Research Article

Regularised Model Identification Improves Accuracy of Multisensor Systems for Noninvasive Continuous Glucose Monitoring in Diabetes Management

Mattia Zanon,[1] Giovanni Sparacino,[1] Andrea Facchinetti,[1] Mark S. Talary,[2] Andreas Caduff,[2] and Claudio Cobelli[1]

[1] *Department of Information Engineering, University of Padova, Via Gradenigo 6B, 35131 Padova, Italy*
[2] *Biovotion AG, Technoparkstrasse 1, 8005 Zurich, Switzerland*

Correspondence should be addressed to Claudio Cobelli; cobelli@dei.unipd.it

Continuous glucose monitoring (CGM) by suitable portable sensors plays a central role in the treatment of diabetes, a disease currently affecting more than 350 million people worldwide. Noninvasive CGM (NI-CGM), in particular, is appealing for reasons related to patient comfort (no needles are used) but challenging. NI-CGM prototypes exploiting multisensor approaches have been recently proposed to deal with physiological and environmental disturbances. In these prototypes, signals measured noninvasively (e.g., skin impedance, temperature, optical skin properties, etc.) are combined through a static multivariate linear model for estimating glucose levels. In this work, by exploiting a dataset of 45 experimental sessions acquired in diabetic subjects, we show that regularisation-based techniques for the identification of the model, such as the least absolute shrinkage and selection operator (better known as LASSO), Ridge regression, and Elastic-Net regression, improve the accuracy of glucose estimates with respect to techniques, such as partial least squares regression, previously used in the literature. More specifically, the Elastic-Net model (i.e., the model identified using a combination of l_1 and l_2 norms) has the best results, according to the metrics widely accepted in the diabetes community. This model represents an important incremental step toward the development of NI-CGM devices effectively usable by patients.

1. Introduction

Diabetes consists of a malfunction of the glucose-insulin regulatory system leading to the onset of long and short term complications, like retinopathy, neuropathy, and cardiovascular and hearth diseases, due to sustained blood glycaemic levels exceeding the normal range of 70–180 mg/dL [1]. According to the International Diabetes Federation, diabetes is estimated to currently affect more than 350 million people worldwide, and this number is rapidly growing [2]. Not surprisingly, in the last few decades, diabetes has received an increasing attention both for its social and economic implications [3].

Standard therapy of type 1 diabetes is based on a combination of diet, physical activity, and exogenous insulin injections, modulated on the basis of individual patient levels of glucose concentration in the blood. Accurate and frequent monitoring of glucose concentrations by portable sensor devices plays a crucial role in the diabetes therapy [4]. Self-monitoring blood glucose (SMBG) sensors require the collection of a blood sample by pricking the skin with a lancet device. An external pocket device is then used to analyze the blood and determine glucose concentration for instance by the glucose oxidase principle [5]. These sensors are uncomfortable for the patient and are typically used no more than 3-4 times per day. Such sparseness of sampling does not allow the observation of glucose dynamics and glucose excursions outside the safety range, and dangerous nocturnal hypoglycaemic swings are often not detected. To overcome these problems, portable continuous glucose monitoring (CGM) sensors, measuring blood glucose values every 1 to 5 minutes for up to 7 consecutive days, have been proposed in the

early 2000s and are now of great interest for the tuning and optimization of diabetes therapy [6, 7]. In particular, online applications of CGM sensors include prevention of hyper-/hypoglycaemic events; see for example [8, 9], and closed-loop glucose control aimed at determining optimal automatic insulin infusion in the so-called artificial pancreas systems; see [10–14]. Notably, dealing with these online applications requires facing nontrivial signal processing issues connected to denoising [15], calibration [16], and prediction [17–21] (see [22, 23] for reviews) calling for the development of the so-called "smart" CGM sensor architecture [24].

Most of the CGM sensors requiring the placement of a small needle in the subcutaneous tissue use an enzyme-based glucose-oxidase electrode and thus are invasive, albeit minimally. To limit patients' discomfort, in the last decade, several noninvasive continuous glucose monitoring (NI-CGM) sensors have been prototyped. Many different physical principles have been considered to pursue NI-CGM, but none of them has clearly outperformed the others so far; see for example [25–27]. One major difficulty with NI-CGM is the fact that environmental (e.g., temperature) and physiological (e.g., sweat, blood oxygenation, etc.) processes act as perturbing factors and often allow blood glucose changes to be tracked only in highly controlled conditions (e.g., during in-hospital studies) [28–30]. To tackle this issue, an approach gaining increasing attention in the last years is the multisensor approach to NI-CGM; that is, instead of focusing on a single physical principle, these devices resort to a combination of technologies. For instance, the GlucoTrack [31] exploits a mix of thermal, acoustic, and electromagnetic technologies and compares the three measurements, assuming they all reflect a glucose-related measurement. A different, yet recent, multi-sensor prototype [32] employs a combination of dielectric spectroscopy (DS) and mid-infrared-based sensors. A further example of multi-sensor device, proposed by Solianis Monitoring AG (Zürich, Switzerland, technology taken over by Biovotion AG, Zurich, Switzerland), embeds sensors of different nature for the biophysical characterisation of skin and underlying tissue in order to track glucose-related and perturbing factors separately [30, 33].

Multisensor approaches to NI-CGM require a model to connect the physical quantities measured by the sensors with blood glucose concentrations. For instance, in the Solianis Monitoring AG prototype (from now on named Solianis for the sake of readability) considered in this work, a multivariate linear regression model is used to combine ~150 signals recorded noninvasively for estimating a glucose concentration profile (see Section 3.1 for more details). As described in [33, 34], the linear regression model is identified on a dataset collected in a population of subjects and comprising multi-sensor channels measurements and reference blood glucose (RBG) values collected in parallel by a gold standard technique. Previous work [35] has shown that a regularised identification method, based on l_1 norm (least absolute shrinkage and selection operator—LASSO), provided a glucose profile more accurate than that of other models identified with approaches controlling complexity such as subset selection method or partial least squares (PLS). In the present work, by using the same dataset of 45 experimental sessions used in [35], we demonstrate that the accuracy of glucose estimates can be further improved by considering l_2 norm regularization (Ridge regression) and a combination of l_1 and l_2 norms (Elastic-Net—EN regression), providing a further incremental step toward the development of an NI-CGM effectively usable by diabetic patients.

2. Database

The database consists of 45 experimental sessions recorded from 6 type 1 diabetic subjects, during which plasma glucose was induced to vary according to different predetermined profiles to cause different hypo- and hyperglycaemic excursions. During each session, multi-sensor data and RBG were acquired in parallel, with a time sampling of 20 seconds and 10–15 minutes, respectively. The RBG samples were acquired by means of a SMBG device. The study was performed at the University Hospital Zurich according to the requirements of good clinical practice and was approved by the local ethical commission. More clinically related information can be found in [33].

