ARX model for interstitial glucose prediction during and after physical activities

Hector M. Romero-Ugalde\textsuperscript{a,b,f,*}, M. Garnotel\textsuperscript{c}, M. Doron\textsuperscript{a,b}, P. Jallon\textsuperscript{a,b}, G. Charpentier\textsuperscript{d,e}, S. Franc\textsuperscript{d,e}, E. Huneker\textsuperscript{f}, C. Simon\textsuperscript{c}, S. Bonnet\textsuperscript{a,b}

\textsuperscript{a} Univ. Grenoble Alpes, F-38000 Grenoble, France
\textsuperscript{b} CEA, LETI, MINATEC Campus, F-38054 Grenoble, France
\textsuperscript{c} CARMEN INSERM U1060/Université de Lyon 1/INRA U1235, CRNH-Rhône-Alpes, Lyon, France
\textsuperscript{d} Centre Hospitalier Sud-Francilien, Department of Diabetes and Endocrinology, Corbeil-Essonnes, France
\textsuperscript{e} Centre d’Etudes et de Recherche pour l’Intensification du Traitement du Diabète (CERITD), Corbeil-Essonnes, France
\textsuperscript{f} Diabeloop SA, 155 Cours Berriat, F-38000 Grenoble, France

1. Introduction

Type 1 Diabetes (T1D) is a disease, where the insulin-producing cells (beta cells) are destroyed by the autoimmune system, causing a failure on blood glucose (BG) control. BG > 300 mg/dl may lead to fatigue, nausea, abdominal pain, excessive thirst, frequent urination and blurred vision. Chronic hyperglycemia (BG > 180 mg/dl) may lead to long-term complications affecting eyes, kidneys, nerves and particularly the cardiovascular system. BG < 70 mg/dl (hypoglycemia) may lead to seizures, coma, and death. In fact, glucose is absorbed into the bloodstream after digestion of carbohydrates (CHO) in a meal, i.e., meals provoke an increase in BG. Glucose is also produced by the liver. Insulin is a hormone that allows glucose in the bloodstream to enter into cells, providing them with the energy they need to function, i.e., insulin provokes a decrease in BG. In this sense, the challenge for patients with T1D is to correctly dose their insulin administration in order to maintain their BG level into a target range, typically BG ∈ [70, 180] mg/dl.

Interstitial glucose (IG) prediction plays an important role for automatically maintaining BG level of T1D patients into the targeted range. For instance, suspending insulin delivery when predicted IG is lower than a given threshold, allows reduction in hypoglycemic events (Garg et al., 2012; Stenerson et al., 2014). As an other example, Model Predictive Control (MPC) algorithms use predicted IG to optimize insulin delivery (Del Favero et al., 2014; Hovorka et al., 2004). In fact, a recently CE marked artificial pancreas (CE marking is a symbol of free marketability in the European Economic Area), uses IG predictions for regulating BG. For more details on this promising technology see Benhamou et al. (2018) and Benhamou et al. (2019). The reader shall notice, that in this paper we make the difference between IG and BG. IG, which is provided by a continuous glucose monitoring (CGM) system and is highly correlated to BG (Kulcu, Tamada, Reach, Potts, & Lesho, 2003), is the measure used by the artificial pancreas to regulate BG.

There exists a wide variety of models for predicting IG from a variety of input variables. In Daskalaki, Prountzou, Diem, and Mougiaskakou (2012), three different model types: autoregressive (AR) models, AR models with exogenous input (ARX) and models based on an artificial neural network (ANN), were proposed. The AR-based models proposed in Daskalaki et al. (2012) only use IG information to perform IG prediction, whereas ARX and ANN-based models proposed in Daskalaki et al. (2012) use IG and insulin information. In Cescon and Johansson (2009), the predictive models receive as input variables IG, meal and insulin data. The models consist of a state-space model, an ARX model and an ARMAX model (autoregressive moving average with exogenous input). In Cescon, Johansson, and Renard (2013a, 2013b), two continuous-time second-order transfer functions are used with one using IG and...
injected insulin as inputs and the other using IG and amount of CHO of a meal. In Stähl and Johansson (2010), a hybrid model combining physiological (insulin and meal sub-models) and black box models (glucose–insulin interaction model and interstitial-continuous glucose monitoring model), was proposed. In Boiroux et al. (2018) and Eren-Oruklu, Cinar, Quinn, and Smith (2009) autoregressive integrated moving-average (ARIMAX) models are used into an MPC algorithm. The two ARIMAX models use IG and injected insulin as inputs to predict IG. Prediction horizon in Eren-Oruklu et al. (2009) was 100 min. In Boiroux, Dun-Henriksen et al. (2018) prediction horizon was set to 10 h. In Gondhalekar, Dassau, and Doyle (2016), a state-space model, receiving insulin and IG as inputs, is used to predict IG in a horizon of 45 min. Prediction is used by an MPC algorithm to optimize insulin delivery.

