

# Model of glucose sensor error components: identification and assessment for new Dexcom G4 generation devices

Andrea Facchinetti · Simone Del Favero ·  
Giovanni Sparacino · Claudio Cobelli

Received: 31 March 2014 / Accepted: 7 November 2014 / Published online: 23 November 2014  
© International Federation for Medical and Biological Engineering 2014

**Abstract** It is clinically well-established that minimally invasive subcutaneous continuous glucose monitoring (CGM) sensors can significantly improve diabetes treatment. However, CGM readings are still not as reliable as those provided by standard fingerprick blood glucose (BG) meters. In addition to unavoidable random measurement noise, other components of sensor error are distortions due to the blood-to-interstitial glucose kinetics and systematic under-/overestimations associated with the sensor calibration process. A quantitative assessment of these components, and the ability to simulate them with precision, is of paramount importance in the design of CGM-based applications, e.g., the artificial pancreas (AP), and in their *in silico* testing. In the present paper, we identify and assess a model of sensor error of for two sensors, i.e., the G4 Platinum (G4P) and the advanced G4 for artificial pancreas studies (G4AP), both belonging to the recently presented “fourth” generation of Dexcom CGM sensors but different in their data processing. Results are also compared with those obtained by a sensor belonging to the previous, “third,” generation by the same manufacturer, the SEVEN Plus (7P). For each sensor, the error model is derived from 12-h CGM recordings of two sensors used simultaneously and BG samples collected in parallel every  $15 \pm 5$  min. Thanks to technological innovations, G4P outperforms 7P, with average mean absolute relative difference (MARD) of 11.1 versus 14.2 %, respectively, and lowering of about 30 % the error of each component. Thanks to the more sophisticated data processing algorithms, G4AP resulted more reliable than G4P, with a MARD of 10.0 %, and a

further decrease to 20 % of the error due to blood-to-interstitial glucose kinetics.

**Keywords** Continuous glucose monitoring · Measurement noise · Diabetes · Parameter estimation · Sensor calibration

## 1 Introduction

New scenarios in the treatment of type 1 diabetes (T1D) have been opened in the early 2000s by the introduction of continuous glucose monitoring (CGM) sensors [34]. CGM sensors are small portable devices whose core is a transcutaneous electrode, implanted in the subcutaneous tissue of the arm or abdomen, where a glucose oxidase electrochemical reaction occurs generating a current signal reflecting concentration of interstitial glucose (IG). This current signal is then converted, almost continuously (1–5 min sampling period), to glucose concentration levels by using a transformation law with parameters tuned via a calibration procedure exploiting reference glucose concentration values periodically collected by using independent portable fingerstick devices. CGM sensors presently in the market can be used for several consecutive days (up to 7) [9, 19, 24, 26].

CGM sensors can be employed for several purposes. An offline analysis of CGM recordings allows a better understanding of how glucose concentration is managed in a given individual, evidencing dangerous hypo/hyperglycemic events that traditional self-monitoring blood glucose (SMBG), consisting in only 3–4 fingerstick samples per day, could not detect, and a quantitative investigation of the risks associated to altered glucose variability [6, 17, 22, 30, 32]. Real-time CGM applications are even more

A. Facchinetti · S. Del Favero · G. Sparacino · C. Cobelli (✉)  
Department of Information Engineering, University of Padova,  
Via G. Gradenigo 6/B, 35131 Padova, Italy  
e-mail: cobelli@dei.unipd.it

challenging and include the generation of preventive alerts to anticipate hypoglycemic or hyperglycemic threshold crossing [3, 12, 33, 36], and artificial pancreas (AP) systems [7, 8, 14, 18, 28, 31]. In spite of promising results and presence of large room for further applications and developments, CGM devices are still not as accurate as standard blood glucose (BG) meters, as evidenced in studies comparing CGM versus BG references frequently measured via laboratory instruments [11, 21]. Recently, some studies reported methods to describe and quantify CGM sensor error [4, 15, 23, 25]. In particular, in [15], three inaccuracy components were identified: distortion due to blood-to-interstitial glucose (BG-to-IG) kinetics, systematic under/overestimations (with time-varying amplitude) due either to errors in sensor calibration, and random measurement noise. A better understanding and a quantitative assessment of these components are of crucial importance in the design of CGM-based applications and in their *in silico* testing, e.g., in the design of AP control algorithms [35].

In the present paper, we aim to assess the error for two new CGM devices belonging to the “fourth” generation of sensors manufactured by Dexcom, the G4 Platinum (G4P) [5], which appeared in the market in 2012, and the advanced G4 for artificial pancreas studies (G4AP) [16], presently available for investigational use only. G4P and G4AP replace a previous, “third” generation, CGM sensor of the same manufacturer, the SEVEN Plus (7P) released in 2009 [1]. All these three CGM devices consist of a 7-day transcutaneous sensor, a transmitter, and a receiver, and measure glucose in the interstitial fluid every 5 min [1, 5, 16]. The transmitter sends an electrical signal to the receiver, where it is processed into a glucose value and adjusted based on twice-daily calibration using self-monitoring of blood glucose (SMBG). As described in [16], G4P and G4AP utilize a glucose oxidase sensor technology improved with respect to that used in the 7P, as well as updated transmitter and receiver hardware. Compared to G4P, the G4AP employs more sophisticated denoising and calibration algorithms to improve accuracy, day-to-day reliability, and consistency from sensor to sensor [16]. To model sensor error, in the present paper we develop an approach recently proposed in [15] for the 7P and usable in presence of multiple CGM datastreams collected by sensors working in parallel. The methodology will be briefly described in Sect. 2, while in Sect. 3 we will apply it to G4P and G4AP datasets where two sensors were simultaneously used in 36 subjects monitored for 7 days and reference BG samples were collected in parallel every  $15 \pm 5$  min in days 1, 4, and 7 [5, 16]. The error model obtained for G4P and G4AP will be also compared with that obtained from data of 9 subjects wearing the 7P undertaken to a similar monitoring, showing that, thanks to technological innovations, G4P outperforms 7P, with average mean absolute relative difference (MARD)

of 11.1 versus 14.2 %, respectively, and lowering of about 30 % the error of each component. Thanks to the more sophisticated data processing algorithms, G4AP is more reliable than G4P, with a MARD of 10.0 %, and a further decrease to 20 % of the error due to BG-to-IG kinetics. A discussion of the insights emerged from the analysis will be reported in Sect. 4, while some conclusions will end the paper in Sect. 5.

## 2 Methods

### 2.1 Databases

The 7P database is taken from a larger pivotal study conducted in 2008 by Dexcom Inc., see [1] for more details. The considered 18 subjects wore simultaneously two 7P sensors with needles placed on the two different sides of the abdominal region. Subjects were instructed to calibrate their CGM device twice per day and to come to the clinic either on day 1, 4, or 7 for a 12-h period of glucose monitoring in which BG samples were measured approximately every  $15 \pm 5$  min using YSI (Yellow Springs, OH) glucose analyzer. However, only 9 subjects and only day 1 of monitoring can be used for the analysis of the present paper. In fact, as specified in [15] and in the next subsection, the method assumes that the portion of data under analysis does not contain any calibration and this requirement was not satisfied by the other nine subjects (who were hospitalized either in day 4 or in day 7).

