11.8% (with a P value of 0.0411) and for the PAGE value, passing from 78.0% to 87.2% (with a P value of 0.0238).

Figure 3.3 graphically summarizes results of the Bayesian calibration algorithm with one SMBG calibration sample per day, the Bayesian calibration

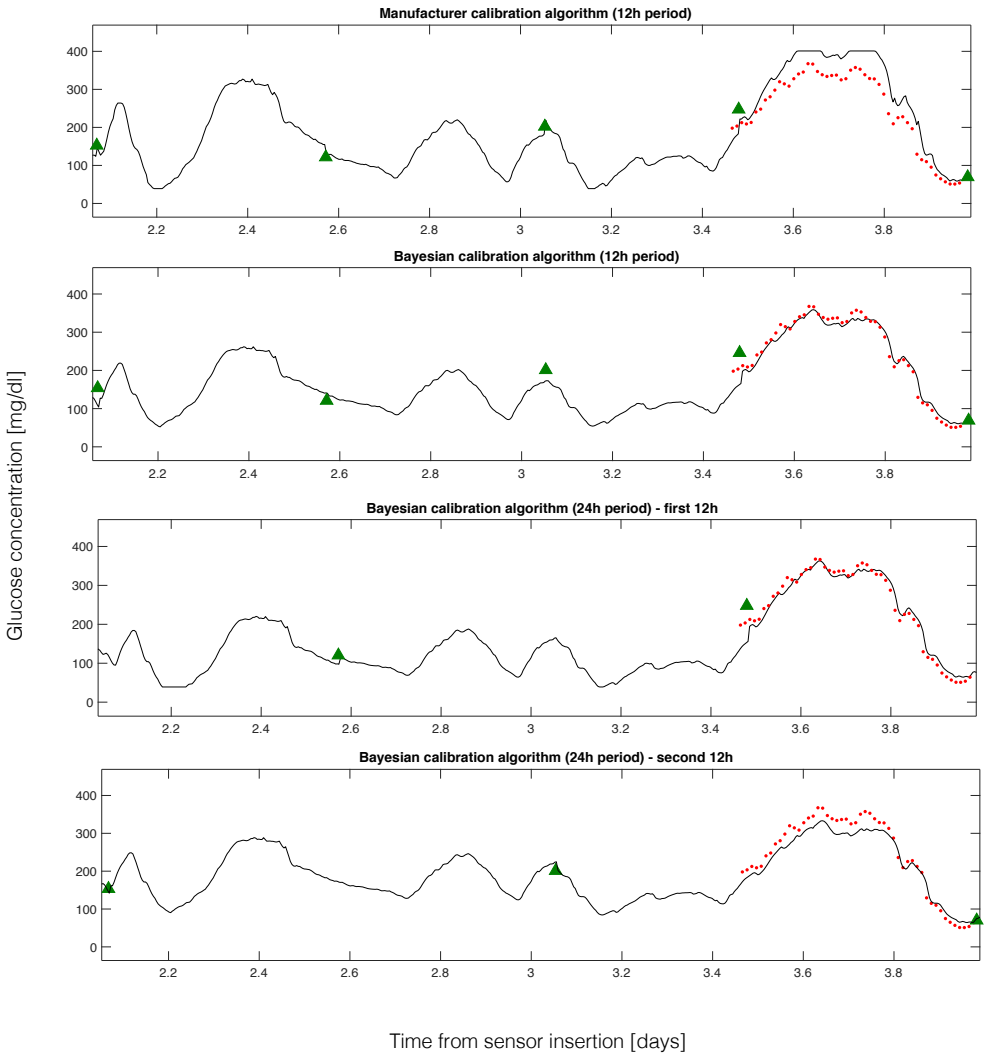

Figure 3.2: Results in a representative subject, day 4. Calibrated profiles (continuous black lines) versus laboratory references (red points) and SMBG samples used for calibration (green triangles). From top to bottom: original manufacturer calibration (12 h periodicity), Bayesian calibration with 12 h periodicity, Bayesian calibration with 24 h periodicity (accuracy assessed in second 12 h), and Bayesian calibration with 24 h periodicity (accuracy assessed in first 12 h).

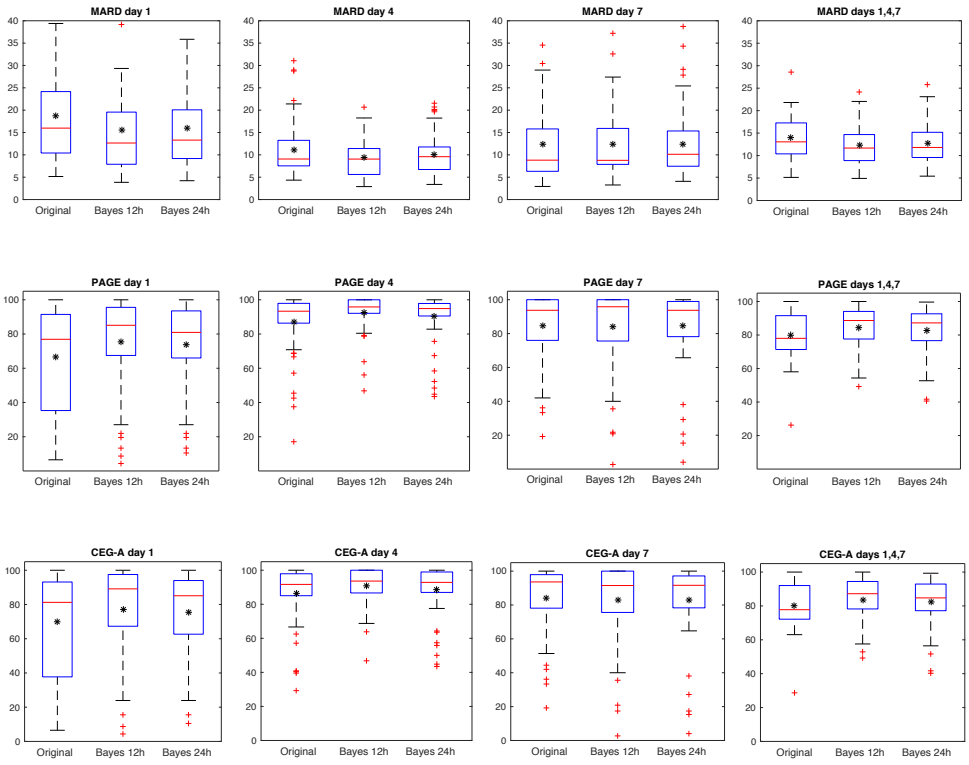

Figure 3.3: Boxplot of performance metrics. From top to bottom: MARD, percentage of accurate glucose estimates (PAGE), and points in zone "A" of the CEG. From left to right: original manufacturer calibration, Bayesian calibration with 12 h period, and Bayesian calibration with 24 h period. In the first three columns, the metrics are computed separately for days 1, 4, and 7; whereas the last column refers to the overall monitoring period.The black star inside each box represents the mean of the distribution.

algorithm with two SMBG samples per day, and the original manufacturer calibration. In particular, a box plot of each metric is reported for the three calibration strategies, separately on day 1, 4, and 7, and globally on all three days.

Of note is that the performance metrics on day 1 are much worse than on the other days of monitoring for both the original manufacturer calibration and the Bayesian calibration algorithm, although the latter shows some improvement compared to the manufacturer calibration. The use of specific prior distributions for calibration model parameters in the first 12 h of monitoring [96] resulted, for most of the sensors, in a significant improvement on day 1. Thus, using the Bayesian method (with both one and two calibrations per day), the overall improvement on days 1, 4, and 7 can be greater than the improve-

ment on the single days. Indeed, if a sensor having an improvement in day 1 accuracy also has an improvement on day 4 and/or on day 7, the overall enhancement is higher than that on the single days.

It is interesting to note that the use of the procedure described in Section 4.1 without the deconvolution step, that is, ignoring the BG-to-IG kinetics, shows only a slightly worse performance. For instance, for the 24 h calibration, the MARD would increase to 11.92%, PAGE would decrease to 87.03%, and, finally, CEG-A would decrease to 84.16%. This allows us to speculate that the deconvolution-based correction within the calibration methodology enhances CGM accuracy, but the relative impact of using reliable priors on calibration parameters is greater.

Finally, since the availability of YSI references on days 1, 4, and 7 allows to assess accuracy of the calibrated profiles only in those time intervals, a global performance evaluation over the entire 7 days of the monitoring period can be done by a direct comparison of CGM-calibrated profiles. In particular, the performance deterioration caused by the reduction in the number of calibrations has been assessed by calculating the relative root mean square error (rRMSE) between CGM profiles that are calibrated using the Bayesian method with two SMBG calibration samples per day versus one SMBG per day. Over the 57 datasets, rRMSE presents a median value of 0.1, thus confirming that performance deterioration caused by the reduction of calibrations from two to one per day is modest.

3.3 Margins of improvement of the algorithm

3.3.1 Accuracy evaluation in the first and second 12 hours

Passing from two to one per day calibration, we have extended the time interval between two consecutive calibrations from 12 to 24 h. Thus, it is of interest to separately assess the accuracy in the first and the second 12 h of the new 24 h inter-calibration windows.

Results, made achievable by the SMBG downsampling strategy described in Section 3.2.2.1, are reported, in terms of performance metrics, in Table 3.2. On average, the first 12 h window shows better results than the second 12 h window on day 1, for all the considered metrics. Regarding day 4, the first 12 h present lower MARD, but almost equal PAGE and CEG-A with respect to the second 12 h. On day 7, the first half-window has better performance than the

Table 3.2: Performance metrics, median values (n=57).

Metric	Day	Bayesian calibration (first 12 h)	Bayesian calibration (second 12 h)
MARD	1	12.66	13.65*
	4	8.99	9.93
	7	9.72	10.39
PAGE	1	85.11	80.75*
	4	95.83	95.45
	7	95.74	93.75
CEG-A	1	89.13	83.46*
	4	95.75	95.56
	7	91.49	91.62

*n=14

second one for metrics MARD and PAGE, whereas CEG-A is almost the same in the two intervals. These differences are expected and can be attributed to the time distance of the last calibration from the YSI window. In fact, accuracy evaluation in the first 12 h window is performed by calibrating the CGM profile just before the beginning of the YSI session (an example is shown in the third panel of Figure 3.2), whereas accuracy evaluation in the second 12 h window is performed by using a CGM profile that is calibrated 12 h before the beginning of the YSI session (an example is shown in the bottom panel of Figure 3.2).

The small deterioration of performance noticed passing from the first to the second 12 h window indicates that, inside an inter-calibration interval, the less is the time distance to the last SMBG sample collected (corresponding to the last update in parameters estimation), the higher is the accuracy of the calibrated profile. Indeed, the linear approximation of the time dependence of calibration parameters (see Section 4.1) becomes critical for large time windows, thus confirming that inter-calibration intervals wider than 24 h cannot be considered with the current calibration model without worsening accuracy. Moreover, the difference in performance noticed in comparing the two 12 h windows is much more visible on day 1: a possible explanation is that, with respect to other days of monitoring, day 1 is characterized by a stronger time dependence of calibration parameters and, therefore, by a higher dependence of accuracy on the time distance from the last parameter update.

3.3.2 Open issues and indications for new investigations

Results reported in this chapter show that the use of a linear time-variant calibration model and the Bayesian estimation procedure allows to reduce the frequency of calibration from two to one per day while, at the same time, improving accuracy compared to the original manufacturer calibration. However, the analysis of Section 3.3.1 evidence that accuracy of the calibrated CGM profiles worsens with time inside the inter-calibration intervals. This result confirms that the calibration model of Equation 3.1 is suitable for describing the time-variability of the relationship between electrical current and IG inside time intervals of limited duration, e.g., at most 24 h.

The use of the Bayesian approach would allow, in principle, to further reduce the frequency of calibrations by deriving more specific prior information on the temporal evolution of calibration parameters. To pursue this objective, the piece-wise linear model of Equation 3.1 needs to be substituted by a more complex model, valid for larger time intervals and, potentially, for the entire monitoring period.

3.4 Summary and concluding remarks

In glucose-oxidase CGM sensors, such as DG4P sensor considered in this chapter, the electrical current signal is transformed to glucose concentration by a calibration model whose parameters are periodically (usually every 12 h) updated by matching SMBG references that are suitably collected by the patient. Previous studies [90], [96] showed that a linear time-varying calibration function, with parameters estimated by a Bayesian procedure embedding suitable priors improves the accuracy of DG4P sensor with respect to the original manufacturer calibration, using the same calibration frequency.

Here, we moved a step further by investigating whether the same Bayesian approach also allows to reduce the number of calibrations from two to one per day. Results show that this is possible, even enhancing accuracy with respect to the original manufacturer calibration (MARD reduced from 13.1% to 11.8%, PAGE increased from 78.0% to 87.2%, and CEGA-A increased from 77.8% to 84.8%). When the same Bayesian calibration algorithm is considered, a negligible and not statistically significant deterioration of the performance is obtained by reducing the number of calibrations to one per day. This indicates that, in a Bayesian framework, the use of well-tuned day-specific priors can surrogate

the information of a second daily SMBG reference, thus allowing to halve the number of calibrations while preserving accuracy.

The separate performance assessment in the first and second 12 h of the 24 h inter-calibration intervals evidences that accuracy deteriorates with time between consecutive calibrations, suggesting that the present piece-wise linear calibration model is not suited to further reduce calibrations to less then one per day. This limitation calls for the development of a new, more complex, calibration model valid for wider time intervals (ideally, the entire monitoring period), capable of incorporating information about the temporal evolution of the calibration parameters. The design, implementation and assessment of a new calibration algorithm satisfying these requirements is described in Chapter 4.

Finally, we highlight that the methodology behind the Bayesian algorithm described in this chapter is general and, in principle, usable with any CGM sensor. It requires the availability of uncalibrated CGM signals and a set of SMBG samples, a sensor-specific calibration model, and a sufficient amount of data to derive reliable priors for model parameters.

Chapter 4

A Bayesian multi-day (BMD) algorithm to calibrate subcutaneous CGM sensors [1]

The calibration algorithm by Vettoretti et al. [90] comprehensively assessed in Chapter 3 showed promising results toward the possibility of reducing the number of calibration by using the Bayesian estimation. Nevertheless, the piece-wise linear calibration function employed in the algorithm is able to approximate the time-variability of the relationship between the current measured by the sensor and IG only for a limited time interval, and, thus, it is not suitable for further reducing the calibrations to less than one per day. In this chapter, we propose a new calibration model capable of describing the time-variability of sensor characteristics over a multiple-day period and a Bayesian estimation procedure to determine its unknown parameters. The new BMD calibration algorithm is implemented in both the DG4P and NGD datasets.

4.1 The BMD calibration algorithm

4.1.1 The calibration model

The BMD calibration algorithm processes the electrical current signal $y_I(t)$, measured by the CGM sensor, and the SMBG measurements, acquired by the fingerprick device, to obtain the IG profile $u_I(t)$. The two measurements, $y_I(t)$ and SMBG, belong to different physical domains, i.e. the current and the glu-

[1] Methods and results described in this chapter have been published in [98] and [99].

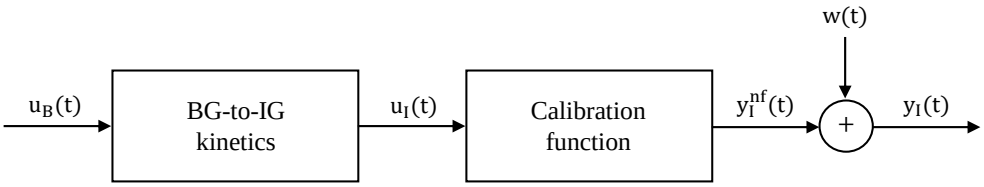

Figure 4.1: Schematic representation of the dynamic system relating the blood glucose $u_B(t)$, whose samples are given by SMBG measurements, to the interstitial current $y_I(t)$, measured by the sensor. The intermediate variables $u_I(t)$ and $y_I^{nf}(t)$ represents, respectively, interstitial glucose and noise-free interstitial current. The variable $w(t)$ represents the measurement noise. (Figure already presented in Chapter 3, reported here for convenience)

cose domain respectively, as well as to different physiological sites. SMBG measures glucose concentration into the blood, while sensor current reflects glucose concentration in the interstitium. Thus, to calibrate a CGM sensor, a model describing the relationship between the two measurements is needed.

