

A Personalized Week-to-Week Updating Algorithm to Improve Continuous Glucose Monitoring Performance

Journal of Diabetes Science and Technology
2017, Vol. 11(6) 1070–1079
© 2017 Diabetes Technology Society
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1932296817734367
journals.sagepub.com/home/dst


Stamatina Zavitsanou, PhD^{1,2}, Joon Bok Lee, PhD¹,
Jordan E. Pinsker, MD², Mei Mei Church, NP²,
Francis J. Doyle III, PhD^{1,2}, and Eyal Dassau, PhD^{1,2}

Abstract

Background: Continuous glucose monitoring (CGM) systems are increasingly becoming essential components in type 1 diabetes mellitus (T1DM) management. Current CGM technology requires frequent calibration to ensure accurate sensor performance. The accuracy of these systems is of great importance since medical decisions are made based on monitored glucose values and trends.

Methods: In this work, we introduce a calibration strategy that is augmented with a weekly updating feature. During the life cycle of the sensor, the calibration mechanism periodically estimates the parameters of a calibration model to fit self-monitoring blood glucose (SMBG) measurements. At the end of each week of use, an optimization problem that minimizes the sum of squared residuals between past reference and predicted blood glucose values is solved remotely to identify personalized calibration parameters. The newly identified parameters are used to initialize the calibration mechanism of the following week.

Results: The proposed method was evaluated using two sets of clinical data both consisting of 6 weeks of Dexcom G4 Platinum CGM data on 10 adults with T1DM (over 10 000 hours of CGM use), with seven SMBG data points per day measured by each subject in an unsupervised outpatient setting. Updating the calibration parameters using the history of calibration data indicated a positive trend of improving CGM performance.

Conclusions: Although not statistically significant, the updating framework showed a relative improvement of CGM accuracy compared to the non-updating, static calibration method. The use of information collected for longer periods is expected to improve the performance of the sensor over time.

Keywords

glucose sensors, weekly updating, continuous glucose monitoring (CGM), type 1 diabetes mellitus

Background

Recent developments in continuous glucose monitoring (CGM) technology have significantly advanced clinical management in type 1 diabetes mellitus (T1DM).¹ CGM sensors provide estimates of glucose concentration based on interstitial glucose measurements at every 1-5 minutes. The sensor is placed on the subcutaneous tissue of the body and measures the electrical signal produced by an electrochemical reaction, taking place on the electroactive surfaces of the sensing region. A calibration process is required to translate the signal measured by the glucose sensor to a meaningful clinical value using a reference glucose value as the standard. Common calibration processes for commercial CGM systems use capillary blood glucose concentration measurements taken by self-monitoring blood glucose (SMBG) as

the reference glucose value.² Recalibration of the currently available sensors is required in frequent time intervals varying from several to one calibration per day, due to changes in sensor sensitivity over its lifespan.^{3,4} Another innovative technological trend in glucose monitoring is the flash glucose monitoring system that is factory calibrated and therefore does not require additional fingerstick SMBG measurements for sensor calibration.⁵

¹Harvard John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA

²William Sansum Diabetes Center, Santa Barbara, CA, USA

Corresponding Author:

Eyal Dassau, PhD, Harvard John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, 02138, USA.
Email: dassau@seas.harvard.edu

However even the latest generation of sensors at the time of this writing have a small degree of error, with reported mean absolute relative difference (MARD) of 9% for the latest Dexcom G5/G4 with 505 algorithm,⁶ 10% for the Medtronic Guardian Sensor 3,⁷ and up to 13% for the Dexcom G4 prior to the 505 upgrade.⁶ CGM systems measure glucose in the interstitial fluid. Therefore, the time required for glucose to diffuse into the interstitial space⁸ results in a time lag between CGM and true blood glucose (BG) readings. As artificial pancreas (AP) technology relies on CGM for adjusting doses of insulin as frequently as every 5 minutes,⁹ dynamic accuracy of the CGM is paramount to the performance of an AP controller and entailing features, for example, fault detection.² Therefore, there is a distinct need to ensure the ability of CGM to provide precise and accurate BG measurements in an effort to enhance the quality of diabetes care.

Several methods aiming to improve glucose sensor accuracy have been reported in the literature, with some performed using batch data analysis offline, while others performed analysis point-to-point that could potentially be used in real-time. In Guerra et al,¹⁰ to account for the interstitial to blood glucose dynamics, deconvolution of the raw sensor signal is performed and a continuous BG profile is obtained that is instead used for calibration. The estimation of the calibration parameters of this method was enhanced by using a Bayesian framework that incorporates a priori statistical knowledge of the parameters.¹¹ A retrofitting algorithm was developed to retrospectively improve CGM accuracy originally tested on Dexcom Seven Plus¹² and then on Dexcom G5 data.¹³ Barceló-Rico et al¹⁴ proposed an adaptive method to compute on-line data normalization parameters of a dynamic calibration model that consists of a set of local models. Other approaches using dynamic calibration models have also been considered.¹⁵⁻¹⁷

However, these methods do not capture the essential cyclic nature of CGM use in people with T1DM. That is, current CGM sensors are replaced after several days (5-7^{6,18} or 14 days¹⁹) of continuous use. This continuous usage provides rich, personalized, time-paired data of raw sensor signal values and SMBG measurements prior to new sensor operation. As a result, continuous CGM wear can be considered within an updating framework to successively improve the accuracy of the CGM over time.²⁰ Such algorithms have been previously used for adaptation of diabetes care regimens to individuals through modification of critical clinical parameters, such as insulin basal rate²¹ or insulin-to-carbohydrate ratios.^{22,23} In addition, AP systems have also been employed within an adaptive framework.^{24,25}

We hypothesized that errors in CGM accuracy could be minimized by exploiting the observed bias of sensor readings in individual patients, as some individuals with T1DM report consistent patterns of error at given times (for example, CGM frequently reporting lower values than their BG meter when glucose is in the lower range). To attempt to

account for this trend, we designed and evaluated a novel week-to-week updating calibration strategy to improve sensor accuracy after each week of successive CGM use in individuals with T1DM, personalized to the sensor pattern observed in each individual. Subsequently, we evaluated the calibration scheme on a new cohort of subjects to show generalizability.