For the analysis, the database was split into two data subsets of 22 and 23 experimental sessions, respectively. If the first data subset is used for identifying the models with the different techniques, the second is used to test the models over "unseen" multi-sensor data (1–>2) and vice versa (2–>1). For the sake of space, we will not discuss results of the application of the model on the same dataset used for their identification, and only model test results will be considered.

Multi-sensor data in the identification data subset undergoes a preprocessing step described in detail in [35]. Briefly, each multi-sensor channel is normalized to have zero mean and unitary standard deviation. These values are then used to standardize the same channels in the test data subset to permit simulation of real-time glucose monitoring. Moreover, the first RBG value available at the beginning of each recording session is used to calibrate the estimated glucose profiles by the model setting a baseline adjustment, to allow for a real-time consideration/implementation.

3. Methods

3.1. Glucose Determination by a Multisensor System. In stating the problem, we deliberately make only a brief description of the framework we are working on and we refer the reader to the quoted bibliography for details.

Rather than focusing on a single physical principle, the Solianis multi-sensor device resorts to a combination of technologies, embedded into the substrate in contact with the skin for the biophysical characterisation of the skin and underlying tissue in order to account for confounding factors, which can significantly deteriorate the accuracy of glucose measurements [30, 36–38]. In particular, glucose-related signals are obtained from DS fringing field capacitive electrodes, with different geometrical properties, providing a spectrum of the frequency-dependent complex dielectric properties of skin and underlying tissue including blood, which can be easily parameterised by its magnitude and phase [39]. Environmental and physiological processes that can interfere
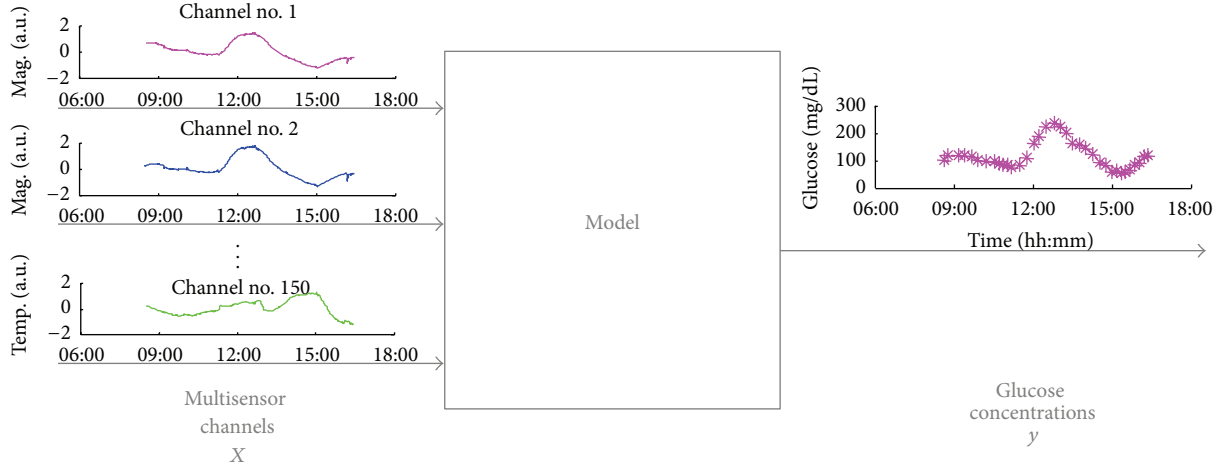
FIGURE 1: A model (center) is needed to properly combine the 150 multisensor channels (left, depicted with lines of different colours) for estimating blood glucose concentration profile (right, magenta stars).

with the measurements of the main glucose-related signals are measured with temperature, optical, humidity, accelerometer, and additional DS sensors incorporated into the same device substrate [30]. About 150 channels are thus provided by the multi-sensor device (Figure 1, left), which have to be properly combined through a mathematical model (Figure 1, middle) for estimating glucose concentrations in the blood (Figure 1, right). Since a mechanistic description comprising the physical principles, linking measured channels with physical/physiological processes, in particular, related to glucose changes, has not yet been fully developed, a "black-box" multivariate linear regression model is used.

Formally, if N is the number of data points available and p is the number of measured channels, the model to identify from the identification dataset is described as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\nu}, \tag{1}$$

where $\mathbf{y}$ represents the $(N \times 1)$ target vector containing RBG values, $\mathbf{X}$ is the $(N \times p)$ matrix collecting the multi-sensor channels, $\boldsymbol{\beta}$ is the $(p \times 1)$ vector containing the coefficients of the linear model, and $\boldsymbol{\nu}$ is the $(N \times 1)$ term representing the error.

The vector $\boldsymbol{\beta}$ in (1) can be identified by minimizing the cost function $F(\boldsymbol{\beta})$ given by the residual sum of squares (RSS):

$$F(\boldsymbol{\beta}) = \mathrm{RSS}(\boldsymbol{\beta}) = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 = \sum_{i=1}^{N} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^2, \tag{2}$$

measuring the distance between data and model predictions. Since this cost function has a quadratic form, a closed-form solution, the so-called ordinary least squares (OLS) estimate, can be obtained. In the case under consideration, OLS suffers from overfitting due to the high dimensionality of the measurement space and to the correlation between subsets of input channels (well visible in the channels showed in Figure 1, left). Recent work [34] showed that further improved results can be obtained by exploiting subset selection techniques. Then, further work [35] investigated the use

of other methods controlling "model complexity," including PLS (widely used in chemometrics and related fields dealing with spectroscopy data) and the LASSO regularization technique. It has been shown that regularization techniques, in particular, the LASSO, outperform PLS since it sets many coefficients to zero being less sensitive to occasional noise in the multi-sensor channels.

Formally speaking, regularized model identification techniques add a term $L(\boldsymbol{\beta})$ to the cost function of (2), leading to

$$F(\boldsymbol{\beta}) = \mathrm{RSS}(\boldsymbol{\beta}) + L(\boldsymbol{\beta}), \tag{3}$$

where $L(\boldsymbol{\beta})$ is a function of $\boldsymbol{\beta}$ reflecting complexity. Depending on the form and on the parameters of $L(\boldsymbol{\beta})$ in (2), the resulting model will present different well-known features, as will be briefly reviewed in the following section.

Thus, the $\boldsymbol{\beta}$ minimizing (3) is identified establishing a trade off between adherence to the data and model complexity, usually by a cross-validation procedure as better discussed next.