Previous works are very interesting, but in their IG prediction models, no physical activity (PA) information (level, type, or sensor data) was considered as input variable. However, it is well-known that PA has a considerable effect on BG (Riddell, Zahariab, Tavelberg, Cinar, & Jamnik, 2015). Authors in Reddy et al. (2018) demonstrated that the effect of PA on BG depends on the type (aerobic or resistance) and the intensity of the exercise. While aerobic physical activities induce a decrease on BG, resistance exercises induce an increase. Authors in Boiroux, Jorgensen, Patek and Breton et al. (2018) use a PA tracking watch to identify the “net” effect of idle, mild, moderate and intensive PA on BG. The aim of these studies was not to physiologically model the effect of PA into BG due to the complexity of this task. In fact, PA provokes an increase of blood flow in heart, lungs and peripheral tissue and a decrease of flow of kidneys and splanchnic organs (Hernández-Ordoñez & Campos-Delgado, 2008; Patton, Fuchs, Hille, Scher, & Steiner, 1989). Peripheral glucose, insulin uptake, and liver’s glucose production are increased during PA (Hernández-Ordoñez & Campos-Delgado, 2008). Glycogen depletion and replenishment are also affected by the intensity of PA (Hernández-Ordoñez & Campos-Delgado, 2008). Moreover, it is well-known that insulin sensitivity is also affected during and after PA (Deroiuch & Boutayeb, 2002; Schiavon et al., 2013). Two physiological models considering most of these effects were proposed in Hernández-Ordoñez and Campos-Delgado (2008) and Man, Breton, and Cobelli (2009). In Hernández-Ordoñez and Campos-Delgado (2008) PA level is measured as a volume percentage of the maximum oxygen consumption (VO2MAX). Depending on the VO2MAX, redistribution of blood flow, peripheral glucose uptake, hepatic glucose production, and peripheral insulin uptake are modulated in their model. In Man et al. (2009), PA level is measured indirectly using the heart rate (HR). Depending on the HR, insulin-independent glucose clearance, insulin sensitivity (up to 22 h), and glucose uptake are modulated in their model. Although these physiological models are very interesting, validation on real patients was never performed. The reader shall notice that, it is very difficult to quantify the effect of PA on the physiological variables affecting the BG behavior. In fact, this effect depends on a large variety of factors such as body weight, age, sex, physiological condition, patient training level, PA type and intensity (Deroiuch & Boutayeb, 2002; Riddell et al., 2015).

System identification is an alternative solution already used for considering the effect of PA in IG prediction. For instance, in Cescon and Renard (2011), a subspace-based patient-specific model is proposed for IG prediction on T1D patients during 30 min of exercise. The model receives CHO, insulin, HR, and respiration rate as inputs. In their model, PA is estimated by using HR, however, it is well known that HR is also modulated by stress (Hildebrandt, Mehlsen, Sestoft, & Nielsen, 1985). This fact may affect IG prediction accuracy in some situations. In Dasanayake, Seborg, Pinsker, Doyle, and Dassau (2015), Dasanayake et al. proposed, a state-space model, which only receives IG and accelerometer signals as inputs. However, their model is only accurate, on IG prediction during PA, when heart rate is higher than 30% of the heart rate reserve (HR’). In Balakrishnan, Samavedham, and Rangaiah (2013), an hybrid model uses as inputs the meal and insulin information, and rate of perceived exertion (to consider PA). Since PA is considered through the patient perception, model performance may be affected. In Turksoy, Bayrak, Quinn, Littlejohn, and Cinar (2013), a model using insulin on board, energy expenditure (computed from accelerometer and HR signals) and galvanic skin response as inputs was proposed. The model, consisting of an ARMAX model, does not receive meal information as input. Therefore, IG prediction accuracy after meals may decrease.

In this paper, we propose an ARX model that uses energy expenditure (EE), insulin on board (IOB), and carbohydrates on board (COB), as inputs for predicting IG. EE, computed from both accelerometers and HR signals (Romero-Ugalde et al., 2017), is used to better consider the effect of PA on IG prediction, as demonstrated in Turksoy et al. (2013) and Turksoy et al. (2018). IOB is computed from the output of an insulin pump. COB is computed from the CHO declared by the patients. Differently to the ARMAX model used in Turksoy et al. (2013), the ARX model proposed in our paper includes the COB as input. This fact, allows to consider the effect of CHO, usually ingested before and during PA to prevent hypoglycemia, in order to improve prediction. Another difference between the ARMAX model presented in Turksoy et al. (2013) and our ARX model is the intended use. While the ARMAX model proposed in Turksoy et al. (2013) was designed to be used in an artificial pancreas that does not require meal announcement, the ARX model proposed in our paper is designed to be used in a hybrid closed-loop artificial pancreas. We could discuss the advantages and disadvantages of both approaches (meal announcement vs unannounced meal), but this is not the aim of this paper. The aim of this paper is to improve IG prediction during and after physical activities.

