ReplayBG: A Digital Twin-Based Methodology to Identify a Personalized Model From Type 1 Diabetes Data and Simulate Glucose Concentrations to Assess Alternative Therapies

Giacomo Cappon ⬤, Martina Vettoretti ⬤, Giovanni Sparacino ⬤, Simone Del Favero ⬤,
and Andrea Facchinetti ⬤

Abstract—Objective: Design and assessment of new therapies for type 1 diabetes (T1D) management can be greatly facilitated by in silico simulations. The ReplayBG simulation methodology here proposed allows "replaying" the scenario behind data already collected by simulating the glucose concentration obtained in response to alternative insulin/carbohydrate therapies and evaluate their efficacy leveraging the concept of digital twin. **Methods:** ReplayBG is based on two steps. First, a personalized model of glucose-insulin dynamics is identified using insulin, carbohydrate, and continuous glucose monitoring (CGM) data. Then, this model is used to simulate the glucose concentration that would have been obtained by "replaying" the same portion of data using a different therapy. The validity of the methodology was evaluated on 100 virtual subjects using the UVa/Padova T1D Simulator (T1DS). In particular, the glucose concentration traces simulated by ReplayBG are compared with those provided by T1DS in five different scenarios of insulin and carbohydrate treatment modifications. Furthermore, we compared ReplayBG with a state-of-the-art methodology for the scope. Finally, two case studies using real data are also presented. **Results:** ReplayBG simulates with high accuracy the effect of the considered insulin and carbohydrate treatment alterations, performing significantly better than state-of-art method in almost all considered situations. **Conclusion:** ReplayBG proved to be a reliable and robust tool to retrospectively explore the effect of new treatments for T1D on the glucose dynamics. It is freely available as open source software at https://github.com/gcappon/replay-bg. **Significance:** ReplayBG offers a new approach to preliminary evaluate new therapies for T1D management before clinical trials.

Index Terms—Type 1 diabetes, simulation, method evaluation, digital Twin.

I. INTRODUCTION

T YPE 1 diabetes (T1D) is a chronic autoimmune disease characterized by the destruction of insulin-producing pancreatic beta-cells [1]. As a result, to keep blood glucose (BG) within the safe target range (i.e., BG $\in$ [70, 180] mg/dl), people affected by T1D are required to assume exogenous insulin, to avoid hyperglycemic episodes (i.e., BG $>$180 mg/dl) that can cause serious complications in the long-term. On the other hand, insulin overdosing inevitably leads to dangerous hypoglycemic events (i.e., BG $<$70 mg/dl) that must be promptly tackled by means of rescue carbohydrates intakes [2].

Assessing (possibly new) therapies for the treatment of T1D is a critical task for obvious safety reasons and can also be expensive and time consuming. Indeed, the clinical assessment requires the enrollment of a sufficient number of subjects to prove that the new therapy is actually effective with respect to the state-of-art.

In this context, in silico clinical trials (ISCT)s played a key role in the last decades [3]. ISCTs leverage complex mathematical models of glucose-insulin regulation (see, e.g. [4], [5], [6], [7]) to set up computer simulation environments to preliminary design and evaluate, in vast virtual cohorts representing the T1D population, new methodologies in a safe and cost-effective way. Moreover, ISCTs allow exploring the impact of new treatments by running multiple tests on the same virtual subject maintaining the same identical conditions, which is something impossible to replicate in clinical trials. One great possibility provided by ISCT tools is using them to tune therapies in a specific real individual. This is fostered by the fact that gathering real-life data is becoming easier than in the past, so it would be ideal to exploit such large datasets to further validate or even personalize new therapies.

The aim being exploring this perspective, Patek et al. [8] proposed a novel data-driven, model-based simulation approach which consists of two steps. Firstly, the patient's glucose
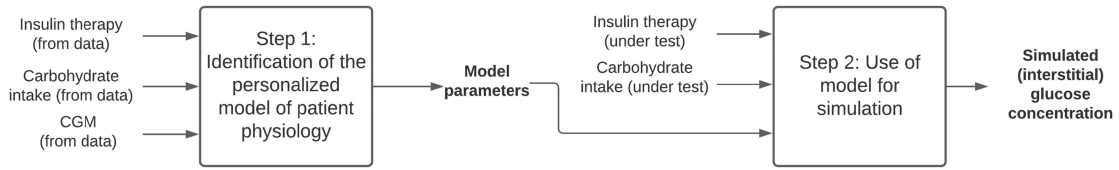
Fig. 1.   Overview of the two steps of ReplayBG. Step 1: model parameters are identified from CHO, insulin and CGM data already collected by the individual. Step 2: the model parameters identified in Step 1 are given in input to model (6) to simulate the (interstitial) glucose concentration that should have been obtained, in the same individual and in the same time window, by adopting an alternative insulin and/or CHO therapy.

concentration measured via continuous glucose monitoring (CGM) devices, and insulin infusion data are used as inputs of a linear model of glucose-insulin regulation of T1D to retrospectively estimate, through deconvolution on a specified time window, a signal, called "net-effect", representing all the unmodeled phenomena affecting glucose dynamics. Secondly, the net effect is used as a forcing input to run the model at hand and test the new therapy under assessment in the same time window with insulin and/or meal inputs modified by the treatment under test. This second step allows to obtain the hypothetical patient glucose concentration that would have been resulted from its adoption.

However, even if the methodology proposed by Patek et al. is a very promising tool to evaluate the effect of alternative insulin therapies on already acquired datasets, its domain of applicability is limited as demonstrated by Vettoretti et al. [9]. As such, recently Hughes et al. [10] proposed and tested an improved version of the methodology, the aim being surpassing such limitations showing better performance and accuracy than [8] in a set of benchmark simulated scenarios.