Letting $u_B(t)$ be the glucose concentration in blood (whose samples are observable by SMBG), a widely established description of the relation between BG and IG profiles is based on a two-compartment model [69] (see Section 1.3.2), in which the glucose concentration in the interstitial compartment can be described by:

$$\tau \cdot \frac{d}{dt} u_I(t) = -u_I(t) + u_B(t) \tag{4.1}$$

where τ is the equilibrium time-constant between plasma and interstitium. In the literature, the value of τ exhibits significant inter-subject variability but, in a given subject, is commonly assumed to be time-invariant [69], [100], [71], [101]. According to Equation 4.1, the IG profile can be interpreted as the output of a first-order linear dynamic system, having the BG profile as input (see first block in the schematic diagram of Figure 4.1, already presented in Chapter 3 and reported here for convenience) and $h(t)$ as impulse response:

$$h(t) = \frac{1}{\tau} \cdot e^{-\frac{t}{\tau}}. \tag{4.2}$$

The low-pass filtering nature of $h(t)$ causes $u_I(t)$ to be a distorted version of $u_B(t)$, presenting both amplitude attenuation and phase delay. Following the cascade of blocks of Figure 4.1, the IG profile $u_I(t)$ is the input of the calibration function which produces as output a current signal referred to the interstitium, $y_I^{nf}(t)$, which once corrupted by additive noise $w(t)$ finally produces the sensor signal $y_I(t)$. To describe the transformation between glucose and current

profiles, the following model is used:

$$y_I(t) = [u_I(t) + b] \cdot s(t) + w(t) \tag{4.3}$$

where b is the baseline of the glucose profile and $s(t)$ represents the sensitivity of the sensor, described by the model:

$$s(t) = s_1 \cdot \alpha(t) + s_2 \cdot \beta(t) + s_3. \tag{4.4}$$

In Equations 4.3 and 4.4, $\alpha(t)$ and $\beta(t)$ are fixed time domain functions that describe the drift of the specific sensor (see Section 1.3.2), whereas b, s_1, s_2, and s_3 are model parameters, undergoing the following constraints:

$$s_1, s_2, s_3 > 0; \quad \frac{s_1}{s_2} = \varphi \tag{4.5}$$

where φ is a fixed value (see Section 4.2.1 for details). As a result, considering the dependence between sensitivity parameters ($s_1 = \varphi \cdot s_2$), the final parameters vector is:

$$\mathbf{p} = [b, s_2, s_3]^T. \tag{4.6}$$

In comparison with the other approaches proposed in the literature, the peculiarity of the model described by Equation 4.3 is its temporal domain of validity, which is the entire monitoring period, at variance with the model used in Chapter 3, whose domain of validity was restricted to the time window between two consecutive calibrations.

Remark

For sake of method generality, in Equation 4.4 the functional form and corresponding parameters of $\alpha(t)$ and $\beta(t)$ are intentionally treated as being device dependent. Indeed, optimizing them is done as part of industrial device development and is not part of the algorithm we propose. In the specific case of the data analyzed in this thesis, details are documented in two patent applications deposited by the sensor manufacturer [102], [103], which also provide numerous examples of sensitivity curves and how they can be derived from clinical and bench data. In general, any time-domain function (such as linear, logarithmic, polynomial, and exponential) able to capture the sensor drift could be embedded in Equation 4.4.

4.1.2 Estimation of model parameters

Each time a new SMBG is acquired for calibration at time t_i, $i = 1, 2, ..., M$ (where M represents the total number of SMBG samples used for calibration), the set of parameters $\mathbf{p}$ is updated by exploiting the new measure $u_B(t_i)$ and all previously acquired SMBG samples. In particular, let us consider the following relation, expressed in vector form:

$$\mathbf{u_B} = \hat{\mathbf{u}}_\mathbf{B}(\mathbf{p}) + \mathbf{w} \tag{4.7}$$

where $\mathbf{u_B}$ is the i x 1 vector containing the SMBG samples acquired at calibration times t_j, $j = 1, ..., i$ ($i = 1, ..., M$):

$$\mathbf{u_B} = [u_B(t_1), ..., u_B(t_{i-1}), u_B(t_i)]^T \tag{4.8}$$

$\hat{\mathbf{u_B}}$ is the i x 1 vector (function of the model parameters), obtained transforming the i x 1 vector $\mathbf{y_I}$, containing $y_I(t_j)$, $j = 1, ..., i$, into BG values. Both the BG-to-IG kinetics and the calibration model are considered in the transformation (as discussed below). The i x 1 vector w represents the error. Note that the length i of the vectors increases of one unit each time a new SMBG is acquired for calibration, since all previous measurements are anyway considered.

The unknown parameters vector $\mathbf{p}$ is estimated by exploiting the data contained in $\mathbf{u_B}$ and $\mathbf{y_I}$ in addition to some a priori knowledge on the distribution of $\mathbf{p}$, derived from a data training set, as described in Section 4.2.1. In particular, the a priori distribution of the parameters vector $\mathbf{p}$ has mean $\boldsymbol{\mu_p}$ and covariance matrix $\boldsymbol{\Sigma_p}$. The error vector $\mathbf{w}$ is assumed to contain white noise samples, uncorrelated from $\mathbf{p}$, with zero mean and diagonal covariance matrix $\boldsymbol{\Sigma_w}$. The error variance is assumed constant over time, i.e., $\boldsymbol{\Sigma_w} = \sigma_w^2 \cdot \boldsymbol{I}$ and σ_w^2 is estimated from the training set (details in Section 4.2.1).

The Bayesian maximum a posteriori (MAP) estimate of $\mathbf{p}$ is obtained by solving the following optimization problem:

$$\hat{\mathbf{p}} = \boldsymbol{argmin}_\mathbf{p}[\mathbf{u_B} - \hat{\mathbf{u}}_\mathbf{B}(\mathbf{p})]^T {\boldsymbol{\Sigma_w}}^{-1}[\mathbf{u_B} - \hat{\mathbf{u}}_\mathbf{B}(\mathbf{p})] + (\boldsymbol{\mu_p} - \mathbf{p})^T {\boldsymbol{\Sigma_p}}^{-1}(\boldsymbol{\mu_p} - \mathbf{p}) \tag{4.9}$$

which, given the presence of nonlinearities, does not have a closed form solution. Thus, the estimate $\hat{\mathbf{p}}$ is found by looking iteratively into the parameter space. The iterative procedure, schematically described by the diagram of Figure 4.2, is summarized in five steps:

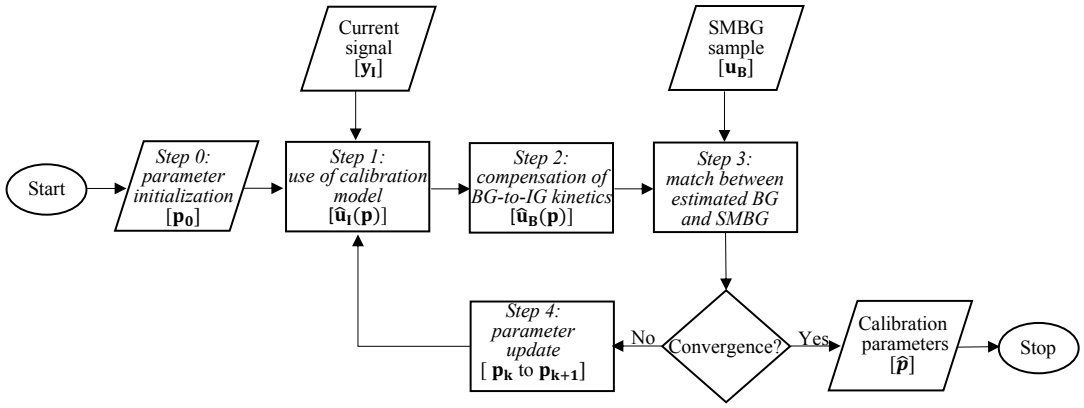

Figure 4.2: Flowchart of the iterative procedure for the estimation of the calibration model parameters (parallelograms denote input/output blocks, whereas the diamond denotes the decision block). Starting from initial vector $\mathbf{p_0}$, at each iteration k the parameter vector is updated, according to the input values $\mathbf{y_I}$ (current signal) and $\mathbf{u_B}$ (blood glucose samples), using the calibration model of Equation 4.3 and compensating for the blood-to-interstitial glucose kinetics.

- i) initialization of parameter vector $\mathbf{p}$;
- ii) estimation of IG profile $\hat{u}_I(t)$ according to the calibration model of Equation 4.3;
- iii) estimation of BG profile, $\hat{u}_B(t)$, accounting for the distortion introduced by the BG-to-IG kinetics, through nonparametric deconvolution;
- iv) matching between SMBG measurements and $\hat{u}_B(t)$;
- v) update of parameter vector for the next iteration.

The following subsections describe individually each of these five steps.

Step 0: parameter initialization

At the first iteration of the i-th calibration ($i = 1, ..., M$) the parameter vector is initialized to the mean value of the prior distribution, $\mathbf{p_0} = \mu_p$.

Step 1: use of calibration model

At each iteration k ($k = 0, 1, ..., N_{iter}$), interstitial glucose $u_I(t)$ is estimated from the current signal $y_I(t)$ by inverting Equation 4.3, which depends from

parameter vector $\mathbf{p_k}$:

$$\hat{u}_I(t, \mathbf{p_k}) = \frac{y_I(t)}{s(t, s_2, s_3)} - b. \tag{4.10}$$

Step 2:compensation of BG-to-IG kinetics

The IG profile obtained at Step 1, $\hat{u}_I(t, \mathbf{p_k})$, cannot be directly matched to SMBG measures, $u_B(t)$. Indeed, the distortion induced by BG-to-IG kinetics needs to be compensated first. As already discussed in Section 4.1.1, the IG profile can be seen as the output of a first-order linear dynamic system whose impulse response is given by Equation 4.2 and whose input is the BG profile. Thus, the estimation of $\hat{u}_B(t, \mathbf{p_k})$ from $\hat{u}_I(t, \mathbf{p_k})$ is an inverse problem that can be solved by deconvolution. In particular, to reconstruct the BG profile from the IG profile a nonparametric stochastic approach [89] is used.

For computational reasons, the deconvolution is applied to temporal windows containing the time instant t_j ,$j = 1, ..., i$ at which each SMBG is acquired. Practically, for each of the i BG measures collected in vector $\mathbf{u_B}$, a time window $\mathbf{\Lambda}$ from $t_i - \mathbf{100}$ min to $t_i + \mathbf{5}$ min is considered. Letting $\mathbf{u_I}(\mathbf{\Lambda})$ be the $nx\mathbf{1}$ vector containing the IG measures estimated at previous step at the sampling instants lying in $\mathbf{\Lambda}$, a uniform sampling grid, with 5-min step, can be defined: $\mathbf{\Omega_s} = t_1, t_2, ..., t_n$. In addition, $\mathbf{w}$ is defined as the $nx\mathbf{1}$ vector of measurement error $w(t)$ at time instants in $\mathbf{\Omega_s}$, assumed to have zero mean and covariance matrix $\mathbf{\Sigma_w} = \sigma^2 R$, with σ^2 unknown constant and $\mathbf{R}$ nxn known matrix whose structure reflects expectations on measurement error variance (here,$\mathbf{R} = \mathbf{I_n}$, since the error samples are assumed to be uncorrelated from the current signal and with variance constant over time). The vector $\mathbf{u_B}(\mathbf{\Lambda})$ is defined as the $Nx\mathbf{1}$ unknown vector containing samples of $u_B(t)$ at time instants on a virtual grid $\mathbf{\Omega_v} = t_{v1}, t_{v2}, ..., t_{vN}$, which is independent from and usually denser than $\mathbf{\Omega_s}$ (here a uniform 1-min step is used). The virtual grid allows us to obtain a denser profile, which can be more easily matched with SMBG samples (see Step 3). Moreover, $\mathbf{\Sigma_v}$ starts from $t_1 - \mathbf{100}$ min, in order to allow initial condition transient to vanish, so that the reconstruction of $u_B(t)$ is not altered in the window of interest $\mathbf{\Lambda}$.

Once all variables have been defined, having $\mathbf{\Omega_s}$ and $\mathbf{\Omega_v}$ both uniform, with $\mathbf{\Omega_s} \subseteq \mathbf{\Omega_v}$, the following matrix equation can be written:

$$\mathbf{u_{I(\Lambda)}} = \mathbf{H} \cdot \mathbf{u_{B(\Lambda)}} + \mathbf{w} \tag{4.11}$$

where $\mathbf{H}$ is the nxN matrix obtained downsampling the NxN transfer matrix $\mathbf{H_v}$ of the BG-to-IG system, maintaining only the rows correspondent to sampling instants in $\mathbf{\Sigma_s}$. Since vector $\mathbf{u_{B(\Lambda)}}$ contains samples of a BG profile, which is a biological signal expected to have a certain smoothness, a double integrated white noise model [89] of unknown variance $\boldsymbol{\lambda^2}$ is chosen to describe entries of $\mathbf{u_{B(\Lambda)}}$. Thus, its covariance matrix is:

$$\mathbf{\Sigma_{u_{B(\Lambda)}}} = \boldsymbol{\lambda^2}(\mathbf{F}^T\mathbf{F})^{-1} \tag{4.12}$$

with $\mathbf{F}$ NxN Toeplitz lower-triangular matrix having $[\mathbf{1, -2, 1, 0, ..., 0}]^T$ as first column. Assuming that $\mathbf{u_{B(\Lambda)}}$ and $\mathbf{w}$ are uncorrelated, the following quadratic optimization problem corresponds to the linear minimum error variance Bayesian estimate of $\mathbf{u_{B(\Lambda)}}$:

$$\begin{aligned}\mathbf{\hat{u}_{B(\Lambda)}} = \boldsymbol{argmin}_{\mathbf{u_{B(\Lambda)}}}(\mathbf{u_{I(\Lambda)}} - \mathbf{Hu_{B(\Lambda)}})^T\mathbf{R}^{-1}(\mathbf{u_{I(\Lambda)}} - \mathbf{Hu_{B(\Lambda)}}) + \\ + \gamma\mathbf{u}^T_{\mathbf{B(\Lambda)}}\mathbf{F}^T\mathbf{F}\mathbf{u_{B(\Lambda)}}\end{aligned} \tag{4.13}$$

where parameter $\gamma = \frac{\sigma^2}{\lambda^2}$, estimated by Maximum Likelihood [89], represents the regularization term that balances the data fit with the smoothness of the estimated profile. The optimization problem of Equation 4.13 admits a closed form solution, expressed as:

$$\mathbf{\hat{u}_{B(\Lambda)}} = (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \gamma\mathbf{F}^T\mathbf{F})^{-1}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{u_{I(\Lambda)}}. \tag{4.14}$$

For every SMBG measurement in vector $\mathbf{u_B}$ (see Equation 4.8), the BG profile $\mathbf{\hat{u}_{B(\Lambda)}}$, which depends on parameter vector $\mathbf{p_k}$, is estimated inside the window $\mathbf{\Lambda}$ that contains the time instant at which the SMBG sample $\boldsymbol{t_j}$, $\boldsymbol{j = 1, ..., i}$ is acquired.

Step 3: match between estimated BG and available SMBG

For each SMBG sample in vector $\mathbf{u_B}$, acquired at time $\boldsymbol{t_j}$, $\boldsymbol{j = 1, ..., i}$, the corresponding estimated value of $\mathbf{\hat{u}_{B(\Lambda)}}$ at time $\boldsymbol{t_j}$ is considered, by exploiting the vector $\mathbf{\hat{u}_B(p_k)}$ used in Equation 4.9:

$$\mathbf{\hat{u}_B(p_k)} = [\hat{u}_B(t_1, \mathbf{p_k}), ..., \hat{u}_B(t_{i-1}, \mathbf{p_k}), \hat{u}_B(t_i, \mathbf{p_k})]^T. \tag{4.15}$$

Step 4: parameter update

At each iteration k, the parameter vector $\mathbf{p_k}$ is updated to a new set of values, $\mathbf{p_{k+1}}$, using the Nelder-Mead simplex algorithm, as described in [104].

Steps 1–4 are reiterated until one of the following stopping criteria occurs: i) the step size in parameters update is smaller than a fixed tolerance (e.g., $\mathbf{10^{-6}}$); ii) the relative change in the value of the objective function is lower than a fixed tolerance (e.g., $\mathbf{10^{-6}}$); iii) the algorithm reaches the maximum number of iterations (e.g., $\mathbf{10^4}$). This last stopping condition is never reached in the datasets analyzed in this thesis, since the iterative procedure always converges.

4.1.3 Calibration of the current signal

For each of the M SMBG samples used for calibration, the parameter vector $\hat{\mathbf{p}}$, estimated from Equation 4.9 by following the five-step procedure of Section 4.1.2, is used to calibrate in real-time the electrical current signal $y_I(t)$, by inverting the model of Equation 4.3:

$$z(t, \hat{\mathbf{p}}) = \frac{y_I(t)}{s(t, \hat{s}_2, \hat{s}_3)} - \hat{b}. \tag{4.16}$$

In particular, since the SMBG samples are acquired at times t_i, $i = 1, 2, ..., M$, the parameter estimated at the i-th calibration are used to calibrate the current signal from $t_i + 5$ min to $t_{i+1} + 5$ min. Indeed, the deconvolution window $\mathbf{\Lambda}$ is defined to end 5 min after the reference time of the BG measurement (in order to avoid edge effects), thus introducing the need to wait 5 min from any t_i before starting a new calibration.

4.2 Implementation

The BMD algorithm is assessed using a 9-fold cross-validation technique. The database under analysis is divided into 9 groups of 12 datasets each. Iteratively, one group was the test set used to evaluate the method, whereas all other groups formed the training set on which the priors for the calibration parameters are derived.

