

Individualized Models for Glucose Prediction in Type 1 Diabetes: comparing black-box approaches to a physiological white-box one

Giacomo Cappon*, Francesco Prendin*, Andrea Facchinetti, Giovanni Sparacino, Simone Del Favero

Abstract—Objective: Accurate blood glucose (BG) prediction are key in next-generation tools for type 1 diabetes (T1D) management, such as improved decision support systems and advanced closed-loop control. Glucose prediction algorithms commonly rely on black-box models. Large physiological models, successfully adopted for simulation, were little explored for glucose prediction, mostly because their parameters are hard to individualize. In this work, we develop a BG prediction algorithm based on a personalized physiological model inspired by the UVA/Padova T1D Simulator. Then we compare white-box and advanced black-box personalized prediction techniques.

Methods: A personalized nonlinear physiological model is identified from patient data through a Bayesian approach based on Markov Chain Monte Carlo technique. The individualized model was integrated within a particle filter (PF) to predict future BG concentrations. The black-box methodologies considered are non-parametric models estimated via gaussian regression (NP), three deep learning methods: long-short-term-memory (LSTM), gated recurrent unit (GRU), temporal convolutional networks (TCN), and a recursive autoregressive with exogenous input model (rARX). BG forecasting performances are assessed for several prediction horizons (PH) on 12 individuals with T1D, monitored in free-living conditions under open-loop therapy for 10 weeks.

Results: NP models provide the most effective BG predictions by achieving a root mean square error (RMSE), RMSE = 18.99 mg/dL, RMSE = 25.72 mg/dL and RMSE = 31.60 mg/dL, significantly outperforming: LSTM, GRU (for PH=30 minutes), TCN, rARX, and the proposed physiological model for PH=30, 45 and 60 minutes.

Conclusions: Black-box strategies remain preferable for glucose prediction even when compared to a white-box model with sound physiological structure and individualized parameters.

This manuscript has been submitted on December 2022. Part of this work was partially supported by Italian Ministry of Education, Universities and Research (MIUR) through the projects “Learn4AP: Patient-Specific Models for an Adaptive, Fault-Tolerant Artificial Pancreas” under the initiative “SIR: Scientific Independence of young Researchers”, project ID: RBSI14JYM2, the project “A noninvasive tattoo-based continuous GLUCOSE Monitoring electronic system FOR Type-1 diabetes individuals (GLUCOMFORT)” (initiative “PRIN: Progetti di Rilevante Interesse Nazionale”, call 2020, project ID: 2020X7XX2P) and through the initiative “Departments of Excellence” (Law 232/2016).

G. Cappon, F. Prendin, A. Facchinetti, G. Sparacino, S. Del Favero are with the Department of Information Engineering, University of Padova, Via G. Gradenigo 6/B, 35131 Padova, Italy. (email: [cappongi, prendin, facchine, gianni, sdfavere]@dei.unipd.it)

correspondence email: sdfavere@dei.unipd.it.

* These authors contributed equally

Index Terms—Type 1 Diabetes, Individualized Glucose Prediction, Nonlinear Physiological Model, Black-box methods, Particle Filter.

I. INTRODUCTION

Type 1 diabetes (T1D) is a chronic autoimmune disease caused by the progressive destruction of beta cells in the pancreas, which leads to the inability of producing endogenous insulin by the organism, [1]. As a result, blood glucose concentration (BG) tends to exceed the hyperglycemic threshold (BG > 180 mg/dL), a situation that, if frequent and prolonged, could lead to several serious cardiovascular long-term complications, as well as nephropathy and neuropathy, [2]. To reduce BG levels, administration of exogenous insulin several times a day is necessary. Unfortunately, excessive exogenous insulin dosing could lead patients to hypoglycemia, i.e., BG < 70 mg/dL, which is dangerous even in the short-term since it could cause fainting, light-headedness, coma and even death, [3].

Effective T1D treatment relies on BG frequent monitoring, made through either the classic fingerstick device, [4], or more modern minimally invasive continuous glucose monitoring (CGM) sensors, [5]–[7], and is far from being trivial, [8]. Indeed, T1D management represents, from a patient perspective, a life-long learning process to understand how several everyday factors (e.g., illness, diet, and physical activity) affect BG levels and how interventions (e.g., rescue carbohydrate intake and, of course, insulin administration) can be used to keep BG in the safe range. In this context, many efforts have been made by the research community to provide new tools able to help patients with T1D, [9], [10]. Among them, CGM-based algorithms able to predict future BG concentration in real-time have the potential to significantly improve T1D therapy efficacy, [11], by enabling proactive therapeutic decisions based on the expected future glucose levels, rather than the current one.

Despite several different methodologies have been proposed for real-time BG prediction, [11], the problem remains open due to several unknown disturbances altering BG levels and to the high inter-/intra-patient variability in glucose physiology. The existing prediction techniques rely on a mathematical model which describes the relationship between a certain set of input features and BG. The choice of a suitable model of glucose-insulin regulation is a critical step and the options

spans from “black-box” models, completely data-driven, to models based on mechanistic/semi-mechanistic nonlinear description of metabolic physiology.

A rich literature has focused on the first option, [11], [12], exploring a broad spectrum of black-box approaches that ranges from linear techniques (auto-regressive, auto-regressive moving average, etc.), typically used in time-series analysis, [13], [14], and system identification, [15]–[17], to the nonlinear approaches commonly adopted in machine learning (support vector regression, random forests, gaussian process, feed forward neural networks, autoregressive networks, etc.), [18]–[20], and including the powerful deep learning techniques, [21]–[23].

Among the white-box models, the nonlinear physiological models available in the T1D literature, there are the so-called minimal models, [24], that proposed simplified descriptions of the physiology with a few equations and model parameters. This parsimonious parametrization grants identifiability in pre-defined experimental conditions. Unfortunately, these models have proved too rigid and simplistic to allow accurate prediction, [11]. A possible white-box alternative are maximal models, commonly used in computer simulations, [25]–[28]. They provide a more realistic physiological description by using several equations with many parameters. Despite their appealing feature of having a clear and solid physiological ground, their use for glucose prediction were substantially less investigated, since their many parameters are hard to be estimated from easily accessible patient data (i.e. CGM, meal and insulin data), making them hard to personalize and thus limiting their predictive effectiveness. Moreover, they have a nonlinear structure, requiring sophisticated tools both for parameters estimation and for the computation of glucose prediction.

In this work, we face the above mentioned challenges and explore the potential of using a white-box maximal-model based methodology for glucose prediction, comparing it to black-box alternatives. The white-box model adopted is inspired by the UVA/Padova T1D Simulator, [28], accepted by the US Food and Drug Administration (FDA) as a replacement of animal preclinical testing of closed-loop drug delivery systems. A Bayesian approach, implemented by Markov Chain Monte Carlo (MCMC), [29], is used to estimate the large number of parameters in the presence of complex nonlinear dynamics. The obtained personalized model is then used within a nonlinear prediction scheme based on a particle filter methodology, [30].

The so derived white-box glucose prediction approach is compared to four black-box algorithms: a nonparametric techniques that recently proved effective in glucose prediction, [17], three deep learning algorithms (long short term memory (LSTM), gated recurrent unit (GRU), temporal convolutional network (TCN)), and a recursive autoregressive with exogenous input (rARX) model. Algorithms’ assessment has been performed for different prediction horizon (PH) on a dataset which comprises 12 subjects monitored for about 10 weeks in daily-life conditions.

The paper is organized as follows. Section II describes the nonlinear physiological model of glucose-insulin regulation

used for BG prediction. Section III describes the black-box methodologies considered for this work. Section IV illustrates the dataset and the evaluation metrics. Section V reports the results and discussion. Finally, Section VI summarizes the main findings, draws some conclusions and proposes possible future developments.

II. WHITE-BOX GLUCOSE PREDICTION MODEL

In this work we explored the use of a white-box maximal models for glucose prediction. The approach is based on two components: i) a nonlinear model of glucose-insulin regulation, which is personalized to capture patient-specific physiology using a Bayesian methodology, and ii) a particle filter that leverages on such model and, handling the non-linearity, allows to formulate glucose predictions.

In the following, Section II.A describes the structure of the physiological model, Section II.B discusses the model identification procedure, and Section II.C reports details on the employed prediction scheme.