3.2. Model Identification by Regularisation Techniques

3.2.1. l_1 Norm: LASSO.
In the $L(\boldsymbol{\beta})$ term of (3), the l_1 norm can be used. In the literature, this l_1 norm has been proposed in signal processing (under the name of basis pursuit) [40] and in statistics [41] for its main feature of inducing sparse solutions. Formally, in (3), the l_1 norm leads to define $L(\boldsymbol{\beta})$ as the sum of the absolute values of the coefficients of the model

$$L(\boldsymbol{\beta}) = \lambda \|\boldsymbol{\beta}\|_1 = \lambda \sum_{j=1}^{p} |\boldsymbol{\beta}_j| \tag{4}$$

multiplied by the scalar nonnegative parameter, hereafter referred to as regularization parameter for the sake of reasoning. Thus, the solution is found by minimizing

$$\widehat{\boldsymbol{\beta}}^{\mathrm{LASSO}} = \arg\min_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^{N} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^2 + \lambda \sum_{j=1}^{p} |\boldsymbol{\beta}_j| \right\} \tag{5}$$

and is known as the least absolute shrinkage and selection operator (LASSO), for its properties of shrinking many coefficients to zero selecting only few input variables. In particular, in our application documented in [35], the LASSO was shown to outperform PLS since it sets many coefficients to zero being less sensitive to occasional noise in multi-sensor channels. Equation (5) does not have a closed-form solution because the cost function is not differentiable when some coefficients $\boldsymbol{\beta}_j$ are zero, and a plethora of methods have been developed to calculate approximate solutions numerically; see [42, 43] for reviews. In particular, an efficient technique for computing the LASSO solution is obtained by modifying the least angle regression algorithm [44]. With this technique, the parameter that has to be fixed represents the number of input variables allowed to enter the model, indicated with j in Section 3.

Remark 1. The regularization parameter weighting the $L(\boldsymbol{\beta})$ term in (5) is obtained by means of a standard procedure known as K-fold cross-validation [45]. Briefly, the identification dataset is split into K approximately equal parts. Then, the model is identified on $K - 1$ parts and tested over the portion of data not considered for deriving the model, calculating the mean squared error (MSE):

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N_{K-1}} (\mathbf{y}_i - \widehat{\mathbf{y}}_t)^2, \qquad (6)$$

where $\mathbf{y}$ and $\widehat{\mathbf{y}}$ represent the true and estimated output, respectively, and N_{K-1} represents the number of data points in the $K - 1$ portion of data. This procedure is repeated K times and for a range of values of the parameter that has to be determined. Then, the cross-validation curve is plotted, presenting the MSE as a function of the regularization parameter. Empirical evidence suggests choosing its value in correspondence with a clear drop of the cross-validation curve.

3.2.2. l_2 Norm: Ridge Regression. The l_2 norm involves the penalization of the sum of squares of the coefficients of the model multiplied by a scalar nonnegative parameter:

$$L(\boldsymbol{\beta}) = \lambda \|\boldsymbol{\beta}\|_2^2 = \lambda \sum_{j=1}^{p} \boldsymbol{\beta}_j^2. \qquad (7)$$

The so-called Ridge regression solution, from now on ridge, is thus given by

$$\widehat{\boldsymbol{\beta}}^{\text{ridge}} = \arg\min_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^{N} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^2 + \lambda \sum_{j=1}^{p} \boldsymbol{\beta}_j^2 \right\}. \qquad (8)$$

The quadratic nature of the cost function in (7) entails a closed-form solution for $\boldsymbol{\beta}$ dependent on the parameter λ:

$$\widehat{\boldsymbol{\beta}}^{\text{ridge}} = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I}_{\text{pxp}})^{-1} \mathbf{X}^T \mathbf{y}. \qquad (9)$$

Also, in this case, the regularization parameter λ can be fixed by means of cross-validation. According to [45], as an indication of the model complexity, the degrees of freedom (df) can be used:

$$\text{df}(\lambda) = \text{tr}\left[\mathbf{X}(\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \right] = \sum_{j=1}^{p} \frac{d_j^2}{d_j^2 + \lambda}, \qquad (10)$$

where the d_js are the singular values of $\mathbf{X}$. Thus, to determine the regularization parameter by cross-validation, the MSE is plotted against df in (10).

3.2.3. $l_1 + l_2$ Norms: Elastic Net-Regression. The so-called Elastic-Net regression, from now on EN, resorts to a weighted sum of the two previously described norms, defining $L(\boldsymbol{\beta})$ as

$$L(\boldsymbol{\beta}) = \lambda \left(\alpha \|\boldsymbol{\beta}\|_1 + (1 - \alpha) \|\boldsymbol{\beta}\|_2^2 \right)$$
$$= \lambda \left(\alpha \sum_{j=1}^{p} |\boldsymbol{\beta}_j| + (1 - \alpha) \sum_{j=1}^{p} \boldsymbol{\beta}_j^2 \right), \qquad (11)$$

where α weighs the contribution of the two norms while λ sets the amount of regularization [46]. Hence, the EN model parameters are obtained solving the following:

$$\widehat{\boldsymbol{\beta}}^{\text{EN}}$$
$$= \arg\min_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^{N} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^2 + \lambda \left(\alpha \sum_{j=1}^{p} |\boldsymbol{\beta}_j| + (1 - \alpha) \sum_{j=1}^{p} \boldsymbol{\beta}_j^2 \right) \right\}. \qquad (12)$$

Problem (12) does not have a closed-form solution, and several numerical algorithms have been proposed to compute an approximate solution (some of them are simple adaptations of those developed for the LASSO [46]). The algorithm that has been used in this work is the one based on cyclical coordinate descent originally developed for the LASSO [47] and adapted to problem (12) following [48–50]. The parameters λ and α are determined by cross-validation, following a procedure similar to that of Remark 1.

3.3. Criteria for Model Assessment. The accuracy of estimated glucose profiles in the model test is measured through a set of indexes widely used in the diabetes research community. In particular, we consider the root mean squared error (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left((\mathbf{y}_i - \widehat{\mathbf{y}}_i)^2 \right)}, \qquad (13)$$

the mean absolute difference (MAD), indicating how much estimated glucose values are lower or higher than the reference,