Methods

The proposed calibration strategy consists of a calibration model, an online calibration mechanism and an updating feature. The strategy was refined by evaluating 6 weeks of Dexcom G4 Platinum CGM data collected by 10 adults with T1DM (over 10 000 hours of CGM use), with seven SMBG data points measured per day by each subject in an unsupervised outpatient setting. The calibration mechanism periodically estimates the calibration model parameters to fit SMBG measurements taken approximately every 12 hours. After evaluating each week of data, an optimization problem that minimizes the sum of squared residuals between reference and predicted blood glucose values of the past weeks is solved remotely to identify personalized calibration parameters. The calibration model is then updated with the newly identified parameters for the following week of CGM use (Figure 1). In the following sections, we present the details of the proposed framework.

Calibration Model

A linear calibration model that is widely considered in the literature^{4,16} is used to translate raw sensor signal to glucose predictions. A linear drift of the sensor signal is considered to approximate the change of sensor sensitivity over time.^{4,11}

$$R = my^{BG} + c\Delta t + b, \quad (1)$$

We choose the independent variable y^{BG} to be the reference blood glucose as suggested in Panteleon et al,²⁶ R is the raw sensor signal, Δt the time from sensor insertion in days, m the calibration curve slope, b the calibration curve intercept and c the sensitivity coefficient. The model parameters m , b and c have to be estimated using time paired reference glucose data and the corresponding raw sensor signal data. The produced raw sensor signal is proportional to glucose concentration in the interstitial fluid and the reference BG is measured in the capillary. The different glucose dynamics as well as the time lag of glucose between the two compartments –interstitial and capillary – should be considered.³ Therefore, it is commonly instructed that calibration should be performed when glucose is not changing at a significant rate (>2 mg/dL/min),⁶ for example, before a meal, to approximate that the two compartments are in equilibrium and the difference in glucose concentrations is minimal.

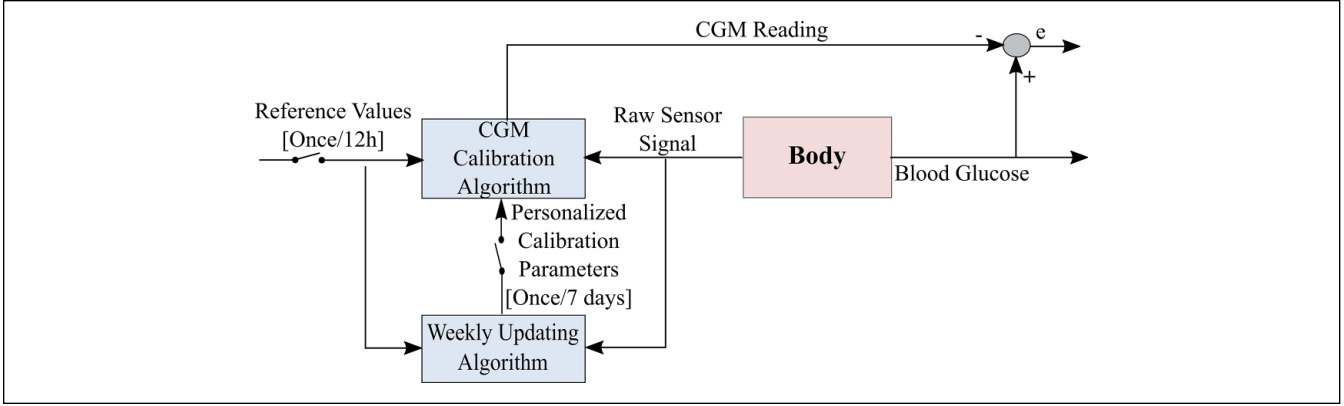


Figure 1. A schematic demonstrating the weekly operation of the proposed calibration strategy. The proposed calibration method is applied directly on the raw sensor signal and therefore there is no learning from the original calibration.

Calibration Mechanism: Recursive Least Squares Parameter Estimation

The sensor takes measurements at discrete times with constant sampling intervals $\tau = 5$ min. Approximately every 12 hours (for Dexcom G4 sensors) the sensor is calibrated and a new time pair of reference glucose value and sensor signal is obtained $(y^{BG}(i), R(i), \Delta t(i)), i = 1, 2, \dots, t$. Let θ denote the calibration parameters m, b and c . At every new calibration point, equation 1 can be written as:

$$\mathbf{Y}^{BG} = \Phi\theta + \mathbf{e}, \quad (2)$$

where

$$\mathbf{Y}^{BG}(t) = [y^{BG}(1) y^{BG}(2) \dots y^{BG}(t)]^T$$

$$\Phi = \begin{bmatrix} \varphi^T(1) \\ \varphi^T(2) \\ \vdots \\ \varphi^T(t) \end{bmatrix}, \text{ with } \varphi^T = [R \ \Delta t \ 1], \theta = \left[\frac{1}{m} - \frac{c}{m} - \frac{b}{m} \right]^T$$

and

$\mathbf{e} = [e(1) e(2) \dots e(t)]^T$, while the superscript \top denotes transposition.