A. Physiological Model of Glucose-Insulin Regulation

The physiological model of glucose-insulin regulation used in this work has two inputs, insulin infusion $I(t)$, and carbohydrate intake $CHO(t)$, and one output, the interstitial glucose concentration $IG(t)$. The model is composed of three subsystems describing subcutaneous insulin absorption, oral glucose absorption, and glucose-insulin kinetics. As for a preliminary version of the model presented in [31], we started from the maximal physiological model implemented in the UVA/Padova T1D Simulator (T1DS), [32], and to simplify it as much as possible to reduce the number of parameters to be estimated for individualization while retaining its ability to achieve good glucose prediction. The details about model structure and its parameters can be found in Appendix I. The overall physiological model is a nonlinear time-invariant state-space model:

$$\begin{cases} \dot{\mathbf{x}}_{phy}(t) = \mathbf{f}_{phy}(\mathbf{x}_{phy}, \mathbf{u}_{phy}, t, \boldsymbol{\theta}_{phy}) \\ y(t) = IG(t) \end{cases} \quad (1)$$

where $\mathbf{x}_{phy}(t)$ is the state vector, defined as

$$\mathbf{x}_{phy}(t) := [x_{ins}, | x_{oral}, | x_{glu}]^T$$

where x_{ins} is the state vector of the subcutaneous insulin absorption subsystem, x_{oral} is the state vector of the oral glucose absorption subsystem, and x_{glu} is the state vector of the glucose-insulin kinetics subsystem;

$$\mathbf{u}_{phy}(t) := [I(t), CHO(t)]$$

is the input vector; $\mathbf{f}_{phy}(\cdot)$ is the state update function obtained combining (6) (7), and (9); $\boldsymbol{\theta}_{phy}$ is the set of unknown model parameters (formally defined in Appendix I) whose estimation will be discussed in the next section.

B. Offline Bayesian Estimation of Personalized Model Parameters by Markov Chain Monte Carlo

Model personalization has been performed by identifying for each patient the unknown model parameters θ_{phy} using the training data $Y := \{CGM(t_k), t_k = k \cdot T_s, k = 1, \dots, D\}$ and $U := \{u_{phy}(t_k), t_k = k \cdot T_s, k = 1, \dots, D\}$ where D is the number of data points available.

The identification has been performed by adopting a Bayesian approach and specifically, in this work θ_{phy} is estimated through its posterior mean defined as

$$\hat{\theta}_{phy} = E[\theta_{phy}|Y, U] = \int \theta p_{\theta|Y, U}(\theta|Y, U) d\theta \quad (2)$$

In fact the posterior mean is known to be the minimum variance unbiased estimator of θ_{phy} .

The Bayes theorem allows to obtain the a posteriori density function $p_{\theta|Y, U}(\theta|Y, U)$ as:

$$p_{\theta|Y, U}(\theta|Y, U) = \frac{p_{Y|\theta, U}(Y|\theta, U)p_{\theta}(\theta)}{\int p_{Y|\theta, U}(Y|\theta, U)p_{\theta}(\theta)d\theta} \quad (3)$$

where $p_{Y|\theta, U}(Y|\theta, U)$ is the likelihood function, i.e., the probability of observing a certain Y given the parameter vector θ and the input U .

Even using (3), the integral in (2) is analytically intractable, therefore it has to be approximated by resorting to MCMC [29]. In particular, we generate N samples $\theta_i, i = 1, \dots, N$ from the posterior distribution $p_{\theta|Y, U}(\theta|Y, U)$, by creating a Markov Chain whose stationary distribution is exactly this posterior (target distribution). Then, these samples θ_i are used to perform Monte Carlo integration to obtain a point estimate of θ_{phy} :

$$\hat{\theta}_{phy} = \frac{1}{N} \sum_i^N \theta_i. \quad (4)$$

To build such a chain, the Single Component Metropolis-Hastings (SCMH) algorithm has been used [29]. Implementative details about the implemented SCMH procedure can be found in Appendix II. An open-source software implementation of the proposed offline Bayesian estimation approach can be found at <https://github.com/gcappon/replay-bg>.

C. Real-time Glucose Prediction through Particle Filter

Up to this point, we focused on estimating the parameters of the physiological model of a T1D subject, to capture the patient-specific dynamics. This process can be done "offline" using the available information obtained from retrospective patient data.

As mentioned above, in this section, we will present how we used such a personalized model to predict, in real-time, future glucose concentrations. This task has to be performed "online", so we resorted to a sequential algorithm that at each timestep t_k , when a new measurement $y(t_k) = CGM(t_k)$ becomes available, updates the current estimate of the model state $x(t_k)$ and uses it to infer future glucose concentration. In particular, we employ the particle filter (PF), [30], the state-of-the-art sequential Bayesian prediction technique capable of handling the nonlinear structure of the model.

PF is based on the recursive update of the posterior probability function $p(x(t_k)|y(t_{1:k}), u(t_{1:k}))$ where $y(t_{1:k})$ is a shorthand for the variables $y(t_1), \dots, y(t_k)$ and $u(t_{1:k})$ indicates $u(t_1), \dots, u(t_k)$.

The recursive update of $p(x(t_{k-1})|y(t_{1:k-1}), u(t_{1:k-1}))$ is performed through two fundamental steps, i.e., one step-ahead prediction and measurement update.

The one step-ahead prediction step assumes that the posterior probability $p(x(t_{k-1})|y(t_{1:k-1}), u(t_{1:k-1}))$ is available at time t_{k-1} and uses such a posterior probability to infer

$$p(x(t_k)|y(t_{1:k-1}), u(t_{1:k}))$$

Then, when at time t_k a new measurement, $y(t_k)$, becomes available, in the measurement update step such a measurement is used to compute the posterior probability

$$p(x(t_k)|y(t_{1:k}), u(t_{1:k})).$$

The two steps are then repeated for each available measurement in the dataset.

PF performs these steps using a sampled approximation of the probability functions at play:

$$p(x(t_{k-1})|y(t_{1:k-1}), u(t_{1:k-1})) \approx \sum_{p=1}^P w^p(t_{k-1}) \delta(x(t_{k-1}) - x^p(t_{k-1})).$$

where $\{x^p(t_{k-1})\}_{p=1}^P$ is a set of P points, called "particles", in the support of $p(x(t_{k-1})|y(t_{1:k-1}), u(t_{1:k-1}))$. Each particle is associated to a weight $\{w^p(t_{k-1})\}_{p=1}^P, \sum_p w^p(t_{k-1}) = 1$, and

$$p(x(t_k)|y(t_{1:k-1}), u(t_{1:k})) \approx \sum_{p=1}^P w^{*p}(t_k) \delta(x(t_k) - x^p(t_k)).$$

where $\{x^p(t_k)\}_{p=1}^P$ are the P particles representing $p(x(t_k)|y(t_{1:k-1}), u(t_{1:k}))$, each associated to a weight $\{w^{*p}(t_k)\}_{p=1}^P, \sum_p w^{*p}(t_k) = 1$.

The one step-ahead prediction step of the PF is particularly convenient. It can be shown that

$$p(x(t_k)|y(t_{1:k-1}), u(t_{1:k})) = \int p(x(t_k)|x(t_{k-1}))p(x(t_{k-1})|y(t_{1:k-1}), u(t_{1:k-1}))dx(t_{k-1})$$

where $p(x(t_k)|x(t_{k-1}))$ is fully described by the state update equations of (1).

Regarding the measurement update step, it is possible to demonstrate that it holds

$$p(x(t_k)|y(t_{1:k}), u(t_{1:k})) \propto p(y(t_k)|x(t_k), u(t_{1:k}))p(x(t_k)|y(t_{1:k-1}), u(t_{1:k}))$$

where $p(y(t_k)|x(t_k), u(t_{1:k}))$ is the likelihood function that is fully specified by (1).

Details on how these quantities are calculated in practice are reported in Appendix III.

As an additional result, the posterior probability $p(x(t_k)|y(t_{1:k}), u(t_{1:k}))$ is furtherly used by the PF to compute the posterior probabilities

$$p(x(t_{k+i})|y(t_{1:k}), u(t_{1:k+i})), \forall i = 1, \dots, PH$$

describing the state distribution predicted i steps-ahead in time up to PH steps ahead (with PH the prediction horizon).

In particular, it is possible show that:

$$p(y(t_{k+i})|y(t_{1:k}), u(t_{1:k+i})) = p(y(t_{k+i})|x(t_{k+i})), \forall i = 1, \dots, PH$$

that is completely specified by equation (1).

Finally, a point estimate of future CGM values at time t_{k+i} , $i = 1, \dots, PH$ can be derived using the expectation of the posterior:

$$\hat{y}(t_{k+i}|t_k) = E[p(y(t_{k+i})|x(t_{k+i}))], \forall i = 1, \dots, PH.$$

An open-source software implementation of the proposed real-time glucose prediction approach based on PF can be found at <https://github.com/checoisback/phy-predict>.

III. BLACK-BOX GLUCOSE PREDICTION MODELS

A. Linear Non-parametric Models

Although the metabolic physiology is nonlinear, several contributions have shown that its approximation with linear models is an appealing option for BG prediction, [17], [33]. For this reason, our analysis considered the use of an advanced nonparametric (NP) identification techniques for linear models, [34]. In particular, this approach estimates the unknown impulse response related to insulin, meal and glucose, from noisy measurements. Unlike the standard approach that constraints the unknown functions to a parametric structure, [35], the nonparametric approach estimates the unknown impulse responses over a infinite-dimensional set given by a Reproducing Kernel Hilbert Space (RKHS). Such a space is completely specified by the choice of the kernel. In this case, the Stable Spline kernel is used as it incorporates key prior knowledge, such as smoothness and stability of the predictor impulse responses to be estimated. In [17], this approach proved to be the most effective among several linear and nonlinear methods for glucose prediction.