$$\text{MAD} = \frac{1}{N} \sum_{i=1}^{N} |\mathbf{y}_i - \widehat{\mathbf{y}}_i|, \qquad (14)$$
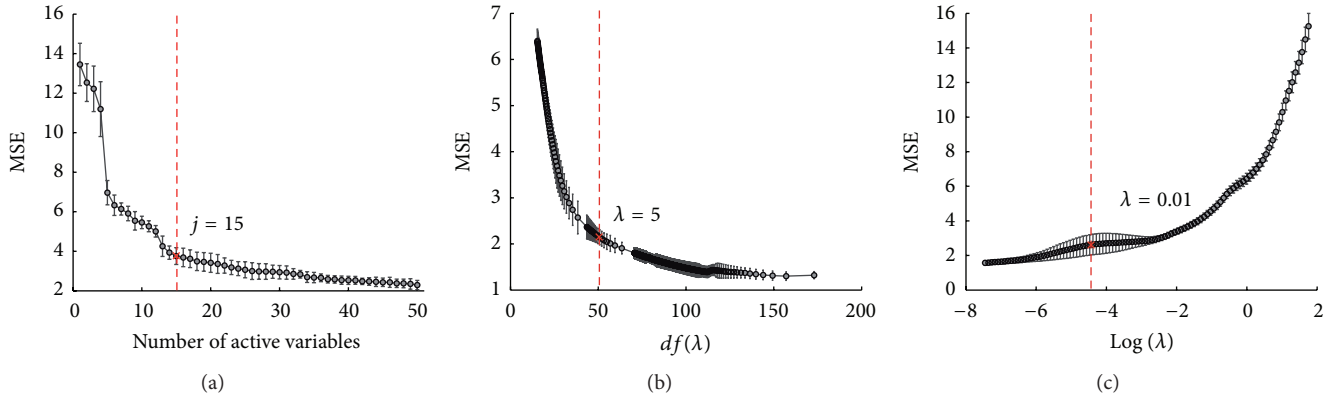
FIGURE 2: 10-fold cross-validation curves for the choice of the "optimal" complexity parameters for LASSO (a), Ridge (b), and EN for $\alpha = 0.4$ (c). The MSE (mean value and one standard deviation) is represented as a function of the model complexity parameter for each method. The red crosses represent the values of the complexity parameter chosen according to the drop in the error curve. The vertical red dashed line is for a better visual identification of the complexity parameters.

and the mean absolute relative difference (MARD), which characterizes the relative errors (in %) of the estimated glucose:

$$\text{MARD} = \frac{1}{N}\sum_{i=1}^{N}\frac{|\mathbf{y}_i - \widehat{\mathbf{y}}_i|}{\mathbf{y}_i}, \tag{15}$$

where $\mathbf{y}_i$ and $\widehat{\mathbf{y}}_i$, for $i = 1,\dots,N$, are, respectively, the N reference RBG samples and the glucose estimates provided by the models (all the experimental sessions are simultaneously considered). Finally, a popular method used in the diabetes community to judge on the point accuracy of glucose sensors is the error grid analysis (EGA) proposed by Clarke and coworkers [51]. The area where estimated glucose by the model and RBG values are displayed as a scatterplot is broken down into five regions (labeled from A to E). Zone A represents those glucose values within 20% of the RBG values and so on. The most dangerous situations are those where estimated glucose values fall into zones C/D/E because, from a clinical point of view, they will lead to unnecessary or even wrong and potentially dangerous treatments. An evolution of EGA developed for CGM sensors is the continuous EGA (CEGA) that also measures the accuracy of estimated glucose trends by creating a grid which is broken down into regions labeled from A_R to E_R; see [52] for details.

To give an idea of the values of these indexes for the current state-of the-art, minimally invasive (needle-based) commercial CGM sensors, a recent study [11] reported mean MARD levels ranging from 11.8 to 20.2% and a percentage of data points in the A+B region ranging from 98.9 to 96.9 under controlled conditions, comparing CGM measurements to gold standard blood glucose sampling.

4. Results

4.1. Regularization Parameter Determination by Cross-Validation. The first step in the analysis is setting the values for λ in (5), (8), and (12) and for α in (12). These were determined by finding where the cross-validation curve presents a clear drop in slope, as explained in Remark 1. Figure 2 shows the values obtained when identification data subset "part 1" is considered, and the red cross, together with a vertical red dashed line for visualization purpose, highlights the selected "optimal" value (cross-validation plots for identification data subset "part 2" are not showed for the sake of space). Specifically, a K-fold cross-validation strategy has been applied, with $K = 10$, for having a good compromise between bias and variance of the estimated error [45]. The left subplot shows the error curve as a function of the number of latent variables for the LASSO model, indicating a drop of the cross-validation curve around 15. The choice of the regularization parameter for ridge followed a similar approach, with the cross-validation curve (middle panel) presenting a drop when the degree of freedom, defined in (10), approximately equals 50, corresponding to $\lambda = 5$. Similarly for EN, the ending part of the drop in the error curve can be noticed for $\log(\lambda) \cong -4.5$ (Figure 2(c)), corresponding to $\lambda = 0.01$. For EN, different cross-validation curves for different values of α were examined. The most reasonable choice seemed to be that obtained with $\alpha = 0.4$. Indeed, this combination of parameters is the one providing a good trade off between the l_1 and l_2 norms allowing a reasonable compromise for the EN model to be achieved. A value of $\alpha = 0.4$ can suggest that, although it is important to shrink channel weights to zero in order to lower the probability of occasional jumps or spikes entering the model, allowing a grouping effect over correlated predictors is also important for a more robust estimation of glucose profiles.