This problem can be solved using recursive least squares with forgetting factor, such that θ is obtained from:

$$\hat{\theta} = \operatorname{argmin} V(\theta, t), \quad (3)$$

with

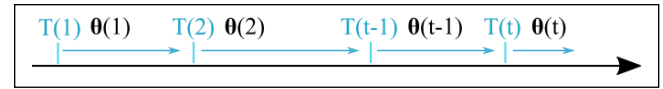


Figure 2. Calibration process. The calibration parameters denoted with θ are updated at $T(i)$ time instants. The interval between parameter updates is usually 12 hours.

$$V(\theta, t) = \frac{1}{2} \sum_{i=1}^t \lambda^{t-i} (y^{BG}(i) - \varphi^T(i)\theta)^2. \quad (4)$$

Here, the weighting factor $\lambda \in (0, 1]$ is used to weight the most recent BG measurements more heavily. A recursive solution of equation 3 can be obtained as:

$$\hat{\theta}(t) = \hat{\theta}(t-1) + K(t)(y^{BG}(t) - \varphi^T(t)\hat{\theta}(t-1)) \quad (5a)$$

$$K(t) = P(t-1)\varphi(t)(\lambda I + \varphi^T(t)P(t-1)\varphi(t))^{-1} \quad (5b)$$

$$P(t) = \frac{1}{\lambda} (I - K(t)\varphi^T(t))P(t-1), \quad (5c)$$

where matrix P is referred to as the covariance matrix. Details on the algorithm derivation can be widely found in the literature.²⁷ At every new calibration point, the estimates on parameters θ are updated using equation 5a, based on the difference between the real blood glucose value and the predicted glucose value. The optimal set of calibration parameters $\theta(t-1)$ obtained at $T(t-1)$, where T is the time of calibration, is applied for the interval $[T(t-1), T(t))$ until the next calibration point, when the new set of parameters is obtained (see Figure 2).

Blood glucose SMBG samples lying outside the reporting range of the sensor, that is, $y^{BG} \notin [40, 400]$ mg/dL are

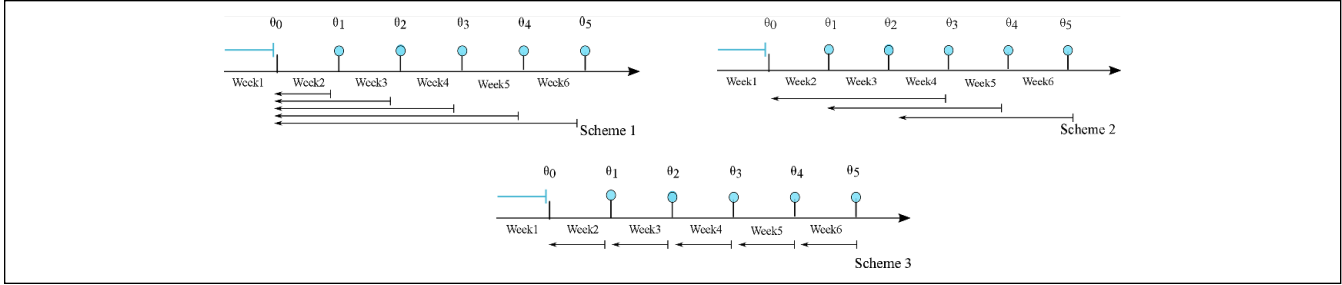


Figure 3. Updating schemes. The blue filled circles denote the end of each week and the implementation of the weekly updating algorithm. The arrows show the history of data used in each scheme. Week 1, in all cases, is not considered in the process, and it is only used to obtain an initial estimate of the parameters.

excluded from calibration. In addition, during standard CGM use, two consecutive reference BG values, taken after two hours from insertion, are required to initialize the sensor session based on the internal calibration algorithm. We initialize our calibration algorithm using the second BG value.

Week-to-Week Update

Herein, we introduce an updating framework that utilizes each subject's past calibration data points to identify personalized calibration parameters that initialize the calibration process of the following week. Thus, starting from population values, as explained in the *Initialization* section below, every week the calibration process is initialized using the newly identified calibration parameters. It is important to note that the weekly updating computations can be performed remotely and therefore no additional resources are required for commercial devices.

At the end of every week of continuous use, the following cost function is minimized:

$$J = \sum_{w=1}^M \mu^{w-1} \left(\mathbf{Y}_w^{BG} - \hat{\mathbf{Y}}_w^{BG} \right)^2, \quad (6)$$

Where $w=1,2,3,\dots,M$ denote the number of weeks. The vector \mathbf{Y}_w^{BG} contains past weekly calibration data until the current week w , such that $\mathbf{Y}_w^{BG} = \{ \mathbf{Y}_1^{BG}, \mathbf{Y}_2^{BG}, \dots, \mathbf{Y}_w^{BG} \}$ with $\mathbf{Y}^{BG}(t) = [y^{BG}(1) y^{BG}(2) \dots y^{BG}(t)]^T$ as specified in equation 2, and $\hat{\mathbf{Y}}_w^{BG}$ is the respective vector of estimated glucose values obtained from equation 1. The parameter μ is a forgetting factor that takes values between $0.9 \leq \mu \leq 0.99$, and controls the trade-off between current and previous data points. In more detail, the current week is denoted with $w=1$, the previous week with $w=2$ and so forth. This indicates that calibration data of the current week is emphasized more compared to calibration data from previous weeks.

The solution of equation 6 gives the set of parameters θ_w that initializes the calibration process for the following week.

The weekly updating feature is applied and evaluated for three windowing schemes (Figure 3).

Scheme 1: In the first scheme the entire history of the calibration data is used for every subject.

Scheme 2: In the second scheme a moving window of the past 3 weeks of calibration data is used for every subject.

Scheme 3: The third scheme is a special case of scheme 2 with one-week horizon.

The idea of using past weekly calibration data to identify individualized calibration parameter sets over time, is based on the assumption that there is an individual consistent pattern of body response to a foreign object. Indeed, it is anticipated that different body reactions can take place after sensor insertion depending on the insertion procedure itself, the insertion site, the chemistry of the biomaterial surface and the transient provisional matrix that forms around it,²⁸ any of which could have an effect on the calibration procedure. As summarized in Anderson et al,²⁹ there are two distinct characteristics that can influence the host response interaction. The first is the material variables, such as composition, surface chemical composition, porosity, and so on of the biomaterial and the second is the generic host response to biomaterials such as protein absorption, macrophages activation and foreign body giant cell production. We assume that since the commercial sensors come in a box of four and all sensors used were from the same factory batch (same lot numbers), the first source of variability of the interactions is minimized and therefore, we focus only on characterizing patient-specific response to the sensor, as reflected through individualized calibration parameters. For the case when sensors from different factory batches are used, the algorithm will gradually adapt to the new calibration settings over time when more weekly sets of data are collected.