B. Nonlinear Deep Learning Approaches

As described in [11], [36], there are a growing number of deep learning methodologies to forecast BG levels. In particular, due to their ability to handle time-series and sequential data, there is an increasing trend to develop both recurrent and convolutional neural networks (RNN and CNN, respectively) for BG forecasting, [21], [22], [37]. Unlike traditional feed-forward neural networks in which the information flows from the input towards the output layer, RNN are characterized by recurrent units with loops propagating the information back to the same unit, [38]. So, each learning step takes into account not only the current input, but also what was learnt from the previous inputs, [38]. In this work we implemented two multi-input RNN based on: i) Long Short Term Memory (LSTM) cells and ii) Gated Recurrent Units (GRU). Specifically, they

are composed by a single hidden layer which comprises 30 units, as in [21]. Similarly to [23], [39], we also investigated the use of CNN-based algorithms, by developing a Temporal Convolutional Network (TCN) which combines 3 causal and dilated convolutional layers (equipped with 8, 16, and 32 filters) that are used for extracting features from the inputs. Finally, all the proposed deep learning algorithms are fed by three input channels (past history of glucose data, meal intake, and insulin injections) and equipped with an output layer (dense), comprising a number of units corresponding to the future BG levels to forecast (in this work 12, corresponding to a $PH = 60$ minutes, sampling time is 5 minutes), as in [21], [40]. As shown in Figure 1, once fed by input data, the algorithms provide as output a trajectory of 12 future consecutive glucose samples. Of note, the deep learning models are developed within Python (Keras library) and trained using a Nvidia Titan RTX.

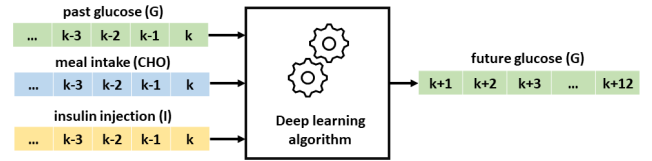


Fig. 1: Schematic representation of BG forecasting with LSTM, GRU and TCN.

C. Recursive Models

Finally, we explored adaptive strategies that have the potential to accommodate the changes over time in patient metabolism. In particular, we implemented the personalized recursive autoregressive model with exogenous input (rARX) proposed by Finan et al. [41]. As the other methodologies, this approach takes in input past CGM values, carbohydrate intakes, and insulin recordings. Beside using this data to predict future CGM values, each time a new CGM sample is collected the model parameters are adapted according to a well-established recursive estimation scheme [35].

IV. METHODOLOGY ASSESSMENT

A. Dataset

The dataset used in this study is the Ohio Type 1 Diabetes Mellitus dataset updated on the 2020 release, [42], from now on referred as the OhioT1DM. The OhioT1DM dataset, comprises 12 subjects with T1D monitored with a Medtronic Enlite CGM system for 8 weeks. Participants wore an insulin pump (Medtronic 530G or 630G) and a wearable system (Basis Peak fitness or Empatica Embrace) to measure physiological variables, such as skin temperature, skin conduction, and heart rate. In addition, subjects reported information on meals: timing, amount, and type (that is, breakfast, lunch, dinner, snack, hypoglycemia treatment). Each subject in the OhioT1DM data set is split into a training set (about the initial 6 monitoring weeks) and into a test set (roughly the last 10 days). This dataset represents a challenge for BG predictive algorithms: glucose dynamics recorded in daily-life conditions are much more complex to describe than

Methods	RMSE [mg/dL]			TG (min)		
	PH = 30 min	PH = 45 min	PH = 60 min	PH = 30 min	PH = 45 min	PH = 60 min
PHY	24.00	33.03	41.14	15	22.5	30
	[20.77-27.14]	[29.17-36.06]	[36.85-44.28]	[12.5-20]	[15-35]	[22.5-50]
NP	18.99	25.72	31.60	15	20	20
	[16.90-21.65]	[23.44-30.52]	[28.99-36.82]	[10-20]	[10-22.5]	[15-27.5]
LSTM	19.75	26.80	32.54	15	20	25
	[17.82-22.01]	[24.55-30.30]	[30.08-37.07]	[12.5-22.5]	[15-25]	[22.5-32.5]
GRU	19.81	26.77	32.78	17.5	20	25
	[17.75-21.93]	[24.62-29.83]	[30.31-36.67]	[12.5-20]	[20-25]	[25-32.5]
TCN	20.11	27.28	33.54	15	20	25
	[17.97-22.39]	[25.46-30.47]	[30.83-37.48]	[10-20]	[17.5-25]	[20-32.5]
rARX	20.19	27.86	33.97	15	20	20
	[18.74-21.90]	[26.22-29.67]	[32.64-36.08]	[10-20]	[10-20]	[10-25]

TABLE I: Comparison between performance metrics (median [25th-75th]) obtained using the PHY, NP, LSTM, GRU, TCN, and rARX models for PH = 30, 45, 60 minutes.

those generated by simulation tools or those recorded during well-controlled in-hospital trial sessions, since in the former the patient is exposed to substantially larger disturbances to glucose homeostasis. Handling data recorded under free-living conditions raises some technical issues mainly about synchronization and completeness of the recorded information. In particular, the OhioT1DM dataset presents long portion of missing CGM readings and the sampling time is not homogeneous. Therefore, all signals were aligned into a uniform time grid with a sampling period of $T_s = 5$ minutes. Any CGM gap in the training set shorter than 30 minutes was interpolated with a first order polynomial while a simple and causal zero-order-hold imputation was performed on the test set. A preliminary investigation indicates that alternative real-time extrapolation techniques (e.g. linear extrapolation) do not lead to a significant improvement in the overall prediction performance.

B. Algorithms Assessment Metrics

Evaluation of prediction accuracy has been performed in the test set by comparing the obtained glucose predictions with the actual CGM values using different PH . Two, frequently used, performance metrics are considered: the root mean square error (RMSE) and the time gain (TG), defined as follows:

$$RMSE(i) = \frac{1}{\sqrt{m}} \|y(t_{k+i}) - \hat{y}(t_{k+i}|t_k)\|_2, \forall i = 1, \dots, PH.$$

where $\|\cdot\|_2$ denotes the ℓ_2 norm. The larger the $RMSE$ the worst the prediction.

$$TG(i) = i \cdot T_s - \text{delay}(y(t_{k+i}), \hat{y}(t_{k+i})), \forall i = 1, \dots, PH.$$

where $\text{delay}(s_1(t_k), s_2(t_k))$ [min] quantifies the delay between two signals $s_1(t_k)$ and $s_2(t_k)$ and is based on the cross-correlation (xcorr), a measure of similarity of the two signals. Specifically, the delay is defined as the temporal shift τ that

maximizes the cross-correlation (xcorr) between s_1 and a τ -shifted version of s_2 :

$$\text{delay}(s_1(t_k), s_2(t_k)) = \underset{\tau}{\operatorname{argmax}} \operatorname{xcorr}(s_1(t_k), s_2(t_k - \tau)). \quad (5)$$

The higher TG , the more prompt and useful the prediction and $TG = 0$ means that the model prediction is comparable to looking at the current glucose level.

Performance metrics of Physiological model (within the PF) are compared against NP, LSTM, GRU and TCN models using a paired t-test unless the hypothesis of normal distribution was rejected by a Lilliefors test (p-value < 0.05). In this case, Wilcoxon ranksign test was used. Reported p-values are two-tailed and considered statistically significant when < 0.05.

V. RESULTS AND DISCUSSION

Table I reports the prediction performance in terms of RMSE and TG, for PH = 30, 45, and 60 minutes achieved by the white-box physiological model (hereafter labeled as PHY), and the considered black-box methodologies (i.e., NP, LSTM, GRU, TCN, and rARX). From Table I, three main outcomes can be observed: i) black box algorithms outperform PHY in terms of RMSE, for all prediction horizons, ii) all methodologies seem to achieve similar TG (no statistically significant difference was found when comparing all pairs of prediction approaches with a paired t-test corrected with the Bofferroni method), and iii) that there are no large differences in terms of RMSE between the considered black-box approaches.