In the present work, we propose ReplayBG, a digital twin-based approach which aims to reliably simulate glucose concentration traces in response to modifications of the insulin/carbohydrates therapy in real individuals, without the need of estimating a net effect signal. As depicted in Fig. 1, the methodology consists of two sequential steps, described in details in Section II. In the first step, the aim being capturing the peculiar glucose-insulin dynamics observed within the data, a (nonlinear) model of glucose-insulin dynamics is identified on already collected patient data (consisting of insulin infusion, carbohydrate intake and CGM) using a robust and powerful Bayesian technique. Then, the model is used to simulate the (interstitial) glucose concentration trace obtained by using an alternative insulin/carbohydrate therapy in the same individual and in the same identical scenario. In Section III, the methodology is tested following the same approach of [10], where synthetic data are generated through multiple ad hoc ISCTs, used to identify the model of ReplayBG, and then used as "ground-truth" to evaluate if ReplayBG is able to reproduce specific glucose fluctuations resulting from different (here 5) possible scenarios obtained as modification of meal intake and/or insulin administration. In Section III we quantitatively compare ReplayBG with the state-of-the-art net-effect-based methodology of Hughes et al. [10]. Section IV illustrates two case studies in which the use of ReplayBG on real data is demonstrated by evaluating two state-of-the-art algorithms for T1D management. Finally, Section V concludes the article considering the strengths

and limitations of the study and discussing the margins for its further development.

II. THE REPLAYBG METHODOLOGY

A. Step 1: Identification of the Personalized Model of Patient Physiology

1) Model Structure: As previously done by Cappon et al. [11], the model was built starting from the physiological model available in the UVa/Padova T1D Simulator (T1DS) [12]. Moreover, as in [11], in order to permit model identification at the individual level using carbohydrate intake $CHO(t)$, exogenous insulin $I(t)$, and CGM data, the original T1DS model has been simplified reducing the number of parameters to be identified, but paying attention in maintaining its ability to effectively describe glucose-insulin dynamics. Note that CGM, CHO, and insulin data should be collected in parallel. As documented in the following, the model is composed of three main subsystems: subcutaneous insulin absorption; oral glucose absorption; glucose-insulin kinetics.

a) Subcutaneous Insulin Absorption Subsystem The model of the subcutaneous insulin absorption system is a simplified version of the model incorporated in T1DS [13]. The model is composed of three compartments and describes the absorption dynamics of exogenous insulin infusion to the plasma. Exogenous insulin $I(t)$ is infused to the first compartment, which represents insulin in a non-monomeric state. Then, "non-monomeric" insulin diffuses to the second compartment, representing insulin in a monomeric state, and eventually reaches plasma. Model equations are:

$$\begin{cases} \dot{I}_{sc1}(t) = -k_d \cdot I_{sc1}(t) + I(t - \beta)/V_I \\ \dot{I}_{sc2}(t) = k_d \cdot I_{sc1}(t) - k_{a2} \cdot I_{sc2}(t) \\ \dot{I}_p(t) = k_{a2} \cdot I_{sc2} - k_e \cdot I_p(t) \end{cases} \quad (1)$$

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.

b) Oral Glucose Absorption Subsystem The model of the oral glucose absorption system, taken from [14], describes the gastro-intestinal tract as a three-compartment system: the first

two compartments quantify the glucose in the stomach, while the third compartment models the upper small intestine where CHO is absorbed. Model equations are:

$$\begin{cases} \dot{Q}_{sto1}(t) = -k_{empt} \cdot Q_{sto1}(t) + CHO(t) \\ \dot{Q}_{sto2}(t) = k_{empt} \cdot Q_{sto1}(t) - k_{empt} \cdot Q_{sto2}(t) \\ \dot{Q}_{gut}(t) = k_{empt} \cdot Q_{sto2}(t) - k_{abs} \cdot Q_{gut}(t) \end{cases} \quad (2)$$

where Q_{sto1} (mg/kg) and Q_{sto2} (mg/kg) are the amounts of glucose in the stomach in solid and liquid state, respectively; Q_{gut} (mg/kg) is the glucose concentration in the intestine; 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 (2) allows to estimate the rate of glucose appearance in plasma Ra (mg/kg/min) as:

$$Ra(t) = f \cdot k_{abs} \cdot Q_{gut}(t) \quad (3)$$

where f (dimensionless) is the fraction of the intestinal content absorbed in the plasma.

c) Glucose-Insulin Kinetics Subsystem The core model is based on a modified version of the glucose-insulin kinetics minimal model introduced in [15], [16]. The model is composed of three compartments, the first describing the effect of insulin action and oral glucose rate of appearance on plasma glucose concentration, the second quantifies the impact of plasma insulin concentration on insulin action, and the last one represents the transport of glucose from plasma to the interstitium. Model equations are:

$$\begin{cases} \dot{G}(t) = -[SG + \rho(G)X(t)] \cdot G(t) + SG \cdot G_b + Ra(t)/V_G \\ \dot{X}(t) = -p_2 \cdot [X(t) - SI \cdot (I_p(t) - I_{pb})] \\ \dot{IG}(t) = -\frac{1}{\alpha}(IG(t) - G(t)) \end{cases}$$
$$\quad (4)$$

where G (mg/dl) is the plasma glucose concentration, X (min^{-1}) is the insulin action on glucose disposal and production; IG (mg/dl) is the interstitial glucose concentration; SG (min^{-1}) is the glucose effectiveness that describes the glucose ability 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 (ml/μ U·min) is the insulin sensitivity; I_{pb} (mU/l) is the basal insulin concentration in the plasma; α (min) is the delay between plasmatic and IG compartments; and $\rho(G)$, is a deterministic function, introduced by Dalla Man et al. [17], that allows to better represent glucose dynamics in the hypoglycemic range by increasing insulin action when glucose decreases below a certain threshold:

$$\rho(G) = \begin{cases} 1 & \text{if } G \geq G_b \\ 1 + 10 r_1 \{[ln(G)]^{r_2} - [ln(G_b)]^{r_2}\}^2 & \\ & \text{if } G_{th} < G < G_b \\ 1 + 10 r_1 \{[ln(G_{th})]^{r_2} - [ln(G_b)]^{r_2}\}^2 & \text{if } G \leq G_{th} \end{cases}$$
$$\quad (5)$$

where G_{th} is the hypoglycemic threshold (set to 60 mg/dl) and r_1 (dimensionless) and r_2 (dimensionless) are two model

parameters, without direct physiological interpretation, set to 1.44 and 0.81, respectively.