In this sense, originality of the proposed ARX model is the fact of using EE, IOB, and COB as inputs variables. We consider that the use of these three variables, usually modulated during (EE, IOB, COB) and after (IOB, COB) a PA, may improve IG predictions.

The rest of the paper is organized as follows: Section 2 presents a detailed description of the experimental protocols to acquire the two databases used in this paper. Section 3 describes the proposed ARX model, and the validation tests. Results, presented in Section 4, are discussed in Section 5. Finally, Section 6 presents the conclusions of the study.

2. Database description

Two different databases were used in this paper to estimate and validate the proposed ARX model. The first database was acquired from a clinical protocol where patients performed a single PA, namely “SPA protocol”. The second database was acquired from a clinical protocol where patients performed four PAs, namely, “FPA protocol”. These protocols were approved by the “French Ethics Committee” and the “French National Agency for Medicines and Health Products Safety (ANSM)”.

2.1. SPA database description

TID patients (N = 35, age > 18 years old, HbA1c < 10%) already treated by insulin pump, were included in the clinical protocol, which was performed on 7 centers in France, in 2012.

After two visits, inclusion analysis and installation/calibration of two continuous glucose monitoring systems (Dexcom® SEVEN® PLUS), patients were hospitalized during 25 h. Fig. 1 illustrates the SPA experimental procedure.

Patients arrived at 18:30 in the afternoon. An intravenous catheter, an insulin pump (JewelPUMP®), an accelerometer (hip-worn GT3X+, ActiGraph), and a PA monitoring system (Actiheart, CamNtech) were placed. Meals were taken at fixed hours (20:00, 8:00 and 12:00). Patients performed a PA at 15:00 during 30 min. The required PA, which consists of a step test, was performed at moderate intensity according
to each patient. From the 35 initial patients, fourteen patients wearing the accelerometer and the HR monitoring system were included on this study. According to the proposed PA protocol, patients reduced basal insulin rate during the half hour of PA + 2 h. Moreover, when patient risked hypoglycemia (based on current and previous CGM measures), snacks were ingested and declared by patient.

Fig. 2 shows, for one patient of the SPA database, data set acquired during the second day of visit 3 of this experimentation. We can observe that (1) HR and counts per minutes (CPM), increase during PA (PA started at 15:00), (2) for this patient, HR and CPM also increase between 10:00 and 11:00, which indicates that this patient performed an undeclared PA, (3) meal have an important and delayed effect on IG (IG increases around 40 min after meal), (4) as already mentioned, insulin basal rate is reduced when the patient started the PA in order to prevent hypoglycemia. More precisely, the aim of Fig. 2 is to illustrate data set used in this work to estimate (order selection and parameter estimation) the ARX model, i.e., data set \( [t_{PA} - 360, t_{PA} + 120] \) min, where \( t_{PA} \) is the time at which PA was started. This time interval was used on all the patients of the SPA database.

2.2. FPA database description

T1D patients (\( N = 36, \) age \( > 18 \) years old, 7.5% < \( \mathrm{HbA1c} < 9.5\% \)), already treated by insulin pump, and able to practice at least one PA during 3 days, were included in the clinical protocol, which was performed in 9 centers in France, in 2016.

This study was performed in 3 visits (see Fig. 3). During the first visit, inclusion was performed, CGM system (Dexcom™ Share AP, Dexcom Inc., San Diego, CA) was installed and calibrated, patients were instructed on the CGM system utilization, patients were randomized on 2 groups, and dates for visits 2 and 3 were established.

Visit 2 was done two days before the main visit (V3). An accelerometer (hip-worn GT3X+, ActiGraph), and a PA monitoring system (Actiheart, CamNtech) were placed.

Concerning the visit 3 (see Fig. 3), patients arrived to the research center at 8:00 (after taking breakfast at home), and spent 72 h in the research center. If patients were in group A, they used a closed-loop heart, CamNtech) were placed.

During visit 3, patients performed daily physical activities, but also some imposed physical activities. Meals, of various CHO quantities, were taken at the same hours during the three days. Imposed physical activities were performed at fixed hours. Intensity and duration of physical activities were not the same during the three days of the visit. When PA started 3 h after meal, each patient reduced the insulin basal rate at 50 or 80% of the current basal rate, depending on the PA intensity, 30 to 60 min before starting the PA. When PA started within the 3 h after meal, bolus correcting meal was reduced. When required (hypoglycemia risk), snacks were ingested and declared by the patient. From the 36 initial patients, fifteen patients wearing the accelerometer and the heart rate monitoring system were included on this study.

Fig. 4 displays, for one patient of the FPA database, data set acquired during the 3 days of visit 3. We can observe that patients performed a PA during day 1 (PA1), a PA during day 2 (PA2), and two PAs during day 3 (PA3 and PA4). We can also observe that when patient risked hypoglycemia, snacks were ingested (see small increases in CHO, in the third CHO panel).