4.2.1 Prior derivation

The a priori information on calibration model parameters is derived by identifying the following nonlinear model on the training data set:

$$y_I(t) = [(u_B(t) * h(t)) + b] \cdot s(t) \tag{4.17}$$

where $*$ stands for convolution, $h(t)$ is as in Equation 4.2, $s(t)$ is defined by Equation 4.4 and depends on parameters s_1, s_2, and s_3. The unknown parameters s_1, s_2, s_3, b, and τ are estimated by nonlinear least squares. The input of the identification procedure is a smoothed BG profile $u_B(t)$, obtained from YSI references using a stochastic Bayesian smoother [89] and the output is the electrical current signal $y_I(t)$. The identification process is performed on each of the N_t time series of the training set, thus obtaining N_t values for each parameter. In particular, defining $\mathbf{\Gamma} = \tau_1, ..., \tau_{N_t}$ the values of parameter τ, a Bayesian prior is built assuming a priori distribution with mean μ_τ and variance σ_τ^2, determined from samples in $\mathbf{\Gamma}$:

$$\begin{aligned} \mu_\tau &= \frac{1}{N_t} \sum_{k=1}^{N_t} \tau_k \\ \sigma_\tau^2 &= \frac{1}{N_t - 1} \sum_{k=1}^{N_t} (\tau_k - \mu_\tau)^2 \end{aligned} \tag{4.18}$$

Similarly, we built a Bayesian prior for parameter vector $\mathbf{p}$.

In the online working modality (see Section 4.1), since the number of BG measurements available was not sufficient to estimate an individual value for τ, we fixed its value to the prior mean μ_τ obtained from the training set at each iteration of the cross-validation. With regard to the other parameters, given the stability of the ratio between parameters s_1 and s_2, to facilitate model identifiability in online modality (when only few BG references are available) we decided to fix the ratio $\frac{s_1}{s_2}$ to the constant φ. The value of φ is estimated, at each iteration of the cross-validation, from the mean value of its distribution on the correspondent training set of size N_t. Thus, Equation 4.6 represents the final parameter vector. Defining $\mathbf{P} = \mathbf{p}_1, ..., \mathbf{p}_{N_t}$ the samples of the parameters vector $\mathbf{p}$, a Bayesian prior is built by assuming a distribution with prior mean μ_p and prior covariance matrix $\mathbf{\Sigma_p}$, where the i-th element of μ_p, μ_i, and the

ij-th element of $\mathbf{\Sigma_p}$, σ_{ij}, are defined as follows:

$$\begin{aligned} \mu_i &= \frac{1}{N_t} \sum_{k=1}^{N_t} p_{k,i} \\ \sigma_{ij} &= \frac{1}{N_t - 1} \sum_{k=1}^{N_t} (p_{k,i} - \mu_i)(p_{k,j} - \mu_j) \end{aligned} \tag{4.19}$$

Training set data have been used also to estimate the error covariance matrix $\mathbf{\Sigma_w}$, used in Equation 4.9. The error variance is assumed constant over time and its value is estimated from the distribution of the differences between SMBG measurements and the correspondent calibrated values (in mg/dl) given by the manufacturer. In particular, for all SMBG samples available in the training set the correspondent CGM calibrated value is matched following the procedure described in Section 2.1.2.1 and the difference between the two measurements is computed. The error variance is thus obtained, at each cross-validation iteration, from the distribution of the SMBG–CGM error on the training set.

4.2.2 Calibration scenarios

The BMD calibration algorithm is applied on the test set by simulating an online working modality. In particular, for each sensor in the test group, the raw current signal $y_I(t)$ is calibrated by exploiting a set of BG references provided by SMBG measurements. Different calibration schedules are tested, in order to assess the accuracy of the calibrated profiles using a different number of SMBG samples per day, i.e., varying the frequency at which parameters of the calibration model are updated. In particular, apart from the first calibration, which is always performed about 2 hours after sensor insertion (exploiting a pair of SMBG samples acquired a few minutes of distance from each other), the following schedules are tested:

- one calibration about every day;
- one calibration about every two days;
- one calibration about every four days;
- zero calibrations (except the initial one).

4.3 Results

The following subsections analyze the results of the implementation of the BMD calibration algorithm in the DG4P dataset (Section 4.3.1) and in the NGD dataset (Section 4.3.2). Then, in Section 4.3.3, we compare the performance of the BMD calibration algorithm in the two datasets.

4.3.1 Results on the DG4P sensor

The calibrated profiles obtained with the three scenarios listed in Section 4.2.2 and the original manufacturer CGM output are shown, for a representative subject of the database, in Figure 4.3, where also YSI references (not used by the calibration procedure) are reported. Top panel refers to the original manufacturer calibration (performed twice per day), whereas the other panels to the new BMD calibration algorithm and, in particular, from top to bottom, to the one-every-day, one-every-two-days and one-every-four-days calibration scenarios. The calibrated profiles obtained with the BMD calibration algorithm are, independently from the number of calibrations, more accurate than the original CGM output given by the manufacturer. In particular, MARD, PAGE and CEGA-A resulted, respectively, 15.04%, 74.26%, 66.18% (original manufacturer calibration, on average two calibrations per day), 8.85%, 97.06%, 94.12% (BMD algorithm, one calibration per day), 8.94%, 97.06%, 92.65% (BMD algorithm, one calibration every two days), and 10.15%, 93.38%, 87.50% (BMD algorithm, one calibration every four days).

The results on the entire dataset are reported via boxplot in Figure 4.4, where the distributions of the three performance metrics, both for the original calibration and for the BMD algorithm (with different calibration frequencies), are shown. In general, the BMD algorithm appears to be more accurate than the original manufacturer calibration, for all the considered metrics and independently from the frequency of calibrations. Indeed, the metrics obtained for the BMD calibration algorithm show distributions concentrated at lower MARD values and higher PAGE and CEGA-A values with respect to the original manufacturer calibration. In particular, the BMD algorithm with one calibration every four days compared to the original manufacturer calibration (on average two calibrations per day) shows 11.62% MARD (vs 12.83%), 89.20% PAGE (vs 80.62%) and 87.5% CEGA-A (vs 81%).

Numeric values of performance metrics and relative statistical analysis re-

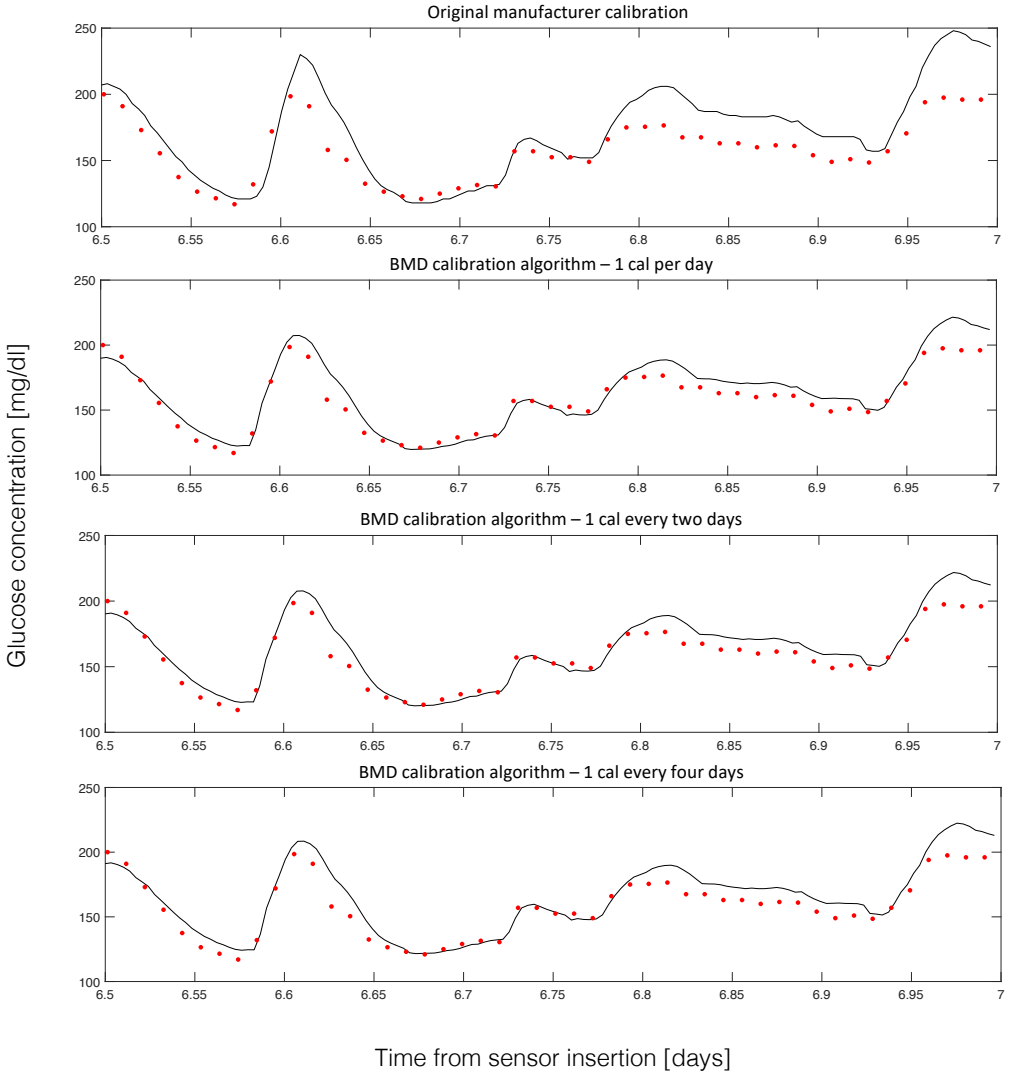

Figure 4.3: Calibrated CGM profiles (continuous lines) and laboratory references for accuracy assessment (points) for a representative subject during in-clinic session on day 7. From top to bottom: original manufacturer calibration (performed two times per day), Bayesian multi-day (BMD) algorithm with one calibration per day, one calibration every two days and one calibration every four days.

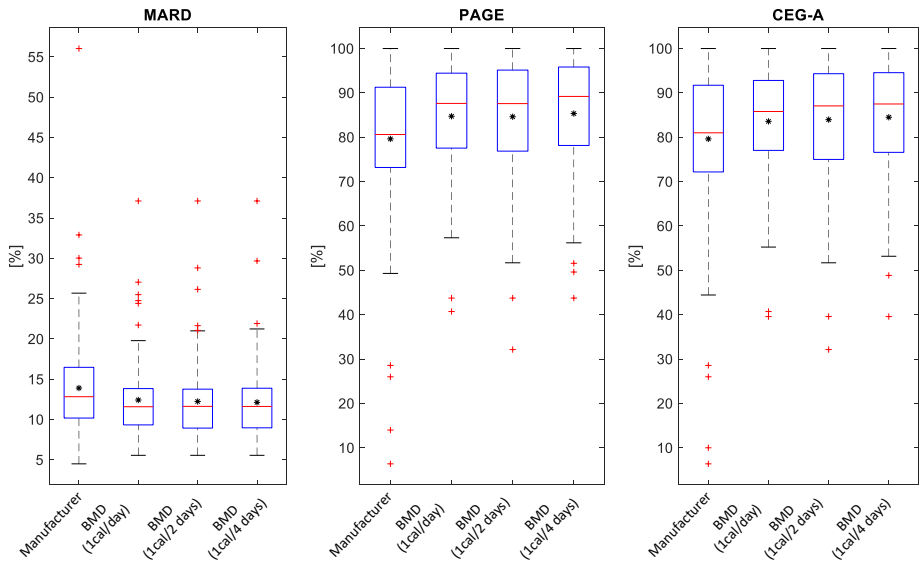

Figure 4.4: Boxplot of performance metrics. From left to right, MARD, PAGE and percentage of points in A zone of the CEG obtained on the test sets for the original calibration (performed twice per day) and for the BMD calibration algorithm with three different calibration frequencies (one calibration per day, one calibration every two days and one calibration every four days). The stars represent the mean values of each distribution.

sults are reported in Table 4.1. Columns 2–5 report, respectively, the median values of the performance metrics obtained with the original calibration, the BMD algorithm with one calibration per day, one calibration every two days and one calibration every four days. Columns 6–8 report the correspondent p-values by the Wilcoxon signed-rank test. The indexes reported in Table 4.1 show that the BMD calibration algorithm improves sensor accuracy compared to manufacturer calibration, independently from the frequency of calibrations. The improvement achieved with the BMD algorithm is statistically significant ($P<0.05$) for all the considered metrics and for all the calibration frequencies tested. In addition, no statistically significant difference ($P>0.05$) is found between the performance metrics distributions obtained with the BMD algorithm by using different calibration frequencies, i.e., one per day vs one every two days vs one every four days (P not shown).

It is interesting to consider also an extreme scenario where no calibration, except the initial one, is performed. In this zero calibrations scenario, the calibration parameters are estimated about two hours after sensor insertion and used for the entire monitoring session without any further update. Since the estimation is performed in the Bayesian setting exploiting only a pair a SMBG samples, it strongly relies on prior information. In this case, the following me-

Table 4.1: Performance metrics (median values).

Metric	Original calibration	BMD algorithm (1cal/day)	BMD algorithm (1cal/2days)	BMD algorithm (1cal/4days)	P (original vs BMD 1cal/day)	P (original vs BMD 1cal/2days)	P (original vs BMD 1cal/4days)
MARD	12.83	11.59	11.63	11.62	0.0122	0.0025	0.0058
PAGE	80.62	87.63	87.59	89.20	0.0017	0.0010	0.0012
CEG-A	81.00	85.81	87.07	87.50	0.0442	0.0092	0.0112

dian values are obtained: 12.03% MARD, 85.67% PAGE and 85.62% CEG-A. Although these indexes are still slightly better than those of the original manufacturer calibration (see Table 4.1), no statistical significant differences are observed. Indeed, the larger the number of calibration references (as moving from day 1 to day 7), the more the posterior distributions differentiate from the priors. This phenomenon is depicted in Figure 4.5, where the posterior distributions of the three calibration parameters are reported, centered with respect to the a priori expected values. We can observe that, not surprisingly, as moving from day 1 to day 7, i.e., as the number of calibration references increases, the posterior distributions differentiate from the priors (this is more evident for parameters $\boldsymbol{b}$ and $\boldsymbol{s_2}$). Notably, the prior distributions are quite flat, much more than the posterior distributions. This suggests that the information brought into the Bayesian estimation process by the calibration references plays a key role in determining the parameter values and that the use of only one calibration reference would let the estimate to excessively rely on prior information.

4.3.2 Results on the NGD sensor

Figure 4.6 shows the result of the application of the BMD calibration algorithm to the raw current measured by the next-generation Dexcom CGM sensor prototype on day 10 of a representative individual selected from the database. YSI reference values acquired during in-clinic session are also reported for accuracy assessment. From top to bottom, the panels refer to the one-per-day, one-every-two-days, one-every-four-days, and the zero calibration scenarios, respectively. The calibrated profiles match YSI references with good degree of accuracy, independent from the frequency of calibrations. Slightly higher values of MARD are observed when passing from the one-per-day calibration scenario to the one-every-two-days scenario (from 9.8% MARD to 11.3% MARD), whereas performance remains substantially stable when further re-

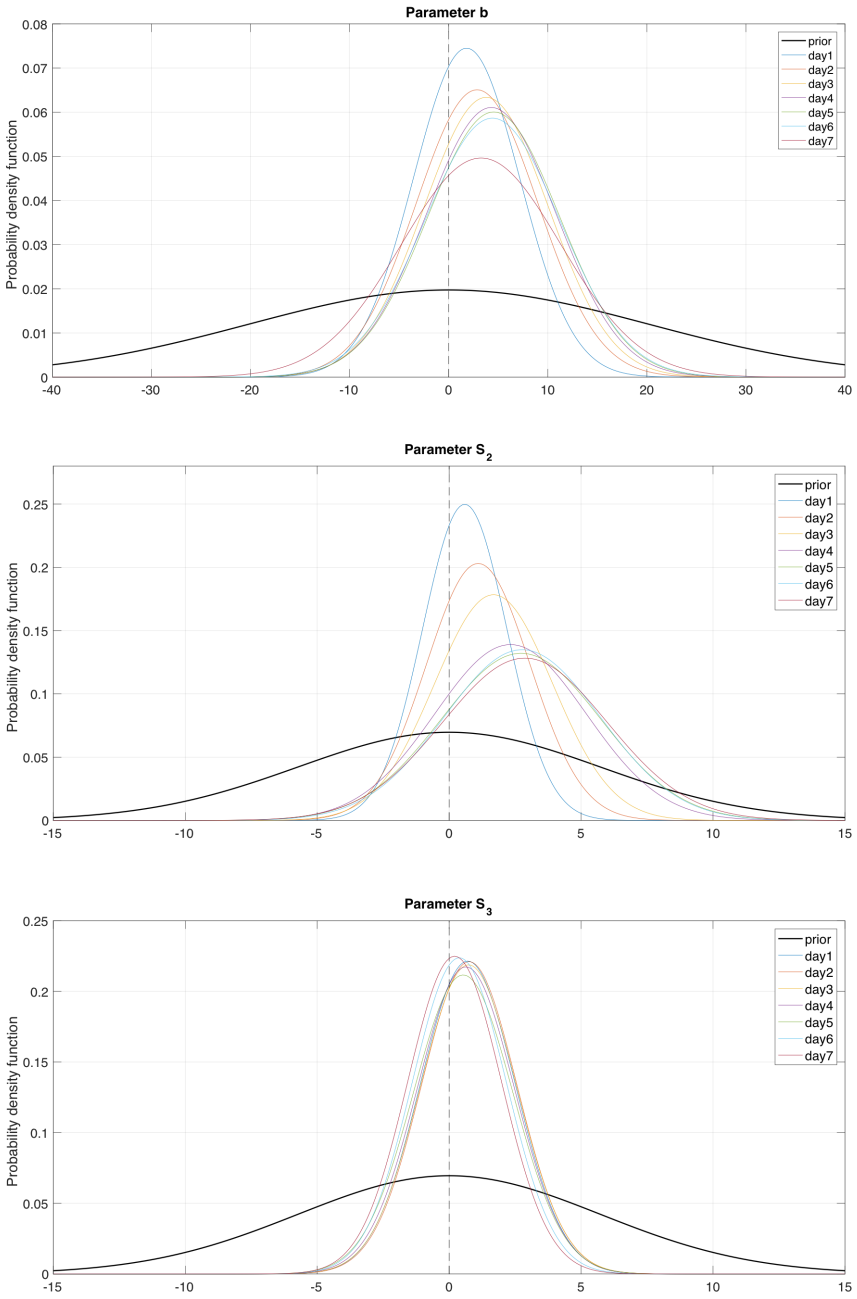

Figure 4.5: Posterior distributions of the three calibration parameters as moving from day 1 to day 7 (each day is depicted with a different color, as reported in the legend) and respective a priori distributions (black curve). From top to bottom: parameter b, s_2, and s_3. All distributions have been centered with respect to the a priori expected values.