Note that in the model of (4) we are assuming, without any loss of generality, the gain between $G(t)$ and $IG(t)$ to be equal to one [18] [19]. With this assumption, in practice, $IG(t)$ represents a noise-free version of the CGM data, allowing us to fit it versus $CGM(t)$ in the identification step described in the following section.

2) Model Identification

For each model subsystem, we reduced as much as possible the number of parameters to be identified, i.e., retaining only the parameters that have the greatest impact on model output, while setting the other parameters to population values. This selection step facilitates the identifiability of the model maintaining capability of describing the glucose dynamics observed in the data. In particular, we fixed 6 parameters to population values, i.e., $V_I = 0.126$ l/kg, $k_e = 0.127$ min^{-1}, $\beta = 8$ min, $f = 0.9$, $V_G = 1.45$ dl/kg, and $\alpha = 7$ min [13] [14], [15]. The remaining parameters in (1)–(4) must be derived, at the individual level, from the data. Therefore, the vector $\boldsymbol{\theta}$ of the unknown parameters is composed by a total of eight variables, two related to the subcutaneous insulin subsystem, i.e., k_{a2} and k_d, two associated to the oral glucose absorption subsystem, i.e., k_{empt} and k_{abs}, and four related to the glucose-insulin kinetics subsystem, i.e, S_G, S_I, p_2, and G_b. Formally, the parameter estimation problem can be stated as the problem of determining $\boldsymbol{\theta}$ from the equations:

$$\begin{cases} \dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}, t, \boldsymbol{\theta}) \\ y(t) = IG(t) \end{cases} \quad (6)$$

where $\boldsymbol{x}(t)$ is the state vector defined as

$$\boldsymbol{x}(t) := [I_{sc1}, I_{sc2}, I_p, Q_{sto1}, Q_{sto2}, Q_{gut}, G, X, IG]^T;$$

$\boldsymbol{u}(t) := [I(t), CHO(t)]$ is the input vector; $\boldsymbol{f}(\cdot)$ is the state update function combining (1) (2), and (4). $\boldsymbol{f}$ depends on the set of unknown parameters $\boldsymbol{\theta}$.

As documented in the literature, (1), (2), given the data at hand (i.e., CGM, I, and CHO) and (4) are *a priori* non identifiable [13], [14], [16]. As such, the identification of the vector $\boldsymbol{\theta}$ at the individual level from insulin and carbohydrate intake and CGM is not trivial, since the resulting model (6) is not *a priori* identifiable as well.

To mitigate *a priori* non-identifiability, we adopted the same Bayesian approach of Pillonetto et al.,giapiPls, i.e., Markov Chain Monte Carlo (MCMC). In fact, MCMC allows overcoming improper model parameters estimates that are usually achieved with maximum-likelihood-based approaches. In details MCMC, allows to obtain a point estimate of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}$, by performing Monte Carlo integration over a set of N samples $\boldsymbol{\theta}_i, i = 1, \ldots, N$ generated from the posterior distribution

$$p_{\boldsymbol{\theta}|Y,U}(\boldsymbol{\theta}|Y,U) = \frac{p_{Y|\boldsymbol{\theta},U}(Y|\boldsymbol{\theta},U)p_{\boldsymbol{\theta}}(\boldsymbol{\theta})}{\int p_{Y|\boldsymbol{\theta},U}(Y|\boldsymbol{\theta},U)p_{\boldsymbol{\theta}}(\boldsymbol{\theta})d\boldsymbol{\theta}} \quad (7)$$

where $p_{Y|\boldsymbol{\theta},U}(Y|\boldsymbol{\theta},U)$ is the likelihood function, that is, the probability of observing a certain sequence of CGM measurements $Y := \{y(t_k), t_k = k \cdot T_s, k = 1, \dots, D\}$ given the parameter vector $\boldsymbol{\theta}$ and the input $U := \{\boldsymbol{u}(t_k), t_k = k \cdot T_s, k = 1, \dots, D\}$ with D the number of available data points and T_s the sampling period, while $p_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ is the *a priori* information on the distributions of unknown parameters, which was obtained from previous studies [13], [14], [21].

The successful application of MCMC to address the *a priori* non-identifiability of the ReplayBG model was already extensively discussed in [11]. As such, for the sake of the article conciseness, we decided not to report again such an analysis in the present article and we refer [11] for further details. Anyway, the reader can find additional analyses and results in the attached Supplementary Material (Sections 2, 3, and 7), where we provided some insights on the practical model identifiability (confidence intervals of parameter estimates, prior vs. posterior probability density distributions of model parameters in a representative subject, and accuracy in capturing real-world CGM dynamics).

In this work, we implemented an improved variant of the identification approach proposed in [11], in which the Monte Carlo integration step is replaced by a new step where the *a posteriori* distribution $p_{\boldsymbol{\theta}|Y,U}(\boldsymbol{\theta}|Y,U)$ is fit and represented in sampled form.