In both protocols (SPA and FPA), CGM calibration was performed (1) at the installation phase by two BG measurements, (2) when instructed by the 12-hour CGM calibration prompt, and (3) when the CGM reading was inaccurate. In SPA protocol and the main visit of FPA protocol, BG was measured, by a glucose meter, at least every hour, but also every 15 min during meals, PA, hypoglycemia, and hyperglycemia.

In both protocols (SPA and FPA), IG was acquired every 5 min, declared PA, declared CHO, insulin basal rate, bolus, and HR (computed from electrocardiogram) were sampled every min. Accelerometer signals were converted in counts per minute. Finally all the signals were preprocessed and re-sampled to a sampling period of 10 min, which is the sampling period used in the proposed ARX model.

The 14 patients of the SPA protocol and the 15 patients of the FPA protocol, used in this study (good quality of CGM, CPM, and HR signals), were different. In fact, SPA and FPA protocols were performed on different years (2012 and 2016).

In both studies (SPA and FPA), CHO and PA type, were declared by the patients.

3. The proposed ARX model

This section presents the ARX model proposed in this paper. Different to the black box models found in the literature (for instances Cescon & Renard, 2011; Dasanayake et al., 2015; Turksoy et al., 2013, 2018), the proposed ARX model uses EE, IOB, and COB as inputs to improve IG prediction. These inputs allow to consider (1) the intensity and duration of a PA, (2) the delivered insulin which is modulated before and during PA to reduce the risk of hypoglycemia, and (3) the CHO, usually ingested before and during PA to prevent hypoglycemia. Notice that these are important factors affecting BG dynamic.

3.1. ARX structure

In system identification, ARX models are among the most used black box structures due to their simplicity (Farzin Piltan, Sulaiman, & Wouters, 2017; Soltanieh & Ogun, 2018). The ARX model is given by (1).

\[
y[k] = a^T y_{k-1} + b^T u_{k-nu} + e[k]
\]

where \( k \) is the current sample. Both \( a = [a_1, \ldots, a_{na}] \) and \( b = [b_1, \ldots, b_{nb}] \) are the model parameters. The regressors \( y_{k-1} = [y[k-1], \ldots, y[k-na]] \) are the previous outputs on which the current output \( y[k] \) depends. The regressors \( u_{k-nu} = [u[k-nk], \ldots, u[k-nk-nb+1]] \) are the delayed inputs on which the current output depends. The parameters \( na \) and \( nb \) are the orders of the ARX model, \( nk \) is the time delay (expressed in samples) before the input affects the output, also called the dead time of the system. Finally, \( e[k] \) is a noisy value.

3.2. Proposed ARX model structure

ARX models may be used for representing multiple-input and single-output (MISO) systems. In this paper, the proposed ARX model, given by (2), receives as inputs the IOB computed according to (3), the COB computed according to (4), and the EE computed according to (6). The black box model in (2) will determine the temporal relations that
may exist between the inputs IOB, COB, EE and the output IG. Model parameters are estimated from MISO data in order to dynamically represent the effect of inputs into IG.

\[
y^{IG}[k] = a^T_k y_{k-1} + b^T_k u^{EE}_{k-nk_1} + c^T_k u^{IOB}_{k-nk_2} + d^T_k u^{COB}_{k-nk_3} 
\]

(2)

The ARX model, proposed in this paper, used a sampling period of 10 min, according to previous works found in the literature (Cescon & Johansson, 2009; Turksoy et al., 2013). However, predictions may be performed every 5 min (sampling period of IG), i.e., every 5 min the ARX model resamples the past signals to 10 min in order to predict IG \([10, 20, 30, \ldots, N]\) min ahead. In this paper we present examples of 30, 60 and 120 min ahead IG predictions.

### 3.3. Model inputs

#### 3.3.1. IOB input

The IOB refers to the injected insulin (bolus and basal), that is still to have an effect on the BG. The IOB is computed as a convolution:

\[
u^{IOB}[n] = \sum_{k=0}^{K} I[n-k] h_{IOB}[k]
\]

(3)

where \(I[k]\) is the quantity of insulin in mU delivered by the insulin pump at the \(k\)th time index.

#### 3.3.2. COB input

In the same spirit, the COB refers to the portion of the meal that is still to have an effect on the IG. The COB is computed as a similar
Convolution:

$$u_{COB}^{\text{conv}}[n] = \sum_{k=0}^{K} \text{CHO}(n-k)h_{COB}[k]$$  \hspace{1cm} (4)

where $\text{CHO}(k)$ is the quantity of CHO in g ingested at the $k$th time index. As already mentioned, CHO were declared by the patients.