Table 4.2: Performance metrics, subject-level analysis, as median[iqr] or mean(sd) for non-normal/normal distributed metrics, respectively.

Metric, unit	One/day calibration	One/two days calibration	One/four days calibration	Zero calibrations
Pairs, #	24 [1.7]	24 [1.7]	24 [1.7]	24 [1]
MAD, mg/dl	16.7 (6.6)	16.7 (7.8)	15.6 [9.5]	16.2 [12.7][a,b]
MARD, %	8.5 [8]	8.4 [6.7]	9.2 [7.9]	9.3 [8.9][a,b]
MRD, %	0.4 (9.4)	0.6 (9.6)	0.7 (10.6)	-0.4 (12.1)
15/15%, %	87.5 [32.2]	85.1 [29.2]	85 [38.7]	82.2 [41.5][a,b]
20/20%, %	95.8 [22.7]	95.8 [17.4]	95.7 [17.4]	91.7 [29.2][a,b,c]
30/30%, %	100 [0]	100 [4.2]	100 [4.2]	100 [4.2]
CEG—A, %	95.8 [22.8]	95.7 [17.4]	95.7 [18.2]	91.5 [29.2][a,b,c]

[a]Statistically different from one-per-day calibration scenario.
[b]Statistically different from one-every-two-days calibration scenario.
[c]Statistically different from one-every-four-days calibration scenario.

ducing calibrations to one-every-four-days or even to zero (11.5% MARD and 11.6% MARD, respectively).

Summary performance of the subject-level analysis on the entire data set is reported in Table 4.2. All metrics are computed for each subject, and then the distribution among the population in considered. In particular, from column two to column five, performance refers to the one-per-day, one-every- two-days, one-every-four-days, and the zero calibration scenarios, respectively. Performance appears substantially stable when progressively reducing the frequency of calibrations from one-per-day to one-every-two-days and one-every-four-days, with no statistically significant differences (P values >0.05) between the three scenarios for any of the considered metrics. In particular, the very small performance deterioration between the first two scenarios shows that the use of a priori knowledge compensates for the lack of SMBG data. When the frequency of calibrations is further reduced, this compensation is progressively less effective, resulting in slightly lower accuracy.

Although median MAD, MARD, and 30/30% remain stable even in the zero calibration scenario, percentage of values within 15/15%, 20/20%, and CEG-A slightly worsen of 3–4% when moving to the zero calibration scenario, where we observe some statistically significant differences.

Distributions of performance metrics MARD, ISO 20/20%, and CEG-A are also reported by boxplot in Figure 4.7, where the slight accuracy deterioration in the zero calibration scenario is evidenced. In general, the indexes re-

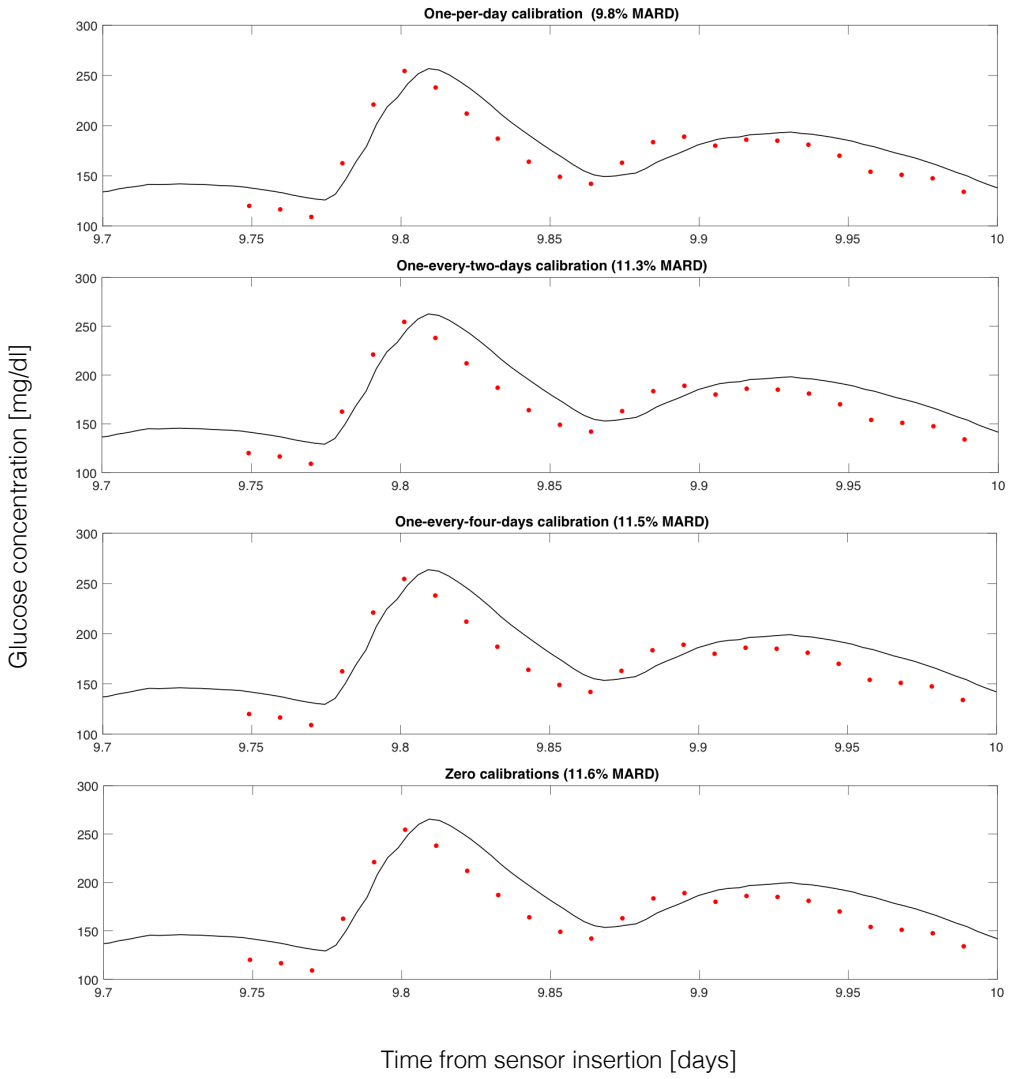

Figure 4.6: Calibrated CGM profiles (continuous lines) and laboratory references for accuracy assessment (points) for a representative subject during in-clinic session on day 10. From top to bottom, calibrated profiles refer to the one-per-day, one-every-two-days, one-every-four-days, and zero calibration scenarios.

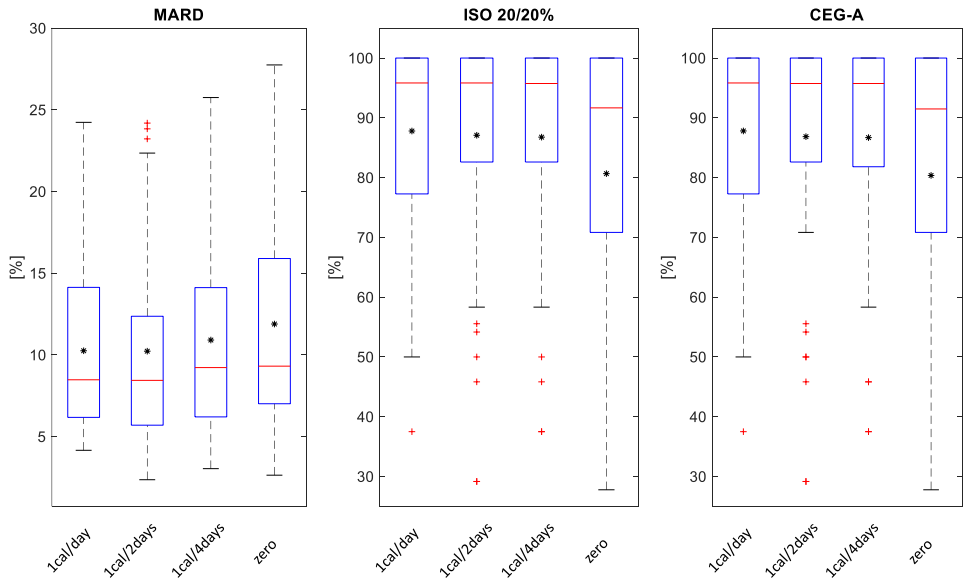

Figure 4.7: Boxplot of performance metrics. From left to right, MARD, percentage of points matching the International Organization for Standardization (ISO) 20/20%, and percentage of points in A zone of the CEG obtained with one calibration per day, one calibration every two days, one calibration every four days and zero calibrations. The stars represent the mean values of each distribution.

ported in Table 4.2 and relative distributions of Figure 4.7 show that the new BMD calibration algorithm insures good sensor accuracy, independent from the frequency of calibrations. The median MARD is lower than 10% for all the calibration scenarios assessed, including the extreme scenario where zero calibrations are performed. Notably, all the scenarios tested present the median value of 100% data points within 30/30%.

Moreover, when looking at the calibration error with its sign (MRD), we notice that there is no systematic error introduced by the calibration process. Indeed, as shown in Figure 4.8, boxplots of the error distribution in the different calibration scenarios are well centered around zero, with no statistically significant differences when tested for the null hypothesis that their median/mean (for non-normal/normal distributions) is equal to zero.

Table 4.3 reports the overall accuracy analysis performed on all CGM-YSI data pairs available. From left to right, metrics refer to the one-per-day, one-every-two-days, one-every-four-days, and to the zero calibration scenarios, respectively. For each scenario, we analyze the overall performance on the entire monitoring period and the performance in different phases of the CGM session, that is, initial phase (day 1), middle phase (days 4–6), and end phase (day

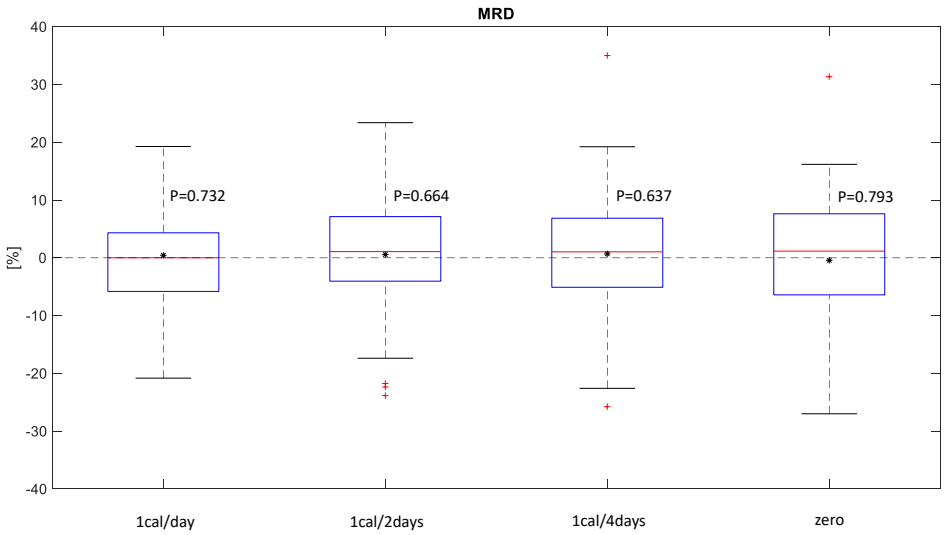

Figure 4.8: Boxplot of calibration error in terms of mean relative difference (MRD). Left to right: one-per-day, one-every-two-days, one-every-four-days, and zero calibrations. The asterisk inside each box represents the mean of the distribution, whereas crosses represent outliers. For each box, the P value refers to the result of a Wilcoxon signed-rank test for non-normally distributed data and a t-test for normally distributed data with the null hypothesis that data come from a distribution whose median/mean (for non-normal/normal distribution) is zero.

10), as documented in Section 2.2.2. Overall, median AD and ARD do not increase significantly when reducing the frequency of calibrations. In particular, median AD and ARD in the zero calibration scenario differs from median AD and ARD in the one-per-day calibration scenario for less than 1 mg/dl and 1%, respectively. No significant differences are observed in 15/15%, 20/20%, and 30/30% metrics and CEG-A when reducing calibrations from one-per-day to one-every-two and one-every-four-days, whereas a slightly worsening of 4–5% occurs when moving to the zero calibration scenario. In addition, no systematic error is found when looking at the signed error RD (P values not shown).

Performance is stable among the different phases of the CGM session (day 1, days 4–6, and day 10) for all the considered metrics in the one-per-day, one-every-two-days, and one-every-four-days calibration scenarios. In particular, no performance deterioration is observed on day 1, which is usually the most critical day for what concerns sensor accuracy [16], [94], [105]. In the zero calibration scenario, the best accuracy is obtained, not surprisingly, in days 4-6. These days correspond to the middle of the session, where the sensor behavior is expected to be more stable. Day 1 performance is still comparable with that

Table 4.3: Performance metrics, overall accuracy analysis.

Metric, unit	Session	One-per-day calibration	One-every-two-days calibration	One-every-four-days calibration	Zero calibrations
Pairs, #	Overall	1198	1197	1198	1264
	Day 1	557	556	557	626
	Day 4–6	331	331	331	328
	Day 10	310	310	310	310
AD, mg/dl[a]	Overall	14.2 [17.6]	13.9 [17.7]	14.9 [18]	15 [19.6]
	Day 1	14.5 [17.8]	13.4 [18.7]	14.9 [18.2]	15 [18.7]
	Day 4–6	13.6 [16.2]	14.3 [15.1]	13.7 [16.6]	12.6 [19.2]
	Day 10	14.1 [18.9]	14.2 [19.6]	16.2 [18.5]	16.9 [22.9]
ARD, %[a]	Overall	8 [10.1]	8 [11.2]	8.9 [11.2]	8.7 [13.5]
	Day 1	8.2 [10.5]	7.8 [12.2]	8.3 [11.9]	9.2 [12.5]
	Day 4–6	7.9 [8.8]	7.9 [9.5]	8.4 [11.8]	7.5 [14.1]
	Day 10	7.9 [10.3]	8.3 [10.8]	9.6 [9.1]	9.8 [14.4]
RD, %[a]	Overall	[16.0] 1.2	[15.2]	1.5 [17.5]	0.5 [17.6]
	Day 1	-0.9 [16.7]	-1.2 [15.1]	-1.4 [16.5]	0.6 [19.2]
	Day 4–6	4.4 [14.6]	5.6 [13.4]	5.3 [14.6]	2.5 [15.2]
	Day 10	-0.5 [16.2]	0.4 [16.6]	2.2 [19.6]	-3.2 [19.5]
15/15%, %	Overall	78.3	76.2	74.5	70.6
	Day 1	77.6	74.1	73.1	71.6
	Day 4–6	80.1	79.5	73.7	72.9
	Day 10	77.7	76.4	78.1	66.4
20/20%, %	Overall	87.3	86.9	86.2	81.1
	Day 1	88.5	86.7	87.1	81.5
	Day 4–6	86.4	87.3	84.6	83.8
	Day 10	86.1	86.8	86.5	77.4
30/30%, %	Overall	96.8	96.5	96.3	94.3
	Day 1	97.5	97.1	97.7	94.6
	Day 4–6	96.1	95.8	95.5	93.3
	Day 10	96.4	96.1	94.8	94.8
CEG—A, %	Overall	87.2	86.4	86.1	80.7
	Day 1	88.5	86	86.7	81.1
	Day 4–6	86.1	87	84.6	83.5
	Day 10	86.1	86.4	86.4	76.8

[a]Median [iqr].

obtained overall, besides including more data points since there is no need to wait for the start-up calibration instead required for the other three scenarios, whereas the end of the session (day 10) shows slightly lower accuracy.

4.3.3 Comparison between DG4P and NGD performance

In this section, we compare the performance of the BMD calibration algorithm on the NGD dataset with that obtained in the present generation DG4P sensor.