From the practical point-of-view, $\boldsymbol{\theta}_i$ samples are generated from a Markov Chain whose stationary distribution is exactly the posterior in (7) (target distribution). Such a chain is built in MATLAB (Mathworks Inc., Natick, MA, USA) using an Adaptive Single Component Metropolis-Hastings scheme [22] described in details in the Supplementary Material (Section 4).

The obtained samples $\boldsymbol{\theta}_i$ are then used to fit a multivariate t copula distribution able to represent $p_{\boldsymbol{\theta}|Y,U}(\boldsymbol{\theta}|Y,U)$ and capture the underneath dependence between unknown parameters [23]. Finally, such a copula distribution is used to generate 1000 realizations of $\boldsymbol{\theta}$, i.e., $\hat{\boldsymbol{\theta}}_r, r = 1, \dots, 1000$, representing in sampled form the target distribution $p_{\boldsymbol{\theta}|Y,U}(\boldsymbol{\theta}|Y,U)$. These samples are stored for their later use in the second step of ReplayBG.

2) Remark: The proposed identification scheme of ReplayBG does not estimate the states of model (6) and, more importantly, the corresponding initial conditions. Identifying such initial conditions is crucial to correctly estimating the unknown model parameter vector $\boldsymbol{\theta}$ and, by product, reliably replaying the glucose profile with new, altered, inputs, i.e., step 2 of ReplayBG described below. To circumvent this issue, ReplayBG assumes all model state initial conditions at steady state. This assumption is granted to be valid when the actions of exogenous insulin and carbohydrate intake are "exhausted", i.e., when the starting point of the portion of data used for model identification and replay is reasonably distant from the last meal and insulin boluses. For this reason, as later described in Sections III.A and IV, in this article we designed simulations and selected portions of real data that guarantee to satisfy these assumptions. The potential ReplayBG user should be aware of this aspect and be careful when selecting the portion of data to work with. Furthermore, as discussed in details in the Supplementary Material (Section 5), as a rule of thumb we suggest to use portions of data that span at least 6

hours. This ensures to obtain both reliable parameter estimates and simulation results.

B. Step 2: Use of Model for Simulation

The second step of ReplayBG uses the model identified in Step 1 to predict the hypothetical glucose concentration profile (which corresponds, in the model, to the variable IG, which we remind to be a noise-free version of CGM) that would have been obtained, in the same individual and in the same time window, from the adoption of an alternative therapy (Fig. 1, right). Specifically, for each r-th realization of the model parameters, $\hat{\boldsymbol{\theta}}_r$, a corresponding $IG(t)$ profile, $\hat{IG}_r(t)$, is obtained by simulating the model (6) using as input the carbohydrate intakes and insulin administrations provided by the alternative therapy to be evaluated. As results, a total of 1000 predicted IG profile realizations are obtained. These realizations are ultimately used to infer the median IG profile, $\hat{IG}(t)_{50}$, and the boundaries of the corresponding $25th - 75th$ percentile confidence interval, $\hat{IG}(t)_{25}$, $\hat{IG}(t)_{75}$, of the therapy under assessment.

III. Assessment of ReplayBG From Synthetic Data

ReplayBG evaluation is performed in silico, using as reference the T1DS [12], a software to generate ISCT that incorporates a population of 100 virtual adult subjects, each characterized by different physiological parameters to capture the inter-/intra-subject variability of T1D population. Briefly, for each virtual subject of the T1DS, a baseline scenario has been set up, which allowed generating CGM, CHO, and insulin data to identify the ReplayBG model. This allowed us to assess the ability of ReplayBG to correctly estimate IG in a T1D individual given a set of data. Then, several different modifications of the baseline scenario have been created, simulating alternative therapies. These therapies have been given as input to both the T1DS and ReplayBG, generating two IG traces: the one obtained with the T1DS served as a reference to assess the ability of ReplayBG to simulate it. This procedure, also adopted by [8] and [10], allowed setting a "ground truth" term of comparison used to quantify the ability of ReplayBG to reconstruct the glucose trace resulting from the therapy modification under evaluation.

A. Synthetic Dataset

Synthetic CGM, CHO, and insulin data of 100 virtual adult subjects were generated using T1DS.

In detail, we setup a 12-hour long single-meal ISCT, from 7:00 AM to 7:00 PM where each subject had a meal of 50 g of carbohydrates at 7:00 AM. The corresponding insulin bolus was computed through the standard formula commonly used in clinical practice [24]:

$$B = \frac{CHO}{CR} + \frac{G_C - G_T}{CF} \tag{8}$$

where B (U) is the calculated insulin amount, CHO (g) is the meal carbohydrate amount, G_C (mg/dl) is the current IG concentration, measured by CGM, G_T (mg/dl) is the target BG concentration, and CR (g/U) and CF (mg/dl/U) are the

carbohydrate-to-insulin ratio and the correction factor, two patient-specific therapy parameters which quantify the patient sensitivity to insulin. Throughout ISCT, patients were treated with the optimal basal insulin infusion rate provided by T1DS. Then, for each subject, we identified the ReplayBG model (6) as described in Section II-A2. Finally, as remarked in Section II-A2, it is important to use data that ensure that the model of ReplayBG is not affected by the choice of the starting point. This is granted since data are generated from the physiological model of T1DS whose states all start from the steady state.