4.2. Model Test. Figure 3 shows examples of estimated blood glucose concentration profiles (continuous line) versus reference RBG samples (open bullets) for two representative experimental sessions. Visual inspection suggests that the profiles provided by the EN model (bottom panels) outperform, in terms of accuracy, those provided by LASSO (top) and Ridge (middle). The same observation applies for the examples depicted in Figure 4, which represents a more challenging situation because of the presence of wider
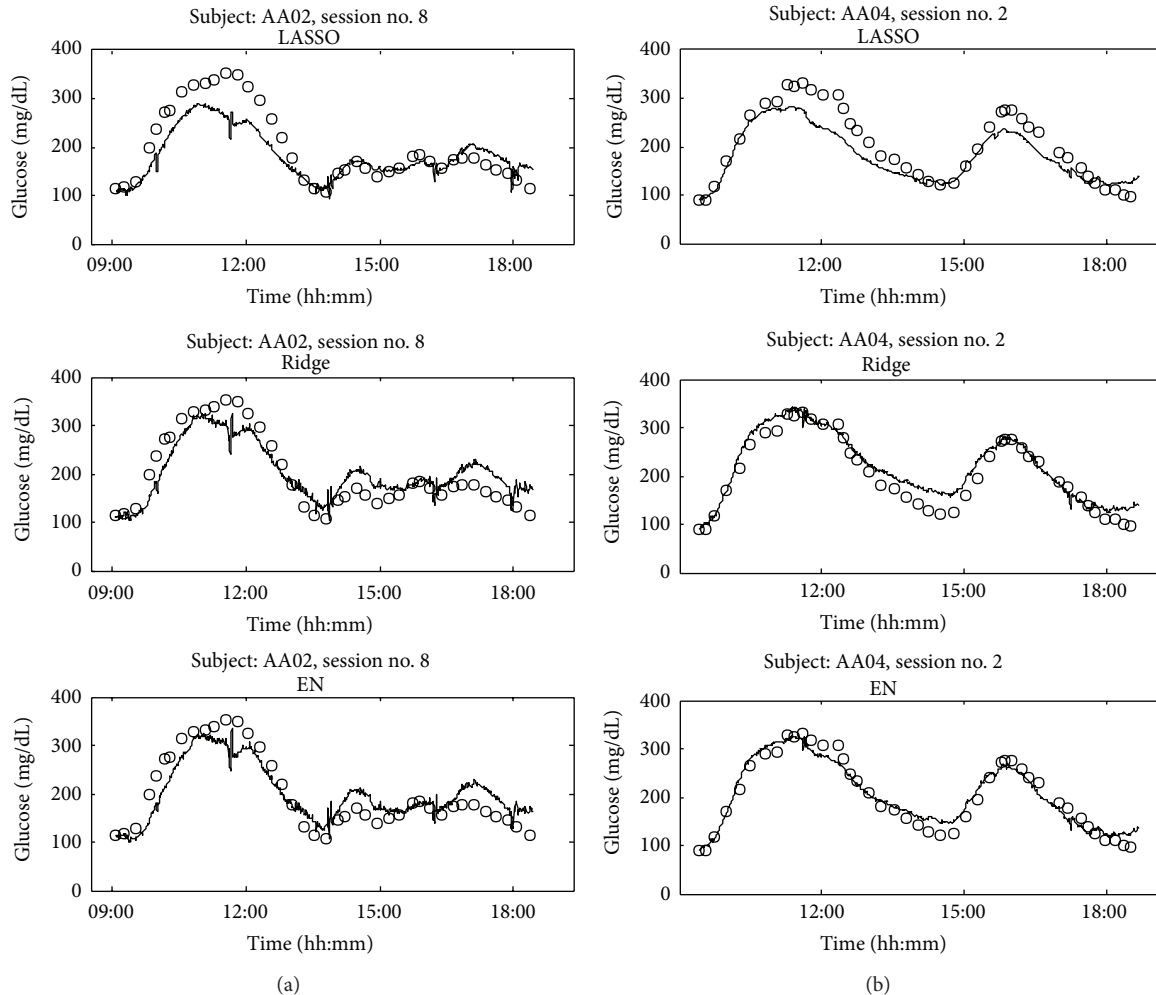
FIGURE 3: Representative recording sessions of Subjects AA02 (a) and AA04 (b). LASSO, Ridge, and EN models test over independent test data subset (continuous lines) versus RBG levels (open bullets). MARD values for the experimental session on the right are of 16.7% (LASSO), 16.9% (Ridge), and 16.5% (EN), while for the experimental session on the left of 12.7% (LASSO), 11.2% (Ridge), and 9.1% (EN).

disturbances, as witnessed by the spikes and jumps affecting the representative multi-sensor raw channels displayed in an additional fourth row of panels. This qualitative observation is supported from the analysis of the MARD values obtained for the representative sessions in Figure 3, that is, 16.7%, 16.9%, and 16.5% (experimental session depicted in the left panels) and 12.7%, 11.2%, and 9.1% for the LASSO, Ridge, and EN models, respectively. When occasional noise is affecting some of the multi-sensor channels, the MARD values slightly worsen, as can be seen for the experiment depicted in Figure 4(a) presenting MARD of 53% for the LASSO, 24.5% for Ridge, and 20.6 for EN. However, the EN model still provides blood glucose estimation profiles more accurately than Ridge and LASSO. This is confirmed, in general, by Table 1, which shows the aggregate results over the test data subset.

By analysing the results in more detail, the LASSO model seems to have a slight advantage in estimating glucose trends (last column of Table 1). The reason is twofold: first,

the regularization performed by the l_1 norm prevents the model coefficients from assuming large values thus predicting glucose profiles that are more flat than the other models, as it happens for example in Figures 3(b) and 4(b); second, channels more sensitive to noise that contain also glucose-related information are considered by Ridge and EN exploiting the effect of the l_2 norm but are less probable to be selected by LASSO, thus yielding smoother estimates. This fact can clearly be seen from Figure 4, where artifacts are present (e.g., in channel no.2) for the session of left data and in channel no.3 for the session of the right data.

Interestingly, the LASSO model seems more robust than the other models to these jumps in the data, preserving the glucose profile with elevated smoothness and reasonably accurate trend. Indeed the l_1 norm shrinks many coefficients to zero allowing an easier interpretation of the results with a reduced number of original variables, representing the strongest effects, considered important for estimating glucose profiles. This is a typical feature of the LASSO to act as
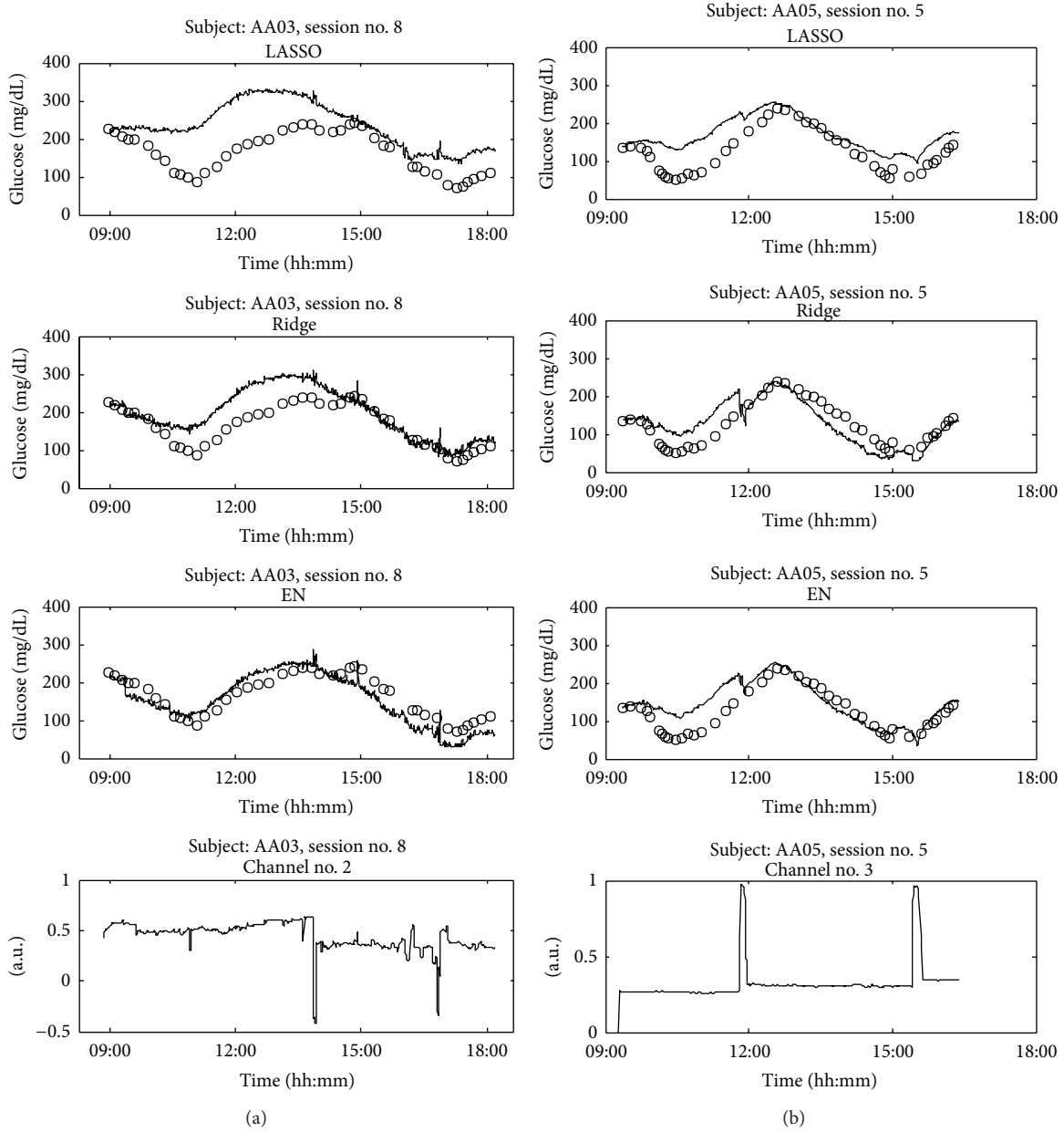
FIGURE 4: Representative recording sessions of Subjects AA03 (a) and AA05 (b). LASSO, Ridge, and EN models test over independent test data subset (continuous lines) versus RBG levels (open bullets). Bottom panels display two representative channels (no.2 and no.3 for subject on the left and on the right, resp.) entering the models, where occasional spikes and jumps are evident. MARD values for the experimental session on the right are of 53% (LASSO), 24.5% (Ridge), and 20.6% (EN), while for the experimental session on the left of 55.7% (LASSO), 34.8% (Ridge), and 34.7% (EN).