Two representative calibrated profiles are reported in Figure 4.9. Top panel refers to the present generation and bottom panel to the next generation. In both panels, the continuous line represents the glucose profile obtained with one-per-day calibration and the dashed line represents the glucose profile when zero calibrations are performed. YSI measurements are also reported for visual accuracy evaluation. In the present generation, we can observe a significant accuracy deterioration when passing from one-per-day calibration to zero calibrations (from 10.9% to 18% MARD). Contrarily, the next-generation sensor appears more stable in maintaining good sensor accuracy when moving from the one-per-day to the zero calibration scenario (from 9.9% to 11.1% MARD). Notably, while systematically lower glucose values are observable in this specific subject, in others, glucose concentration results overestimated.

A global performance comparison is reported in Figure 4.10, where population median MARD values are reported for the present generation (plain circles) and for the next generation (solid circles) when progressively reducing calibrations from one-per-day to one-every-two-days, one-every-four-days, and zero. We observe that the next-generation sensor prototype always performs better than the present-generation sensor, independent from the frequency of calibrations. Indeed, the median MARD values always remain below 10%, whereas, in the present generation, values are always above 10%. Moreover, while the present-generation performance significantly worsens when moving to the calibration-free scenario, the next generation appears to reach an asymptotic MARD value, which does not deteriorate neither in the calibration-free scenario.

4.4 Summary and concluding remarks

In this chapter, we proposed a new calibration algorithm which employs a calibration model valid for the entire monitoring period and a Bayesian estimation

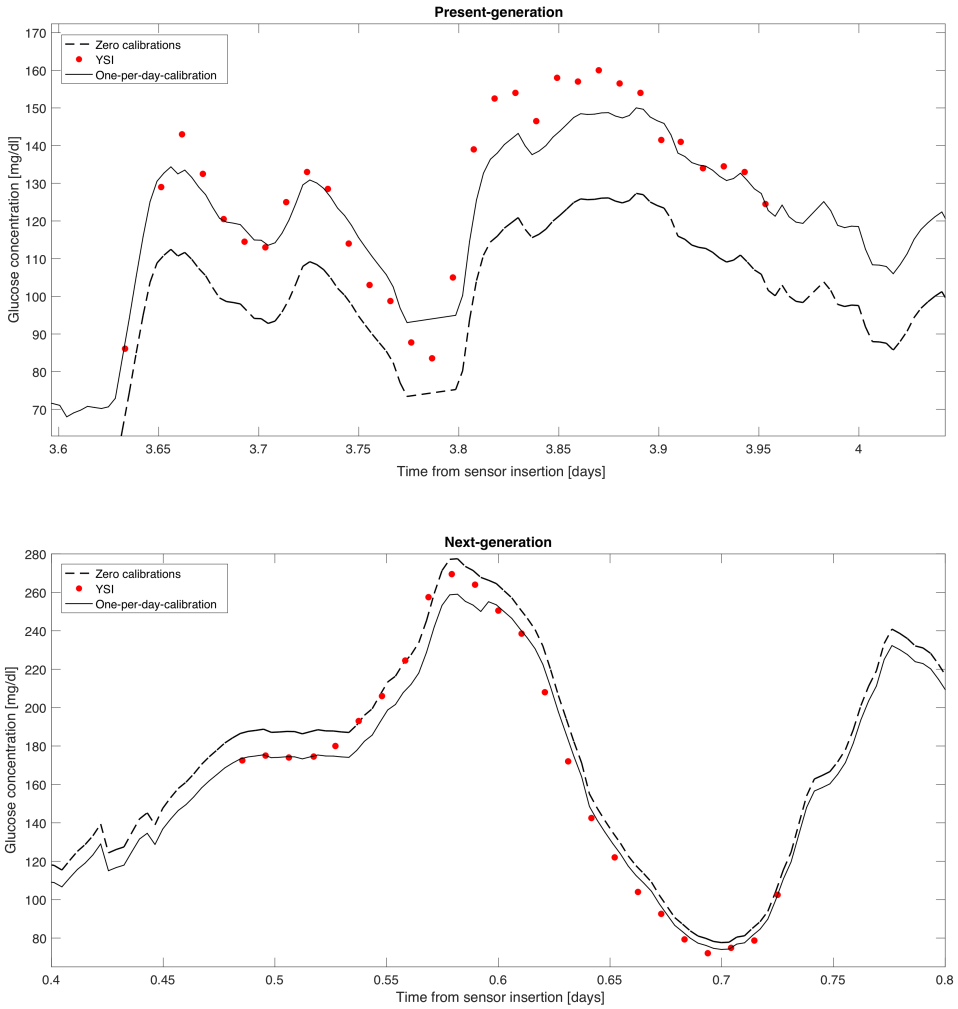

Figure 4.9: Calibrated CGM profiles in the one-per-day calibration scenario (continuous line) and in the zero calibration scenario (dashed line), and reference measurements (dots). Top panel refers to the DG4P CGM sensor and bottom panel refers to the NGD CGM sensor prototype, in two subjects representative of the mean behavior in the population.

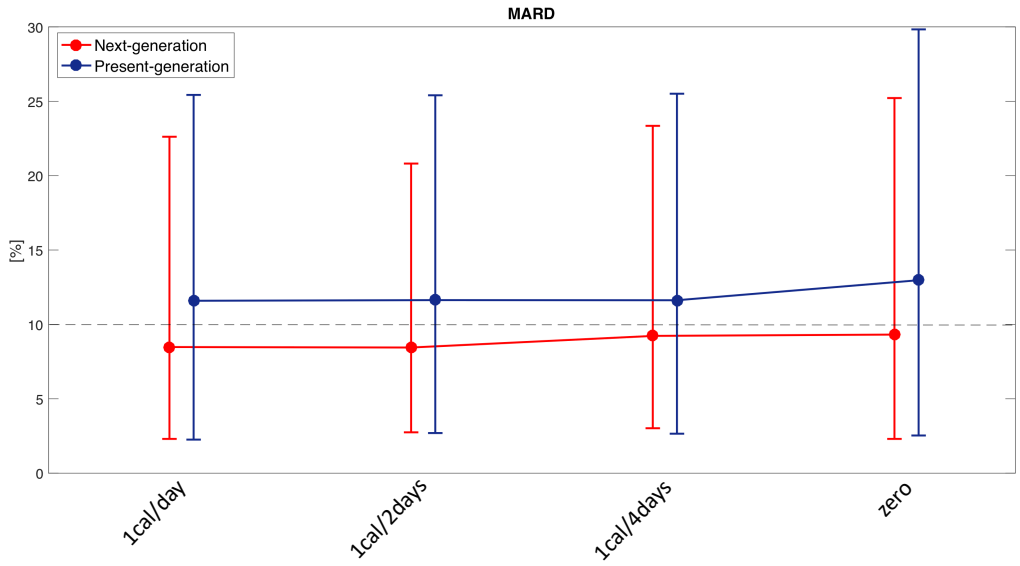

Figure 4.10: MARD, median values and relative interquartile ranges in the NGD dataset (solid circles) and in the DG4P dataset (plain circles). From left to right, values refer to the one-per-day, one-every-two-days, one-every-four-days, and the zero calibration scenarios.

procedure determining the values of the unknown parameters.

The proposed BMD calibration algorithm provides real-time estimation of glucose concentration by processing the raw electrical current signal and a few BG measurements, taking into account the BG-to-IG kinetics and employing a nonlinear model to describe the variability of sensor characteristics over time. The estimation of calibration model parameters is set in the Bayesian framework, where priors derived from a training set data ensure identifiability, even using only a few BG calibration references.

Performance of the BMD calibration algorithm was assessed on two different datasets, one of the present generation DG4P sensor, and one of a next generation CGM sensor prototype (NGD dataset). We tested different calibration scenarios, where we progressively reduced the frequency of calibrations from one per day until zero.

Results show that, in the DG4P data, the BMD algorithm provides a statistically significant accuracy improvement compared to the original manufacturer calibration (required at least every 12 h), independently from the number of calibrations performed (one per day, one every two days, one every four days). In particular, with the BMD algorithm, two calibrations over a period of 7 days are sufficient to guarantee a good sensor accuracy, i.e., 11.6% MARD

(vs. 12.8% MARD obtained by the original calibration by exploiting at least 14 calibrations over the 7-day period). The more challenging calibration-free scenario showed instead poorer performance, with no statistically significant difference compared to the original calibration.

In the NGD data, performance of the BMD algorithm is even better, with MARD always lower than 10%, independently from the frequency of calibrations, and even when zero calibrations are performed. In particular, this result is in line with a recent achievement obtained by the company producing the NGD sensor, Dexcom Inc. (San Diego, CA), which in 2018 has launched the new-generation G6 device (already mentioned in Section 1.2.3) showing as well lower than 10% MARD with zero calibrations [59].

The comparison between present and next generation sensors shows that the BMD calibration algorithm allows obtaining more accurate CGM traces when applied to the next generation technology compared to its application to the present generation (8.5% vs. 11.59% MARD with one-per-day calibration, 8.4% vs. 11.63% MARD with one-every-two-days calibration, 9.2% vs. 11.62% MARD with one-every-four-days calibration, and 9.3% vs. 12.97% MARD with zero calibrations). In particular, the NGD technology seems more suited for a calibration-free scenario than the present generation.

To conclude, we would highlight possible limitations and relative margins of improvement of the proposed calibration technique. One important aspect to be considered is, for instance, the investigation of how accuracy is impacted when considering larger datasets with more variability in sensor-to-sensor characteristics. Indeed, the Bayesian approach here proposed relays on prior information on calibration model parameters to reduce or actually substitute (in the zero calibration scenario) the BG calibration references. In principle, the lower the inter-sensor variability is, the higher is the accuracy of the method, especially when only few, or zero, calibration references are used. When instead potential higher inter-subject or inter-sensor variability is considered, the method would require additional flexibility to efficiently cope with the different sensor behaviors. To address this issue, in the following chapter we propose a multi-model approach and a Bayesian model selection framework which aim at guaranteeing wide margins of flexibility for both the determination of the most appropriate calibration model and the numerical values of its unknown parameters.

Chapter 5

Dealing with CGM sensors-subject variability: a multi-model Bayesian framework (MMBF) [1]

The calibration algorithms discussed so far rely on a unique, predetermined, calibration model, whose parameters are periodically updated in individual patients by matching the reference blood glucose values. However, in presence of high variability of sensor-subject characteristics, the use of a unique calibration model could be not sufficient to guarantee optimal performance in the entire population of sensor/subjects. In this chapter, we propose a Bayesian model selection approach aimed at enhancing calibration performance by allowing the choice of the most appropriate calibration model among a finite set of candidates, each having a given a priori probability. Ideally, the MMBF can be applied as preliminary step to determine the calibration model to any calibration algorithm. Here, we apply the MMBF to the BMD calibration algorithm of Chapter 4 and assess performance on synthetic data.

As we have seen so far, a calibration model can be described by the generic function f, which depends from parameters θ:

$$y_I(t) = f(\theta, u_B(\cdot)) + w(t) \tag{5.1}$$

where y_I is the raw electrical signal measured in the interstitial fluid, u_B the BG profile, and w the measurement error. To improve effectiveness of this model during sensor functioning, its parameters θ are periodically updated by using

[1] The methods described in this chapter, published in [106], have been of inspiration for a US Provisional Application (details subject to confidentiality).

reference SMBG values collected by the patient through finger prick devices.

All the algorithms proposed so far rely on a unique, i.e., population, calibration model described by the function f, whose parameters θ are tuned "in vivo" to adapt to the specific sensor and to the specific subject in which the sensor is used. This process assumes that the relationship between electrical current measured by the sensor and BG concentration can be described by a standard function f in the entire population, requiring both the sensor-to-sensor and the inter-subject variability to be minimized. Although sensors' manufacturers can reduce "in factory" the sensor-to-sensor variability, there is a second source of "in vivo" variability caused by the specific body-sensor interactions that influence the calibration function f, determining possible different behaviors in the population [107], [74].

In the following, we propose a new calibration framework to enhance flexibility in the choice of the calibration function f with the aim of compensating for the "in vivo" variability of sensor characteristics.

5.1 Calibration framework incorporating calibration model selection

5.1.1 Model selection in a Bayesian embedding

The Bayes theory offers a straightforward way to deal with model uncertainty, in particular when different models are under consideration to describe data [108]. In particular, let D be the data and $M = M_1, ..., M_K$ be a set of K models under consideration for D. The probability density function of D under the specific model $M_i, i = 1, ..., K$ is given by $p(D|\theta_i, M_i)$, where θ_i is the vector of unknown parameters for M_i. A priori distributions are introduced for all the unknowns, i.e., the parameters of each model, $p(\theta_i|M_i)$, and the models themselves, $p(M_i), i = 1, ..., K$. In this probabilistic setup, the data D can be seen as generated by the following three-step hierarchical procedure [108]:

- the model M_i is generated from $p(M_i), i = 1, ..., K$, where $\sum_{i=1}^{K} p(M_i) = 1$;
- given the model M_i, the relative parameter vector θ_i is generated from $p(\theta_i|M_i)$;
- finally, the data D are generated from $p(D|\theta_i, M_i)$.

Given the observed data D, the model selection problem consists in finding the model in M that actually generated the data. From Bayes' theorem, the probability that a specific model $M_i, i = 1, ..., K$ was the model generated at step 1 is:

$$p(M_i|D) = \frac{p(D|M_i) \cdot p(M_i)}{\sum_{i=1}^{K} p(D|M_i) \cdot p(M_i)} \tag{5.2}$$

where $p(M_i|D)$ is the posterior probability of model M_i, $p(M_i)$ is the a priori probability of model M_i, and $p(D|M_i)$ is the integrated likelihood of model M_i, which is expressed as follows:

$$p(D|M_i) = \int p(D|M_i, \theta_i) \cdot p(\theta_i|M_i) d\theta_i \tag{5.3}$$

The posterior distributions $p(M_i|D), i = 1, ..., K$, which can be computed by different numerical integration strategies [109], [110], are the fundamental objects for model selection, providing a full post-data representation of model uncertainty.

In the following sections, we contextualize this general Bayesian model selection approach to the specific framework of CGM sensors' calibration.

5.1.2 Incorporation of Bayesian model selection in the calibration problem

Given a set of measures D, i.e., a set of SMBG samples acquired through finger-prick devices and correspondent raw sensor data measured at the same time instants, the calibration model is determined among the candidates in M by choosing the model with the highest a posteriori probability. The two major challenges for the practical implementation of this procedure are: i) the definition of all a priori probabilities (i.e., $p(M_i)$ and $p(\theta_i|M_i)$ for $i = 1, ..., K$), and ii) the numerical computation of the integrated likelihood of Equation 5.3. These two aspects are discussed in the following subsections.

Definition of a priori probabilities

A priori probabilities for the candidate models and a priori probabilities for each model parameter can be derived by retrospective off-line analysis of CGM sensor data. For instance, given a training data set DT containing N_T raw CGM signals and reference BG values, the model M_i that best fits the data can be determined by ad hoc criteria (e.g., Akaike criterion) for each record in DT.

Then, the a priori probability of a specific model M_i will be derived from the number of times N_i that the model was chosen as the best fit in DT, i.e., $\frac{N_i}{N_T}$.

In an analogous way, a priori probabilities for model parameters can be derived from DT by off-line identification. Specifically, let θ be the parameters vector for the generic model M, which was the best fit for N records in DT. We assume that θ is distributed with prior mean μ_θ and prior covariance matrix Σ_θ, where the i-th element of μ_θ, μ_i, and the ij-th element of Σ_θ, σ_{ij}, are defined as follows:

$$\begin{aligned} \mu_i &= \frac{1}{N}\sum_{n=1}^{N}\theta_{n,i} \\ \sigma_{ij} &= \frac{1}{N-1}\sum_{n=1}^{N}(\theta_{n,i}-\mu_i)(\theta_{n,j}-\mu_j) \end{aligned} \tag{5.4}$$

Integrated likelihood computation

The integrated likelihood expressed in Equation 5.3, except for few elementary cases, cannot be evaluated analytically and requires the use of numerical integration methods [109]. Using the simple Monte Carlo integration, the marginal likelihood in Equation 5.3 is approximated by averaging the likelihoods values over a set of N_{iter} iterations:

$$\hat{p}(D|M_i) = \frac{1}{N_{iter}}\sum_{n=1}^{N_{iter}} p(D|M_i,\theta_n) \tag{5.5}$$

where, at each iteration, a set of model parameters θ_n is sampled from its prior distribution and the probability of the observed data given that specific set of parameters is computed. The estimated integrated likelihood is the average of the likelihoods of the sampled parameters values $\theta_n : n = 1, ..., N_{iter}$. A possible difficulty with this type of estimation is that most of the θ_n may have small likelihood values so that the simulation process would be inefficient requiring a large number of iterations to reach convergence. Eventually, the precision of the simple Monte Carlo integration can be improved by the use of variance reduction techniques such as importance sampling [109], [110], [111].

Given the prior probabilities and the integrated likelihoods for each candidate model, the model with higher a posteriori probability is chosen and used to calibrate the sensor signal, as described in the following section.