B. The Five Simulated Scenarios

To assess the efficacy of ReplayBG, we simulated some alternative therapies in the 100 virtual patients of Section III-A (used for ReplayBG model identification), and then applied ReplayBG step 2 to evaluate the ability of ReplayBG to replicate them. In particular, first, synthetic data were generated through T1DS using the simulation setup used for the identification of ReplayBG but with different insulin and meal input profiles. Then, we fed ReplayBG with the same identical alternative therapies and compared the median IG profile predicted by ReplayBG with the "ground-truth" provided by T1DS. Specifically, we simulated a total of five scenarios, each of them representing a possible reliable therapy adjustment whose effect on glucose could be of clinical interest to be evaluated:

- **Scenario 1: Meal insulin bolus dose modulation**. This scenario simulates the situation in which we would be interested in evaluating, for example, the effect of a different insulin dosing formula. To simulate such a scenario, the original dose of insulin bolus, calculated using (8), was modulated by a multiplying factor ranging from -50% to 50% with a step of 10%. The original optimal basal insulin rate and meal carbohydrate amount were not altered;
- **Scenario 2: Meal CHO amount modulation**. Here, the aim is to simulate a scenario in which the amount of CHO at meal time is different from the nominal one used to calculated the insulin bolus. For example, this scenario can be useful if we want to test the impact of an error in CHO counting. The original meal carbohydrate amount was modulated by a multiplying factor ranging from -100%, which corresponds to the complete elimination of the meal intake, to $+100\%$, which, on the other hand, doubled its amount, with a step of 20%. Insulin inputs were not altered;
- **Scenario 3: Basal insulin infusion rate modulation**. This third simulation scenario is similar to scenario 1, except that here we want to test a new therapy in which only the basal insulin infusion rates are modified. To simulate this, the optimal basal infusion rate was modulated by a -50% to $+50\%$ factor, with an increasing step of 10%. The original meal insulin bolus dose and meal intake were not altered;
- **Scenario 4: Snack addition**. This scenario was devised to test what would have happened if the subject had either applied an hypotreatment strategy, e.g., in presence of hypoglycemia, or had taken a snack (without the corresponding insulin bolus). To realize this, five hours after

the start of the ISCT, i.e., at 5:00 PM, a snack of 5 g to 25 g of carbohydrates, with an increasing step of 5 g, was added to the carbohydrate input intake regime. The original insulin administrations were unaltered.
- **Scenario 5: Bolus addition**. Finally, the fifth scenario represents the case in which we are interested in testing new methodologies that advise the assumption of insulin boluses. Practically, such a scenario is simulated by adding in the original insulin administration regimen, an insulin bolus of $10/CF$ U to $50/CF$ U, with an increasing step of $10/CF$, was administered at 2:00 PM. In particular, rather than using a fixed amount of insulin for each subject, we decided to modulate this amount by the correction factor CF in order to account for the specific subjects' insulin sensitivity. Original meal intake was not altered.

C. The Chosen Performance Metrics

For each scenario 1–5, we evaluated the performance of ReplayBG by considering the average Mean Absolute Relative Difference (MARD) (%) between the "true" IG profile provided by T1DS, $IG(t_k)_p$, and the median IG profile simulated through ReplayBG, $\hat{IG}(t_k)_{50_p}$, computed over the virtual cohort:

$$MARD = \frac{1}{100 \cdot D} \sum_p^{100} \sum_k^D \left| \frac{IG(t_k)_p - \hat{IG}(t_k)_{50_p}}{IG(t_k)_p} \right| \quad (9)$$

where D is the number of data points available. To evaluate whether the ReplayBG methodology is capable of reliably simulating variations in the therapy of a given individual, we considered a MARD threshold of 10%. Indeed, MARD $<10\%$ can be considered a "good" glucose match since it approximates the accuracy found in currently available CGM devices. We also considered, as additional performance metric, the root mean square error (RMSE) (mg/dl) between $IG(t_k)_p$ and $\hat{IG}(t_k)_{50_p}$:

$$RMSE = \frac{1}{100} \sum_p^{100} \sqrt{\frac{1}{D} \sum_k^D (IG(t_k)_p - \hat{IG}(t_k)_{50_p})^2} \quad (10)$$

Another evaluation tool that is useful to assess the clinical acceptability of the simulations performed by ReplayBG is the Clarke Error Grid Analysis (CEGA) [25]. However, for the sake of article conciseness, we decided to report CEGA results the Supplementary Material (Section 1).

D. Comparing ReplayBG With State-of-The-Art Techniques

To further support the novelty and the potential of ReplayBG, we compared it against the methodology of Hughes et al. [10]. Of note, we did not compare ReplayBG also vs. the approach of Patek et al. [8], since [10] is an updated, improved version of [8], thus representing the most challenging approach to compete with. Hereafter, [10] will be referred to as "Net-Effect" for the sake of simplicity.