In (3) and (4), $K = 144$ to consider 24 h of data, and $h[k] = h[k\tau_s]$ with $\tau_s = 10$ min, is given by (5):

$$h[k] = \left(1 + \frac{k}{\tau_s}\right) e^{-\frac{k}{\tau_s}}.$$  \hspace{1cm} (5)

$\tau_s = 50$ min in (3) was set from a population-based study (SPA protocol), where insulinemia was measured every 10 min during a given period. $\tau_s = 40$ min in (4) was empirically set. We considered the fact that CHO have usually a faster effect on IG than insulin.

3.3.3. EE input

The EE is computed from accelerometer and HR signals, according to (6).

$$u^{EE} = \begin{cases} 
\alpha_1 \text{HR}^t + \beta_1, & \text{if HRT} \geq S_{\text{HR}}^t \\
\alpha_2 \text{LC} + \beta_2, & \text{if HRT} < S_{\text{HR}}^t \text{ and LC} < S_{\text{LC}}\\
\alpha_3 \text{LC} + \beta_3, & \text{if HRT} < S_{\text{HR}}^t \text{ and LC} \geq S_{\text{LC}}
\end{cases}$$  \hspace{1cm} (6)

where $\alpha_1 = 5.45$, $\beta_1 = -66.09$, $\alpha_2 = 256.09$, $\beta_2 = -0.13$, $\alpha_3 = 85.99$, $\beta_3 = 82.39$ are the model parameters. $S_{\text{HR}}^t = 40$ bpm and $S_{\text{LC}} = 0.5$ are the cut points obtained from a population-based approach. LC is a linear combination of the normalized values of HRT and CPM, computed as:

$$\text{LC} = \theta_1 \text{CPM} + \theta_2 \text{HR}^t.$$  

$$\text{HR}^t = \text{HR} - \text{resting HR}, \text{CPM are the counts per minute (a quantity derived from the accelerometer signal Sandroff, Riskin, Agiovlasitis, & Motl, 2014).}$$
Notice that, in this model, PA information declared by the patients is not used to compute EE. For more details on model (6) see Romero-Ugalde et al. (2017).

3.4. Model orders and delays

The orders ($n_a$, $n_b$, $n_c$, and $n_d$) and delays ($n_k_1$, $n_k_2$, and $n_k_3$) of the ARX model proposed in this paper, were obtained by a standard system identification methodology, described in the following:

1. $n_a$, $n_b$, $n_c$, and $n_d$ are changed among a given range of values,
2. each time that $n_a$, $n_b$, $n_c$, $n_d$, $n_k_1$, $n_k_2$, or $n_k_3$ changes, model parameters are estimated, by the classical least squares algorithm (7), and an associated Akaike final prediction error ($FPE$) is computed by (8).

$$\hat{\theta} = (X^TX)^{-1}X^Ty,$$  \hfill (7)

where $X$ is the regression matrix, $y$ is an $N \times 1$ vector of outputs (IG in this case), and $\hat{\theta}$ represents the estimated parameters.

$$FPE = \frac{1 + d/N}{1 - d/N} \left( \frac{1}{N} \sum \epsilon(t, \hat{\theta})^2 \right),$$  \hfill (8)

where $N$ is the number of values in the estimation data set, $\epsilon(t, \hat{\theta})$ is a vector of prediction errors, and $d$ is the number of estimated parameters ($\hat{\theta}$).

3. the model structure (orders and delays) leading the lowest final prediction error is chosen as the best candidate. Notice that by using this selection criteria (8) the overparametrization is penalized.

Database used for choosing the model structure (orders and delays) was the SPA database. Datasets into the span $[t_{PA} - 360 \text{ min}, t_{PA} + 120 \text{ min}]$ sampled at $r = 10 \text{ min}$, were used for training.

Accordingly with the system identification procedure described above: (1) model orders and delays were changed in the ranges: $n_a \in [1, 10]$, $n_b \in [7, 8]$, $n_c \in [10, 15]$, $n_d \in [10, 15]$, $n_k_1 \in [1, 2]$, $n_k_2 \in [1, 2]$, $n_k_3 \in [1, 3]$; (2) each time that a given set of orders and delays was set: (a) a regression matrix was constructed for each patient, (b) the regression matrix of all patients was concatenated, (c) model parameters were estimated, by the least squares algorithm, and (d) a final prediction error associated with such model is computed; (3) finally, the model structure (orders and delays $n_a = 3$, $n_b = 5$, $n_c = 11$, $n_d = 7$, and $n_k_1 = n_k_2 = n_k_3 = 2$), yielding the lowest final prediction error was chosen as the best candidate. It is interesting to notice that the selected model structure allows to consider the effect of insulin (Hinshaw et al., 2013) and CHO (American Diabetes Association, 2001) have the most important effect on BG.

3.5. ARX validation tests

The goals of the validation tests presented in this paper are (1) to evaluate the possibility of proposing a population-based ARX model, (2) to verify the hypothesis on ARX models stating that performance of personalized ARX models may increase as the number of quality-training-data increases, and (3) to test the improvement achieved by using the three regressors.