a variable selection method. If, from one side, smoother estimates of glucose profiles are obtained with the shrinking properties of the LASSO, sometimes this can lead to biased estimates (see Figure 4(a)).

The Ridge model is identified minimizing the RSS cost function subject to a bound on the l_2 norm of the coefficients. This norm does not have the ability of inducing sparseness on the coefficients of the multivariate linear regression model; thus a parsimonious model is not identified and all the predictors are kept in the model. This might cause the

estimated glucose profiles by the Ridge model to be sensitive to occasional spikes or jumps in the multi-sensor channels, as can be seen in Figure 4(b), where the Ridge model is the one more sensitive among the three. However, estimated glucose profiles by the Ridge model show accuracy indicators slightly better than those of LASSO. This might indicate that (a) channels discharged by the l_1 norm because sensitive to occasional spikes or jumps actually contain useful glucose-related information and (b) that retaining information from all the input channels may help in compensating noisy

TABLE 1: Model test performance when "part 1" of the data set is used for model identification and "part 2" for model test. In brackets is the complexity model parameter chosen by means of cross-validation. Mean and standard deviation (in brackets) over the experimental sessions for root mean squared error (RMSE), mean absolute difference (MAD), mean absolute relative difference (MARD), error grid analysis (EGA (Clarke)) (A + B (A) C/D/E regions whose sum accounts for 100% of data points), continuous error grid analysis (CEGA) ($A_R + B_R$ (A_R) $C_R/D_R/E_R$ regions whose sum accounts for 100% of data points).

	RMSE (mg/dL)	MAD (mg/dL)	MARD (%)	EGA A + B (A) C/D/E	CEGA $A_R + B_R$ (A_R) $C_R/D_R/E_R$
LASSO	57.9	48.6	37.8	89.4 (42.2)	89.2 (62.1)
($j = 15$)	(27.1)	(23.7)	(20)	0.9/9.6/0.1	6.3/2.5/2
Ridge	52.3	44.1	35	91 (58.7)	88 (63)
($\lambda = 5$)	(22.8)	(19.2)	(17.7)	0.1/8.9/0	4.9/4.8/2.3
EN	51.8	43.9	34.1	92.3 (59.9)	88.6 (65)
($\lambda = 0.01$; $\alpha = 0.4$)	(24.3)	(20.5)	(17.2)	0.1/7.6/0	4.9/4.4/2.1

channels thanks to the grouping effect induced by the l_2 norm. Thus, it is reasonable to expect that a combination of the l_1 and l_2 norms could identify a model sharing both properties of sparseness and grouping effect.

Indeed, as mentioned before, the EN model results outperform those of Ridge and LASSO allowing a reasonably accurate estimation of the glucose profile concentrations also when occasional noise is affecting some multi-sensor channels (see Figure 4), presenting lower MARD values than LASSO and Ridge. Thus, EN is the model presenting the best indicators and is only slightly worse than LASSO in accuracy for glucose trends (see CEGA results). Moreover, its clinical accuracy in terms of Clarke error grid, with a percentage of points within the A + B of 92.3 (see Table 1), is substantially close to that of minimally invasive devices, spanning from 98.9 to 96.9 [53].

The good results obtained by the EN model are likely due to the combination of the l_1 and l_2 norms, giving to this model both the advantages of LASSO and Ridge. Indeed, a limitation of the LASSO is that, if there is a group of correlated variables, then it tends to select only one variable from the group and does not care which one is selected, thus lacking the ability of revealing grouping information. On the opposite, the l_2 norm allows all coefficients to enter the model, resulting in more sensitive to noisy channels. Thus, the l_1 norm shrinks channel weights to zero (eliminating multi-sensor channels not useful for predicting glucose), while the l_2 norm encourages a grouping effect (automatically including whole groups into the model once one channel among them is selected). This combination results in indicators outperforming those of the other models and in estimated glucose profiles with a good trade off between sparseness of the model coefficients and robustness due to the grouping effect (see, e.g., Figure 4(a)).

Model test results when data subset "part 2" is used for model identification and data subset "part 1" for model test

are comparable with those in Table 1 (not shown here for the sake of space).