The G4P and G4AP database is taken from a larger pivotal study conducted in 2011 by Dexcom Inc., see [5, 16] for details. The considered 36 subjects wore simultaneously two G4P sensors at the same time on the abdominal region. Similarly to the 7P protocol, these subjects were instructed to calibrate their CGM device twice per day and to come to the clinic three times, on days 1, 4, and 7 for a 12-h period in which BG samples were measured approximately every  $15 \pm 5$  min using YSI. All data of 36 subjects were suitable for the analysis. Note that the G4AP database is obtained processing the raw sensor data of the G4P database with more sophisticated signal processing algorithms, using the simulated prospective procedure documented in [16].

All study protocols were reviewed by the FDA through the Investigational Device Exemption process and approved by the institutional review boards of all participating centers. All subjects provided witnessed, written informed consent prior to enrollment. Table 1 summarizes subjects' demographic characteristics in terms of disease category (type 1 or type 2 diabetes), sex, age, body mass index (BMI), and glycosylated hemoglobin (HbA1c). More details can be found in [1, 5, 16].

**Table 1** Demographic information for the datasets used in this study

	SEVEN Plus	G4 Platinum/G4AP
Diabetes type	9 T1D	32 T1D, 4 T2D
Sex	6 m, 3 f	19 m, 17 f
Age (years)	48.2 ± 10.3	39.5 ± 14.4
HbA1c	7.6 ± 1.6 %	7.8 ± 1.4 %
BMI	29.4 ± 3.1 kg/m <sup>2</sup>	28.7 ± 6.1 kg/m <sup>2</sup>

Age, HbA1c and BMI are reported as mean ± SD value

2.1.1 Remark 1

Note that the 7P database used here is not the one adopted in [15], which is different, in terms of number of sensors used, protocol, and data acquisition, to the available G4P and G4AP datasets. For sake of easiness of comparison, a 7P dataset collected with modalities similar to those employed for the G4P and G4AP datasets has been chosen.

2.2 The sensor error model

2.2.1 Overall model description

The adopted sensor error model and the identification strategy used to estimate its unknown parameters are those already described in [15] and applied therein to the 7P. Briefly, a schematic general representation of how sensor error affecting the output of  $n$  different CGM sensors working in parallel is generated is shown in Fig. 1 ( $n = 2$  for the data of this paper), where each of the three components, BG-to-IG kinetics, calibration error, and random measurement noise, is described by a dedicated block. As it will be

discussed later in Sect. 2.2.3, the availability of  $n$  different CGM sensors working in parallel is key for the identification of the random measurement noise components [15]. The specific equations usable to describe these components are reported in the subsection below, while the identification of the unknown parameters is discussed in Sect. 2.2.3.

2.2.2 Model of the three error components

As shown in Fig. 1, the BG signal is transformed into the IG signal according to the BG-to-IG kinetics process, which can be modeled using the linear time-invariant two-compartment model of [29]. This model allows relating BG and IG by the following differential equation:

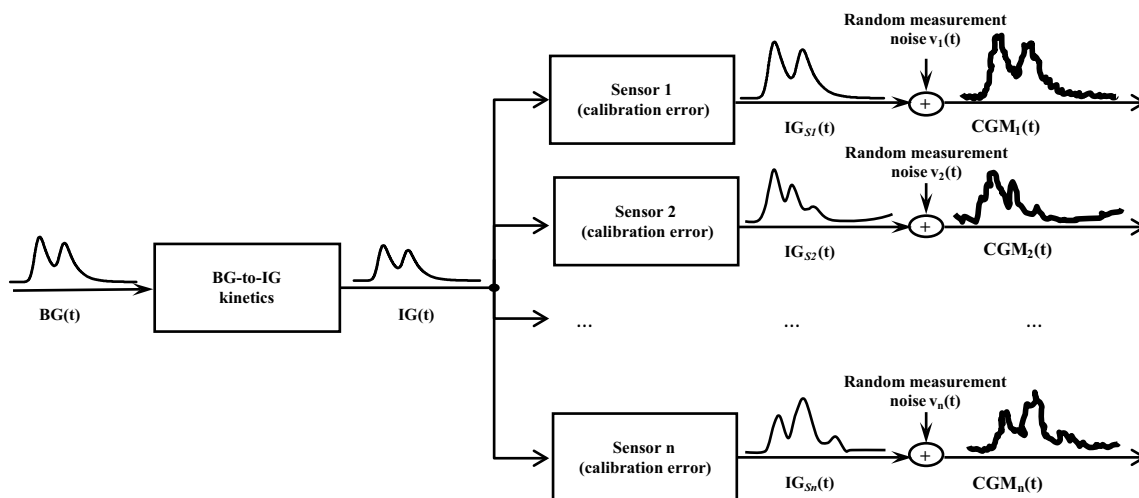
$$\dot{IG}(t) = \frac{1}{\tau}(BG(t) - IG(t)) \tag{1}$$

where  $\tau$  is a time constant. We assume that, in each individual, the influence of BG-to-IG kinetics is the same in all the  $n$  parallel sensor recordings. The time constant  $\tau$  is unknown and is estimated in each subject as described in Sect. 2.2.3.

Then, as displayed in the second layer of the block-scheme of Fig. 1, the IG signal is measured independently by the  $n$  sensors generating the  $IG_{Si}$ ,  $i = 1 \dots n$ , signals. The relationship between IG and  $IG_{Si}$  is modeled as:

$$IG_{Si}(t) = a_i(t)IG(t) + b_i(t) \tag{2}$$

where  $a_i(t)$  and  $b_i(t)$  represent, respectively, the gain and offset of the  $i$ -th sensor. As apparent from the chosen notation, sensor gain and offset are time-varying; this allows taking into account time-variant under-/overestimations produced by the sensor. In order to obtain a flexible



**Fig. 1** Schematic description of how  $n$  simultaneous CGM data-streams are modeled. From left to right, the BG signal is transformed into IG signal through the BG-to-IG kinetics; the IG signal is measured

by each of the  $n$  CGM sensors, generating for the  $i$ -th sensor the  $IG_{Si}$  profile, affected by a calibration error; finally, the measured  $CGM_i$  is affected by additive random measurement noise  $v_i$

description for gain and offset, polynomial models are used:

$$a_i(t) = \sum_{k=0}^m a_{ik} t^k \quad (3a)$$

$$b_i(t) = \sum_{k=0}^l b_{ik} t^k \quad (3b)$$

where  $\{a_{ik}, k = 0 \dots m\}$  and  $\{b_{ik}, k = 0 \dots l\}$  are unknown model parameters. Also, the degrees of the polynomials,  $m$  and  $l$ , are treated as unknown parameters. The identification procedure is described in the next subsection.