3.5.1. Test 1: The population-based ARX model obtained from the SPA database is evaluated on the FPA database

As a first approach, we evaluated the possibility of using an ARX model for predicting IG in any T1D adult patient. In this sense, ARX parameters (a, b, c, and d in (2)) are obtained from data sets on which model orders and delays were chosen. Similar to the model structure selection, the concatenated regression matrix was used to obtain the SPA population-based ARX parameters. Then, the population-based (PB) model was used for predicting IG on the 15 patients of the FPA database, during 30, 60 and 120 min, after the physical activity (PA3) was started (see Fig. 4).

3.5.2. Test 2: Increasing ARX performance by increasing training data

We hypothesize, according to literature (Herpe et al., 2006), that black box models performance may increase, if the number of quality-available-training data increases. In this sense, patient-specific ARX models were obtained on the FPA database from:

(a) a single data set (T1), i.e., data around PA1, PA2, or PA4 is used separately for training;
(b) on two data sets (T2), i.e., data around PA1 and PA2, PA1 and PA4, or PA2 and PA4 is used for training;
(c) on three data sets (T3), i.e., data around PA1, PA2, and PA4 is used for training.

We refer to data around a given PA, as a set of 8 h of data into the span $[t_{PA} - 60, t_{PA} + 360 \text{ min}]$, where $t_{PA}$ is the time at which one of the physical activities PA1, PA2 or PA4 was started. In this sense models T1 are trained on 8 h of data, T2 models are trained on 16 h of data, and T3 models are trained on 24 h of data.

Finally, T1, T2, and T3 ARX models were used for predicting IG, into the span $[t_{PA3} - t_{PA3} + 30 \text{ min}]$, $[t_{PA3} - t_{PA3} + 60 \text{ min}]$, and $[t_{PA3} - t_{PA3} + 120 \text{ min}]$.

3.5.3. Test 3: Improvement achieved by the use of the three variables (insulin, meal, and EE)

On the FPA database, we obtained (1) T3 models using COB, IOB, and EE as inputs, (2) T3 models using only COB and IOB as inputs (NEE), (3) T3 models using only EE and IOB as inputs (NCOB), and (4) T3 models using only EE and COB as inputs (NIOB). Models T3, NEE, NCOB, and NIOB, are compared in order to show the improvement reached by the use of three simultaneous variables as inputs of the ARX models.

3.6. Performance indicator

Performance indicator used for measuring model accuracy is the root-mean-square error (RMSE), given by:

$$\text{RMSE} = \left\{ \frac{1}{N} \sum_{n=1}^{N} \left[ y^{IG}[n] - \hat{y}^{IG}[n] \right]^2 \right\}^{1/2},$$  \hfill (9)

which is a standard indicator used in IG prediction (Cescon & Johansson, 2009; Cescon et al., 2013a, 2013b; Daskalaki et al., 2012; Stähl & Johansson, 2010). In (9), $y^{IG}[n]$ and $\hat{y}^{IG}[n]$, are the measured and predicted IG at instant $n$, respectively. Since in tests described above the ARX models are used for predicting IG during 30, 60, and 120 min, and the sampling period was 10 min, then $N = 3$, 6, and 12, respectively.

$P$-value, computed with Wilcoxon Matched-Pairs signed-rank test (MacFarland & Yates, 2016), which is a non-parametric statistical hypothesis test, was used for validation on test 3, described above.

4. Results

Parameters of the proposed PB ARX model were $a_1 = 1.67$, $a_2 = -0.74$, $a_3 = 0.06$, $b_1 = -1.03$, $b_2 = 1.87$, $b_3 = -0.71$, $b_4 = 1.35$, $b_5 = -0.12$, $c_1 = 3.39e^{-4}$, $c_2 = -1.09e^{-4}$, $c_3 = -0.34e^{-4}$, $c_4 = -3.02e^{-4}$, $c_5 = 3.82e^{-2}$, $c_6 = -5.73e^{-2}$, $c_7 = -1.19e^{-2}$, $c_8 = 2.63e^{-4}$, $c_9 = -1.26e^{-4}$, $c_{10} = 3.37e^{-4}$, $c_{11} = -1.74e^{-4}$, $d_1 = 0.57e^{-3}$, $d_2 = -0.73e^{-2}$, $d_3 = 0.45e^{-2}$, $d_4 = -0.08e^{-2}$, $d_5 = -0.39e^{-2}$, $d_6 = 0.98e^{-2}$, and $d_7 = -0.54e^{-2}$.

Fig. 5 displays the parameters of the 15 patient-specific ARX models (boxplots) obtained for T3 models. On the same figure, parameters of the population-based ARX model (PB-model) are also represented (single blue points).
Fig. 5. Parameters of the personalized ARX models derived from T3 models. Since parameters were personalized for 15 patients, each box is composed of 15 points, i.e., each parameter can take 15 values. Parameters of the PB model are also displayed (blue points), representing the value of each parameter.