5. Conclusions

In diabetes management, tight monitoring of glycaemic levels by CGM sensors is important for avoiding both long and short term complications related to hyper-and hypo-glycaemic excursions. NI-CGM devices are potentially more appealing than the minimally invasive sensors based on needle electrodes, but their development is challenging for several technological and methodological reasons. In the last years, the idea of embedding sensors of different nature within the same device in order to obtain a better biophysical characterisation of the skin and underlying tissues has gained particular attention to develop NI-CGM. In these multi-sensor approaches, a model linking the measured multi-sensor channels to glucose is needed, together with a set of techniques that can be used to identify its parameters. In this work, we investigated the use of regularisation-based methods to identify the linear regression model employed in the multi-sensor device for NI-CGM proposed in [30]. Results on 45 experimental sessions indicate that the EN model generally outperforms the other models: thanks to the combination of the l_1 and l_2 norms, it allows to take the advantage of the LASSO—shrinking many model weights to zero being more robust to possible occasional jumps or spikes occurring on the multi-sensor data—and of the Ridge model—averaging the contribution of correlated channels allowing a more robust estimation of glucose profiles. With respect to the previous sensor literature, where PLS represents the current state of the art (see [34, 54, 55] to mention just a few), we showed that EN can become very useful with multi-sensor data. While retaining information from a group of variables (as PLS does), EN also automatically selects those channels representing the strongest effects, giving more insight into the specific problem at hand.

To conclude, in this work, we showed that further increased point accuracy can be obtained through suitable techniques for the identification of the multivariate model, representing an important incremental step towards the development of NI-CGM devices. While most of the accuracy indices of Table 1 have not yet reached a fully comparable level with those of current enzyme-based needle sensors [53], glucose trends estimated by the considered NI-CGM device exhibit an acceptable accuracy (last column of Table 1). This result could be potentially important in the treatment of diabetes since the glucose trend can be valid adjunctive information to complement standard SMBG devices that measure glucose by fingerstick, for example, helping the diabetic patient in preventing the occurrence of critical events, such as hypoglycaemia, by exploiting the dynamic risk concept recently developed in [56].

References

[1] M. Brownlee, "The pathobiology of diabetic complications: a unifying mechanism," *Diabetes*, vol. 54, no. 6, pp. 1615–1625, 2005.

[2] *IDF Diabetes Atlas*, 5th edition, 2013, http://www.idf.org/diabetesatlas/.

[3] J. E. Shaw, R. A. Sicree, and P. Z. Zimmet, "Global estimates of the prevalence of diabetes for 2010 and 2030," *Diabetes Research and Clinical Practice*, vol. 87, no. 1, pp. 4–14, 2010.

[4] B. W. Bode and T. Battelino, "Continuous glucose monitoring in 2010," *International Journal of Clinical Practice, Supplement*, no. 170, pp. 10–15, 2011.

[5] G. McGarraugh, "The chemistry of commercial continuous glucose monitors," *Diabetes Technology & Therapeutics*, vol. 11, supplement 1, pp. S17–S24, 2009.

[6] S. Garg, H. Zisser, S. Schwartz et al., "Improvement in glycemic excursions with a transcutaneous, real-time continuous glucose sensor: a randomized controlled trial," *Diabetes Care*, vol. 29, no. 1, pp. 44–50, 2006.

[7] T. Battelino and J. Bolinder, "Clinical use of real-time continuous glucose monitoring," *Current Diabetes Reviews*, vol. 4, no. 3, pp. 218–222, 2008.

[8] G. Sparacino, A. Facchinetti, A. Maran, and C. Cobelli, "Continuous glucose monitoring time series and hypo/hyperglycemia prevention: requirements, methods, open problems," *Current Diabetes Reviews*, vol. 4, no. 3, pp. 181–192, 2008.

[9] C. Zecchin, A. Facchinetti, G. Sparacino, and C. Cobelli, "Reduction of number and duration of hypoglycemic events by glucose prediction methods: a proof-of-concept in silico study," *Diabetes Technology and Therapeutics*, vol. 15, no. 1, pp. 66–77, 2013.

[10] R. Hovorka, J. M. Allen, D. Elleri et al., "Manual closed-loop insulin delivery in children and adolescents with type 1 diabetes: a phase 2 randomised crossover trial," *The Lancet*, vol. 375, no. 9716, pp. 743–751, 2010.

[11] S. J. Russell, F. H. El-Khatib, D. M. Nathan, K. L. Magyar, J. Jiang, and E. R. Damiano, "Blood glucose control in type 1 diabetes with a bihormonal bionic endocrine pancreas," *Diabetes Care*, vol. 35, no. 11, pp. 2148–2155, 2012.

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

[13] A. Dauber, L. Corcia, J. Safer, M. S. Agus, S. Einis, and G. M. Steil, "Closed-loop insulin therapy improves glycemic control in children aged <7 years: a randomized controlled trial," *Diabetes Care*, vol. 36, no. 2, pp. 222–227, 2013.

[14] M. Phillip, T. Battelino, E. Atlas et al., "Nocturnal glucose control with an artificial pancreas at a diabetes camp," *New England Journal of Medicine*, vol. 368, no. 9, pp. 824–833, 2013.

[15] A. Facchinetti, G. Sparacino, and C. Cobelli, "Online denoising method to handle intraindividual variability of signal-to-noise ratio in continuous glucose monitoring," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 9, pp. 2664–2671, 2011.

[16] S. Guerra, A. Facchinetti, G. Sparacino, G. de Nicolao, and C. Cobelli, "Enhancing the accuracy of subcutaneous glucose sensors: a real-time deconvolution-based approach," *IEEE Transaction on Biomedical Engineering*, vol. 59, no. 6, pp. 1658–1669, 2012.

[17] A. Gani, A. V. Gribok, Y. Lu, W. K. Ward, R. A. Vigersky, and J. Reifman, "Universal glucose models for predicting subcutaneous glucose concentration in humans," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 1, pp. 157–165, 2010.

[18] M. Eren-Oruklu, A. Cinar, and L. Quinn, "Hypoglycemia prediction with subject-specific recursive time-series models," *Journal of Diabetes Science and Technology*, vol. 4, no. 1, pp. 25–33, 2010.

[19] A. Facchinetti, G. Sparacino, E. Trifoglio, and C. Cobelli, "A new index to optimally design and compare continuous glucose monitoring glucose prediction algorithms," *Diabetes Technology and Therapeutics*, vol. 13, no. 2, pp. 111–119, 2011.

[20] V. Naumova, S. V. Pereverzyev, and S. Sivananthan, "A meta-learning approach to the regularized learning-case study: blood glucose prediction," *Neural Networks*, vol. 33, pp. 181–193, 2012.

[21] C. Zecchin, A. Facchinetti, G. Sparacino, G. de Nicolao, and C. Cobelli, "Neural network incorporating meal information improves accuracy of short-time prediction of glucose concentration," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 6, pp. 1550–1560, 2012.

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

[23] B. W. Bequette, "Continuous glucose monitoring: real-time algorithms for calibration, filtering, and alarms," *Journal of Diabetes Science and Technology*, vol. 4, no. 2, pp. 404–418, 2010.