Finally, we assume that the output of each of the  $n$  CGM sensors,  $\text{CGM}_i(t)$ , is affected by additive measurement noise  $v_i(t)$ :

$$\text{CGM}_i(t) = \text{IG}_{\text{Si}}(t) + v_i(t) \quad (4)$$

where  $v_i(t)$  is assumed to be composed by two signals: (i) a common component  $\text{cc}(t)$ , assumed to be present in the same realization in all the  $n$  residual profiles of the single subject, largely imputable to possible suboptimal modeling of previous steps (e.g., BG-to-IG kinetics) and other common effects not directly taken into account; (ii) a sensor-specific component  $\text{ssc}_i(t)$ , which is specific to the  $i$ -th sensor and not correlated with  $\text{ssc}_j(t)$ ,  $j \neq i$ . Both  $\text{cc}(t)$  and  $\text{ssc}_i(t)$  are modeled as an autoregressive (AR) process:

$$\text{cc}(t) = \sum_{k=0}^r \beta_k \text{cc}(t-k) + w_1(t) \quad (5a)$$

$$\text{ssc}_i(t) = \sum_{k=0}^q \alpha_{ik} \text{ssc}_i(t-k) + w_{i2}(t) \quad (5b)$$

where the model parameters  $\{\alpha_{ik}, k = 0 \dots q\}$  and  $\{\beta_{ik}, k = 0 \dots p\}$  and the orders  $q$  and  $r$  are unknown and identified as described below, and  $w_1(t)$  and  $w_{i2}(t)$  are zero-mean white noise processes.

### 2.2.3 Identification of the unknown parameters

An a priori identifiability analysis performed analytically (computations not reported for sake of paper readability) demonstrated that all the considered models are uniquely identifiable [27]. In practice, the estimation of the unknown parameters is performed in two steps.

In step #1, we estimate the polynomial degrees  $m$  and  $l$  of  $a_i(t)$  and  $b_i(t)$  using the Bayesian information criterion (BIC). Via nonlinear least squares, we obtain the values

of  $\tau$  and of the coefficients of  $a_i(t)$  and  $b_i(t)$ . Formally, we denote as:

$$\text{BIC}_{m=M, l=L} = d \ln(\text{RSS}) + p \ln(d) \quad (6)$$

the BIC value for the combination of  $m = M$  and  $l = L$ , where  $d$  is the total number of CGM data available in each subject from all the  $n$  sensors;  $p = (M + L + 3)$ , the number of parameters; and RSS, the residual sum of squares calculated as:

$$\text{RSS} = \sum_{i=1}^n \sum_{j=1}^{d_i} \eta_{ij}^2 \quad (7)$$

in which  $d_i$  is the number of CGM samples of the  $i$ -th sensor and  $\eta$  is the uncorrelated version of the measurement noise  $v$  of (5), obtained using a population whitening filter and needed to apply BIC. As suggested in [15], we will consider only the following candidate combinations of  $m$  and  $l$ :  $m = l = 0$  (“constant”), where parameters are assumed to be time-invariant (i.e., no time-variance);  $m = l = 1$  (“linear”), i.e., both the gain and the offset vary in time following a linear trajectory;  $m = l = 2$  (“quadratic”), i.e., a quadratic time-evolution is considered; and  $m = l = 3$  (“cubic”), i.e., time variations obey to a cubic law. In the specific, to assess the optimal orders  $m$  and  $l$ , we will look at the differences in BIC value:

$$\Delta \text{BIC}_{(m=l=P, m=l=P+1)} = \text{BIC}_{m=P, l=P} - \text{BIC}_{m=P+1, l=P+1} \quad (8)$$

Note that a positive difference in  $\Delta \text{BIC}$  means that the higher-order model performs better than the lower one (for instance,  $\Delta \text{BIC}_{(m=l=0, m=l=1)} > 0$  means that the “linear” performs better than the “constant” model).

In step #2, as demonstrated in [15], thanks to the availability of multiple sensors, we can dissect the random measurement noise into the common component  $\text{cc}(t)$  and the sensor-specific component  $\text{ssc}_i(t)$ , and estimate their unknown parameters. Starting from the residual profile  $\text{res}_i(t)$ :

$$\text{res}_i(t) = \text{CGM}_i(t) - \text{IG}_{\text{Si}}(t) \quad (9)$$

the  $\text{cc}(t)$  is derived as follow:

$$\widehat{\text{cc}}(t) = \frac{1}{n} \sum_{i=1}^n \text{res}_i(t) \quad (10)$$

while the realization of the  $\text{ssc}_i(t)$  is extracted as:

$$\widehat{\text{ssc}}_i(t) = \text{res}_i(t) - \widehat{\text{cc}}(t) \quad (11)$$

and both fitted by the AR models of (5a) and (5b), where the orders  $r$  and  $q$  and their coefficients  $\{\alpha_{ik}, k = 0 \dots q\}$  and  $\{\beta_{ik}, k = 0 \dots r\}$  are estimated using the BIC criterion as well.

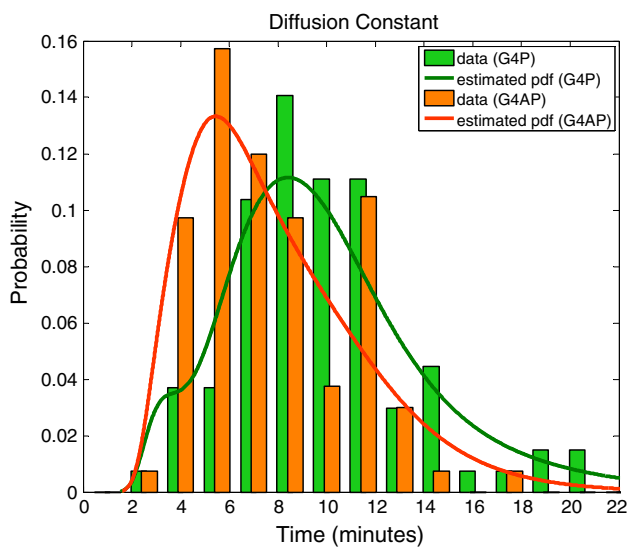
### 3 Results

#### 3.1 G4P and G4AP error components

##### 3.1.1 BG-to-IG kinetics

Figure 2 depicts the distribution of estimated  $\tau$  values in G4 (green) and G4AP (orange) datasets. The continuous line is the probability density function derived from the data of the histogram via nonparametric kernel density estimation. It is apparent how the estimated  $\tau$  values vary among individuals, in line with the results already reported in [15] for the 7P. With regard to the numerical values of  $\tau$ , for the G4P dataset the average value resulted 9.7 min (standard deviation SD = 3.6 min), while for the G4AP dataset the average value was 7.7 min (SD = 3.0 min), which is significantly lower ( $p < 0.01$ ). Since the G4AP dataset is obtained from the G4 one using, prospectively, different signal processing algorithms [16], this reduction in  $\tau$  values should be interpreted as result of a better filtering of raw sensor data performed in G4AP. In fact, for each individual the effect of BG-to-IG kinetics on  $\tau$  does not change between the two datasets.