Initialization

For the calibration algorithm initialization, two variables need to be specified, the initial values of the parameters θ_0 and an initial choice of the covariance matrix P_0 . To obtain

an initial estimate of the range of the parameter values, equation 6 was solved for all subjects and for week 1, using all available data points. The average of the parameter sets obtained for each subject was used to initialize the algorithm at week 2. Week 1 can be considered as a pre-processing week and not part of the weekly updating procedure. The following weeks, the calibration process was initialized with the parameters θ_w obtained from equation 6 using only past calibration data (Figure 3). The initial covariance matrix of the parameter estimates indicates the level of uncertainty on the initial parameters. With the assumption that we have confidence in the initial estimate of the parameter set that was obtained as the solution of equation 6, the covariance matrix $P_0 \in \mathbb{R}^{3 \times 3}$ is a diagonal matrix with diagonal elements chosen to be $\delta \cdot \theta_w$ for every week. The values of the design parameters δ , λ and μ are specified in a following section.

Performance Evaluation Metrics

The following outcome metrics were used for evaluation:

- 1) Percentage improvement of sum of squared residuals

$$\%RSS = 100 \cdot \frac{RSS_{default} - RSS_{new}}{RSS_{default}}, \text{ with } RSS = \sum_{i=1}^n \left(y_i^{BG} - \hat{y}_i^{BG} \right)^2$$

- 2) Root mean square error

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(y_i^{BG} - \hat{y}_i^{BG} \right)^2}$$

- 3) Mean absolute difference

$$MAD = \frac{\sum_{i=1}^n \left| y_i^{BG} - \hat{y}_i^{BG} \right|}{n}$$

where n is the number of data pairs.

Mean absolute relative difference (MARD) is a commonly considered metric to assess sensor accuracy.³⁰ However, the objective of this study was to learn the personalized calibration parameters of each subject using SMBG data usually collected in everyday life. Therefore, the SMBG dataset collected is not appropriate to compute meaningful MARD values since it is small, and it consists of infrequent SMBG measurements obtained under unsupervised setting. Hence, we decided to exclude MARD for performance assessment since it could result in misleading conclusions when compared to the usual ranges commonly reported in the literature.

Data Sets

We use two sets of data consisting of 6 weeks and 10 subjects. We refer to set 1 as the training set and to set 2 as the

Table 1. Demographics of Study Subjects for the First Data Collection Period.

Characteristic	Mean \pm SD
Age, years	49 \pm 10.1
Number of subjects	10 (3F)
Weight, kg	84.3 \pm 18.0
Body mass index, kg/m ²	27.3 \pm 4.9
HbA1c, %	7.8 \pm 0.92
Duration of diabetes, years	25 \pm 11.4
Total daily insulin, U/day	47.3 \pm 19.7

validation set, since the first set is used to develop the proposed mechanism, and the second set as a completely novel and independent set that we use to evaluate the performance of the proposed mechanism. However, there is a second level of data differentiation. We divide the SMBG data in both sets to data labeled as “calibration data points” and to “test-set data points.” We use the calibration data to apply the algorithm (recursive least square and weekly updating feature) and the test-set data is used exclusively for performance evaluation. The SMBG data obtained in both datasets was labeled as either calibration data or SMBG data; therefore, the distinction between data points was inherent in the dataset and not performed by us.

Training Data Set

To evaluate the proposed method, 10 subjects with T1DM collected 6 weeks of CGM data per subject, with instructions to check at least seven fingerstick glucose measurements per day. Demographics from this first clinical cohort who completed the protocol are presented in Table 1. The data set collected by this cohort was treated as the training set to finalize the design details of the proposed method.

First, we evaluated the performance of the proposed calibration process. We refer to this calibration process as *static* since no week-to-week updating of the parameters is performed yet. We consider the *static* calibration as an intermediate to compare the proposed updating scheme to the default calibration. Since the details of the default calibration are proprietary and, hence, not available, the weekly updating scheme cannot be directly applied to the CGM internal calibration algorithm. The static calibration is initialized for all subjects and all weeks with the average parameter set obtained from week 1, as described in the *Initialization* section. Therefore, week 1 for all subjects is not included in the analysis. Furthermore, for a fair comparison of the proposed algorithm with the original CGM, test BG points that correspond to default CGM values that are outside the displayed range of 40 to 400 mg/dL are excluded from the performance evaluation analysis. Finally, if 80% of the available test-data for a particular week is missing, this week is not considered in the evaluation process, to prevent introduction of bias in the computed performance metrics.

Table 2. Demographics of Study Subjects for the Second Data Collection Period.

Characteristic	Mean \pm SD
Age, years,	58.1 \pm 11.1
Number of subjects	10 (7F)
Weight, kg	77 \pm 16.7
Body mass index, kg/m ²	26.1 \pm 6.5
HbA1c, %	7.2 \pm 0.72
Duration of diabetes, years	40.6 \pm 15.9
Total daily insulin, U/day	39.3 \pm 8.9

Hence, 16 of the total 60 weeks were not included in the analysis and a total of 1531 SMBG samples over all subjects and all weeks were used for evaluation. The performance metrics were computed for all data points in every week and for every subject and a weekly median interquartile range (IQR) over all subjects is reported.

Validation Data Set

To test the generalizability of the algorithm, a second cohort of subjects with T1DM collected 6 weeks of CGM data per subject, with instructions to check at least 7 fingerstick glucose measurements per day. Demographic characteristics of the 10 subjects who completed this second data collection protocol are shown in Table 2.

Similarly to the evaluation process considered for the training set cohort, data labeled as calibration are excluded from the performance analysis and only test-set data points are included.