[24] A. Facchinetti, G. Sparacino, S. Guerra et al., "Real-time improvement of continuous glucose monitoring accuracy: the smart sensor concept," *Diabetes Care*, vol. 36, no. 4, pp. 793–800, 2013.

[25] A. Tura, "Advances in the development of devices for noninvasive glycemia monitoring: who will win the race?" *Nutritional Therapy and Metabolism*, vol. 28, no. 1, pp. 33–39, 2010.

[26] S. K. Vashist, "Non-invasive glucose monitoring technology in diabetes management: a review," *Analytica Chimica Acta*, vol. 750, pp. 16–27, 2012.

[27] G. Sparacino, M. Zanon, A. Facchinetti, C. Zecchin, A. Maran, and C. Cobelli, "Italian contributions to the development of continuous glucose monitoring sensors for diabetes management," *Sensors*, vol. 12, no. 10, pp. 13753–13780, 2012.

[28] K. V. Larin, M. S. Eledrisi, M. Motamedi, and R. O. Esenaliev, "Noninvasive blood glucose monitoring with optical coherence tomography: a pilot study in human subjects," *Diabetes Care*, vol. 25, no. 12, pp. 2263–2267, 2002.

[29] M. A. Arnold and G. W. Small, "Noninvasive glucose sensing," *Analytical Chemistry*, vol. 77, no. 17, pp. 5429–5439, 2005.

[30] A. Caduff, M. S. Talary, M. Mueller et al., "Non-invasive glucose monitoring in patients with type 1 diabetes: a multisensor system combining sensors for dielectric and optical characterisation of skin," *Biosensors and Bioelectronics*, vol. 24, no. 9, pp. 2778–2784, 2009.

[31] I. Harman-Boehm, A. Gal, A. M. Raykhman, E. Naidis, and Y. Mayzel, "Noninvasive glucose monitoring: increasing accuracy by combination of multi-technology and multi-sensors," *Journal of Diabetes Science and Technology*, vol. 4, no. 3, pp. 583–595, 2010.

[32] C. F. Amaral, M. Brischwein, and B. Wolf, "Multiparameter techniques for non-invasive measurement of blood glucose," *Sensors and Actuators B*, vol. 140, no. 1, pp. 12–16, 2009.

[33] A. Caduff, M. Mueller, A. Megej et al., "Characteristics of a multisensor system for non invasive glucose monitoring with external validation and prospective evaluation," *Biosensors and Bioelectronics*, vol. 26, no. 9, pp. 3794–3800, 2011.

[34] M. Mueller, M. S. Talary, L. Falco, O. de Feo, W. A. Stahel, and A. Caduff, "Data processing for noninvasive continuous glucose monitoring with a multisensor device," *Journal of Diabetes Science and Technology*, vol. 5, no. 3, pp. 694–702, 2011.

[35] M. Zanon, G. Sparacino, A. Facchinetti et al., "Non-invasive continuous glucose monitoring: improved accuracy of point and trend estimates of the multisensor system," *Medical and Biological Engineering and Computing*, vol. 50, no. 10, pp. 1047–1057, 2012.

[36] A. Caduff, M. S. Talary, and P. Zakharov, "Cutaneous blood perfusion as a perturbing factor for noninvasive glucose monitoring," *Diabetes Technology and Therapeutics*, vol. 12, no. 1, pp. 1–9, 2010.

[37] P. Zakharov, F. Dewarrat, A. Caduff, and M. S. Talary, "The effect of blood content on the optical and dielectric skin properties," *Physiological Measurement*, vol. 32, no. 1, pp. 131–149, 2011.

[38] D. Huber, M. Talary, F. Dewarrat, and A. Caduff, "The compensation of perturbing temperature fluctuation in glucose monitoring technologies based on impedance spectroscopy," *Medical and Biological Engineering and Computing*, vol. 45, no. 9, pp. 863–876, 2007.

[39] P. Åberg, *Skin cancer as seen by electrical impedance [Ph.D. thesis]*, Division of Medical Engineering, Department of Laboratory Medicine, Karolinska Institutet, Stockholm, Sweden, 2004.

[40] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[41] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society B*, vol. 58, no. 1, pp. 267–288, 1996.

[42] M. Schmidt, G. Fung, and R. Rosales, "Optimization methods for l1-regularization," Tech. Rep. TR-2009-19, University of British Columbia, 2009.

[43] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, New York, NY, USA, 2004.

[44] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Annals of Statistics*, vol. 32, no. 2, pp. 407–499, 2004.

[45] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Series in Statistics, Springer, New York, NY, USA, 2nd edition, 2009.

[46] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society B*, vol. 67, no. 2, pp. 301–320, 2005.

[47] J. Friedman, T. Hastie, H. Höfling, and R. Tibshirani, "Pathwise coordinate optimization," *The Annals of Applied Statistics*, vol. 1, no. 2, pp. 302–332, 2007.

[48] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1200–1224, 1995.

[49] A. J. van der Kooij, "Prediction accuracy and stability of regression with optimal scaling transformations," Tech. Rep., Department of Data Theory, Leiden University, 2007.

[50] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *Journal of Statistical Software*, vol. 33, no. 1, pp. 1–22, 2010.

[51] W. L. Clarke, D. Cox, L. A. Gonder-Frederick, W. Carter, and S. L. Pohl, "Evaluating clinical accuracy of systems for self-monitoring of blood glucose," *Diabetes Care*, vol. 10, no. 5, pp. 622–628, 1987.

[52] W. L. Clarke, S. Anderson, and B. Kovatchev, "Evaluating clinical accuracy of continuous glucose monitoring systems: continuous glucose—error grid analysis (CG-EGA)," *Current Diabetes Reviews*, vol. 4, no. 3, pp. 193–199, 2008.

[53] E. R. Damiano, F. H. El-Khatib, H. Zheng, D. M. Nathan, and S. J. Russell, "A comparative effectiveness analysis of three continuous glucose monitors," *Diabetes Care*, vol. 36, no. 2, pp. 251–259, 2013.

[54] A. M. K. Enejder, T. G. Scecina, J. Oh et al., "Raman spectroscopy for noninvasive glucose measurements," *Journal of Biomedical Optics*, vol. 10, no. 3, Article ID 031114, 2005.

[55] M. A. Arnold, L. Liu, and J. T. Olesberg, "Selectivity assessment of noninvasive glucose measurements based on analysis of multivariate calibration vectors," *Journal of Diabetes Science and Technology*, vol. 1, no. 4, pp. 454–462, 2007.

[56] S. Guerra, G. Sparacino, A. Facchinetti, M. Schiavon, C. Dalla Man, and C. Cobelli, "A dynamic risk measure from continuous glucose monitoring data," *Diabetes Technology and Therapeutics*, vol. 13, no. 8, pp. 843–852, 2011.