**3.1.1.1 Remark 2** The distribution of the estimated  $\tau$  values reported in Fig. 2 is relative to the orders  $m$  and  $l$  of the polynomials  $a_i(t)$  and  $b_i(t)$  identified as best according to the BIC criterion, i.e.,  $m = l = 1$  for day 1, and  $m = l = 0$  for both day 4 and day 7 (see next section for details). No



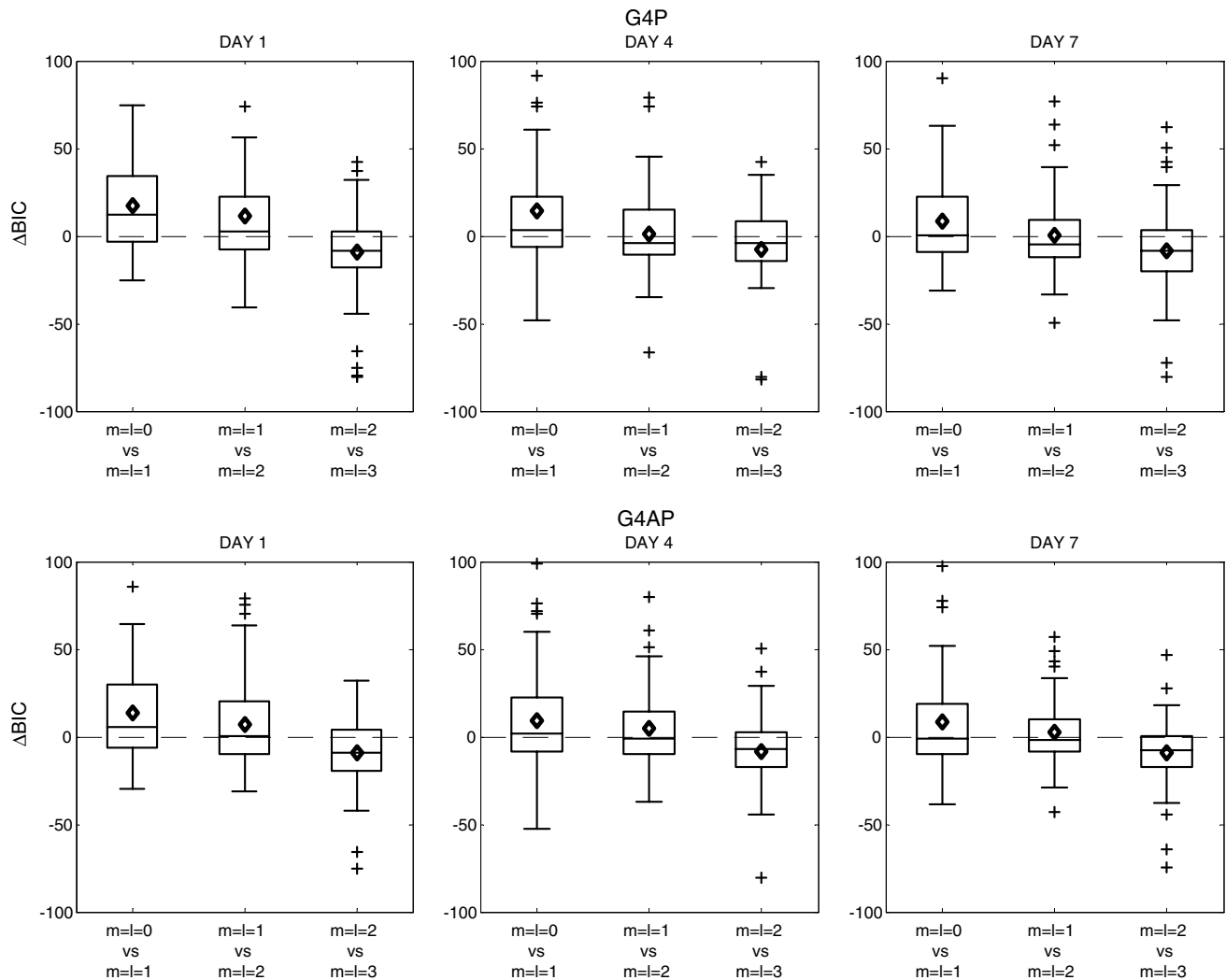
**Fig. 2** Distribution of estimated  $\tau$  values for G4P (green) and G4AP (orange) datasets relative to the orders  $m$  and  $l$  of the polynomials  $a_i(t)$  and  $b_i(t)$  identified as best according to the BIC criterion, i.e.,  $m = l = 1$  for day 1, and  $m = l = 0$  for both day 4 and day 7. The continuous line is the probability density function derived from the data of the histogram via nonparametric kernel density estimation (color figure online)

significant differences on the distribution of the estimated  $\tau$  values have been observed changing  $m$  and  $l$ .

##### 3.1.2 Calibration error

Differently from [15], in which only a 12-h time window was available, the datasets employed in this work allow characterizing the sensor error in three 12-h time windows located in days 1, 4, and 7, where YSI reference BG measurements are available. The characterization of the sensor error in these time windows is important since the 3 days are representative of three specific pictures of sensor behaviors: just after the insertion (day 1), standard working modality (day 4), and close to end of life (day 7). Figure 3 shows a summary of the results (top G4P, bottom G4AP) for the selection of the orders  $m$  and  $l$ , respectively, for the polynomials  $a_i(t)$  and  $b_i(t)$  of the (3a) and (3b) describing the sensor gain and offset, in day 1 (left), day 4 (middle), and day 7 (right) of monitoring. For each day, the box-plots of  $\Delta BIC$  values of  $m = l = 0$  versus  $m = l = 1$  (left),  $m = l = 1$  versus  $m = l = 2$  (middle), and  $m = l = 2$  versus  $m = l = 3$  (right) models are shown. The diamond represents the mean value.

With regard to the G4P sensor (top line panels in Fig. 3), starting from day 1 (left panel), the use of polynomials  $a_i(t)$  and  $b_i(t)$  of order 1 ( $m = l = 1$ ) is necessary for a correct modeling of the calibration error. In fact, on average, the mean BIC value for  $m = l = 1$  is significantly lower than that for  $m = l = 0$  ( $p = 0.001$ ), with  $BIC_{m=l=1} \leq BIC_{m=l=0}$  (i.e.,  $\Delta BIC_{(m=l=0, m=l=1)} > 0$ ) in 70 % of the cases. The incremental improvement passing to  $m = l = 2$  is not so evident, being the reduction on average BIC not significant ( $p = 0.188$ ) and  $BIC_{m=l=2} \leq BIC_{m=l=1}$  in about 50 % of the subjects. This means that the use of a second-order polynomial could be necessary to describe the sensor error in some cases, but not on average. Finally,  $m = l = 3$  is not convenient, being the average  $BIC_{m=l=3}$  significantly higher than the lower-order model ( $p < 0.01$ ). In conclusion, results evidence that G4P sensor presents a time-variant calibration error during day 1 and that the choice of  $m = l = 1$  for the polynomials  $a_i(t)$  and  $b_i(t)$  is preferable. This result is expected, being known that, during day 1, CGM sensors usually suffer a relative lack accuracy due to sensor variability. Focusing now on day 4 (middle panel), none of the BIC distributions is significantly lower than the others. In fact, starting from the  $m = l = 0$  versus  $m = l = 1$  model comparison, even if  $\Delta BIC_{(m=l=0, m=l=1)} > 0$  in 59 % of the cases, evidencing that in some subjects time-variance of gain and offset is still present, the use of  $m = l = 0$  for the polynomials  $a_i(t)$  and  $b_i(t)$  is sufficient for a satisfactory description of the calibration error, being the average  $BIC_{m=l=0}$  values not significantly different from those of  $BIC_{m=l=1}$  ( $p = 0.19$ ).