Selection of Design Parameters

The design parameters that need to be specified are the forgetting factors λ (equation 4) and μ (equation 6) and the parameter δ of the covariance matrix P_0 (equation 5c). To determine parameters λ and δ , we consider that their values vary in the following ranges: $\lambda = \{0.1, 0.2, 0.3, \dots, 0.9, 0.92, 0.94, \dots, 0.99\}$ and $\delta = \{0.0001, 0.001, \dots, 1000\}$. For each combination of the parameters, the static calibration for the training set is solved and the median of root mean square error (RMSE) for all 10 subjects and each week and the median of all 5 weeks is calculated to obtain an insight of the weekly performance. The values of the parameters λ and δ that minimize the median of the performance metric RMSE are selected. The resulting contour plot is illustrated in Figure 4. The optimal values for λ and δ that result in the minimal value of median RMSE are 0.9 and 0.01 respectively. Similarly, the parameter μ is obtained after solving the calibration with the updating feature for μ varying within $\{0.9, 0.91, 0.94, \dots, 0.99\}$ and parameters λ and δ being fixed to their optimal values. The value of μ that minimizes the mean of RMSE for the entire training set is 0.95. Minimization of the mean of RMSE was selected to

obtain μ instead of the median to investigate the averaged performance of the entire training set.

Results

Evaluation of the Proposed Method

The performance metrics calculated for the training set are presented in Table 3. We can observe that the proposed calibration performs similarly to the default CGM calibration and the difference in the reported values is not statistically significant (P value $>$.05). The median RMSE compared to the default calibration is decreased for all weeks except week 2. Indeed, the maximum decrease in median RMSE is 8.10 mg/dL in week 6. This improvement is consistent with the reported %RSS, that for week 6 the median sum of squared residuals of all subjects is decreased by 8.05%. However, for weeks 2 to 5 the median value of %RSS is negative, indicating that the sum of squared residuals is greater for the static calibration compared to the default calibration. The two performance metrics do not show a consistent pattern of accuracy improvement for all weekly sets of data. Nevertheless, the focus of this article is to evaluate the augmentation of the internal calibration algorithm with the weekly updating feature. Since the details of the calibration mechanism employed by commercial CGM are unknown, the proposed calibration algorithm can be considered as an acceptable alternative to default calibration that can be used for the evaluation of the week-to-week updating schemes. The performance of the static calibration is presented in Figure 5 and compared to the default CGM for an exemplar subject. For this particular subject the %RSS is computed to be -8.6% .

The performance of the calibration algorithm augmented with the three weekly updating schemes is compared to the static calibration, and presented in Figure 6. As depicted from the figure, the reduction in sum of square residuals for all schemes is apparent even after one week of implementation, with %RSS ranging from 4.63% to 5.45%. Scheme 1 that uses all past calibration data shows a continuous reduction of %RSS that is maintained across the continuous weeks of use between 0.65 and 5.45%. Scheme 2 does not show the same improvement in %RSS as scheme 1 at week 6, with %RSS being -0.99% compared to 0.65% for scheme 1. This leads to the conclusion that using a window of four consecutive weeks of data (weeks 2, 3, 4, and 5) gives better parameter estimates than when using 3 weeks of data. However, to strengthen this conclusion further studies using more weeks of data are required to determine the optimal window length. Typically, a window of 4 weeks of consecutive CGM use will consist of a well distributed SMBG dataset.

The effect of the quality of the dataset on the parameter estimates is more pronounced for scheme 3 as demonstrated in Figure 6. Although, an initial reduction of 4.63% in %RSS is observed for the first week of implementation, the following weeks the improvement is not consistent and further

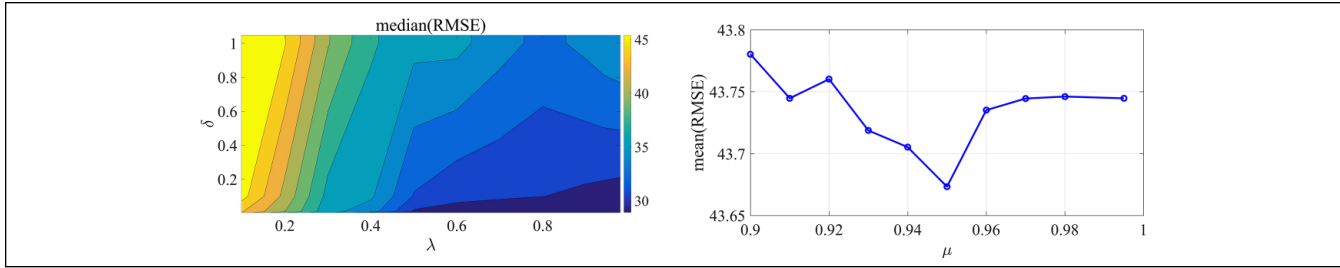


Figure 4. (Left) Contour plot zoomed in the area of interest ($\delta \in [0.000,1]$), indicating the range of parameters that the median of RMSE is minimized. (Right) Mean RMSE of the training set for different forgetting factors μ .

Table 3. Performance Metrics: Comparison Between Default CGM Calibration and Proposed Static Calibration.

	Median %RSS [IQR]	Median RMSE [IQR], default calibration	Median RMSE [IQR], static calibration
Week 1	—	—	—
Week 2	-18.85 [-31.57, 2.51]	40.27 [22.97, 60.61]	42.04 [26.08, 64.14]
Week 3	-4.17 [-15.43, 12.61]	37.84 [25.34, 59.97]	33.66 [24.58, 62.96]
Week 4	-4.74 [-9.14, 2.84]	57.80 [28.83, 63.86]	57.71 [28.54, 65.32]
Week 5	-0.89 [-21.92, 18.93]	41.89 [25.05, 68.03]	39.46 [28.30, 62.69]
Week 6	8.05 [-3.43, 23.49]	36.45 [26.17, 53.07]	28.39 [23.14, 48.90]

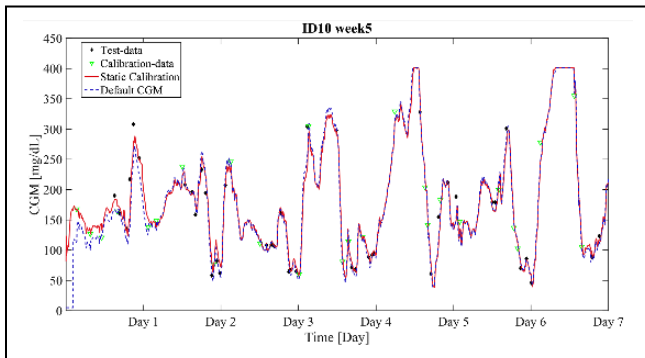


Figure 5. Default calibration versus static calibration for adult no. 10 during week 5. The %RSS is computed to -8.6%.

iteration of the algorithm actually decreased the sensor performance, as for example in week 5 that static calibration outperformed scheme 3 for 50% of the subjects. The median RMSE at week 6 for calibration with scheme 1 is reduced to 29.87 mg/dL compared to scheme 2 with RMSE 30.38 mg/dL and scheme 3 with median RMSE 32.72 mg/dL. While MAD for the same week is reduced from 25.62 mg/dL for scheme 3 to 23.99 mg/dL for scheme 2 and to 23.57 mg/dL for scheme 1.