Specifically, the RMSE obtained by PHY is larger than the one of the other methodologies by approximately 5 mg/dL, 7 mg/dL, and 9 mg/dL for PH = 30, 45, and 60 minutes, respectively. This corresponds to a performance deterioration of about 25%. The NP approach allows to achieve the lowest RMSE: median RMSE = 18.99 mg/dL, RMSE = 25.72 mg/dL and RMSE = 31.60 mg/dL for PH = 30, 45 and 60 minutes, respectively. The performance improvement is found to be

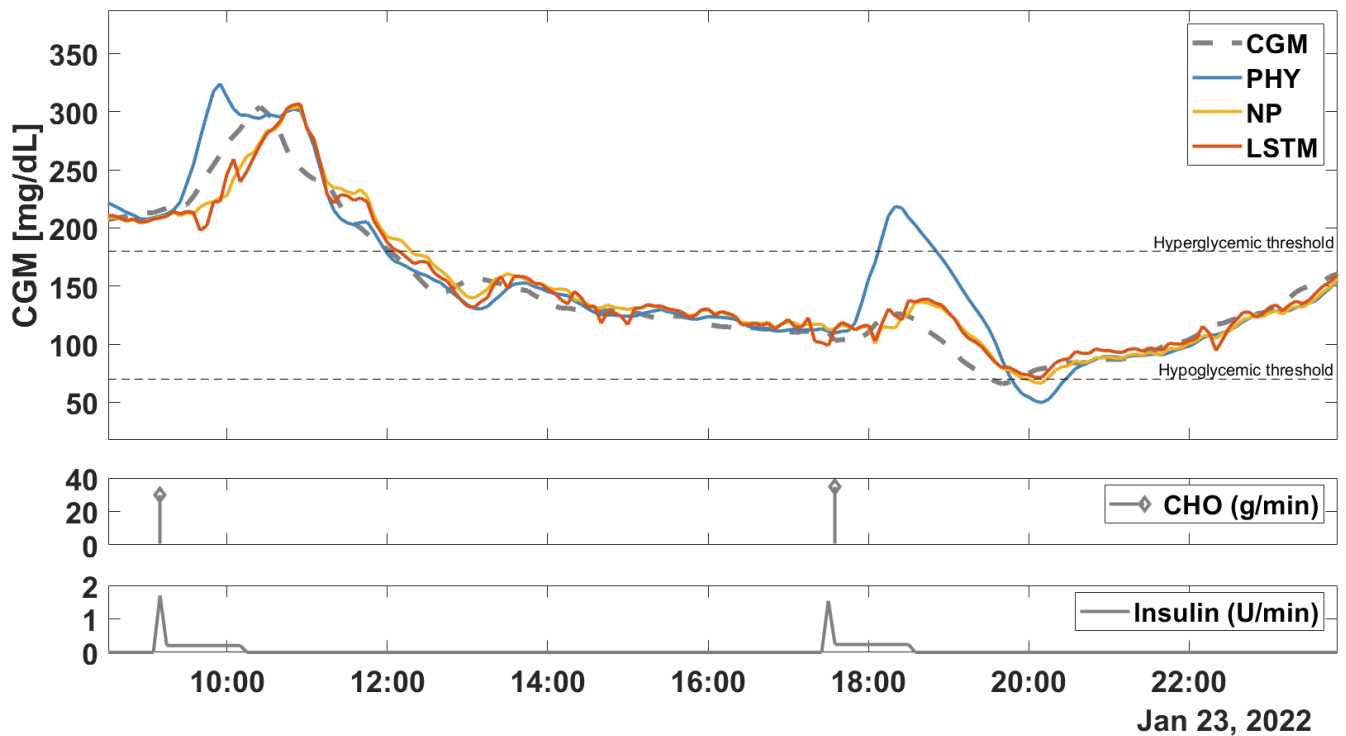


Fig. 2: Representative subject (ID:570) of the OhioT1DM dataset. The upper panel shows CGM data (grey dashed line) and the 30-min ahead prediction obtained using PHY (blue line), NP (yellow line) and LSTM (red line). Middle panel shows the CHO content of the meal, in g/min. Bottom panel shows injected insulin (U/min)

statistically significant with respect to LSTM and GRU (p -value = 0.04 and p -value = 0.03) for PH = 30 minutes, while no significant difference is found for longer prediction horizons. Furthermore, NP model showed to be significantly better than TCN, with p -value = 0.001, p -value = 0.006 and p -value = 0.009, for PH = 45 and PH = 60 minutes, respectively. The NP approach achieves the largest median improvement with respect to PHY, decreasing RMSE by 26.5%, 26.2% and 24.9%, for PH = 30, 45 and 60, respectively (p -values $< 10^{-4}$ for all the considered PHs). The numerical results obtained in this work by LSTM, GRU and TCN are comparable with what has been obtained in other literature contributions dealing with the assessment of individualized deep learning algorithms for BG forecasting in the OhioT1DM dataset, [22], [23], [33]. In [22], a recurrent convolutional network granted a mean RMSE = 20.6 mg/dL, 26.8 mg/dL and 33.9 mg/dL for PH = 30, 45 and 60 minutes, respectively. Similarly, the TCN and the LSTM tested in [23] achieved a mean RMSE = 20.23 mg/dL and RMSE = 20.11 mg/dL for PH = 30 minutes, and RMSE = 34.21 mg/dL and 33.10 mg/dL for PH = 60 minutes. Finally, despite permitting model adaptation, rARX performs slightly worse than the other black-box models employed in this work but still better than PHY, by achieving a median RMSE = 20.18 mg/dL, 27.91 mg/dL and 34.14 mg/dL for PH = 30, 45 and 60 minutes, respectively.

To the best of our knowledge, there are only two other literature contributions investigating maximal model individualization, [43], [44]. However, both in [43], [44], rich datasets of frequent plasma glucose and insulin measurements collected

in a clinical setting were required, thus strongly limiting the applicability of these methodologies to be used on data collected in free-living conditions. Furthermore, none of two models derived was used for glucose prediction.

Regarding the use of particle filters to obtain glucose predictions, the work of [45] showed the potential of this approach for the scope. However, in the prediction scheme of [45], they adopted an average model, thus not personalizing the methodology at the patient level.

With the aim of better understanding the consistent difference between PHY and black-box models in glucose prediction, we consider an illustrative example on a one-day long data portion extracted from the test set, depicted in Figure 2. In the top panel, CGM data (grey dashed line) of a subject of the OhioT1DM dataset (ID:570) is shown with the 30-minute ahead-in-time prediction of PHY, NP, and LSTM models, in blue, yellow and red, respectively. Middle and bottom panels show meal and insulin data, respectively. Analyzing the extracted portion of data, there are two meals with similar amount (30 g/min at 9:10, 35 g/min at 17:35) and two similar corresponding insulin boluses delivered with the so-called dual-wave mode, [46]. It is interesting to note that the two corresponding postprandial responses are very different. In the first case, glucose increases from 200 mg/dL to 300 mg/dL within an hour after the meal, whereas, in the second case, glucose remains almost flat (about 20 mg/dL excursion) and then decreases after an hour, reaching hypoglycemia. The first postprandial excursion is aligned with the physiological expectation that a meal should be followed by a glucose

increase, while in the second postprandial excursion this does not happen, for causes that are hard to guess and that might span from an high fat meal composition to psychological stress slowing down digestion. The (unavoidably simplified) physiological model structure has not the flexibility to cover both types of responses and the model imposes a similar shape to the two postprandial glucose excursions leading, as a consequence, to an extremely large prediction error observed during the second meal. On the contrary, the black-box approaches prove more flexible and are able to produce two different postprandial shapes despite the similar inputs.

As a final consideration, we report a preliminary investigation of the computational power required to compute a prediction using each method. In fact, in the final deployment of a glucose prediction algorithm, this step might have been performed in real-time on portable hardware. As a proxy, we considered the computation time required on a DELL desktop PC equipped with an Intel(R) Core(TM) i7-9700 CPU @ 3.00GHz, 16.0 GB RAM. The code was not optimized, PHY (considering $N=5000$ particles), NP and rARX are implemented in Matlab while LSTM, GRU and TCN were implemented using Python (Keras library). Therefore, this result offers only an early exploration of this issue. The average time required to PHY to compute the 30-minutes ahead prediction is about 30 millisecond and is one order of magnitude larger than the time required to the other methods: the NP method needs 2 millisecond, rARX 0.05 milliseconds and the non-linear methods (less than 3 milliseconds). Moreover, it should be noted that the computation time of PHY is strongly affected by the number of particle considered (here 5000), whereas the other approaches exhibit a computational time less sensitive to the method hyperparameters. The time required to train the models is not considered, as this computationally demanding step could be performed on a remote server.

Future work will attempt to increasing the flexibility of white-box models. This will include considering time-varying model parameters estimated in real-time by the PF, to track patient-specific intraday variability and meal-to-meal differences in CHO absorption. Moreover, further investigations will compare white-box and black-box modeling approaches for control purposes, i.e. when they are used within a model-based closed-loop controller or as core element of a decision support systems. In fact, it is unclear if superior prediction performance will translate into superior glycemic control or if the physiologically grounded, albeit simplified, structure of white-box model could lead to more robust and effective control actions. Finally, future work will analyze the impact of newer and more accurate CGM sensors on prediction accuracy. In fact, the CGM sensor used in the dataset considered in this work was less accurate than the newer generation devices currently available in the market [47], [48].