5.1.3 Use of the selected calibration model

Assume that the model selection procedure resulted in the selection of model M, described by the calibration function f as in Equation 5.1. Given a set of data, i.e., a set of SMBG and correspondent raw sensor measurements contained in vectors u_B and y_I, respectively, the parameter vector θ is estimated as described in Chapter 4 by Bayesian MAP estimation:

$$\hat{\theta} = argmin_{\theta}[u_B - \hat{u}_B(\theta)]^T \Sigma_w^{-1}[u_B - \hat{u}_B(\theta)] + (\mu_\theta - \theta)^T \Sigma_\theta^{-1}(\mu_\theta - \theta) \tag{5.6}$$

where $\hat{u}_B$ is the vector of estimated BG values obtained using the calibration model M (i.e., $\hat{u}_B = f^{-1}(\hat{\theta}, y_I)$), σ_w is the error covariance matrix, and the prior mean μ and covariance matrix Σ_θ are as in Equation 5.4. The complete parameter estimation procedure, which embeds the deconvolution step [89] to compensate for the BG-to-IG kinetics [69], was already described in detail in Chapter 4. We summarize here the main steps:

- according to the selected model M, initialization of parameters vector θ to its prior mean μ_θ;
- estimation of IG profile $\hat{u}_I(t)$ according to the selected calibration model M;
- estimation of BG profile, $\hat{u}_B(t)$, accounting for the distortion introduced by the BG-to-IG kinetics [69] by means of deconvolution [89];
- matching between SMBG measurements and $\hat{u}_B(t)$;
- update of parameters vector for the following iteration.

The estimated parameter vector $\hat{\theta}$ is used to obtain, in real-time, the calibrated CGM profile. The procedure is updated each time a new SMBG measurement is collected.

5.2 Assessment on simulated data

The MMBF is tested on a set of synthetic data generated by the UVA/Padova type 1 diabetes simulator (T1DS) with the patient behavior model of [112]. The T1DS is a widely established simulation tool which describes glucose, insulin, and glucagon kinetics in type 1 diabetes patients. It is accepted by the FDA

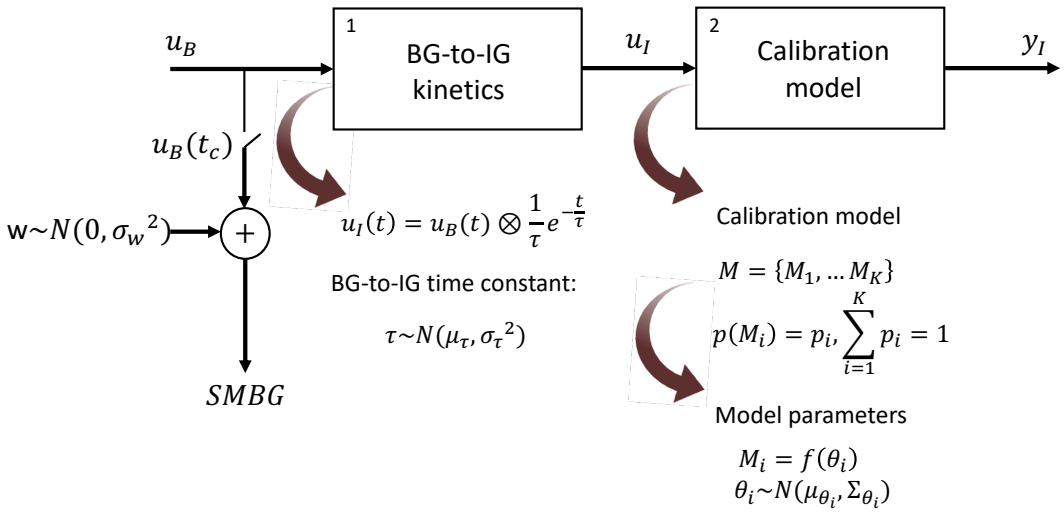

Figure 5.1: Schematic representation of the simulation setup. The blood glucose profile $\boldsymbol{u_B}$ is simulated using the UVA/Padova type 1 diabetes simulator (T1DS). SMBG measurements are obtained from BG samples at calibration times $\boldsymbol{t_c}$ by adding the measurement noise $\boldsymbol{w}$. The electrical current signal $\boldsymbol{y_I}$ is obtained from $\boldsymbol{u_B}$ by i) convolution with the BG-to-IG system response with time-constant $\boldsymbol{\tau}$ and ii) application of the calibration model $\boldsymbol{M}$ with parameter vector $\boldsymbol{\theta}$.

to test diabetes therapies. Here, we utilize its extended version which embeds modules describing glucose monitoring devices, e.g., SMBG and CGM devices, the patient's behavior in making treatment decisions, and insulin administration [112], allowing the simulation of realistic CMG time-series.

5.2.1 Generation of the simulated dataset

To generate the simulated dataset used to test the MMBF, we utilized the scheme of Figure 5.1.

For each subject, a BG concentration profile $\boldsymbol{u_B}$ is simulated using the T1DS [112]. From the BG profile, random SMBG measurements are generated at calibration times $\boldsymbol{t_c}$ (every 12h) according to the following model:

$$\boldsymbol{SMBG = BG(t_c) + w, \quad w \sim N(0, \sigma_w^2)} \tag{5.7}$$

where σ_w^2 is the noise variance (we assumed $\sigma_w = \mathbf{0.05} \cdot \boldsymbol{BG(t_c)}$). As per block 1 of Figure 5.1, the IG profile $\boldsymbol{u_I}$ is obtained from $\boldsymbol{u_B}$ accounting for the

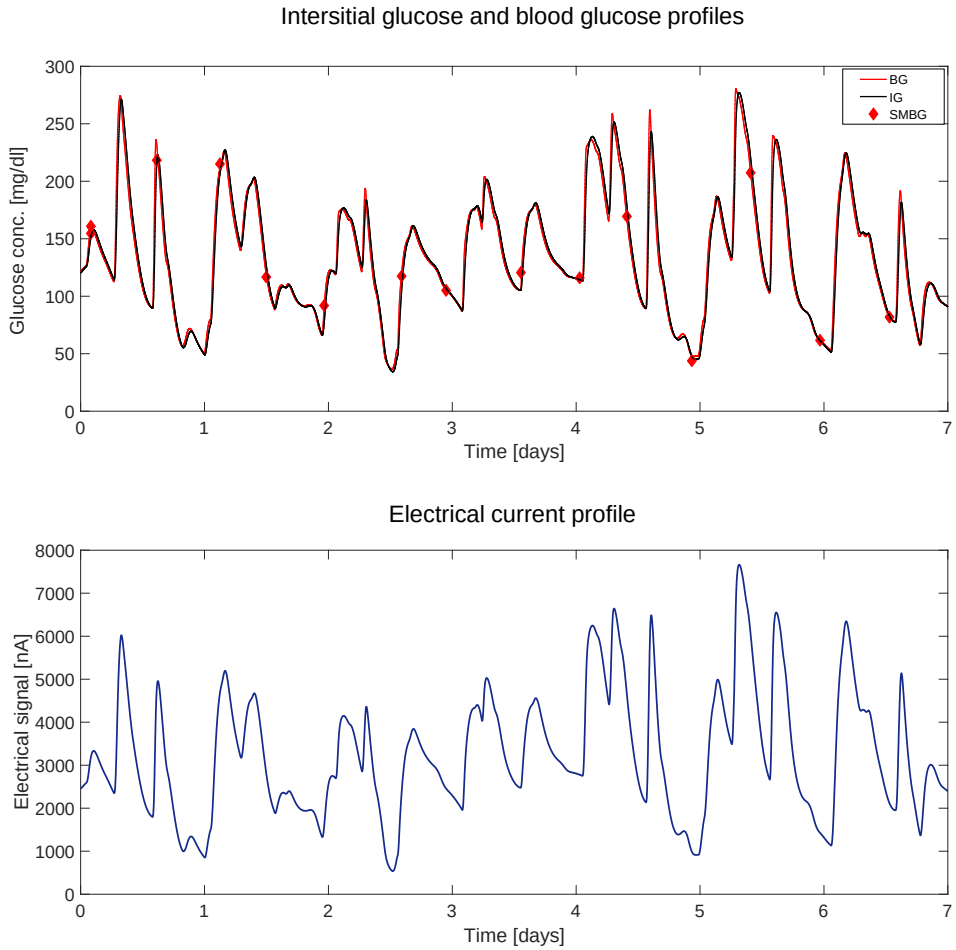

Figure 5.2: Example of simulated dataset. Top panel: blood glucose profile (continuous line), interstitial glucose profile (dashed line) and SMBG samples (diamonds). Bottom panel: electrical current profile.

BG-to-IG kinetics by means of convolution:

$$u_I(t) = u_B(t) * \frac{1}{\tau} \cdot e^{-\frac{t}{\tau}}, \quad \tau \sim N(10, 1.5) \tag{5.8}$$

Top panel of Figure 5.2 shows an example of simulated IG profile (dashed line) together with the BG profile (continuous line) and SMBG measurements (diamonds).

As far as block 2 of Figure 5.1 is concerned, here we simulate a basic scenario in which the calibration model is determined among a set of two candi-

dates, $M = \{M_a, M_b\}$:

$$\begin{aligned} M_a: \quad & u_I(t) = \frac{y_I(t)}{a_1 \cdot t + a_2} - a_3 \\ M_b: \quad & u_I(t) = \frac{y_I(t)}{b_1 + b_2 \cdot e^{-2t} - b_3} \end{aligned} \tag{5.9}$$

The candidate models differ for the mathematical formulation of the denominator, referred to as the sensitivity of the sensor, which is a linear function for M_a and an exponential function for M_b. Parameters a_3 and b_3 represent, in both cases, the offset of the models.

A priori probabilities of the candidate models, M_a and M_b, and relative parameters, $\theta_a = [a_1, a_2, a_3]$ and $\theta_b = [b_1, b_2, b_3]$ were arbitrary fixed as follows:

$$\begin{aligned} p(M_a) = 0.7; \quad & p(\theta_a | M_a) \sim N(\mu_a, \Sigma_a) \\ p(M_b) = 0.3; \quad & p(\theta_b | M_b) \sim N(\mu_b, \Sigma_b) \end{aligned} \tag{5.10}$$

Specific mean and covariance values used in the simulations were:

$$\begin{aligned} \mu_a = [0.3, 28, -10]; \quad & \Sigma_a = \begin{bmatrix} 25 & 11 & -6.6 \\ 11 & 16 & -9 \\ -6.6 & -9 & 9 \end{bmatrix} \\ \mu_b = [30, 5, -10]; \quad & \Sigma_b = \begin{bmatrix} 9 & 3 & -2.4 \\ 3 & 25 & 6.6 \\ -2.4 & 6.6 & 9 \end{bmatrix} \end{aligned} \tag{5.11}$$

which are qualitatively chosen in accordance to correlations between parameters observed in real data (for instance, a negative correlation between sensitivity and baseline parameters is commonly observed in real data).

For each virtual subject, the electrical current signal is simulated by sampling from the prior distributions one of the two models according to their a priori probabilities and a relative set of parameters (as described in Section 5.1.1). An example of simulated current signal, obtained applying the model M_b to the data of Figure 5.2, top panel, is depicted in Figure 5.2, bottom panel.

In total, we simulated 15 different virtual datasets, each comprising the true BG profile as generated by the T1DS, the noisy SMBG measurements needed for calibration, and the electrical current signal as it would be measured by the sensor. In addition, for each virtual dataset, the calibration model used to convert the glucose concentration into electrical current values is, of course, known.

5.2.2 Model selection and performance assessment

To assess the MMBF using the synthetic data generated in Section 5.2.1, a real-time calibration scenario is simulated. Each time a new SMBG is available for calibration, i.e., about every 12h, the calibration model is chosen among the candidates in $\boldsymbol{M}$ by maximizing Equation 5.2. As a result, we obtain a calibrated CGM profile that can be evaluated, in terms of accuracy, against the true BG profile available in the dataset.

The number of Monte Carlo iterations for the computation of the marginal likelihood, i.e., parameter $\boldsymbol{N_iter}$ in Equation 5.5, was fixed to $\mathbf{10^4}$.

To assess performance of the MMBF we looked at two indexes:

- the number of simulations for which the model that actually generated the data corresponded to the model selected by the Bayesian model selection;

- the root mean square error (RMSE) between the true BG profile and the calibrated CGM profile obtained with the selected model vs. the CGM profile obtained by using a unique population model, i.e., model $\boldsymbol{M_a}$, which is the most probable in the simulated population.

Statistical significance of the differences in RMSE is assessed by Wilcoxon test.

5.3 Results

In the 15 analyzed datasets, the model chosen by the Bayesian selection method, i.e., the model that showed the higher a posteriori probability, was equal to the model that actually generated the data in 100% of the cases.

When the calibration model selected by the Bayesian model selection procedure is used for calibrating the data, the median RMSE between the calibrated CGM output and the true BG profile was 17.69 mg/dl. If a population model approach is used instead, i.e., the most probable model in the considered population, $\boldsymbol{M_a}$, is used with no previous model selection, then the median RMSE between calibrated CGM output and the true BG profile was 19.11 mg/dl. Results are schematized in Table 5.1, where the result of the statistical test evidences a significant difference in favor of the MMBF (p-value of 0.0059).

Table 5.1: Calibration accuracy as root mean square error (RMSE) for the MMBF vs. the population model approach.

Calibration accuracy	MMBF approach	Population approach	p-value
RMSE (mg/dl)	17.69 [3.41]	19.11 [5.07]	0.0059

Results reported as median [interquartile range].

5.4 Work to be done for validation on real data

In this Chapter we have assessed the MMBF on a simulated dataset, proving the suitability of such model-selection method for the calibration problem. To strength our findings a validation of the method on a real dataset is certainly needed. To do so, a large amount of data characterized by high inter-sensor and inter-subject variability is required for identifying each possible candidate model and for deriving the a priori information on models and model parameters. However, such data are not yet available, at the time of writing, for our use. Indeed, the datasets described in Chapter 2 are not suitable for this application, since the amount of data we would need is on the order of thousands of sensors. In principle, such data could be available to many CGM sensors manufacturers through their cloud platforms.

5.5 Summary and concluding remarks

In minimally invasive CGM sensors, the measured electrical signal is transformed to BG values by a calibration model whose parameters are periodically updated by matching SMBG references collected by finger prick devices. State-of-art calibration techniques rely on a predetermined calibration model, with parameters that can be tuned "in vivo" to adapt to the specific sensor and to the specific subject in which the sensor is used. Retrospective analysis of CGM recordings suggests that this approach may be not sufficient to explain the inter-sensor and inter-subject variability of the relationship between raw sensor measurements and glucose concentrations. On the one hand, the sensor-to-sensor variability can be reduced "in factory" by sensor manufacturers. On the other hand, a second source of "in vivo" variability, which is caused by the specific body-sensor interactions occurring after sensor insertion, may result in different relationships between the raw measurements and glucose concentrations that cannot be explained by a unique population model.

In this chapter, we proposed a MMBF that aims at offering flexibility for both the choice of the calibration model and the numerical values of its unknown parameters. In particular, we hypothesized that the calibration function does not necessarily obey to a unique model. Rather, the most convenient shape of the calibration function is determined in real-time among a set of candidate models, each one depending on a set of unknown parameters for which a priori statistical description is available. The method was preliminary assessed in simulation, in which only two candidate models were considered. Results show that the proposed approach is able to identify the correct calibration model with consequent improvement of sensor accuracy compared to the use of a unique population approach.

The MMBF here described has a general statistical background and could be applied, in principle, to any CGM sensor for which a suitable training set is available. Future work will include a more extensive test of the proposed approach on simulated datasets and the validation on real datasets.

Chapter 6

Conclusions

6.1 Summary of the main achievements

In the last twenty years, CGM devices have revolutionized the treatment of diabetes by providing high frequency measurement of glucose concentration. Among the different sensing technologies, the minimally invasive CGM devices based on the glucose-oxidase electrochemical principle represent the most successful commercial products. In glucose-oxidase CGM sensors, the electrical current signal measured by the sensor is in real time transformed to glucose concentration by a calibration procedure, which requires some BG references to be used as ground truth.

The two fundamental aspects that determine the efficacy of a calibration process are the accuracy of the calibrated profile and the number of calibrations that are needed to ensure such performance. Indeed, on the one hand, frequent calibrations would help guaranteeing good sensor accuracy, since more BG measurements are available as reference. On the other hand, each calibration requires the patients to acquire a BG sample through fingerprick devices, thus adding uncomfortable extra actions to the many already needed in the routine of diabetes management. The review of the past and existing calibration techniques proposed in the literature and employed by CGM sensor manufacturers conducted in Chapter 1 evidences that some critical aspects (i.e., the BG-to-IG kinetics and the time-variability of sensor sensitivity) are often neglected by many algorithms, with consequent suboptimal performance. The primary aim of this thesis was to develop a calibration algorithm able to provide accurate glucose estimations while, at the same time, reducing the frequency of calibrations.

As first finding of the present Ph.D. thesis, we demonstrated that the calibration algorithm of Vettoretti et al. [90], when implemented with day-specific prior information on calibration model parameters, is able to reduce the calibration frequency from two to one per day, while still improving accuracy if compared to the manufacturer calibration, which requires two calibrations per day. As possible margin for improvement, we demonstrated that the linear calibration model of [90] has domain of validity limited to certain time-intervals, of, e.g., 24 hours, not allowing to further reduce calibrations to less than one per day. The limited temporal domain of validity of the calibration model called for the development of a new calibration algorithm employing a calibration model valid for a multiple-day period.