Table 1

RMSE (mean ± standard deviation) computed for IG prediction during 30, 60 and 120 min, performed by models T3, T2, T1, and PB.

<table>
<thead>
<tr>
<th>Model</th>
<th>30 min (mg/dL)</th>
<th>60 min (mg/dL)</th>
<th>120 min (mg/dL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T3</td>
<td>7.75 ± 4.51</td>
<td>15.86 ± 9.61</td>
<td>35.24 ± 19.52</td>
</tr>
<tr>
<td>T2</td>
<td>8.94 ± 6.17</td>
<td>19.11 ± 12.24</td>
<td>41.22 ± 28.01</td>
</tr>
<tr>
<td>T1</td>
<td>21.82 ± 17.02</td>
<td>81.78 ± 120.61</td>
<td>327.04 ± 473.29</td>
</tr>
<tr>
<td>PB</td>
<td>16.70 ± 15.56</td>
<td>31.67 ± 25.84</td>
<td>44.50 ± 30.45</td>
</tr>
</tbody>
</table>

We observe in Fig. 5 that all the parameters of the PB model are inside the boxplots, i.e., min and max values of each parameter of the T3 models. Moreover, the parameters $a_1$, $b$, $c$, and $d$ are inside the interquartile range, i.e., first and third quartiles. We see that the coefficients $a_1$, $a_2$ are roughly opposite, which means that, as expected, the prediction is sensitive to the current IG slope. Since the proposed ARX model is a black box model, more explanation on the meaning of the rest of parameters may not be given.

Fig. 6 (left) shows results of tests 1 and 2. Performance reached by the ARX models obtained from different training data sets are represented by boxplots T3, T2, and T1, respectively. Performance of the SPA population-based ARX model is represented by boxplot PB. Table 1 presents (mean ± standard deviation) RMSE on IG prediction during 30, 60 and 120 min, performed by each model.

Fig. 6 (right) displays results of test 3, i.e., RMSE reached by models T3, NEE, NCOB, and NIOB. Table 2 displays $p$-values (obtained by Wilcoxon Matched-Pairs signed-rank test) of the comparison between RMSE reached by the model T3 and the models NEE, NCOB, and NIOB, on IG prediction during 30, 60 and 120 min.

Table 2

P-values, by Wilcoxon Matched-Pairs signed-rank test, from comparisons between RMSE reached by the model T3 and the models NEE, NCOB, and NIOB, on IG prediction during 30, 60 and 120 min.

<table>
<thead>
<tr>
<th>Model</th>
<th>30 min</th>
<th>60 min</th>
<th>120 min</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEE</td>
<td>0.03</td>
<td>0.04</td>
<td>0.28</td>
</tr>
<tr>
<td>NCOB</td>
<td>0.35</td>
<td>0.30</td>
<td>0.04</td>
</tr>
<tr>
<td>NIOB</td>
<td>0.52</td>
<td>0.21</td>
<td>0.08</td>
</tr>
</tbody>
</table>

5. Discussion

Fig. 5 shows that parameters of the population-based ARX model, obtained from the SPA database, are within the ranges of parameters of the patient-specific ARX models obtained from the FPA database. Therefore, performance reached by the PB model is good (compared with the obtained patient-specific models) for IG predictions (during 30, 60, and 120 min) on the FPA database. In fact, in Dasanayake et al. (2015), authors reported an average mean absolute error (MAE), over 15 patients, during 30 min of IG predictions (during PA) of 19.7 mg/dL. The PB model proposed in this paper reached an average MAE of 14.93 mg/dL, over 15 patients, during 30 min of IG prediction. Considering these results we can conclude that the proposed PB model is more accurate than the model presented in Dasanayake et al. (2015) on 30-min IG predictions. But the two models were not developed under the same conditions and for the same goal: (1) the model that we propose in this paper allows IG prediction during and out of PA periods, whereas the model proposed in Dasanayake et al. (2015) is only accurate on PA periods; (2) the model proposed in this paper uses EE (from HR and accelerometer signals), IOB, and COB (from CHO declared by the patient) as exogenous inputs, whereas the model proposed in Dasanayake et al. (2015) only uses accelerometer signals; (3) experiments in Dasanayake et al. (2015) were different of those performed in this paper. In fact, this is the first time that SPA and FPA
protocols are reported in a paper. For these reasons, a true comparison may not be established.

Fig. 6 (left) allows us to demonstrate that performance of patient-specific ARX models increases if the number of training data increases. We can observe that, for T3 models trained on 24 h of data, the RMSE on IG prediction during 30, 60 and 120 min is lower than RMSE obtained by models T2 and T1, i.e., models trained on 16 and 8 h of data, respectively. Table 1 confirms these results quantitatively. Mean
RMSE reached by T3 models is three times lower than the one obtained by models T1 on IG prediction during 30 min, five times lower on IG prediction 60 min ahead and nine times lower on IG predictions 120 min ahead. Comparing T3 and T2 models, difference in performance is less important than that observed while comparing T3 and T1 models, but T3 models remain more accurate.