VI. CONCLUSION

Real-time glucose prediction algorithms are key for developing next-generation tools to improve diabetes care. The diabetes research community intensively focused on the use of black-box prediction approaches, investigating many techniques spanning from linear models to deep learning-based

approaches. On the other hand, white-box maximal models for glucose predictions are less investigated, due a number of technical difficulties they pose.

This work compared five black-box methodologies (a non-parametric technique that recently proved effective in glucose prediction, [17], a LSTM, [33], a GRU, a TCN, [23], and a rARX [41]), with a newly developed white-box technique based on a nonlinear physiological model of glucose-insulin dynamics, whose parameters are individualized through a MCMC approach and embedded in a PF to predict future glucose values. On the data under study, collected by T1D patients in free-living conditions, the considered black-box methodologies significantly outperform the white-box approach for all the PH under study. Moreover, among the data-driven algorithms, the best performance are achieved by the linear NP approach, that grants statistically significant improvement in the performance with respect to LSTM and GRU for $PH = 30$ minutes, and to TCN for all the considered PH. One possible reason for the differences in performance between white-box and black-box models might reside in the fact that the first are less flexible in accommodating the large variety of patterns observed in the data and that might be caused by multiple unmodeled factors, including variability in meal absorption, different meal compositions, stress, illnesses, physical activity, inaccuracy in estimating carbohydrate content of a meal. Future works will aim at increasing the flexibility of white-box models and at comparing white-box and black-box model for control purposes.

APPENDIX I PHYSIOLOGICAL MODEL OF GLUCOSE-INSULIN REGULATION

This appendix reports a detailed description of the nonlinear physiological model of glucose-insulin regulation used by the white-box glucose prediction approach presented in Section II. The model is composed of three subsystems describing subcutaneous insulin absorption, oral glucose absorption, and glucose-insulin kinetics, to each presented in one of the following sections.

A. Subcutaneous Insulin Absorption Subsystem

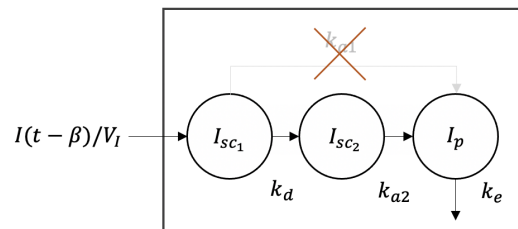


Fig. 3: Subcutaneous insulin absorption subsystem scheme.

The subcutaneous insulin absorption model is a slightly simplified version of the one used in the T1DS model, described in [49] and illustrated in Figure 3. The model is composed of three compartments. Exogenous insulin I is infused to the

first compartment where it appears after a delay β . In the first compartment, representing insulin in a non-monomeric state, insulin is transformed in a monomeric state and then diffused to the plasma. The model equations are:

$$\begin{cases} \dot{I}_{sc1}(t) = -k_d \cdot I_{sc1}(t) + I(t - \beta)/V_I \\ \dot{I}_{sc2}(t) = k_d \cdot I_{sc1}(t) - k_{a2} \cdot I_{sc2}(t) \\ \dot{I}_p(t) = k_{a2} \cdot I_{sc2} - k_e \cdot I_p(t) \end{cases} \quad (6)$$

where I_{sc1} (mU/kg) and I_{sc2} (mU/kg) represent the insulin in a non-monomeric and monomeric state, respectively; I_p (mU/l) is the plasma insulin concentration; k_d (min^{-1}) is the rate constant of diffusion from the first to the second compartment; k_{a2} (min^{-1}) is the rate constant of subcutaneous insulin absorption from the second compartment to the plasma; k_e (min^{-1}) is the fractional clearance rate; V_I (l/kg) is the volume of insulin distribution; β (min) is the delay in the appearance of insulin in the first compartment. A priori information on model parameters has been obtained from the literature, [49]. Specifically, V_I and β have been set to population values, i.e. 0.126 l/kg and 8 min, respectively. Furthermore, k_d has been constrained to $k_d \geq k_{a2}$ since the two combinations are interchangeable. Unknown model parameters, identified via the MCMC procedure presented in Section II.B, are $\theta_{ins} = [k_{a2}, k_d]$.

B. Oral Glucose Absorption Subsystem

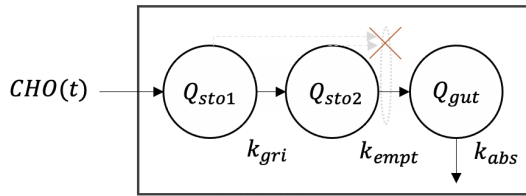


Fig. 4: Oral glucose absorption subsystem scheme.

The oral glucose absorption subsystem model, taken from [50], represents a simplified version of the model used in the T1DS and is illustrated in Figure 4. It describes the gastro-intestinal tract as three-compartment system: the first two compartments account for food in the stomach (solid and grinded state), while the third compartment models the upper small intestine where CHO is absorbed. Model equations are:

$$\begin{cases} \dot{Q}_{sto1}(t) = -k_{gri} \cdot Q_{sto1}(t) + CHO(t) \\ \dot{Q}_{sto2}(t) = k_{gri} \cdot Q_{sto1}(t) - k_{empt} \cdot Q_{sto2}(t) \\ \dot{Q}_{gut}(t) = k_{empt} \cdot Q_{sto2}(t) - k_{abs} \cdot Q_{gut}(t) \end{cases} \quad (7)$$

where Q_{sto1} (mg/kg) and Q_{sto2} (mg/kg) are the glucose amount in the stomach in a solid and liquid state, respectively; Q_{gut} (mg/kg) is the glucose concentration in the intestine; k_{gri} (min^{-1}) is the rate constant of grinding; k_{empt} (min^{-1}) is the rate constant of gastric emptying; k_{abs} (min^{-1}) is the rate constant of intestinal absorption; CHO (mg/kg/min) is the ingested carbohydrate rate. Model (7) allows to estimate the rate of glucose appearance in plasma Ra (mg/kg/min) as:

$$Ra(t) = f \cdot k_{abs} \cdot Q_{gut}(t) \quad (8)$$

where f (dimensionless) is the fraction of the intestinal content absorbed in the plasma. A priori information on model (7) has been obtained from the literature, [50]. In particular, we set f equal to 0.9 and we constrained $k_{gri} = k_{empt}$. Furthermore, k_{abs} has been constrained to $k_{abs} \leq k_{empt}$ since the two combinations are interchangeable. As such, the unknown model parameters, estimated using the MCMC approach presented in Section II.B, are $\theta_{oral} = [k_{abs}, k_{empt}]$.

C. Glucose-Insulin Kinetics Subsystem

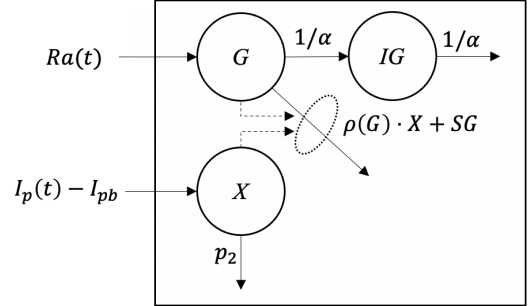


Fig. 5: Glucose-insulin kinetics subsystem scheme.

The subsystem of glucose-insulin kinetics is based on a well-known two-compartment model that describes the impact of the plasmatic insulin action and glucose rate of appearance in plasma glucose concentration introduced in [51]. The model is further equipped with a third compartment to describe the transport of glucose from plasma to the interstitium where it is measured by the sensor. The model is illustrated in Fig. 5. Model equations are:

$$\begin{cases} \dot{G}(t) = -[SG + \rho(G)X(t)] \cdot G(t) + SG \cdot G_b + Ra(t)/V_G \\ \dot{X}(t) = -p_2 \cdot [X(t) - SI \cdot (I_p(t) - I_{pb})] \\ \dot{IG}(t) = -\frac{1}{\alpha} (IG(t) - G(t)) \end{cases} \quad (9)$$

where G (mg/dl) is the plasma glucose concentration, X (min^{-1}) is the insulin action on glucose disposal and production; SG (min^{-1}) is the glucose effectiveness describing glucose ability, per se, to promote glucose disposal and inhibit glucose production; G_b (mg/dl) is the basal glucose concentration in the plasma; V_G (dl/kg) is the volume of glucose distribution; p_2 (min^{-1}) is the rate constant of insulin action dynamics; SI ($\text{ml}/\mu\text{U} \cdot \text{min}$) is the insulin sensitivity; I_{pb} (mU/l) is the basal insulin concentration in the plasma; IG (mg/dl) is the interstitial glucose concentration; α (min) is the delay between the plasmatic and interstitial glucose concentration compartments; and $\rho(G)$ is a function, introduced by [52], that allows to better describe the insulin action in the hypoglycemic range by increasing its impact when glucose concentration decreases below a certain level. Furthermore, to account for patient-specific intraday insulin sensitivity variability, [53], the parameter SI is considered time-varying over the day:

$$SI = \begin{cases} SI_B & \text{if } 4 \text{ AM} < t \leq 11 \text{ AM} \\ SI_L & \text{if } 11 \text{ AM} < t \leq 5 \text{ PM} \\ SI_D & \text{otherwise} \end{cases} \quad (10)$$

A priori information on parameter distributions has been obtained from the literature, [54]. Particularly, V_G has been fixed to population value, i.e. 1.45 dl/kg. Unknown model parameters of glucose-insulin subsystem, identified via the MCMC procedure presented in Section II.B, are $\theta_{glu} = [SG, SI_B, SI_L, SI_D, G_b, p_2]$.