The development of a new calibration algorithm with the aforementioned characteristics represents a major achievement of this Ph.D thesis, described in Chapter 4. The new BMD algorithm is based on a multiple-day model of sensor time-variability with second order statistical priors on its unknown parameters. The numerical values of model parameters are determined by a Bayesian estimation procedure exploiting a few BG references sparsely collected by the patient. In addition, the distortion introduced by the BG-to-IG kinetics is compensated during parameter identification via nonparametric deconvolution. The BMD calibration algorithm was applied to two datasets acquired with the present generation DG4P sensor and a NGD sensor prototype. In the DG4P data, results show that the BMD calibration algorithm significantly improves sensor accuracy compared to the manufacturer calibration while reducing calibration frequency from 2 to 0.25 per day. In the NGD data, performance of the BMD algorithm is even better than that of present generation, allowing to further reduce calibration frequency till zero.

However, higher inter-subject or inter-sensor variability could potentially characterize large sets of data, requiring additional flexibility to efficiently cope with the different sensor behaviors. To address this issue, in Chapter 5 we further developed the newly proposed BMD calibration algorithm in a multi-model scenario. In particular, to cope with the inter-sensor and inter-subject variability, we proposed a Bayesian framework also considering calibration model selection, in which the most likely calibration model is chosen among a finite set of candidates. A preliminary assessment of the MMBF was conducted on synthetic data. Results show a statistically significant accuracy improvement compared to the use of a unique calibration model.

6.2 Possible applications and future developments

In this thesis, the proposed calibration techniques have been retrospectively applied to a commercial CGM sensor and a next generation CGM sensor prototype of a specific manufacturer, i.e., Dexcom Inc. (San Diego, CA), thanks to the availability of some specific datasets.

A first, straightforward, application of the BMD calibration algorithm described in Chapter 4 would be the implementation of the algorithm directly on the DG4P or the NGD sensor's hardware. Indeed, the BMD algorithm and its implementation are already intended for real-time use in a portable device, given also the modest computational complexity of the involved operations. Of course, the use of the Bayesian estimation would require the preliminary estimation of a priori distributions of calibration model parameters from a training set of data.

Secondly, in more general terms, the proposed BMD calibration algorithm can be potentially applied to other CGM devices for which a raw measurement taken from the interstitial compartment needs to be converted into glucose concentration by matching some sparse BG references. In this case, the application of the method would require: i) to derive specific sensitivity and baseline functions that best describe the characteristics of the given sensor, which could be, in principle, different between different sensor types, and ii) derive a priori information on model parameters from a training dataset of the same sensor type.

An interesting future development, provided that suitable data become available, would be the application and validation of the proposed calibration techniques to a different glucose sensor other than a Dexcom CGM sensor. Indeed, the calibration techniques here proposed are intentionally designed to be flexible and applicable to a large class of glucose sensor devices.

Also, in the present thesis, the analyzed datasets were all acquired in adult populations. However, many CGM sensors currently available on the market are approved by the FDA to be used also in adolescents and pediatrics. One interesting aspect that needs to be analyzed in future work is the adaptability of the proposed calibration techniques to these specific populations. This requires, in particular, investigating whether the prior information on calibration parameters needs to be estimated from training data sets unique for each population or can be derived by leveraging existing distributions.

Finally, the MMBF described in Chapter 5 can be potentially applied to

any CGM sensor to describe the different behaviors caused by either inter-sensor and inter-subject variability. This would require the availability of large training datasets in order to capture the more probable sensor behaviors and estimate a priori probabilities for each model type and for the parameters of each model. The requirement of such amount of training data could appear, in principle, difficult to satisfy. However, in the last five years many CGM sensors manufacturers have started employing dedicated cloud platforms for storing and analyzing the large amount of data produced every day by their users. The availability of such datasets would render the application of the MMBF straightforward.

Provided that real data for such assessment become available for our use, the validation of the MMBF will be object of future work. Such validation would require to demonstrate that the accuracy of the calibrated profiles is enhanced by choosing the model among a set of candidates, rather than using a unique population calibration model.

Bibliography

[1] G. Cappon, G. Acciaroli, M. Vettoretti, A. Facchinetti, and G. Sparacino, "Wearable continuous glucose monitoring sensors: a revolution in diabetes treatment," *Electronics*, vol. 6, no. 4, p. 65, 2017.

[2] G. Acciaroli, M. Vettoretti, A. Facchinetti, and G. Sparacino, "Calibration of minimally invasive continuous glucose monitoring sensors: State-of-the-art and current perspectives," *Biosensors*, vol. 8, no. 1, p. E24, 2018.

[3] American Diabetes Association, "2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2018." *Diabetes care*, vol. 41, no. Suppl 1, pp. S13–S27, 2018.

[4] H. W. van Dijk, F. D. Verbraak, P. H. B. Kok, M. K. Garvin, M. Sonka, K. Lee, J. H. DeVries, R. P. J. Michels, M. E. J. van Velthoven, R. O. Schlingemann, and M. D. Abràmoff, "Decreased Retinal Ganglion Cell Layer Thickness in Patients with Type 1 Diabetes," *Invest Ophthalmol Vis Sci*, vol. 51, no. 7, p. 3660, 2010.

[5] T. M. Vriesendorp, J. H. DeVries, S. van Santen, H. S. Moeniralam, E. de Jonge, Y. B. W. E. M. Roos, M. J. Schultz, F. R. Rosendaal, and J. B. L. Hoekstra, "Evaluation of short-term consequences of hypoglycemia in an intensive care unit," *Crit Care Med*, vol. 34, no. 11, pp. 2714–2718, 2006.

[6] B. Lemkes, J. Hermanides, J. De Vries, F. Holleman, J. Meijers, and J. Hoekstra, "Hyperglycemia: a prothrombotic factor?" *J Thromb Haemost*, vol. 8, no. 8, pp. 1663–1669, 2010.

[7] W.-P. You and M. Henneberg, "Type 1 diabetes prevalence increasing globally and regionally: the role of natural selection and life expectancy at birth," *BMJ Open Diabetes Res Care*, vol. 4, no. 1, 2016.

[8] World Health Organization (WHO), "Diabetes facts sheet." [Online]. Available: http://www.who.int/mediacentre/factsheets/fs312/en/

[9] C. Hayes and A. Kriska, "Role of physical activity in diabetes management and prevention," *J Am Diet Assoc*, vol. 108, no. 4, pp. S19–S23, 2008.

[10] S. H. Ley, O. Hamdy, V. Mohan, and F. B. Hu, "Prevention and management of type 2 diabetes: dietary components and nutritional strategies," *Lancet*, vol. 383, no. 9933, pp. 1999–2007, 2014.

[11] S. F. Clarke and J. R. Foster, "A history of blood glucose meters and their role in self-monitoring of diabetes mellitus." *J Biomed Sci*, vol. 69, no. 2, pp. 83–93, 2012.

[12] J. Wang, "Electrochemical glucose biosensors," *Chem Rev*, vol. 108, no. 2, pp. 814–825, 2008.

[13] P. H. Sönksen, S. L. Judd, and C. Lowy, "Home monitoring of blood-glucose. Method for improving diabetic control." *Lancet*, vol. 1, no. 8067, pp. 729–32, 1978.

[14] S. Walford, E. A. Gale, S. P. Allison, and R. B. Tattersall, "Self-monitoring of blood-glucose. Improvement of diabetic control." *Lancet*, vol. 1, no. 8067, pp. 732–5, 1978.

[15] Y. Ikeda, N. Tajima, N. Minami, Y. Ide, J. Yokoyama, and M. Abe, "Pilot study of self-measurement of blood glucose using the Dextrostix-Eyetone system for juvenile-onset diabetes." *Diabetologia*, vol. 15, no. 2, pp. 91–3, 1978.

[16] M. Christiansen, T. Bailey, E. Watkins, D. Liljenquist, D. Price, K. Nakamura, R. Boock, and T. Peyser, "A new-generation continuous glucose monitoring system: improved accuracy and reliability compared with a previous-generation system," *Diabetes Technol Ther*, vol. 15, no. 10, pp. 881–888, 2013.

[17] E. H. Yoo and S. Y. Lee, "Glucose biosensors: an overview of use in clinical practice," *Sensors*, vol. 10, no. 12, pp. 4558–4576, 2010.

[18] J. E. Lane, J. P. Shivers, and H. Zisser, "Continuous glucose monitors," *Curr Opin Endocrinol Diabetes Obes*, vol. 20, no. 2, pp. 106–111, 2013.

[19] J. C. Pickup, M. Ford Holloway, and K. Samsi, "Real-time continuous glucose monitoring in type 1 diabetes: a qualitative framework analysis of patient narratives," *Diabetes Care*, vol. 38, no. 4, pp. 544–50, 2014.

[20] D. DeSalvo and B. Buckingham, "Continuous glucose monitoring: current use and future directions," *Curr Diabetes Rep*, vol. 13, no. 5, pp. 657–662, 2013.

[21] N. Mauras, L. Fox, K. Englert, and R. W. Beck, "Continuous glucose monitoring in type 1 diabetes," *Endocrine*, vol. 43, no. 1, pp. 41–50, 2013.

[22] N. Haviland, J. Walsh, R. Roberts, and T. S. Bailey, "Update on clinical utility of continuous glucose monitoring in type 1 diabetes," *Curr Diabetes Rep*, vol. 16, no. 11, p. 115, 2016.

[23] C. G. Parkin, C. Graham, and J. Smolskis, "Continuous glucose monitoring use in type 1 diabetes: longitudinal analysis demonstrates meaningful improvements in HbA1c and reductions in health care utilization," *J Diabetes Sci Technol*, vol. 11, no. 3, pp. 522–528, 2017.

[24] E. Toschi and H. Wolpert, "Utility of continuous glucose monitoring in type 1 and type 2 diabetes," *Endocrinol Metab Clin North Am*, vol. 45, no. 4, pp. 895–904, 2016.

[25] W. H. Polonsky, D. Hessler, K. J. Ruedy, R. W. Beck, and DIAMOND Study Group, "The impact of continuous glucose monitoring on markers of quality of life in adults with type 1 diabetes: further findings from the DIAMOND randomized clinical trial," *Diabetes Care*, vol. 40, no. 6, pp. 736–741, 2017.

[26] M. L. Litchman and N. A. Allen, "Real-time continuous glucose monitoring facilitates feelings of safety in older adults with type 1 diabetes: a qualitative study," *J Diabetes Sci Technol*, vol. 11, no. 5, pp. 988–995, 2017.

[27] R. M. Bergenstal, W. V. Tamborlane, A. Ahmann, J. B. Buse, G. Dailey, S. N. Davis, C. Joyce, B. A. Perkins, J. B. Welsh, S. M. Willi, M. A. Wood, and STAR 3 Study Group, "Sensor-augmented pump therapy for A1C reduction (STAR 3) study: results from the 6-month continuation phase," *Diabetes Care*, vol. 34, no. 11, pp. 2403–2405, 2011.

[28] P. Pesl, P. Herrero, M. Reddy, M. Xenou, N. Oliver, D. Johnston, C. Toumazou, and P. Georgiou, "An advanced bolus calculator for type 1 diabetes: system architecture and usability results," *IEEE J Biomed Health Inform*, vol. 20, no. 1, pp. 11–17, 2016.

[29] M. D. Breton, S. D. Patek, D. Lv, E. Schertz, J. Robic, J. Pinnata, L. Kollar, C. Barnett, C. Wakeman, M. Oliveri, C. Fabris, D. Chernavvsky, B. P. Kovatchev, and S. M. Anderson, "Continuous glucose monitoring and insulin informed advisory system with automated titration and dosing of insulin reduces glucose variability in type 1 diabetes mellitus," *Diabetes Technol Ther*, vol. 20, no. 8, pp. 531–540, 2018.

[30] J. Kropff and J. H. DeVries, "Continuous glucose monitoring, future products, and update on worldwide artificial pancreas projects," *Diabetes Technol Ther*, vol. 18, no. S2, pp. 253–63, 2016.

[31] C. Ramkissoon, P. Herrero, J. Bondia, and J. Vehi, "Unannounced meals in the artificial pancreas: detection using continuous glucose monitoring," *Sensors*, vol. 18, no. 3, p. 884, 2018.

[32] C. Toffanin, M. Messori, F. Di Palma, G. De Nicolao, C. Cobelli, and L. Magni, "Artificial pancreas: model predictive control design from clinical experience," *J Diabetes Sci Technol*, vol. 7, no. 6, pp. 1470–1483, 2013.

[33] P. Rossetti, C. Quirós, V. Moscardó, A. Comas, M. Giménez, F. J. Ampudia-Blasco, F. León, E. Montaser, I. Conget, J. Bondia, and J. Vehí, "Closed-loop control of postprandial glycemia using an insulin-on-board limitation through continuous action on glucose target," *Diabetes Technol Ther*, vol. 19, no. 6, pp. 355–362, 2017.

[34] M. Breton, A. Farret, D. Bruttomesso, S. A. Diabetes, and u. 2012, "Fully integrated artificial pancreas in type 1 diabetes: modular closed-loop glucose control maintains near normoglycemia," *Am Diabetes Assoc*, vol. 61, no. 9, pp. 2230–2238, 2012.

[35] Y. M. Luijf, J. H. DeVries, K. Zwinderman, L. Leelarathna, M. Nodale, K. Caldwell, K. Kumareswaran, D. Elleri, J. M. Allen, M. E. Wilinska, M. L. Evans, R. Hovorka, W. Doll, M. Ellmerer, J. K. Mader, E. Renard, J. Place, A. Farret, C. Cobelli, S. Del Favero, C. Dalla Man, A. Avogaro,

D. Bruttomesso, A. Filippi, R. Scotton, L. Magni, G. Lanzola, F. Di Palma, P. Soru, C. Toffanin, G. De Nicolao, S. Arnolds, C. Benesch, L. Heinemann, and AP@home Consortium, "Day and night closed-loop control in adults with type 1 diabetes: a comparison of two closed-loop algorithms driving continuous subcutaneous insulin infusion versus patient self-management," *Diabetes Care*, vol. 36, no. 12, pp. 3882–3887, 2013.

[36] H. P. Chase, F. J. Doyle, H. Zisser, E. Renard, R. Nimri, C. Cobelli, B. A. Buckingham, D. M. Maahs, S. Anderson, L. Magni, J. Lum, P. Calhoun, C. Kollman, and R. W. Beck, "Multicenter closed-loop/hybrid meal bolus insulin delivery with type 1 diabetes," *Diabetes Technol Ther*, vol. 16, no. 10, pp. 623–632, 2014.

[37] C. Cobelli, E. Renard, and B. Kovatchev, "Artificial pancreas: past, present, future," *Diabetes*, vol. 60, no. 11, pp. 2672–2682, 2011.

[38] S. Vaddiraju, D. J. Burgess, I. Tomazos, F. C. Jain, and F. Papadimitrakopoulos, "Technologies for continuous glucose monitoring: current problems and future promises," *J Diabetes Sci Technol*, vol. 44, no. 66, pp. 1540–1562, 2010.

[39] L. B. Mohammadi, T. Klotzbuecher, S. Sigloch, K. Welzel, M. Göddel, T. R. Pieber, and L. Schaupp, "In vivo evaluation of a chip based near infrared sensor for continuous glucose monitoring," *Biosens Bioelectron*, vol. 53, pp. 99–104, 2014.

[40] M. Mortellaro and A. DeHennis, "Performance characterization of an abiotic and fluorescent-based continuous glucose monitoring system in patients with type 1 diabetes," *Biosens Bioelectron*, vol. 61, pp. 227–231, 2014.

[41] H. Zhao, X. Guo, Y. Wang, X. Duan, H. Qu, H. Zhang, D. Zhang, and W. Pang, "Microchip based electrochemical-piezoelectric integrated multi-mode sensing system for continuous glucose monitoring," *Sens Actuators B Chem*, vol. 223, pp. 83–88, 2016.

[42] M. Rumpler, J. Mader, J. Fischer, R. Thar, J. Granger, F. Deliane, I. Klimant, F. Aberer, F. Sinner, T. Pieber, and M. Hajnsek, "First application of a transcutaneous optical single-port glucose monitoring device in patients with type 1 diabetes mellitus," *Biosens Bioelectron*, vol. 88, pp. 240–248, 2017.

[43] C. Chen, X.-L. Zhao, Z.-H. Li, Z.-G. Zhu, S.-H. Qian, and A. Flewitt, "Current and emerging technology for continuous glucose monitoring," *Sensors*, vol. 17, no. 12, p. 182, 2017.

[44] R. Gifford, "Continuous glucose monitoring: 40 years, what we've learned and what's next," *Chem Phys Chem*, vol. 14, no. 10, pp. 2032–2044, 2013.

[45] S. K. Garg, R. O. Potts, N. R. Ackerman, S. J. Fermi, J. A. Tamada, and H. P. Chase, "Correlation of fingerstick blood glucose measurements with GlucoWatch biographer glucose results in young subjects with type 1 diabetes." *Diabetes care*, vol. 22, no. 10, pp. 1708–14, 1999.

[46] W. L. Clarke, "The original Clarke Error Grid Analysis (EGA)," *Diabetes Technol Ther*, vol. 7, no. 5, pp. 776–779, 2005.

[47] J. Mastrototaro, J. Shin, A. Marcus, G. Sulur, and STAR 1 Clinical Trial Investigators, "The Accuracy and Efficacy of Real-Time Continuous Glucose Monitoring Sensor in Patients with Type 1 Diabetes," *Diabetes Technology & Therapeutics*, vol. 10, no. 5, pp. 385–390, 2008.