From Figure 6, we can conclude that scheme 1 presents a consistent reduction in the sum of square residuals for all evaluated weeks. In the rest of the article, we will only consider scheme 1 as the weekly updating scheme. It should be emphasized that scheme 1 is chosen for the particular data set. When more weeks of continuous CGM use are considered, instead of using all-prior data, a moving window with

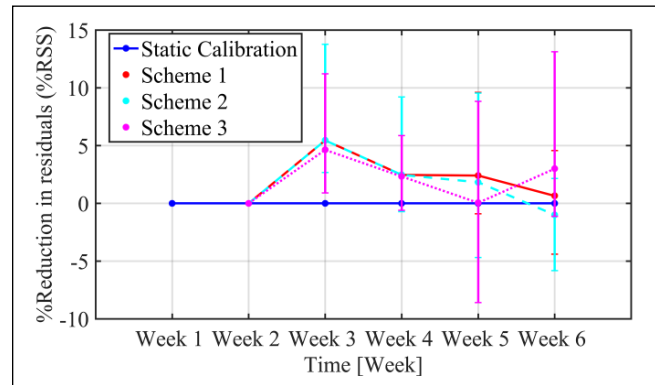


Figure 6. Improvements in percentage reduction of sum of squared residuals (%RSS) over successive weeks of use in terms of median (IQR) static calibration (blue line) versus the three updating schemes, normalized by residuals of the static calibration.

an optimally defined length should be used. In this way, the effect of long-term or seasonal changes in the physiology that could reduce the algorithm’s learning ability is restricted.

Figure 7 presents a boxplot representation of the distribution of RMSE computed on the test-set for separate segments of the week for the default CGM, the static calibration and the weekly-updating calibration with scheme 1. The RMSE distributions are calculated for all subjects and all weeks (except week 1). As demonstrated in the figure, for days 1-3, the weekly-updating feature (median RMSE = 36.98 mg/dL) improves the accuracy of the sensor when compared to the default calibration (median RMSE = 43.37 mg/dL, P value =

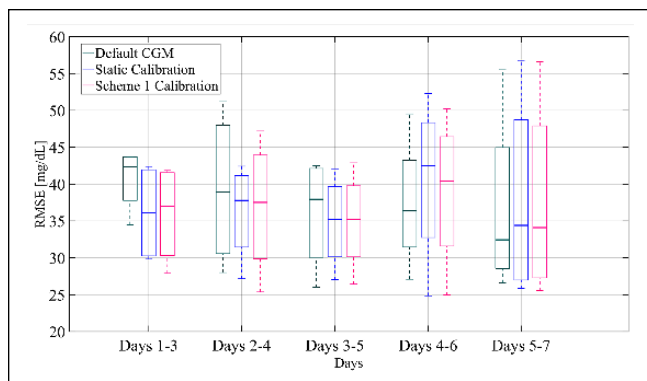


Figure 7. Box-and-whisker plot representation of RMSE distribution over all subjects and all weeks of the training set computed for separate segments of the week for the default CGM, the static calibration and calibration augmented with Scheme 1.

.34) but not for the static calibration (median RMSE = 36.10 mg/dL, P value = .88). In days 2-4 and 3-5, the weekly-updating calibration augmented with scheme 1 demonstrates a decrease in median RMSE compared to the default CGM, by approximately 1.4 mg/dL and 2.7 mg/dL respectively. However, for these segments of the week the RMSE distributions of static and weekly-updating calibration are similar. The same profile is observed for days 4-6 and 5-7, where the weekly updating calibration presents a slight improvement over the static calibration, with the median RMSE being reduced by 2.5 and 0.3 mg/dL respectively.

Validation of the Proposed Method

In this dataset, it was noticed that for some subjects there had been a sensor failure before the end of a week and therefore the sensor had to be replaced. For these cases, the parameter w in equation 6 was not updated and the BG data points for the particular segment of the week had the same weight as the previous week's data points. For a fair comparison of the alternative calibration mechanisms for all subjects, these segments of the week are excluded and only full weeks are considered in the comparison study. Figure 8 presents the boxplot representation of the RMSE distribution over all subjects and all weeks computed for the test-set data pairs for separate segments of the weeks. We can deduce that over all weeks and all subjects days 1-3 after sensor insertion demonstrate a slight increased median RMSE for scheme 1 calibration (27.40 mg/dL) compared to static calibration (26.70 mg/dL, P value = .34). On the contrary, in days 2-4 and 3-5 RMSE presents similar distributions for the static and the weekly updating calibration. However, for days 4-6 after sensor insertion, the median RMSE of calibration with scheme 1 is 23.99 mg/dL compared to 25.43 mg/dL (P value = .88) for static calibration, while for days 5-7, the median RMSE is reduced to 23.57 mg/dL from 23.94 mg/dL (P value = .88) with static calibration.