For the convenience of the reader, here we condensed the principles of the Net-Effect methodology. Briefly, it can be summarized in three sequential steps. First, least-square fitting is
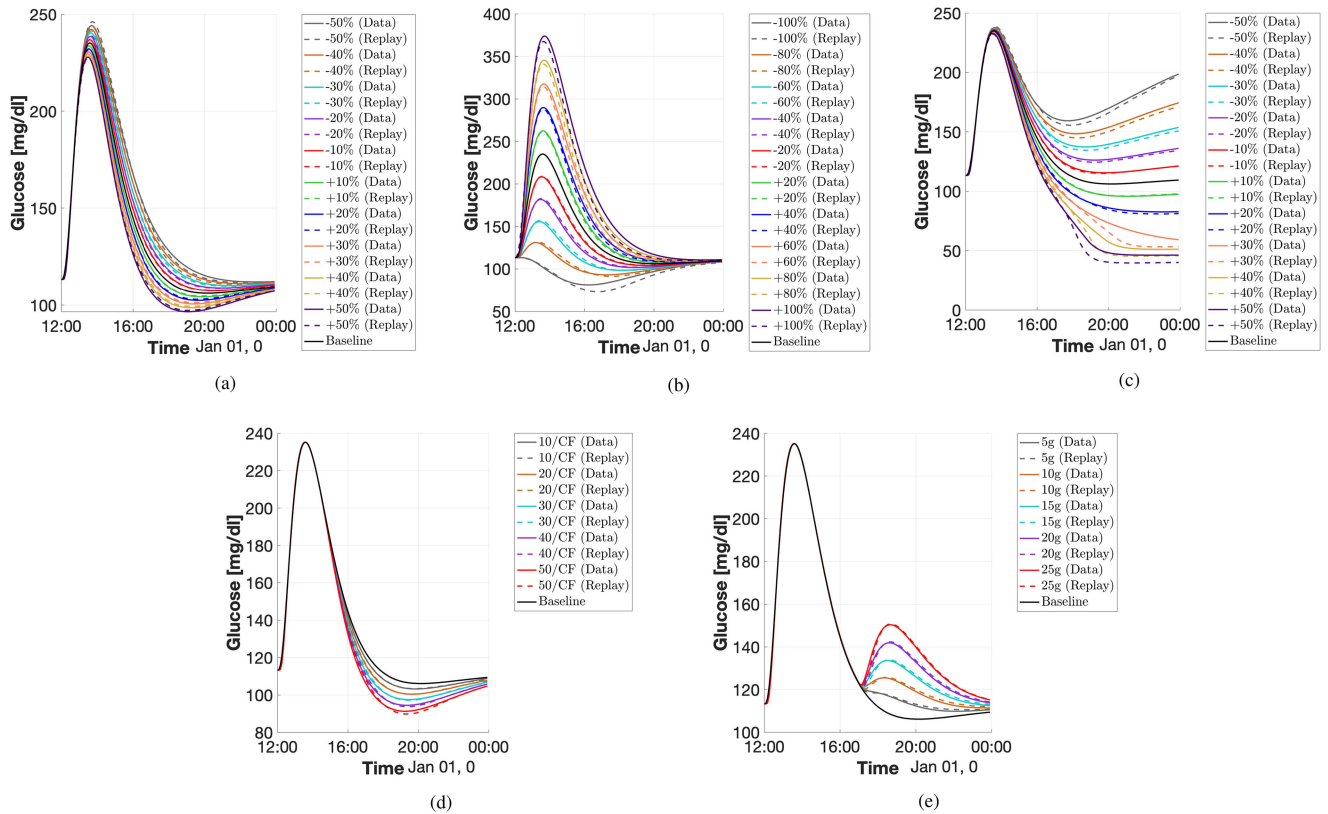
Fig. 2.   Target simulated IG profiles (continuous lines) compared against the respective IG traces obtained from step 2 of ReplayBG (dashed lines) when the original meal and insulin input data are altered as defined in scenario 1 (a), scenario 2 (b), scenario 3 (c), scenario 4 (d), and scenario 5 (e) in a representative subject. Legends indicate the input alteration applied in the specific scenario corresponding to the respective IG profile.

used to identify a subset of parameters of a linear time-invariant (LTI) model of glucose-insulin dynamics to accommodate, thus personalize, the model itself to the patient-specific physiological intravariability. Secondly, regularized deconvolution is used to extract a residual signal representative of all unmodeled phenomena, the so-called "net-effect" signal. Finally, the net-effect signal is fed back into the LTI model, together with the known insulin and meal records, to reproduce the observed original glucose trace. In this setup, similarly to ReplayBG, if the original meal and/or insulin inputs are altered, an hypothetical trace of the expected glucose time-course under the modified inputs can be produced by simulating the LTI using as inputs the net-effect signal and the new meal and insulin records. Further details on Net-Effect can be found in [10].

Net-Effect has been reimplemented as described in [10] and quantitatively evaluated adopting the same procedure used for ReplayBG, i.e., on the same data, the same five scenarios, and the same metrics described in Sections III-A, III-B, and III-C, respectively. Then, for each of the five scenarios, we compared the MARD and RMSE distributions obtained with ReplayBG and Net-Effect both numerically and statistically. Specifically, to assess whether the difference between methods is statistically significant, for each scenario and for each considered input modulation/alteration, we performed the two-tailed Wilcoxon signed rank test with a 5% significance level corrected by the number of comparisons made in each scenario using the Bonferroni approach to account for possible repetition biases.

The Wilcoxon signed rank test has been chosen since the MARD distributions were not Gaussian according to the Lilliefors test.

E. Results

For each of the five scenarios, we graphically discussed the application and the performance of ReplayBG on a representative virtual subject. This will be done resorting to Fig. 2, which is composed by five panels, each of them reporting, for each scenario, the original simulated data vs the IG traces obtained with ReplayBG for the specific meal and insulin input profiles. Then, the results, obtained in the entire virtual cohort, are discussed and reported in Tables I-V as median and 5th–95th percentile range of MARD, and RMSE. Finally, ReplayBG is compared against Net-Effect. This will be done through Tables I-V, which report also the median and 5th–95th percentile range of MARD, and RMSE obtained with Net-Effect, and Fig. 3, which shows and compares, through five panels, the MARD distributions of ReplayBG (in red) vs. Net-Effect (in blue) for each scenario and considered meal/insulin input alteration. RMSE distributions are not visualized in dedicated plots since they are qualitatively similar.