[48] H. C. Zisser, T. S. Bailey, S. Schwartz, R. E. Ratner, and J. Wise, "Accuracy of the SEVEN® continuous glucose monitoring system: comparison with frequently sampled venous glucose measurements," *J Diabetes Sci Technol*, vol. 3, no. 5, pp. 1146–1154, 2009.

[49] R. L. Weinstein, 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: comparison with frequent laboratory reference measurements," *Diabetes Care*, vol. 30, no. 5, pp. 1125–1130, 2007.

[50] T. S. Bailey, A. Ahmann, R. Brazg, M. Christiansen, S. Garg, E. Watkins, J. B. Welsh, and S. W. Lee, "Accuracy and acceptability of the 6-day Enlite continuous subcutaneous glucose sensor," *Diabetes Technol Ther*, vol. 16, no. 5, pp. 277–283, 2014.

[51] M. P. Christiansen, S. K. Garg, R. Brazg, B. W. Bode, T. S. Bailey, R. H. Slover, A. Sullivan, S. Huang, J. Shin, S. W. Lee, and F. R. Kaufman, "Accuracy of a fourth-generation subcutaneous continuous glucose sensor," *Diabetes Technol Ther*, vol. 19, no. 8, pp. 446–456, 2017.

[52] H. Hakami, "Medtronic Diabetes on Medtronic Minimed 670 G FDA approval." [Online]. Available: https://www.medtronicdiabetes.com/blog/

[53] T. S. Bailey, A. Chang, and M. Christiansen, "Clinical accuracy of a continuous glucose monitoring system with an advanced algorithm," *J Diabetes Sci Technol*, vol. 9, no. 2, pp. 209–214, 2015.

[54] Dexcom Inc, "Dexcom newsroom." [Online]. Available: https://www.dexcom.com/news

[55] G. McGarraugh, R. Brazg, and R. Weinstein, "FreeStyle Navigator continuous glucose monitoring system with TRUstart algorithm, a 1-hour warm-up time," *J Diabetes Sci Technol*, vol. 5, no. 1, pp. 99–106, 2011.

[56] T. Bailey, B. W. Bode, M. P. Christiansen, L. J. Klaff, and S. Alva, "The performance and usability of a factory-calibrated flash glucose monitoring System," *Diabetes Technol Ther*, vol. 17, no. 11, pp. 787–94, 2015.

[57] A. Dehennis, M. A. Mortellaro, and S. Ioacara, "Multisite study of an implanted continuous glucose sensor over 90 days in patients with diabetes mellitus," *J Diabetes Sci Technol*, vol. 9, no. 5, pp. 951–956, 2015.

[58] B. W. Bequette, "Continuous glucose monitoring: real-time algorithms for calibration, filtering, and alarms," *J Diabetes Sci Technol*, vol. 4, no. 2, p. 404–418, 2010.

[59] J. Hughes, J. B. Welsh, N. C. Bhavaraju, S. J. Vanslyke, and A. K. Balo, "Stability, accuracy, and risk assessment of a novel subcutaneous glucose sensor," *Diabetes Technol Ther*, vol. 19, no. S3, pp. S–21–S–24, 2017.

[60] G. McGarraugh, "The chemistry of commercial continuous glucose monitors," *Diabetes Technol Ther*, vol. 11, no. s1, pp. S17–S24, 2009.

[61] A. E. Panteleon, K. Rebrin, and G. M. Steil, "The role of the independent variable to glucose sensor calibration," *Diabetes Technol Ther*, vol. 5, no. 3, pp. 401–10, 2003.

[62] J. Mastrototaro, T. Gross, and J. Shin, "Glucose monitor calibration methods," 2002. [Online]. Available: https://patents.google.com/patent/US6424847B1/en

[63] J. Shin, K. Holtzclaw, N. Dangui, J. Kanderian, S., J. Mastrototaro, and P. Hong, "Real time self-adjusting calibration algorithm," 2005. [Online]. Available: https://patents.google.com/patent/US6895263B2/en

[64] A. Kamath, P. Simpson, J. Brauker, and J. Goode, P.V., "Calibration techniques for a continuous analyte sensor," 2013. [Online]. Available: https://patents.google.com/patent/US8428678B2/en

[65] E. Budiman, "Method and device for providing offset model based calibration for analyte sensor," 2013. [Online]. Available: https://patents.google.com/patent/US8532935B2/en

[66] C. Choleau, J. C. Klein, G. Reach, B. Aussedat, V. Demaria-Pesce, G. S. Wilson, R. Gifford, and W. K. Ward, "Calibration of a subcutaneous amperometric glucose sensor implanted for 7 days in diabetic patients: Part 2. Superiority of the one-point calibration method," *Biosens Bioelectron*, vol. 17, no. 8, pp. 647–54, 2002.

[67] Z. Mahmoudi, M. D. Johansen, J. S. Christiansen, and O. Hejlesen, "Comparison between one-point calibration and two-point calibration approaches in a continuous glucose monitoring algorithm," *J Diabetes Sci Technol*, vol. 8, no. 4, pp. 709–19, 2014.

[68] K. Rebrin, G. M. Steil, W. P. van Antwerp, and J. J. Mastrototaro, "Subcutaneous glucose predicts plasma glucose independent of insulin: implications for continuous monitoring." *Am J Physiol*, vol. 277, no. 3 Pt 1, pp. 561–71, 1999.

[69] M. Schiavon, C. Dalla Man, S. Dube, M. Slama, Y. C. Kudva, T. Peyser, A. Basu, R. Basu, and C. Cobelli, "Modeling plasma-to-interstitium glucose kinetics from multitracer plasma and microdialysis data," *Diabetes Technol Ther*, vol. 17, no. 11, pp. 825–831, 2015.

[70] D. B. Keenan, J. J. Mastrototaro, G. Voskanyan, and G. M. Steil, "Delays in minimally invasive continuous glucose monitoring devices: a review of current technology," *J Diabetes Sci Technol*, vol. 33, no. 55, pp. 1207–1214, 2009.

[71] K. Rebrin, N. F. Sheppard, and G. M. Steil, "Use of subcutaneous interstitial fluid glucose to estimate blood glucose: revisiting delay and sensor offset," *J Diabetes Sci Technol*, vol. 4, no. 5, pp. 1087–1098, 2010.

[72] A. Facchinetti, G. Sparacino, and C. Cobelli, "Reconstruction of glucose in plasma from interstitial fluid continuous glucose monitoring data: role of sensor calibration," *J Diabetes Sci Technol*, vol. 1, no. 5, pp. 617–23, 2007.

[73] K. L. Helton, B. D. Ratner, and N. A. Wisniewski, "Biomechanics of the sensor-tissue interface—Effects of motion, pressure, and design on sensor performance and the foreign body response—Part I: theoretical framework," *J Diabetes Sci Technol*, vol. 5, no. 3, pp. 632–646, 2011.

[74] U. Klueh, Z. Liu, B. Feldman, T. P. Henning, B. Cho, T. Ouyang, and D. Kreutzer, "Metabolic biofouling of glucose sensors in vivo: role of tissue microhemorrhages," *J Diabetes Sci Technol*, vol. 5, no. 3, pp. 583–595, 2011.

[75] Diabetes Research In Children Network (Direcnet) Study Group, B. A. Buckingham, C. Kollman, R. Beck, A. Kalajian, R. Fiallo-Scharer, M. J. Tansey, L. A. Fox, D. M. Wilson, S. A. Weinzimer, K. J. Ruedy, and W. V. Tamborlane, "Evaluation of factors affecting CGMS calibration," *Diabetes Technol Ther*, vol. 8, no. 3, pp. 318–325, 2006.

[76] B. Aussedat, V. Thomé-Duret, G. Reach, F. Lemmonier, J. C. Klein, Y. Hu, and G. S. Wilson, "A user-friendly method for calibrating a subcutaneous glucose sensor-based hypoglycaemic alarm." *Biosens Bioelectron*, vol. 12, no. 11, pp. 1061–71, 1997.

[77] E. J. Knobbe, W. L. Lim, and B. A. Buckingham, "Method and apparatus for real-time estimation of physiological parameters," 2003. [Online]. Available: https://patents.google.com/patent/US6575905B2/en

[78] E. J. Knobbe and B. Buckingham, "The extended Kalman filter for continuous glucose monitoring," *Diabetes Technol Ther*, vol. 7, no. 1, pp. 15–27, 2005.

[79] M. Kuure-Kinsey, C. C. Palerm, and B. W. Bequette, "A dual-rate Kalman filter for continuous glucose monitoring," in *Conf Proc IEEE Eng Med Biol Soc 2006*.

[80] A. Facchinetti, G. Sparacino, and C. Cobelli, "Enhanced accuracy of continuous glucose monitoring by online extended Kalman filtering," *Diabetes Technol Ther*, vol. 12, no. 5, pp. 353–63, 2010.

[81] Y. Leal, W. Garcia-Gabin, J. Bondia, E. Esteve, W. Ricart, J.-M. Fernandez-Real, and J. Vehi, "Real-time glucose estimation algorithm for continuous glucose monitoring using autoregressive models," *J Diabetes Sci Technol*, vol. 4, no. 2, p. 391–403, 2010.

[82] Y. Leal, W. Garcia-Gabin, J. Bondia, E. Esteve, W. Ricart, J.-M. Fernández-Real, and J. Vehí, "Enhanced algorithm for glucose estimation using the continuous glucose monitoring system." *Med Sci Monit*, vol. 16, no. 6, pp. 51–8, 2010.

[83] F. Barceló-Rico, J. Bondia, J. L. Díez, and P. Rossetti, "A multiple local models approach to accuracy improvement in continuous glucose monitoring," *Diabetes Technol Ther*, vol. 14, no. 1, pp. 74–82, 2012.

[84] F. Barcelo-Rico, J. L. Diez, P. Rossetti, J. Vehi, and J. Bondia, "Adaptive calibration algorithm for plasma glucose estimation in continuous glucose monitoring," *IEEE J Biomed Health Inform*, vol. 17, no. 3, pp. 530–8, 2013.

[85] Z. Mahmoudi, M. D. Johansen, J. S. Christiansen, and O. K. Hejlesen, "A multistep algorithm for processing and calibration of microdialysis continuous glucose monitoring data," *Diabetes Technol Ther*, vol. 15, no. 10, pp. 825–35, 2013.

[86] H. Kirchsteiger, L. Zaccarian, E. Renard, and L. del Re, "A novel online recalibration strategy for continuous glucose measurement sensors employing LMI techniques," in *Conf Proc IEEE Eng Med Biol Soc 2013*.

[87] H. Kirchsteiger, L. Zaccarian, E. Renard, and L. Del Re, "LMI-based approaches for the calibration of continuous glucose measurement sensors," *IEEE J Biomed Health Inform*, vol. 19, no. 5, pp. 1697–706, 2015.

[88] S. Guerra, A. Facchinetti, G. Sparacino, G. D. Nicolao, and C. Cobelli, "Enhancing the accuracy of subcutaneous glucose sensors: A real-time deconvolution-based approach," *IEEE Trans Biomed Eng*, vol. 59, no. 6, pp. 1658–69, 2012.

[89] G. De Nicolao, G. Sparacino, and C. Cobelli, "Nonparametric input estimation in physiological systems: problems, methods, and case studies," *Automatica*, vol. 33, no. 5, pp. 851–870, 1997.

[90] M. Vettoretti, A. Facchinetti, S. Del Favero, G. Sparacino, and C. Cobelli, "Online calibration of glucose sensors from the measured current by a time-varying calibration function and Bayesian priors," *IEEE Trans Biomed Eng*, vol. 63, no. 8, pp. 1631–1641, 2016.

[91] J. B. Lee, E. Dassau, and F. J. Doyle, "A run-to-run approach to enhance continuous glucose monitor accuracy based on continuous wear," *IFAC-PapersOnLine*, vol. 48, no. 20, pp. 237–242, 2015.

[92] S. Zavitsanou, J. B. Lee, J. E. Pinsker, M. M. Church, F. J. Doyle, and E. Dassau, "A personalized week-to-week updating algorithm to improve continuous glucose monitoring performance," *J Diabetes Sci Technol*, vol. 11, no. 6, pp. 1070–1079, 2017.

[93] International Organization for Standardization, "ISO 15197:2013: In Vitro Diagnostic Test Systems— Requirements for Blood-Glucose Monitoring Systems for Self-Testing in Managing Diabetes Mellitus." [Online]. Available: https://www.iso.org/standard/54976.html

[94] G. Acciaroli, M. Vettoretti, A. Facchinetti, G. Sparacino, and C. Cobelli, "From two to one per day calibration of Dexcom G4 platinum by a time-varying day-specific Bayesian prior," *Diabetes Technol Ther*, vol. 18, no. 8, 2016.

[95] G. Sparacino, A. Facchinetti, and C. Cobelli, "Smart continuous glucose monitoring sensors: on-line signal processing issues," *Sensors*, vol. 10, no. 7, p. 6751–6772, 2010.

[96] G. Acciaroli, M. Vettoretti, A. Facchinetti, G. Sparacino, and C. Cobelli, "Glucose sensors for diabetes management: no need of 24h warm-up time by use of Bayes estimation," in *Conf Proc IEEE Eng Med Biol Soc 2015*.

[97] A. Oppenheim, "Linear time-invariant systems," in *Signal and systems*, 2nd ed., A. Oppenheim, A. Willsky, and Nawab SH, Eds. Upper Saddle River: Prentice Hall, 1997, p. 75–178.

[98] G. Acciaroli, M. Vettoretti, A. Facchinetti, G. Sparacino, and C. Cobelli, "Reduction of blood glucose measurements to calibrate subcutaneous glucose sensors: a Bayesian multi-day framework," *IEEE Trans Biomed Eng*, vol. 65, no. 3, pp. 587–595, 2017.

[99] G. Acciaroli, M. Vettoretti, A. Facchinetti, and G. Sparacino, "Toward calibration-free continuous glucose monitoring sensors: Bayesian calibration approach applied to next-generation Dexcom technology," *Diabetes Technol Ther*, vol. 20, no. 1, pp. 59–67, 2018.

[100] C. Cobelli, M. Schiavon, C. Dalla Man, A. Basu, and R. Basu, "Interstitial fluid glucose is not just a shifted-in-time but a distorted mirror of blood glucose: insight from an in silico study," *Diabetes Technol Ther*, vol. 18, no. 8, pp. 505–511, 2016.

[101] A. Basu, S. Dube, S. Veettil, M. Slama, Y. C. Kudva, T. Peyser, R. E. Carter, C. Cobelli, and R. Basu, "Time lag of glucose from intravascular to interstitial compartment in type 1 diabetes." *J Diabetes Sci Technol*, vol. 9, no. 1, pp. 63–8, 2015.

[102] M. J. Estes, P. C. Simpson, and A. U. Kamath, "Advanced analyte sensor calibration and error detection," 2012. [Online]. Available: https://patents.google.com/patent/US20120265036A1/en

[103] S. J. Vanslyke, N. C. Bhavaraju, L. Bohnett, A. Garcia, A. U. Kamath, and J. Pryor, "Advanced calibration for analyte sensors," 2013. [Online]. Available: https://patents.google.com/patent/US20140278189A1/en

[104] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the Nelder–Mead simplex method in low dimensions," *SIAM J Optim*, vol. 9, no. 1, pp. 112–147, 1998.

[105] A. Garcia, A. L. Rack-Gomer, N. C. Bhavaraju, H. Hampapuram, A. Kamath, T. Peyser, A. Facchinetti, C. Zecchin, G. Sparacino, and C. Cobelli, "Dexcom G4AP: an advanced continuous glucose monitor for the artificial pancreas," *J Diabetes Sci Technol*, vol. 7, no. 6, pp. 1436–45, 2013.

[106] G. Acciaroli, M. Vettoretti, A. Facchinetti, and G. Sparacino, "Bayesian model selection framework to improve calibration of continuous glucose monitoring sensors for diabetes management," in *Conf Proc IEEE Eng Med Biol Soc 2018*.

[107] K. L. Helton, B. D. Ratner, and N. A. Wisniewski, "Biomechanics of the sensor-tissue interface—effects of motion, pressure, and design on sensor performance and foreign body response—Part II: examples and application," *J Diabetes Sci Technol*, vol. 5, no. 3, pp. 647–656, 2011.

[108] H. Chipman, E. George, R. McCulloch, M. Clyde, D. Foster, and R. Stine, "The practical implementation of Bayesian model selection," *Lecture Notes Monograph Series*, vol. 1, pp. 65–134, 2001.

[109] M. Evans and T. Swartz, "Methods for approximating integrals in statistics with special emphasis on bayesian integration problems," *Stat Sci*, vol. 10, no. 3, pp. 254–272, 1995.

[110] R. E. Kass and A. E. Raftery, "Bayes factors," *J Am Stat Assoc*, vol. 90, no. 430, pp. 773–795, 1995.

[111] J. Geweke, "Bayesian inference in econometric models using Monte Carlo integration," *Econometrica*, vol. 57, no. 6, p. 1317, 1989.

[112] M. Vettoretti, A. Facchinetti, G. Sparacino, and C. Cobelli, "Type-1 diabetes patient decision simulator for in silico testing safety and effectiveness of insulin treatments," *IEEE Trans Biomed Eng*, vol. 65, no. 6, pp. 1281–1290, 2018.