APPENDIX II

OFFLINE BAYESIAN ESTIMATION: IMPLEMENTATIVE DETAILS

We partitioned θ_{phy} into five sets $\theta_{phy} := [\theta_1, \theta_2, \theta_3, \theta_4, \theta_5]$, namely $\theta_1 := [SG, SI_B]$, $\theta_2 := [SI_L, G_b]$, $\theta_3 := [SI_D]$, $\theta_4 := [p_2, k_{a2}, k_d]$, $\theta_5 := [k_{empt}, k_{abs}]$.

This partitioning scheme has been chosen since it improves MC mixing and allows to break the correlation between SI and p_2 , known to be critical from the literature, [55]. An iteration i of the algorithm consists of five steps $p = 1, \dots, 5$ and each step updates the p -th partition of θ_{phy} , θ_p , by approval/rejection of a sample ϕ_p extracted from the proposal density function $q_p(\cdot)$. Specifically, as prescribed by the SCM procedure, approval occurs with probability α

$$\alpha = \min\left(1, \frac{\pi(\phi_p|\theta_{i,-p})q_p(\theta_{i-1,p}|\phi_p, \theta_{i,-p})}{\pi(\theta_{i-1,p}|\theta_{i,-p})q_p(\phi_p|\theta_{i-1,p}, \theta_{i,-p})}\right)$$

with $\pi(\theta_p|\theta_{i,-p})$ proportional to the posterior of θ_p given that the other components θ_{-p} assume the value $\theta_{-p} = \theta_{i,-p}$:

$$\pi(\theta_p|\theta_{i,-p}) = p_{Y|U}(Y|\theta_p, \theta_{i,-p}, U)p_\theta(\theta_p|\theta_{i,-p}, U)$$

where $\theta_{i,-p}$ comprises all the other components of θ_{phy} except for θ_p , and $p_\theta(\theta_p|\theta_{i,-p}, U)$ is the prior probability distribution of θ_p given $\theta_{i,-p}$. Precisely, $\theta_{i,-p}$ contains the most updated version of each component as available at the current stage of the algorithm: $\theta_{i,-p} = [\theta_{i,1}, \dots, \theta_{i,p-1}, \theta_{i-1,p+1}, \theta_{i-1,5}]$. Components up to $p-1$ have already been updated when processing the p -th components at iteration i , while other components, from $p+1$ to 5, have not been updated yet, so their value computed in the previous iteration $i-1$ is used.

For what it concerns the proposal distribution, we used a Gaussian centered in the value assumed by θ_p in the previous chain iteration

$$q_p(\cdot) = N(\theta_{i-1,p}, \Sigma_p)$$

where Σ_p is a tuning parameter that regulates the acceptance rate of the chain. We set Σ_p to a diagonal matrix whose components are an estimate of the conditional standard deviation of each element of partition p , $sd(\theta_{phy_p}|Y, U)$, multiplied by a scaling factor $2.4/\sqrt{d}$, where d is the number of parameters to be estimated in p -th partition, as suggested in [56]. This estimates is computed by running two exploratory MCMCs for $nIter = 600$ iterations and updated every 1500 iterations of the algorithm, thus implementing an adaptive SCM.

Finally, the convergence of the MCMC has been verified through the well-known Raftery-Lewis criterion, [29], which provides the number of iterations necessary to ensure the Markov Chain to represent the target posterior distribution.

The Adaptive Single Component Metropolis Hasting is summarized in Algorithm 1.

Algorithm 1: Adaptive Single Component Metropolis Hastings

```

i ← 0;
initialize  $\theta_{phy_0}, nIter$ ;
repeat
  for  $p \leftarrow 1$  to 5 do
    set  $\theta_{phy_{i,-p}} =$ 
       $[\theta_{phy_{i,1}}, \dots, \theta_{phy_{i,p-1}}, \theta_{phy_{i-1,p+1}}, \theta_{phy_{i-1,5}}]$ ;
    sample  $\phi_p \sim q_p(\cdot)$ ;
    set
       $\alpha = \min\left(1, \frac{\pi(\phi_p|\theta_{i,-p})q_p(\theta_{i-1,p}|\phi_p, \theta_{i,-p})}{\pi(\theta_{i-1,p}|\theta_{i,-p})q_p(\phi_p|\theta_{i-1,p}, \theta_{i,-p})}\right)$ 
    sample  $U \sim \text{Uniform}(0,1)$ ;
    if  $U \leq \alpha$  then
      | set  $\theta_{phy_{i,p}} = \phi_p$ ;
    else
      | set  $\theta_{phy_{i,p}} = \theta_{phy_{i-1,p}}$ ;
    end
  end
   $i \leftarrow i + 1$ 
until  $n < nIter$ ;
```

APPENDIX III

REAL-TIME PREDICTION THROUGH PARTICLE FILTER: IMPLEMENTATIVE DETAILS

In the following, we present the numerical scheme implemented by PF to perform the one step-ahead prediction, measurement update, and multiple step-ahead prediction.

One step-ahead prediction step. Recalling that, at time t_{k-1} , $p(x(t_{k-1})|y(t_{1:k-1}), u(t_{1:k-1}))$ is available in a sampled form defined by set of P particles $\{x^p(t_{k-1})\}_{p=1}^P$ with associated weights $\{w(t_{k-1})^p\}_{p=1}^P$, $\sum_p w(t_{k-1})^p = 1$ such that

$$p(x(t_{k-1})|y(t_{1:k-1}), u(t_{1:k-1})) \approx \sum_{p=1}^P w^p(t_{k-1}) \delta(x(t_{k-1}) - x^p(t_{k-1})),$$

PF performs the one step-ahead prediction step by drawing a new set of particles $\{x^p(t_k)\}_{p=1}^P$ from $p(x(t_k)|x(t_{k-1}))$:

$$x^p(t_k) \sim p(x^p(t_k)|x^p(t_{k-1})). \quad (11)$$

This probability is specified by (1):

$$p(x^p(t_k)|x^p(t_{k-1})) = N(f(x^p(t_{k-1}), \mathbf{u}, t_{k-1}, \theta), \Sigma_v).$$

In view of this, to draw the new set of particles it is sufficient to let each particle $x^p(t_{k-1})$ evolve according to model (1), and corrupt it with a realization of the noise v .

Measurement update step. The PF algorithm sets the weight $w^{*p}(t_k)$ of each p -th particle $x^p(t_k)$ to the likelihood function evaluated on $x^p(t_k)$

$$w^{*p}(t_k) = p(y(t_k)|x^p(t_k), u(t_{1:k})). \quad (12)$$

In particular, the statistics of the stochastic modelling error e , $p(y(t_k)|x^p(t_k), u(t_{1:k}))$ is defined as:

$$p(y(t_k)|x(t_k), u(t_{1:k})) = N(y(t_k) - y^p(t_k), SD_\epsilon). \quad (13)$$

where $y^p(t_k)$ is obtained using (1) and SD_ϵ is the constant standard deviation of the error.

Weights are then normalized such that $\sum_p w^{*p}(t_k) = 1$.

This provides a sampled form representation of the posterior density

$$p(x(t_k)|y(t_{1:k-1}), u(t_{1:k})) \approx \sum_{p=1}^P w^{*p}(t_k) \delta(x(t_k) - x^p(t_k)).$$

Resampling step. To improve the accuracy of PF, the measurement update step is completed by updating the set of particles. Specifically, $\{x^p(t_k)\}_{p=1}^P$ are substituted with a new set of P particles, $\{x^{*p}(t_k)\}_{p=1}^P$ generated from the sampled representation of $p(x(t_k)|y(t_{1:k-1}), u(t_{1:k}))$ such that $\Pr(x^{*p}(t_k) = x^p(t_k)) = w^{*p}(t_k)$. This step is performed by a well-established resampling algorithm [57].