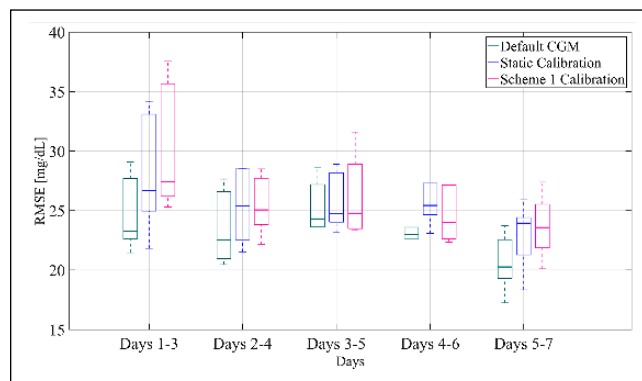


Figure 8. Box-and-whisker plot representation of RMSE distribution over all subjects and all weeks of the validation set computed for separate segments of the week for the default CGM, the static calibration and calibration augmented with Scheme 1.

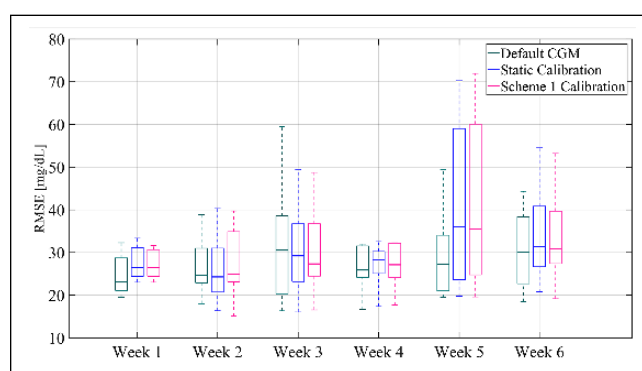


Figure 9. Box-and-whisker plot representation of RMSE distribution averaged over all subjects of the validation set and computed for each week for the default CGM, the static calibration and calibration augmented with Scheme 1.

In conclusion, Figure 9 shows the comparison of RMSE distribution between default CGM, static calibration and calibration with scheme 1 for each week. It can be seen that after week 2 there is a consistent improvement in RMSE for the weekly updating calibration compared to the static calibration. In more detail, the median RMSE of the weekly updating feature for all subjects decreases by 1.89 mg/dL for week 3, by 1.15 mg/dL for week 4, while weeks 5 and 6 present a reduction of 0.49 and 0.52 mg/dL, respectively. It should also be noted that the results are comparable to the default CGM, except for week 5, since the proposed method based on the available data could not capture the sensor performance. Nonetheless, this analysis indicates that the updating of the calibration parameters can improve the accuracy of the CGM over consecutive weeks of use.

Discussion

Improving the accuracy of CGM is very important for diabetes care in general, as well as for the performance of artificial

pancreas technology that may rely exclusively on the CGM for insulin dosing.

Our design aims to improve the accuracy of current CGM sensors by leveraging information that is collected progressively over consecutive CGM use. Unlike studies on sensor accuracy reported in the literature where data sets are collected under supervised settings and at frequent time points,^{11,12,31} the design of this particular study was not intended to evaluate the sensor accuracy. Therefore, the data points collected and used by the proposed algorithm are available during standard CGM use; that is, SMBG calibration points required by the users for satisfactory sensor operation. This, however, can lead to high variability in the number and quality of calibration points collected by the users. Since there are usually few calibration data points per week, it is important to evaluate the distribution of available SMBG data that will be used for parameter estimation; obtaining parameter estimates based on sampling data that is biased for example toward hyperglycemia will not provide reliable parameter values to initialize the calibration process of the following week. Therefore, the sampling scheme selected, is not only a function of the number of consecutive weeks but also a function of the level of adequate data representation. Hence, datasets with varying window lengths can be chosen based on the variance of the calibration data obtained at each consecutive week. A longer study is thus required to evaluate the efficacy of the proposed method. Another factor that could have restricted the effect of the updating feature on the performance of the CGM is the calibration mechanism itself. The sensor calibration method performed by commercial CGM is proprietary, and therefore completely unknown. This implies that the selected calibration model and/or calibration mechanism may differ significantly from the underlying in the default CGM. Even though the *static* calibration demonstrated similar performance to the default CGM, the updating feature did not perform significantly better than the *static* calibration. However, we strongly believe that using past data to identify individual calibration parameters will improve the sensor accuracy over time. Therefore, if the proposed method is applied and further tailored to the specifics of the default calibration method then we can investigate directly its impact on the sensor accuracy without being constrained by errors occurring due to the intermediate step of static calibration.

Conclusions

Accurate CGM performance is of paramount importance for people with T1DM to achieve better glycemic control. The proposed algorithm uses weekly sets of calibration data to identify individual sensor calibration parameters to ultimately improve sensor accuracy over time. Although the results presented in this work do not show statistically significant improvement, the use of information of past calibration data collected on a weekly basis to estimate the

calibration model parameters showed a positive trend toward reducing the weekly RMSE when compared to a static calibration algorithm. Using data points that are collected in an outpatient setting entails the challenge of biased and infrequent sampling that can restrict the performance of the developed algorithms. A dataset comprising of longer duration of continuous CGM use and respective SMBG points is necessary to identify the benefits of using updating features to improve CGM performance over time.

Abbreviations

AP, artificial pancreas; BG, blood glucose; CGM, continuous glucose monitoring; IQR, interquartile range; MARD, mean absolute relative difference; RMSE, root mean square error; MAD, mean absolute difference; SMBG, self-monitoring blood glucose; T1DM, type 1 diabetes mellitus.

Acknowledgments

The authors would like to thank Dr Sunil Deshpande and Dr Ankush Chakrabarty (Harvard John A. Paulson School of Engineering and Applied Sciences, Harvard University) for fruitful discussions, and Tyler Jean and Dr Wendy C. Bevier for helping to collect clinical data (William Sansum Diabetes Center).

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the JDRF grant SRA-2016-147 and National Institutes of Health grant DP3DK104057. Product support was provided by Dexcom, Inc, and the Investigator-Initiated Study Program of LifeScan, Inc.