Fig. 6 (right) shows that models using EE, COB and IOB as inputs are more accurate than those models using only IOB and COB, on 30 and 60 min ahead IG predictions. A p-value lower than 0.05 was found, when evaluating the difference between T3 and NEE models. However, a p-value = 0.28 was found on IG prediction during 120 min. These results may be interpreted as follows. When a T1D patient performs a PA, the use of EE as input in the models is very important (during PA and even 30 min after PA). However, when the effect of the performed PA is reduced (120 min after PA), the use of EE as input is less important, but, as showed in Fig. 6 (right), still allows improvement in performance. Concerning the effect of suppressing the COB input or the IOB input of the models, T3 models reached better performance than NCOB and NIOB models on the three prediction horizons, but difference was not significative.

Finally, Fig. 7 shows an example on 30 min ahead IG prediction, by the T3 model during a day (day 3 of visit 3 of the FPA protocol shown in Fig. 3). We can observe that the proposed model is good for predicting IG during PA but also out of PA. Therefore we consider that this model may be used on any application where 30 min ahead IG prediction is required, regardless of whether or not, the patients is performing a PA.

In fact, other works have reported results on 30 min ahead IG prediction by linear black box models. In Daskalaki et al. (2012) an AR model, using only IG information as input, reached a RMSE ranged between 14.0 and 21.6 mg/dL, whereas an ARX model, using IG and insulin information as inputs, reached a RMSE ranged between 13.3 and 18.8 mg/dL. These results are promising, however, estimation and validation were done on a virtual population, and the model was not confronted to the complex dynamics of IG on real T1D patients. In Cescon and Johansson (2009), a state-space model, receiving meal and insulin data as input variables, reached a RMSE = 18.08 mg/dL on IG prediction 30 min ahead, on a real T1D patient. In Stähl and Johansson (2010) a model combining physiological and black box models, reached RMSE of 19.1, 19.5, and 21.1 mg/dL on IG prediction during 20, 40, and 60 min, respectively. The reader shall notice that similar to our model, in these works insulin and IG information is automatically acquired by an insulin pump and a CGM system, respectively. CHO information is manually reported by the patients. However, experiments performed in these works (PA is not considered) are different to those performed in our study. Therefore, a comparison between results reported by the other works and ours may not be fair. But, RMSE reached by our models, using EE, COB, and IOB as inputs, trained on 24 h of data (i.e., T3 models), on IG prediction during 30 and 60 min (RMSE = 7.75 ± 4.51 and RMSE = 15.86 ± 9.61, respectively), on 15 T1D patients, show the interest of this paper.

5.1. Limitations

The main limitation of the study is that models were performed and validated on two databases composed of adult patients. Then, we cannot assure that proposed models will accurately predict IG on children or adolescent patients. Other limitation of the proposed model is the inability to accommodate for disturbances that may occur in the prediction horizon. In fact, predictions are based on past and current data, then future meals and physical activities will not be considered on the IG prediction. Other limitation is the fact of using IOB and COB time constants at fixed values (one population-based and the other empirically chosen). In fact, these time constants, which are patient dependent variables, may affect IG prediction performed by the proposed models. Another, limitation of the proposed model is that insulin sensitivity, which is also a patient dependent variable that varies during the day, is not considered by the model. This limitation may be overcome by adapting the model parameters during the day (Daskalaki et al., 2012). Finally, a limitation concerns the heart rate and accelerometer sensors errors. For instance, in Fig. 4 (bottom panel) some sensors errors (HR < 50 bpm) are displayed. Although sensors signals were processed on this work, online signal processing may not lead to the same results. This limitation will be overcome with the advance of sensor technologies (measurements and connectivity).

6. Conclusion

This paper presented ARX models for predicting IG during and after PA. We showed that a population-based ARX model may be used for predicting IG 30 min ahead with an acceptable accuracy. We demonstrated that performance of the ARX models increases when the number of training data increases. This result is very interesting, since on T1D patients using an artificial pancreas, the number of quality training data will increase through time. However, patients’ physiology also evolves through time, then training data should be correctly chosen. In fact, data collected a long time ago may not be representative of the patient's BG dynamics. Finally, we showed the interest of using meal, insulin, and physical activity information as inputs to increases performance on IG prediction during and after PA. In fact, current models found in the literature are limited to predicting IG during PA (insulin or meal are not used as inputs), or out of a PA (EE is not used as input). The fact of considering the 3 variables as inputs, allows the proposed model to perform accurate IG predictions during and out of PA.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The present study was supported by grants from the French National Agency ANR TECSAN 2015 (DIABELOOP AP project).

References


Cescon, Marzia, Johansson, Rolf, & Renard, Eric (2013b). Low-complexity mimo models of t1dm glucose metabolism. In 2013 9th asian control conference (ASCC) (pp. 1–6). IEEE.