As a result, all new particles $\{x^p(t_k)\}_{p=1}^P$ are associated to the same weight $w^{*p}(t_k) = 1/P$, thus the approximation of $p(x(t_k)|y(t_{1:k-1}), u(t_{1:k}))$ simplifies to

$$p(x(t_k)|y(t_{1:k-1}), u(t_{1:k})) \approx \frac{1}{P} \sum_{p=1}^P \delta(x(t_k) - x^p(t_k)).$$

Multiple steps-ahead prediction. Multiple steps ahead predictions can be obtained as follows. First, the probabilities $p(x(t_{k+i})|y(t_{1:k}), u(t_{1:k+i}))$, $\forall i = 1, \dots, PH$ are obtained in sampled form starting from $p(x(t_k)|y(t_{1:k}), u(t_{1:k}))$ and propagating the P particles $\{x^p(t_k)\}_{p=1}^P$ i steps ahead as we did in the one step-ahead prediction step. Then, the set of particles $\{y^p(t_{k+i}|t_k)\}_{p=1}^P$ is computed and used to obtain $p(y(t_{k+i})|y(t_{1:k}), u(t_{1:k+i}))$, $\forall i = 1, \dots, PH$ in sampled form.

Finally, a point estimate of glucose i steps-ahead, $y(t_{k+i})$, is obtained as the average computed over the sampled form of $p(y(t_{k+i})|y(t_{1:k}), u(t_{1:k+i}))$, i.e.:

$$\hat{y}(t_{k+i}|t_k) = \frac{1}{P} \sum_{p=1}^P y^p(t_{k+i}|t_k), \forall i = 1, \dots, PH.$$

The implemented PF is summarized in Algorithm 2.

REFERENCES

- [1] American Diabetes Association, "Diagnosis and Classification of Diabetes Mellitus," *Diabetes Care*, vol. 33, no. Suppl 1, pp. S62–S69, 2010.
- [2] D. M. Nathan, "The Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Study at 30 Years: Overview," *Diabetes Care*, vol. 37, no. 1, pp. 9–16, 2014.
- [3] P. E. Cryer, S. N. Davis, and H. Shamon, "Hypoglycemia in Diabetes," *Diabetes Care*, vol. 26, no. 6, pp. 1902–1912, 2003.
- [4] S. F. Clarke and J. R. Foster, "A History of Blood Glucose Meters and Their Role in Self-Monitoring of Diabetes Mellitus," *Br. J. Biomed. Sci.*, vol. 69, no. 2, pp. 83–93, 2012.
- [5] G. Cappon, G. Acciaroli, M. Vettoretti, A. Facchinetti, and G. Sparacino, "Wearable Continuous Glucose Monitoring Sensors: A Revolution in Diabetes Treatment," *Electronics (Switzerland)*, vol. 6, 09 2017.
- [6] M. Vettoretti, G. Cappon, G. Acciaroli, A. Facchinetti, and G. Sparacino, "Continuous Glucose Monitoring: Current Use in Diabetes Management and Possible Future Applications," *J. Diabetes Sci. Technol.*, vol. 12, no. 5, pp. 1064–1071, 2018.
- [7] D. De Salvo and B. Buckingham, "Continuous glucose monitoring: current use and future directions," *Curr. Diab. Rep.*, vol. 13, no. 5, pp. 657–62, Oct 2013.
- [8] American Diabetes Association, "Standards of Medical Care in Diabetes," *Diabetes Care*, vol. 27, no. Suppl 1, pp. S15–S35, 2004.

Algorithm 2: Particle Filter

```

k ← 1;
xt0 = N(xss, Σv);
repeat
  Step 1: One step-ahead prediction;
  for p ← 1 to P do
    sample xp(tk) from
      N(f(xp(tk-1), u, tk-1, θ), Σv);
  end
  Step 2: Measurement update;
  for p ← 1 to P do
    compute w*p(tk) = N(y(tk) - yp(tk), SDε);
  end
  normalize w*p(tk) = w*p(tk) / Σp w*p(tk);
  Step 3: Resampling;
  sample {x*p(tk)}p=1P s.t. Pr(x*p(tk) = xp(tk) =
    w*p(tk);
  set {xp(tk)}p=1P = {x*p(tk)}p=1P;
  Step 4: Multiple steps-ahead prediction;
  for p ← 1 to P do
    compute yp(tk+i|tk), ∀ i = 1, ..., PH;
  end
  compute
    ŷ(tk+i|tk) = 1/P Σp=1P yp(tk+i|tk), ∀ i = 1, ..., PH;
  k ← k + 1
until k ≤ D;

```

- [9] P. J. O'Connor, J. M. Sperl-Hillen, C. J. Fazio, B. M. Averbeck, B. H. Rank, and K. L. Margolis, "Outpatient diabetes clinical decision support: current status and future directions," *Diabet. Med.*, vol. 33, no. 6, pp. 734–741, 2016.
- [10] P. Jia, P. Zhao, J. Chen, and M. Zhang, "Evaluation of clinical decision support systems for diabetes care: An overview of current evidence," *J. Eval. Clin. Pract.*, vol. 25, no. 1, pp. 66–77, 2018.
- [11] S. Oviedo, J. Vehí, R. Calm, and A. J., "A review of personalized blood glucose prediction strategies for T1DM patients," *Int. J. Numer. Method. Biomed. Eng.*, vol. 33, no. 6, p. e2833, 2017.
- [12] A. Z. Woldaregay, E. Årsand, S. Walderhaug, D. Albers, L. Mamykina, T. Botsis, and G. Hartvigsen, "Data-driven modeling and prediction of blood glucose dynamics: Machine learning applications in type 1 diabetes," *Artificial Intelligence in Medicine*, vol. 98, pp. 109–134, 2019.
- [13] G. Sparacino, F. Zanderigo, S. Corazza, A. Maran, A. Facchinetti, and C. Cobelli, "Glucose concentration can be predicted ahead in time from continuous glucose monitoring sensor time-series," *IEEE Transactions on biomedical engineering*, vol. 54, no. 5, pp. 931–937, 2007.
- [14] M. Frandes, B. Timar, R. Timar, and D. Lungeanu, "Chaotic time series prediction for glucose dynamics in type 1 diabetes mellitus using regime-switching models," *Scientific reports*, vol. 7, no. 1, pp. 1–10, 2017.
- [15] J. Yang, L. Li, Y. Shi, and X. Xie, "An arima model with adaptive orders for predicting blood glucose concentrations and hypoglycemia," *IEEE journal of biomedical and health informatics*, vol. 23, no. 3, pp. 1251–1260, 2018.
- [16] X. Yu, M. Rashid, J. Feng, N. Hobbs, I. Hajizadeh, S. Samadi, M. Sevil, C. Lazar, Z. Maloney, E. Littlejohn *et al.*, "Online glucose prediction using computationally efficient sparse kernel filtering algorithms in type-1 diabetes," *IEEE Transactions on Control Systems Technology*, vol. 28, no. 1, pp. 3–15, 2018.
- [17] S. Faccioli, A. Facchinetti, G. Sparacino, G. Pilonetto, and S. Del Favero, "Linear model identification for personalized prediction and control in diabetes," *IEEE transactions on bio-medical engineering*, vol. 69, no. 2, pp. 558–568, 2022.
- [18] E. I. Georga, V. C. Protopappas, D. Ardigo, D. Polyzos, and D. I. Fotiadis, "A glucose model based on support vector regression for the prediction of hypoglycemic events under free-living conditions," *Diabetes technology & therapeutics*, vol. 15, no. 8, pp. 634–643, 2013.