References

1. Pickup JC, Freeman SC, Sutton AJ. Glycaemic control in type 1 diabetes during real time continuous glucose monitoring compared with self monitoring of blood glucose: meta-analysis of randomised controlled trials using individual patient data. *BMJ*. 2011;343:3805.
2. Facchinetti A. Continuous glucose monitoring sensors: past, present and future algorithmic challenges. *Sensors*. 2016;16:2-12.
3. Matysik F-M. *Trends in Bioelectroanalysis*. Cham, Switzerland: Springer; 2016.
4. Lodwig V, Heinemann L. Continuous glucose monitoring with glucose sensors: calibration and assessment criteria. *Diabetes Technol Ther*. 2003;5:573-587.
5. Bailey T, Bode BW, Christiansen MP, Klaff LJ, Alva S. The performance and usability of a factory-calibrated flash glucose monitoring system. *Diabetes Technol Ther*. 2015;17:787-794.
6. Dexcom. G4 Platinum with Share user guide. 2015. Available at: www.dexcom.com/sites/dexcom.com/files/user-guides/G4-PLATINUM-with-Share-US-ifu-Adult_1.pdf.

7. Garg SK, Weinzimer SA, Tamborlane WV, et al. Glucose outcomes with the in-home use of a hybrid closed-loop insulin delivery system in adolescents and adults with type 1 diabetes. *Diabetes Technol Ther.* 2017;19:155-163.
8. Schmelzeisen-Redeker G, Schoemaker M, Kirchsteiger H, Freckmann G, Heinemann L, Del Re L. Time delay of CGM sensors: relevance, causes, and countermeasures. *J Diabetes Sci Technol.* 2015;9:1006-1015.
9. Kropff J, DeVries JH. Continuous glucose monitoring, future products, and update on worldwide artificial pancreas projects. *Diabetes Technol Ther.* 2016;18(suppl 2):S253-S263.
10. Guerra S, Facchinetti A, Sparacino G, De Nicolao G, Cobelli C. Enhancing the accuracy of subcutaneous glucose sensors: a real-time deconvolution-based approach. *IEEE Trans Biomed Eng.* 2012;59:1658-1669.
11. Vettoretti M, Facchinetti A, Favero SD, Sparacino G, Cobelli C. Online calibration of glucose sensors from the measured current by a time-varying calibration function and Bayesian priors. *IEEE Trans Biomed Eng.* 2016;63:1631-1641.
12. Del Favero S, Facchinetti A, Sparacino G, Cobelli C. Retrofitting of continuous glucose monitoring traces allows more accurate assessment of glucose control in outpatient studies. *Diabetes Technol Ther.* 2015;17:355-363.
13. Del Favero S, Facchinetti A, Sparacino G, Cobelli C. Retrofitting real-life Dexcom G5 data. *Diabetes Technol Ther.* 2017;19:237-245.
14. Barceló-Rico F, Diez JL, Rossetti P, Vehi J, Bondia J. Adaptive calibration algorithm for plasma glucose estimation in continuous glucose monitoring. *IEEE J Biomed Health Inform.* 2013;17:530-538.
15. Knobbe EJ, Buckingham B. The extended Kalman filter for continuous glucose monitoring. *Diabetes Technol Ther.* 2005;7:15-27.
16. Bequette BW. Continuous glucose monitoring: real-time algorithms for calibration, filtering, and alarms. *J Diabetes Sci Technol.* 2010;4:404-418.
17. Kuure-Kinsey M, Palerm CC, Bequette BW. A Dual-Rate Kalman filter for continuous glucose monitoring. In: *International Conference of the IEEE Engineering in Medicine and Biology Society.* New York, NY: IEEE; 2006:63-66.
18. Medtronic. Enlite glucose sensor user guide. 2014. Available at: www.medtronicdiabetes.com/download-library/enlite-sensor.
19. Abbott. Abbott FreeStyle Libre user manual. Available at: https://freestylediabetes.co.uk/images/uploads/documents/FreeStyle_Libre_Manual.pdf.
20. Lee JB, Dassau E, Doyle FJ III. A run-to-run approach to enhance continuous glucose monitor accuracy based on continuous wear. *IFAC-PapersOnLine.* 2015;48:237-242.
21. Palerm CC, Zisser H, Jovanović L, Doyle FJ III. A run-to-run control strategy to adjust basal insulin infusion rates in type 1 diabetes. *J Process Control.* 2008;18:258-265.
22. Palerm C, Zisser H, Bevier WC, Jovanović L, Doyle FJ III. Prandial insulin dosing using run-to-run. *Diabetes Care.* 2007;30:1131-1136.
23. Zisser H, Palerm CC, Bevier WC, Doyle FJ III, Jovanović L. Clinical update on optimal prandial insulin dosing using a refined run-to-run control algorithm. *J Diabetes Sci Technol.* 2009;3:487-491.
24. Wang Y, Dassau E, Doyle FJ III. Closed-loop control of artificial pancreatic beta-cell in type 1 diabetes mellitus using model predictive iterative learning control. *IEEE Trans Biomed Eng.* 2010;57:211-219.
25. Toffanin C, Visentin R, Messori M, Palma FD, Magni L, Cobelli C. Towards a run-to-run adaptive artificial pancreas: in silico results. *IEEE Trans Biomed Eng.* 2017:1-11.
26. Panteleon AE, Rebrin K, Steil GM. The role of the independent variable to glucose sensor calibration. *Diabetes Technol Ther.* 2003;5:401-410.
27. Åström KJ, Wittenmark B. *Adaptive Control.* Reading, MA: Addison-Wesley; 1989.
28. Williams DF. On the mechanisms of biocompatibility. *Biomaterials.* 2008;29:2941-2953.
29. Anderson JM, Rodriguez A, Chang DT. Foreign body reaction to biomaterials. *Semin Immunol.* 2008;20:86-100.
30. Reiterer F, Polteraue P, Schoemaker M, et al. Significance and reliability of MARD for the accuracy of CGM systems. *J Diabetes Sci Technol.* 2017;11:59-67.
31. Pleus S, Schmid C, Link M, et al. Performance evaluation of a continuous glucose monitoring system under conditions similar to daily life. *J Diabetes Sci Technol.* 2013;7:833-841.