**Fig. 3** Boxplots of  $\Delta\text{BIC}$  values (top G4P, bottom G4AP) obtained in the determination of the orders  $m$  and  $l$  of the polynomials  $a_i(t)$  and  $b_i(t)$  of the calibration error function in day 1 (left), day 4 (middle), and day 7 (right) of monitoring. For each day,  $m = l = 0$  versus

$m = l = 1$  (left),  $m = l = 1$  versus  $m = l = 2$  (middle), and  $m = l = 2$  versus  $m = l = 3$  (right) models are compared. The diamond represents the mean value. Note that a positive value of  $\Delta\text{BIC}$  means that the higher order is preferable

Also in this case, the use of polynomials of higher orders (i.e., second or third) is not convenient. Finally, conclusions similar to those drawn for day 4 can be applied to day 7, where a constant approximation of  $a_i(t)$  and  $b_i(t)$  is selected as optimal (right panel).

Moving to the G4AP sensor (bottom line panels in Fig. 3) and focusing on day 1 (left), order 1 for the polynomials  $a_i(t)$  and  $b_i(t)$  is still assumed as the most convenient, being  $\text{BIC}_{m=l=1}$  significantly lower than  $\text{BIC}_{m=l=0}$  ( $p = 0.02$ ). However, the percentage of cases in which  $\Delta\text{BIC}_{(m=l=0, m=l=1)} > 0$  (i.e.,  $\text{BIC}_{m=l=1} \leq \text{BIC}_{m=l=0}$ ) reduces to 59 % (from 70 % of G4P), meaning that the enhanced signal processing of G4AP is able to better compensate the time-varying effects of calibration error. On both day 4

(middle) and day 7 (right), order 0 for both polynomials  $a_i(t)$  and  $b_i(t)$  is sufficient for a satisfactory description of the calibration error. Of note that, in day 4,  $\Delta\text{BIC}_{(m=l=0, m=l=1)} > 0$  in 53 % of the cases (from 59 % of G4P), evidencing also in this case an improvement, even if slighter, in compensation of the time-varying effects of calibration error.

Table 2 summarizes, in terms mean  $\pm$  SD, the estimated model parameter values for the selected orders only ( $m = l = 1$  for day 1,  $m = l = 0$  for both day 4 and day 7) for both G4 and G4AP sensors. Of note is that all model parameters are identified with an elevated precision (the percent coefficient of variation, CV, is  $< 20$  % in almost all cases), confirming that the considered models are practically identifiable.

**Table 2** Mean  $\pm$  SD parameter values and the percentage of CV < 20 % for each day analyzed for both G4 and G4AP sensors

Sensor	Day	$a_0$			$a_1$			$b_0$			$b_1$		
		Mean	SD	% of CV < 20 %	Mean	SD	CV < 20 %	Mean	SD	% of CV < 20 %	Mean	SD	% of CV < 20 %
G4	1	1.16	0.31	100	-0.000116	0.000791	97	-9.4	55.6	97	0.0027	0.1289	91
	4	1.04	0.16	100	-	-	-	2.8	15.8	100	-	-	-
	7	1.05	0.18	100	-	-	-	1.9	25.6	100	-	-	-
G4AP	1	1.09	0.26	100	-0.000060	0.000615	94	-6.4	50.1	94	0.0133	0.1090	90
	4	1.05	0.15	100	-	-	-	-2.6	14.9	100	-	-	-
	7	1.07	0.13	100	-	-	-	-0.9	16.2	100	-	-	-

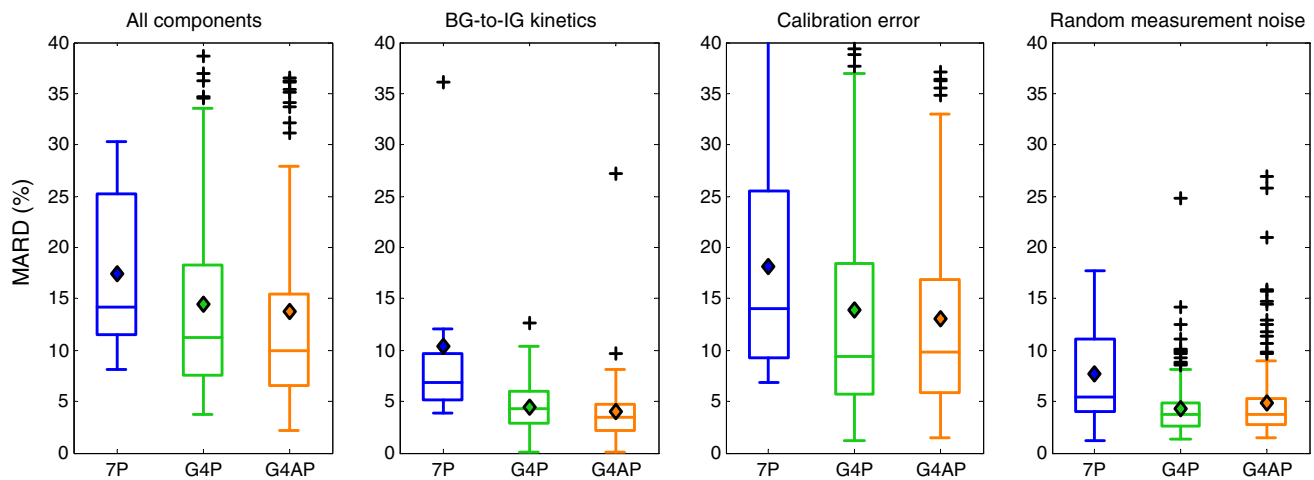
As stated in Sect. 3.1.2,  $m = l = 1$  is selected for day 1, while  $m = l = 0$  for both day 4 and day 7

**Table 3** Median variance values of AR(2) processes describing the common and the sensor-specific components in G4P and G4AP sensor

Day	Component	Median variance (mg <sup>2</sup> /dl <sup>2</sup> )	
		G4P	G4AP
1	Common	45.7	31.8
	Sensor-specific	13.1	12.3
4	Common	36.3	31.0
	Sensor-specific	11.7	8.9
7	Common	26.6	28.2
	Sensor-specific	5.8	5.9