- [19] C. Pérez-Gandía, A. Facchinetti, G. Sparacino, C. Cobelli, E. Gómez, M. Rigla, A. de Leiva, and M. Hernando, "Artificial neural network algorithm for online glucose prediction from continuous glucose monitoring," *Diabetes technology & therapeutics*, vol. 12, no. 1, pp. 81–88, 2010.
- [20] F. D'Antoni, M. Merone, V. Piemonte, G. Iannello, and P. Soda, "Auto-regressive time delayed jump neural network for blood glucose levels forecasting," *Knowledge-Based Systems*, vol. 203, p. 106134, 2020.
- [21] A. Aliberti, I. Pupillo, S. Terna, E. Macii, S. Di Cataldo, E. Patti, and A. Acquaviva, "A multi-patient data-driven approach to blood glucose prediction," *IEEE Access*, vol. 7, pp. 69 311–69 325, 2019.
- [22] J. Daniels, P. Herrero, and P. Georgiou, "A multitask learning approach to personalized blood glucose prediction," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 1, pp. 436–445, 2021.
- [23] T. Zhu, L. Kuang, J. Daniels, P. Herrero, K. Li, and P. Georgiou, "Iomt-enabled real-time blood glucose prediction with deep learning and edge computing," *IEEE Internet of Things Journal*, 2022.
- [24] R. N. Bergman, L. S. Phillips, and C. C., "Physiologic evaluation of factors controlling glucose tolerance in man: measurement of insulin sensitivity and beta-cell glucose sensitivity from the response to intravenous glucose," *J. Clin. Invest.*, vol. 68, no. 6, pp. 1456–1467, 1981.
- [25] R. Hovorka, L. J. Canonico, V. Chassin, U. Haueter, M. Massi-Benedetti, M. Orsini Federici, T. T. Pieber, H. C. Schaller, L. Schaupp, T. Vering, and M. E. Wilinska, "Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes," *Physiol. Meas.*, vol. 25, no. 4, pp. 905–920, 2004.
- [26] M. E. Wilinska, L. J. Chassin, C. L. Acerini, J. M. Allen, D. B. Dunger, and R. Hovorka, "Simulation environment to evaluate closed-loop insulin delivery systems in type 1 diabetes," *J. Diabetes Sci. Technol.*, vol. 3, no. 5, pp. 1047–1057, 2010.
- [27] S. S. Kanderian, S. Weinzimer, G. Voskanyan, and G. M. Steil, "Identification of intraday metabolic profiles during closed-loop glucose control in individuals with type 1 diabetes," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 12, pp. 3524–3533, 2013.
- [28] C. Dalla Man, R. A. Rizza, and C. Cobelli, "Meal simulation model of the glucose-insulin system," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 10, pp. 1740–1749, 2007.
- [29] W. R. Gilks, S. Richardson, and D. Spiegelhalter, "Markov chain monte carlo in practice," *1st edition: Chapman and Hall*, 1996.
- [30] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *IEEE T. Signal Proces.*, vol. 50, no. 2, pp. 174–188, 2002.
- [31] C. Cappon, A. Facchinetti, G. Sparacino, and S. Del Favero, "A Bayesian Framework to Identify Type 1 Diabetes Physiological Models Using Easily Accessible Patient Data," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Jul 2019, pp. 6914–6917.
- [32] R. Visentin, E. Campos-Náñez, M. Schiavon, D. Lv, M. Vettoretti, M. Breton, B. P. Kovatchev, C. Dalla Man, and C. Cobelli, "The uva/padova type 1 diabetes simulator goes from single meal to single day," *J. Diabetes Sci. Technol.*, vol. 12, no. 2, pp. 273–281, 2018.
- [33] J. Xie and Q. Wang, "Benchmarking machine learning algorithms on blood glucose prediction for type 1 diabetes in comparison with classical time-series models," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 11, pp. 3101–3124, 2020.
- [34] G. Pillonetto, T. Chen, A. Chiuso, G. De Nicolao, and L. Ljung, *Regularized System Identification: Learning Dynamic Models from Data*. Springer Nature, 2022.
- [35] D. Ljung, "System identification: Theory for the user," *2nd edition. Upper Saddle River, NJ: Prentice-Hall*, 1999.
- [36] T. Zhu, K. Li, P. Herrero, and P. Georgiou, "Deep learning for diabetes: a systematic review," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 7, pp. 2744–2757, 2020.
- [37] K. Li, J. Daniels, C. Liu, P. Herrero, and P. Georgiou, "Convolutional recurrent neural networks for glucose prediction," *IEEE journal of biomedical and health informatics*, vol. 24, no. 2, pp. 603–613, 2019.
- [38] J. T. Connor, R. D. Martin, and L. E. Atlas, "Recurrent neural networks and robust time series prediction," *IEEE transactions on neural networks*, vol. 5, no. 2, pp. 240–254, 1994.
- [39] K. Li, C. Liu, T. Zhu, P. Herrero, and P. Georgiou, "Glunet: A deep learning framework for accurate glucose forecasting," *IEEE journal of biomedical and health informatics*, vol. 24, no. 2, pp. 414–423, 2019.
- [40] M. Zhang, K. B. Flores, and H. T. Tran, "Deep learning and regression approaches to forecasting blood glucose levels for type 1 diabetes," *Biomedical Signal Processing and Control*, vol. 69, p. 102923, 2021.
- [41] D. A. Finan, F. J. Doyle III, C. C. Palerm, W. C. Bevier, H. C. Zisser, L. Jovanović, and D. E. Seborg, "Experimental evaluation of a recursive model identification technique for type 1 diabetes," *Journal of diabetes science and technology*, vol. 3, no. 5, pp. 1192–1202, 2009.
- [42] C. Marling and R. Bunescu. (2020) The ohiot1dm dataset for blood glucose level prediction: Update 2020. [Online]. Available: <http://smarthealth.cs.ohio.edu/bgplp/OhioT1DM-dataset-paper.pdf>
- [43] R. Visentin, C. Dalla Man, and C. Cobelli, "One-day bayesian cloning of type 1 diabetes subjects: Toward a single-day uva/padova type 1 diabetes simulator," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 11, pp. 2416–2424, 2016.
- [44] A. Haidar, M. E. Wilinska, J. A. Graveston, and R. Hovorka, "Stochastic virtual population of subjects with type 1 diabetes for the assessment of closed-loop glucose controllers," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 12, pp. 3524–3533, 2013.
- [45] Z. Mahmoudi, S. L. Wendt, D. Boiroux, M. Hagdrup, K. Norgaard, N. K. Poulsen, H. Madsen, and J. B. Jorgensen, "Comparison of three nonlinear filters for fault detection in continuous glucose monitors," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2016, pp. 3507–3510.
- [46] S. W. Lee, M. Cao, S. Sajid, M. Hayes, L. Choi, C. Rother, and R. De León, "The dual-wave bolus feature in continuous subcutaneous insulin infusion pumps controls prolonged post-prandial hyperglycaemia better than standard bolus in type 1 diabetes," *Diabetes, nutrition & metabolism*, vol. 17, no. 4, pp. 211–216, 2004.
- [47] J. Kropff, D. Bruttomesso, W. Doll, A. Farret, S. Galasso, Y. M. Luijff, J. K. Mader, J. Place, F. Boscarì, T. Pieber *et al.*, "Accuracy of two continuous glucose monitoring systems: a head-to-head comparison under clinical research centre and daily life conditions," *Diabetes, Obesity and Metabolism*, vol. 17, no. 4, pp. 343–349, 2015.
- [48] V. Matuleviciene, J. I. Joseph, M. Andelin, I. B. Hirsch, S. Attvall, A. Pivodic, S. Dahlqvist, D. Klonoff, B. Haraldsson, and M. Lind, "A clinical trial of the accuracy and treatment experience of the dexcom g4 sensor (dexcom g4 system) and enlite sensor (guardian real-time system) tested simultaneously in ambulatory patients with type 1 diabetes," *Diabetes technology & therapeutics*, vol. 16, no. 11, pp. 759–767, 2014.
- [49] M. Schiavon, C. Dalla Man, and C. Cobelli, "Modeling subcutaneous absorption of fast-acting insulin in type 1 diabetes," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 2079–2086, 2018.
- [50] C. Dalla Man, M. Camilleri, and C. Cobelli, "A system model of oral glucose absorption: Validation on gold standard data," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 12, pp. 2472–2478, 2006.
- [51] R. N. Bergman, C. R. Bowden, and C. Cobelli, "Quantitative estimation of insulin sensitivity," *Am. J. Physiol.*, vol. 236, no. 6, pp. e667–e677, 1979.
- [52] C. Dalla Man, F. Micheletto, D. Lv, M. Breton, B. Kovatchev, and C. Cobelli, "The uva/padova type 1 diabetes simulator: New features," *J. Diabetes Sci. Technol.*, vol. 8, no. 1, pp. 26–34, 2017.
- [53] R. Visentin, C. Dalla Man, Y. Kudva, A. Basu, and C. Cobelli, "Circadian variability of insulin sensitivity: Physiological input for in silico artificial pancreas," *Diabetes Technol. Ther.*, vol. 17, no. 1, pp. 1–7, 2015.
- [54] C. Dalla Man, A. Caumo, R. Basu, R. Rizza, G. Toffolo, and C. Cobelli, "Minimal model estimation of glucose absorption and insulin sensitivity from oral test: validation with a tracer method," *J. Theor. Med.*, vol. 1, no. 4, pp. 313–323, 1999.
- [55] G. Pillonetto, G. Sparacino, and C. Cobelli, "Numerical non-identifiability regions of the minimal model of glucose kinetics: superiority of bayesian estimation," *Math. Biosci.*, vol. 184, no. 1, pp. 53–67, 2003.
- [56] H. Haario, E. Saksman, and J. Tamminen, "Adaptive proposal distribution for random walk metropolis algorithm," *Computational Statistics*, vol. 14, pp. 375–395, 1999.
- [57] G. Kitagawa, "Monte carlo filter and smoother for non-gaussian nonlinear state space models," *J. Comput. Graph. Statist.*, vol. 5, no. 1, pp. 1–25, 1996.