Fig. 2(a) compares, for a representative subject, the reference IG data obtained using T1DS (continuous lines) when the ten different modulations considered in scenario 1 are applied to the original meal insulin bolus dose with the respective IG profiles obtained by step 2 of ReplayBG (dashed lines). Results show
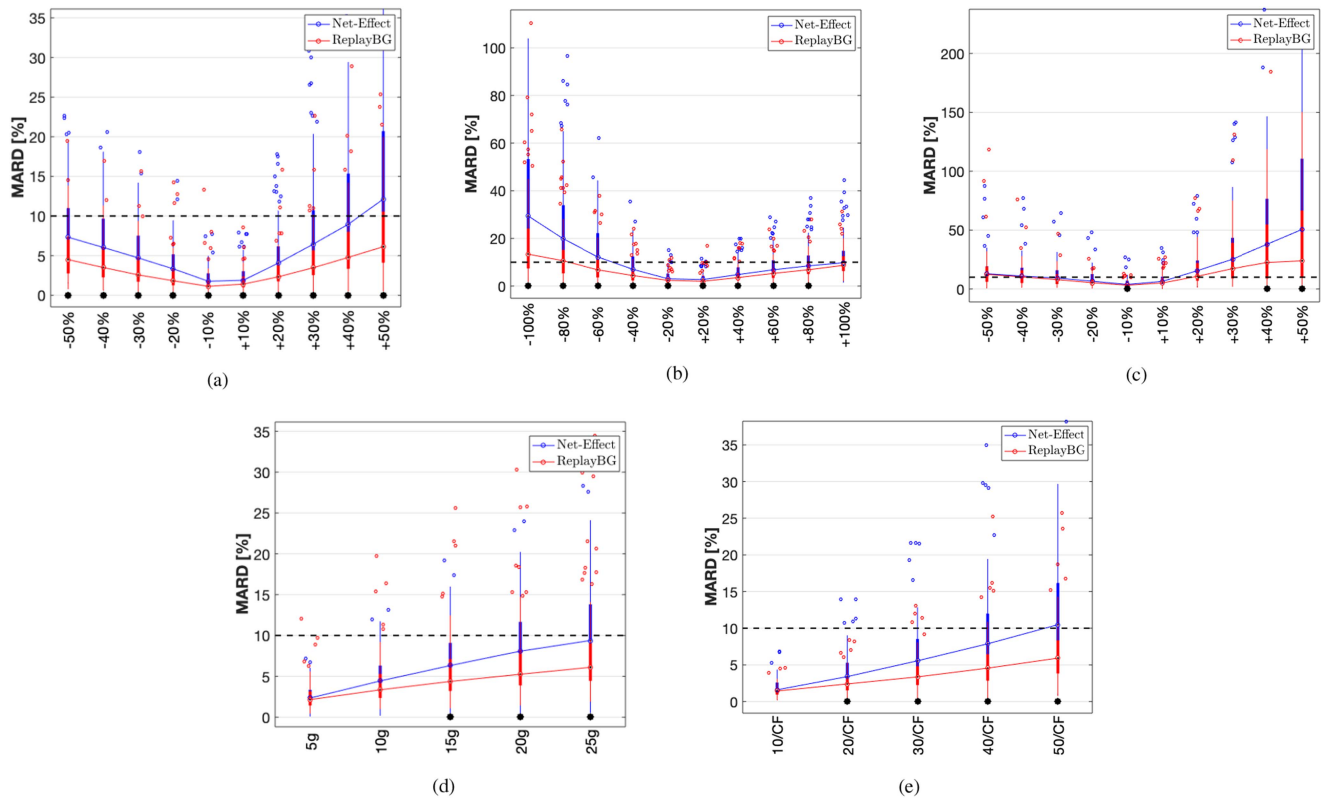
Fig. 3. MARD boxplot distributions obtained using ReplayBG (in red) and the Net-Effect methodology (in blue) when the original meal carbohydrate intake is altered as defined in scenario 1 (a), scenario 2 (b), scenario 3 (c), scenario 4 (d), and scenario 5 (e) in a representative subject. Legends indicate the input alteration applied in the specific scenario corresponding to the respective IG profile. Dashed black line indicates the MARD = 10% threshold.

TABLE I
SCENARIO 1: MEDIAN [5TH–95TH PERCENTILE RANGE] OF MARD$_1$ AND RMSE$_1$ OBTAINED WITH REPLAYBG AND NET-EFFECT

Mod. (%)	Method	MARD$_1$ (%)	RMSE$_1$ (mg/dl)
-50	ReplayBG	**4.51** [2.76, 7.37]	**8.30** [5.25, 12.04]
	Net-Effect	**7.33** [4.89, 11.00]	**13.95** [9.11, 19.82]
-40	ReplayBG	**3.51** [2.26, 6.08]	**6.40** [3.92, 10.30]
	Net-Effect	**6.05** [4.02, 9.68]	**10.92** [7.22, 15.63]
-30	ReplayBG	**2.57** [1.71, 4.81]	**4.68** [2.87, 7.39]
	Net-Effect	**4.72** [3.00, 7.53]	**8.14** [5.41, 11.81]
-20	ReplayBG	**1.84** [1.28, 3.31]	**3.10** [2.07, 5.12]
	Net-Effect	**3.34** [2.07, 5.18]	**5.50** [3.57, 8.15]
-10	ReplayBG	**1.14** [0.82, 1.89]	**1.91** [1.43, 2.73]
	Net-Effect	**1.78** [1.09, 2.77]	**2.84** [1.77, 4.12]
+10	ReplayBG	**1.42** [1.01, 2.35]	**2.08** [1.54, 3.12]
	Net-Effect	**1.90** [1.21, 3.04]	**2.91** [1.88, 4.28]
+20	ReplayBG	**2.33** [1.76, 3.81]	**3.18** [2.33, 4.77]
	Net-Effect	**4.07** [2.44, 6.18]	**5.79** [3.73, 8.57]
+30	ReplayBG	**3.49** [2.54, 5.68]	**4.59** [2.97, 6.87]
	Net-Effect	**6.46** [3.68, 10.72]	**9.09** [5.32, 13.06]
+40	ReplayBG	**4.78** [3.34, 8.01]	**5.94** [3.76, 9.00]
	Net-Effect	**8.97** [5.09, 15.34]	**12.33** [7.10, 17.86]
+50	ReplayBG	**6.13** [4.12, 10.55]	**7.30** [4.63, 10.89]
	Net-Effect	**12.14** [6.57, 20.71]	**15.50** [9.10, 22.27]

Table I shows the overall metrics calculated on the 100 virtual subjects for the ten different meal insulin bolus modulations considered in scenario 1.