3.1.3 Random measurement noise

After having compensated the delay due to BG-to-IG kinetics and sensor filtering, and time-variant under-/over-estimations due to calibration error, the next last step is to consider the residuals, i.e., what remains unexplained after the two previous steps. Thanks to the availability of  $n$  CGM sensors in parallel (here  $n = 2$ ), the modeling technique of [15] estimates if the residuals are realizations of sensor-specific component  $ssc(t)$  only or if there is also a common component  $cc(t)$ , reflecting possible suboptimal modeling of previous steps (e.g., BG-to IG kinetics) and other common effects not directly taken into account. The analysis of the residuals highlights, for both G4P and G4AP, the presence of a common component in all the 3 days considered. This confirms what already concluded in [15], suggesting that the model employed so far is not perfect, e.g., the modeling of BG-to-IG kinetics can be improved using suitable tracer data [2]. With regard to the specific description of these two components, the analysis evidenced that both can be optimally described by an autoregressive (AR) process of order 2 in all the 3 days of monitoring. This is in line with the results previously achieved on 7P [15]. Table 3 summarizes the results of such an investigation, which reports the median variance value of the AR(2) processes identified on the common and the sensor-specific components for each of the 3 days. The first interesting result is that the variance of the common component reduces in G4 (but not significantly) and hardly changes in the G4AP, which is expected, being both suboptimal modeling of previous steps and other effects not directly taken into account by the employed models common to all sensors. The second result is that the variance of the sensor-specific component reduces significantly during the monitoring. More precisely, in the G4P case, the variance of AR(2) process describing the sensor-specific component is reduced significantly passing both from day 1 to day 4 (13.1–11.7 mg<sup>2</sup>/dl<sup>2</sup>,  $p < 0.001$ ) and from day 4 to day 7 (11.7–5.8 mg<sup>2</sup>/dl<sup>2</sup>,  $p = 0.02$ ). Similarly, the variance of



**Fig. 4** Mean absolute relative difference (MARD) of 7P (blue), G4P (green), and G4AP (orange). From left to right, the global MARD, calculated comparing  $CGM_i$  versus YSI signals; the MARD due to BG-to-IG kinetics and sensor data processing, calculated comparing YSI versus IG signals; the MARD of calibration error, obtained com-

paring IG versus  $IG_{Si}$  signals; and the MARD of the random measurement noise, which derives from the comparison of  $IG_{Si}$  versus  $CGM_i$  signals. The diamond represents the mean value (color figure online)

AR(2) process describing the sensor-specific component in G4AP sensor significantly decreases passing both from day 1 to day 4 ( $12.3\text{--}8.9\text{ mg}^2/\text{dl}^2$ ,  $p < 0.001$ ) and from day 4 to day 7 ( $8.9\text{--}5.9\text{ mg}^2/\text{dl}^2$ ,  $p = 0.02$ ). Finally, the median variance of both common and sensor-specific components of the G4AP is always comparable, or lower, than that of G4P sensor, even if the difference is never statistically significant.

### 3.2 Comparison of G4P and G4AP with 7P

#### 3.2.1 Comparison between the components

With regard to the time constant of the BG-to-IG kinetics model, the average  $\tau$  value in the 7P dataset is 12.0 min, which is significantly higher not only than the values found for G4P and G4AP, 9.7 and 7.7 min, respectively ( $p < 0.01$  for both), but also higher than those reported in the literature for 7P, around 7–8 min [15, 20]. The fact that 7P data were available for day 1 only is not critical, since no differences between average  $\tau$  values estimated separately in day 1, 4, and 7 were evidenced in both G4P and G4AP. We can then speculate that the main cause of this discrepancy is the too limited number of subjects of the 7P dataset, which does not allow a robust estimation of the 7P average  $\tau$  value. Therefore, it becomes difficult to draw any solid conclusion on this comparison.

As concerns the calibration error, comparing G4P and G4AP results with that obtained on 7P dataset, on both fourth generation sensors time-variance of gain and offset is still present. However, it is possible to affirm that

the calibration error is time-variant in day 1 of monitoring only, in which a polynomial of degree 1 for  $a_i(t)$  and  $b_i(t)$  is necessary, with G4AP better suited than G4P for time-variance compensation.

Finally, let us focus on the random measurement noise and, in particular, on the common and sensor-specific components. If we compare the distributions of the variances of G4P and G4AP (median value in Table 2) with those obtained for the 7P dataset (median values  $57.9$  and  $28.9\text{ mg}^2/\text{dl}^2$  for the common and sensor-specific components, respectively, aligned with those reported in [15]), there is a significant reduction in all variances (7P vs. G4P and 7P vs. G4AP,  $p < 0.01$ ), the only exception being the variance of common component of 7P versus G4P in day 1. This result shows that, with respect to 7P, both G4P and G4AP present a decrease in the variance of both the components of the random measurement noise, showing an increased general performance of the sensor in terms of accuracy and precision of the output.

#### 3.2.2 MARD evaluation

Figure 4 shows a schematic representation of the distribution of the error of the three sensors (7P, G4P, and G4AP) in terms of mean absolute relative distance (MARD). From left to right: the global MARD, calculated comparing  $CGM_i$  versus YSI signals; the MARD due to BG-to-IG kinetics and sensor data processing, calculated comparing YSI versus IG signals; the MARD of calibration error, obtained comparing IG versus  $IG_{Si}$  signals; and the MARD of the random measurement noise, which derives from the



comparison of  $IG_{Si}$  versus  $CGM_t$  signals. As shown in [5], the MARD of the G4P is lower than that of the 7P sensor, as demonstrated by the significant reduction of the median value of MARD from 14.1 % of 7P to 11.2 % of G4P ( $p = 0.036$ , nonparametric Wilcoxon rank-sum test). Here, thanks to the dissection of the sensor error using the proposed modeling approach, we can further demonstrate that all sensor error components have been significantly reduced: the delay due to the BG-to-IG kinetics and sensor filtering from 6.8 to 4.4 % ( $p < 0.001$ ); the calibration error from 14.1 to 9.4 % ( $p = 0.031$ ); and the random measurement noise from 5.4 to 3.7 % ( $p = 0.003$ ). On average, all the components have been reduced of more than 30 %.

When comparing G4P and G4AP, the improvement achieved by G4AP reported in [16] is also confirmed. The global MARD of the G4AP is significantly lower than that of the G4P sensor, passing from a median value of MARD from 11.2 to 10.0 % ( $p = 0.047$ ). The dissection into the different sensor error components evidences a significant reduction in the error attributed to the BG-to-IG kinetics and sensor filtering, from 4.4 % of G4P to 3.4 % of G4AP ( $p < 0.001$ ). This confirms that the new signal processing algorithms employed in the G4AP are able to filter raw data better than in G4P, significantly reducing the delay due to filtering. With regard to the calibration error component, and the random measurement noise error component, no reduction in median MARD is present.

#### 4 Discussion

Thanks to the availability of BG references measured in parallel to  $n = 2$  CGM sensor datastreams during three different days of monitoring, we have been able to characterize sensor error components in all the key use scenarios of the sensor, i.e., just after the insertion (day 1), during standard working modality (day 4), and close to end of use (day 7). The analysis highlighted that the behavior of the CGM sensor evolves during the monitoring. In day 1, significant variability in both sensor gain and offset is present, while during day 4 and 7 this component tends to be marginal. In addition, the measurement noise component appears to decrease day after day. This result brings new insights into CGM sensor error modeling, since the models proposed so far in [4, 15, 25] were, in a certain sense, “global,” i.e., the CGM sensor error model of day 1 was identical to that of day 7 and insensitive to inter-day sensor variability.

With regard to the analysis of the time constant  $\tau$  of the BG-to-IG kinetics model, results confirm the finding of [15], i.e.,  $\tau$  is subject specific and it varies from subject to subject. It is worth noting that the estimation of  $\tau$  is likely slightly biased because of the delay of CGM with respect to IG due to hardware/software sensor processing, which

however does not change between sensors of the same technology. As far as numerical values are concerned, we showed that, thanks to improved filtering and calibration algorithms for data processing, G4AP is able to achieve a better performance (in terms of accuracy) than G4P sensor, significantly reducing of about 2 min (from 9.7 to 7.7 min) the delay with respect of BG concentration. As concluded in [15], the model of the BG-to-IG kinetics is still the weakest link of the model and needs to be improved when suitable data for model identification will become available [2]. For instance, the distribution of the  $\tau$  value for G4AP presents a second peak at 12 min, which is unexpected and probably due to a suboptimal modeling of BG-to-IG kinetics.

Moving to the analysis of the optimal order  $m$  and  $l$  for polynomials  $a_i(t)$  and  $b_i(t)$  of the calibration error, the results are similar for both G4P and G4AP. The first note is that a third-order polynomial is never necessary. Focusing on day 1, polynomials of order 1 appear to be preferable, even if a quadratic one could be needed in some cases. With regard to days 4 and 7, polynomials of order 0 appear to be sufficient for a satisfactory description calibration error, confirming that the time-variance of calibration parameters tends to decrease during the monitoring.

Other interesting observations can be drawn from the analysis of the variances of the common and sensor-specific components. First, the variance of the common component does not change during the monitoring, confirming that this systematic error due to suboptimal modeling (likely of the BG-to-IG kinetics) or other common effects not directly taken into account is independent from the day of monitoring. Second, for both sensors, the sensor-specific component variance decreases during the monitoring, evidencing that the uncertainty on the output of the sensor is greater just after the insertion.

Furthermore, the dissection of sensor error pointed out that both G4P and G4AP present for each of the three error components a significant reduction with respect to the 7P, while the improvement of G4AP with respect of G4P is less evident, being statistically significant for the overall and BG-to-IG kinetics errors only. This result is expected, because 7P and G4P/G4AP are sensors of different generations and thus based on different technology, while G4P and G4AP share the same technology and differ on software data processing only. The availability of other larger G4P and G4AP datasets (presently not available to us) would be useful to confirm the results. In general, as already discussed in [15], the largest source of CGM inaccuracy remains the error deriving from calibration.

Some observations on 7P dataset and results are also necessary. The first observation concerns the MARD outcomes relative to 7P sensor, which are slightly higher than those obtained, with a different dataset, in [15]. This

difference can be attributed both to the different protocol used during the hospitalized day and, more likely, to the different day of acquisition of YSI reference data with respect to the start of CGM data acquisition. In fact, in the 7P dataset considered in this work, YSI values are available in day 1 only, while in the dataset used in [15] they are available in day 3 only; as demonstrated by several papers regarding the accuracy of 7P sensor, e.g., [1, 5], the accuracy in day 1 is often worse than in the other days of monitoring. The second observation concerns a possible bias in comparison with 7P versus G4 sensors MARD values. In fact, MARD values of the 7P sensor are calculated on day 1 of monitoring only, while those of G4P and G4AP sensors are relative to days 1, 4, and 7. As mentioned before, being the accuracy for day 1 often worse than in the subsequent days of monitoring, a bias could be present. However, in practice, this bias is very limited and does not affect the results of the comparison, being the reported overall median MARD (14.1 %) in line with that calculated on the whole 7-day monitoring and reported in recent studies (e.g., 14.0 % of [5]). Finally, it may seem that the results for the 7P sensor error model are incomplete, being available for day 1 only. We believe that the absence of 7P results for days 4 and 7 is not critical for the comparison with G4 and G4AP for two main reasons: (i) it has been widely demonstrated that G4P sensor performs better than the 7P in all days of monitoring [5] and (ii) since the 7P sensor has been replaced by the G4 generation sensors in all open-loop and close-loop trials involving Dexcom sensors, a more exhaustive analysis of the 7P sensor error would be, at the present time, of minor interest.

## 5 Conclusions

A detailed dissection and quantitative description of CGM sensor error are important when designing new CGM-based applications, e.g., glucose predictors and artificial pancreas control algorithms, and in in silico testing of diabetes technologies and treatments. In this work, we have analyzed and modeled the error components of two recent CGM sensors produced by Dexcom Inc., the G4P and the G4AP, and compared them with what can be obtained for a sensor of previous generation, the 7P. The analysis gave new insights into CGM sensor error description by evidencing its time-variance during the monitoring, and could be considered a basis on which to create a new CGM sensor error model to be included, e.g., in the UVA/Padova T1D simulator [10], in order to simulate longer reliable scenarios. To complete such a model, further investigation is however needed. In fact, at the present time, only models for days 1, 4, and 7 are available, while for the other days are missing. Further work may concern the analysis of correlation of sensor

error before and after a calibration when suitable datasets become available and the application of more sophisticated iterative techniques for the identification of model parameters will allow performing the estimation all model parameters in a single step [13].

## References

- Bailey T, Zisser H, Chang A (2009) New features and performance of a next-generation SEVEN-day continuous glucose monitoring system with short lag time. *Diabetes Technol Ther* 11(12):749–755
- Basu A, Dube S, Slama M, Errazuriz I, Amezcua JC, Kudva YC, Peyser T, Carter RE, Cobelli C, Basu R (2013) Time lag of glucose from intravascular to interstitial compartment in humans. *Diabetes* 62(12):4083–4087
- Bequette BW (2010) Continuous glucose monitoring: real-time algorithms for calibration, filtering, and alarms. *J Diabetes Sci Technol* 4(2):404–418
- Breton M, Kovatchev B (2008) Analysis, modeling, and simulation of the accuracy of continuous glucose sensors. *J Diabetes Sci Technol* 2(5):853–862
- Christiansen M, Bailey T, Watkins E, Liljenquist D, Price D, Nakamura K, Boock R, Peyser T (2013) A new-generation continuous glucose monitoring system: improved accuracy and reliability compared with a previous-generation system. *Diabetes Technol Ther* 15(10):881–888
- Clarke W, Kovatchev B (2009) Statistical tools to analyze continuous glucose monitor data. *Diabetes Technol Ther* 11(Suppl 1):45–54
- Cobelli C, Renard E, Kovatchev B (2011) Artificial pancreas: past, present, future. *Diabetes* 60(11):2672–2682
- Cobelli C, Renard E, Kovatchev BP, Keith-Hynes P, Ben Brahim N, Place J, Del Favero S, Breton M, Farret A, Bruttomesso D, Dassau E, Zisser H, Doyle FJ, Patek SD, Avogaro A (2012) Pilot studies of wearable outpatient artificial pancreas in type 1 diabetes. *Diabetes Care* 35(9):e65–e67
- Cox M (2009) An overview of continuous glucose monitoring systems. *J Pediatr Health Care* 23(5):344–347
- Dalla Man C, Micheletto F, Lv D, Breton M, Kovatchev B, Cobelli C (2014) The UVA/Padova type 1 diabetes simulator: new features. *J Diabetes Sci Technol* 8(1):26–34
- Damiano ER, El-Khatib FH, Zheng H, Nathan DM, Russell SJ (2013) A comparative effectiveness analysis of three continuous glucose monitors. *Diabetes Care* 36(2):251–259
- Daskalaki E, Norgaard K, Zuger T, Proutzou A, Diem P, Mougiakakou S (2013) An early warning system for hypoglycemic/hyperglycemic events based on fusion of adaptive prediction models. *J Diabetes Sci Technol* 7(3):689–698
- Docherty PD, Chase JG, David T (2012) Characterisation of the iterative integral parameter identification method. *Med Biol Eng Comput* 50(2):127–134
- Dua P, Doyle FJ, Pistikopoulos EN (2009) Multi-objective blood glucose control for type 1 diabetes. *Med Biol Eng Comput* 47(3):343–352
- Facchinetti A, Del Favero S, Sparacino G, Castle J, Ward W, Cobelli C (2014) Modeling the glucose sensor error. *IEEE Trans Biomed Eng* 61(3):620–629
- Garcia A, Rack-Gomer AL, Bhavaraju NC, Hampapuram H, Kamath A, Peyser T, Facchinetti A, Zecchin C, Sparacino G, Cobelli C (2013) Dexcom G4AP: an advanced continuous glucose monitor for the artificial pancreas. *J Diabetes Sci Technol* 7(6):1436–1445

17. Guerra S, Sparacino G, Facchinetti A, Schiavon M, Man CD, Cobelli C (2011) A dynamic risk measure from continuous glucose monitoring data. *Diabetes Technol Ther* 13(8):843–852
18. Hovorka R, Allen JM, Elleri D, Chassin LJ, Harris J, Xing D, Kollman C, Hovorka T, Larsen AM, Nodale M, De Palma A, Wilinska ME, Acerini CL, Dunger DB (2010) Manual closed-loop insulin delivery in children and adolescents with type 1 diabetes: a phase 2 randomised crossover trial. *Lancet* 375(9716):743–751
19. Joubert M, Reznik Y (2012) Personal continuous glucose monitoring (CGM) in diabetes management: review of the literature and implementation for practical use. *Diabetes Res Clin Pract* 96(3):294–305
20. Kamath A, Mahalingam A, Brauker J (2009) Analysis of time lags and other sources of error of the DexCom SEVEN continuous glucose monitor. *Diabetes Technol Ther* 11(11):689–695
21. Kovatchev B, Anderson S, Heinemann L, Clarke W (2008) Comparison of the numerical and clinical accuracy of four continuous glucose monitors. *Diabetes Care* 31(6):1160–1164
22. Kovatchev BP, Cox DJ, Gonder-Frederick LA, Clarke W (1997) Symmetrization of the blood glucose measurement scale and its applications. *Diabetes Care* 20(11):1655–1658
23. Laguna AJ, Rossetti P, Ampudia-Blasco FJ, Vehí J, Bondia J (2014) Postprandial performance of Dexcom® SEVEN® PLUS and Medtronic® Paradigm® Veo™: modeling and statistical analysis. *Biomed Signal Process Control* 10:322–331
24. Lane JE, Shivers JP, Zisser H (2013) Continuous glucose monitors: current status and future developments. *Curr Opin Endocrinol Diabetes Obes* 20(2):106–111
25. Lunn DJ, Wei C, Hovorka R (2011) Fitting dynamic models with forcing functions: application to continuous glucose monitoring in insulin therapy. *Stat Med* 30(18):2234–2250
26. McGarraugh G (2009) The chemistry of commercial continuous glucose monitors. *Diabetes Technol Ther* 11(Suppl 1):17–24
27. Ogunfunmi T (2007) Adaptive nonlinear system identification : the Volterra and Wiener model approaches. Springer, New York
28. Phillip M, Battelino T, Atlas E, Kordonouri O, Bratina N, Miller S, Biester T, Stefanija MA, Muller I, Nimri R, Danne T (2013) Nocturnal glucose control with an artificial pancreas at a diabetes camp. *N Engl J Med* 368(9):824–833
29. Rebrin K, Steil GM, van Antwerp WP, Mastrototaro JJ (1999) Subcutaneous glucose predicts plasma glucose independent of insulin: implications for continuous monitoring. *Am J Physiol* 277(3 Pt 1):E561–E571
30. Rodbard D (2011) Glycemic variability: measurement and utility in clinical medicine and research—one viewpoint. *Diabetes Technol Ther* 13(11):1077–1080
31. Russell SJ, El-Khatib FH, Nathan DM, Magyar KL, Jiang J, Damiano ER (2012) Blood glucose control in type 1 diabetes with a bihormonal bionic endocrine pancreas. *Diabetes Care* 35(11):2148–2155
32. Service FJ (2013) Glucose variability. *Diabetes* 62(5):1398–1404
33. Sparacino G, Facchinetti A, Cobelli C (2010) “Smart” continuous glucose monitoring sensors: on-line signal processing issues. *Sensors (Basel)* 10(7):6751–6772
34. Tamborlane WV, Beck RW, Bode BW, Buckingham B, Chase HP, Clemons R, Fiallo-Scharer R, Fox LA, Gilliam LK, Hirsch IB, Huang ES, Kollman C, Kowalski AJ, Laffel L, Lawrence JM, Lee J, Mauras N, O’Grady M, Ruedy KJ, Tansey M, Tsalikian E, Weinzimer S, Wilson DM, Wolpert H, Wysocki T, Xing D, Chase HP, Fiallo-Scharer R, Messer L, Gage V, Burdick P, Laffel L, Milaszewski K, Pratt K, Bismuth E, Keady J, Lawlor M, Buckingham B, Wilson DM, Block J, Benassi K, Tsalikian E, Tansey M, Kucera D, Coffey J, Cabbage J, Wolpert H, Shetty G, Atakov-Castillo A, Giusti J, O’Donnell S, Ghiloni S, Hirsch IB, Gilliam LK, Fitzpatrick K, Khakpour D, Wysocki T, Fox LA, Mauras N, Englert K, Permuy J, Bode BW, O’Neil K, Tolbert L, Lawrence JM, Clemons R, Maeva M, Sattler B, Weinzimer S, Tamborlane WV, Ives B, Bosson-Heenan J, Beck RW, Ruedy KJ, Kollman C, Xing D, Jackson J, Steffes M, Buckska JM, Nowicki ML, Van Hale C, Makky V, O’Grady M, Huang E, Basu A, Meltzer DO, Zhao L, Lee J, Kowalski AJ, Laffel L, Tamborlane WV, Beck RW, Kowalski AJ, Ruedy KJ, Weinstock RS, Anderson BJ, Kruger D, LaVange L, Rodriguez H (2008) Continuous glucose monitoring and intensive treatment of type 1 diabetes. *N Engl J Med* 359:1464–1476
35. Toffanin C, Messori M, Di Palma F, De Nicolao G, Cobelli C, Magni L (2013) Artificial pancreas: model predictive control design from clinical experience. *J Diabetes Sci Technol* 7(6):1470–1483
36. Zecchin C, Facchinetti A, Sparacino G, Cobelli C (2013) Reduction of number and duration of hypoglycemic events by glucose prediction methods: a proof-of-concept in silico study. *Diabetes Technol Ther* 15(1):66–77