The results show that ReplayBG is very accurate in reproducing the target IG fluctuations. Particularly, achieved median MARD is well below the 10% threshold ranging from a minimum of 1.14% to 6.13%. Of note, when the original meal insulin dose is increased by 50%, in a small number of subjects, the obtained MARD results above 10%, however, the overall ReplayBG performance can be considered more than satisfactory. Similar considerations can be made for RMSE$_1$ which ranges from 1.91 mg/dl to 8.30 mg/dl. Comparing ReplayBG with Net-Effect, results highlight that ReplayBG is able to estimate the target IG traces more accurately. This can be also appreciated from Fig. 3(a). In particular, significantly better results are achieved by ReplayBG for all considered meal insulin bolus modulations. Notably, median MARD is always below the 10% threshold using ReplayBG (maximum MARD = 6.13%), while Net-Effect exceeds this limit when the original meal insulin bolus is increased by 50%, (maximum MARD = 12.14%).

Fig. 2(b) shows, for a representative subject, the target IG profiles simulated with T1DS (continuous lines) and the respective replayed IG profiles obtained with ReplayBG (dashed lines) for the ten different modulations of the original meal carbohydrate amount considered in scenario 2. Results are satisfactory and the reference glucose dynamics are correctly captured by

that ReplayBG is able to reconstruct with good accuracy the target reference IG traces capturing the peaks and nadirs as well as the overall glucose dynamics.

TABLE II
SCENARIO 2: MEDIAN [5TH–95TH PERCENTILE RANGE] OF $MARD_2$ AND $RMSE_2$ OBTAINED WITH REPLAYBG AND NET-EFFECT

Mod. (%)	Method	$MARD_2$ (%)	$RMSE_2$ (mg/dl)
-100	ReplayBG	**13.40** [7.39, 24.14]	**11.74** [6.76, 18.57]
	Net-Effect	**29.40** [16.83, 53.28]	**26.72** [15.89, 38.42]
-80	ReplayBG	**10.60** [5.33, 15.17]	**10.94** [6.02, 14.43]
	Net-Effect	**19.85** [12.30, 33.95]	**23.22** [13.01, 32.37]
-60	ReplayBG	**6.82** [3.63, 11.11]	**7.77** [4.74, 11.27]
	Net-Effect	**12.22** [7.09, 22.20]	**16.51** [10.24, 24.71]
-40	ReplayBG	**4.43** [2.37, 6.51]	**5.48** [3.30, 7.84]
	Net-Effect	**7.04** [4.34, 12.41]	**10.83** [6.70, 16.29]
-20	ReplayBG	**2.40** [1.45, 3.26]	**3.08** [2.10, 4.03]
	Net-Effect	**3.07** [2.07, 5.24]	**5.40** [3.40, 8.28]
+20	ReplayBG	**2.40** [1.45, 3.26]	**3.32** [2.44, 4.64]
	Net-Effect	**2.73** [1.91, 4.33]	**5.25** [3.64, 7.83]
+40	ReplayBG	**3.57** [2.34, 5.44]	**6.63** [4.59, 9.16]
	Net-Effect	**4.83** [3.54, 7.86]	**10.59** [6.90, 15.39]
+60	ReplayBG	**5.31** [3.41, 7.64]	**10.34** [7.23, 14.09]
	Net-Effect	**6.83** [4.99, 10.85]	**15.29** [10.44, 22.84]
+80	ReplayBG	**6.85** [5.04, 10.11]	**14.39** [10.59, 19.67]
	Net-Effect	**8.47** [6.30, 12.80]	**20.18** [13.85, 29.64]
+100	ReplayBG	**8.70** [6.29, 12.52]	**18.61** [14.14, 26.17]
	Net-Effect	**9.80** [7.53, 14.80]	**25.24** [17.50, 35.40]

TABLE III
SCENARIO 3: MEDIAN [5TH–95TH PERCENTILE RANGE] OF $MARD_3$ AND $RMSE_3$ OBTAINED WITH REPLAYBG AND NET-EFFECT

Mod. (%)	Method	$MARD_3$ (%)	$RMSE_3$ (mg/dl)
-50	ReplayBG	**12.38** [6.16, 18.98]	**28.20** [17.35 - 53.18]
	Net-Effect	**12.91** [7.81, 19.27]	**31.37** [17.35, 53.18]
-40	ReplayBG	**10.25** [5.08, 15.90]	**20.88** [10.60, 37.31]
	Net-Effect	**11.01** [6.14, 18.03]	**23.81** [12.30, 41.93]
-30	ReplayBG	**7.93** [4.14, 12.40]	**14.36** [7.38, 25.42]
	Net-Effect	**9.26** [4.71, 15.92]	**17.31** [8.83, 31.66]
-20	ReplayBG	**5.31** [2.92, 8.39]	**9.07** [4.84, 15.14]
	Net-Effect	**6.62** [3.29, 12.63]	**11.26** [5.48, 20.76]
-10	ReplayBG	**3.19** [2.05, 4.77]	**4.51** [2.96, 7.21]
	Net-Effect	**3.85** [2.28, 7.03]	**6.13** [3.56, 10.80]
+10	ReplayBG	**5.08** [3.20, 8.87]	**5.97** [3.73, 10.10]
	Net-Effect	**6.45** [3.12, 10.46]	**7.18** [3.82, 11.45]
+20	ReplayBG	**9.90** [6.95, 22.00]	**10.78** [5.93, 19.20]
	Net-Effect	**15.55** [8.63, 24.26]	**12.10** [8.70, 20.65]
+30	ReplayBG	**17.27** [8.95, 39.15]	**14.20** [6.25, 25.63]
	Net-Effect	**25.30** [13.16, 43.42]	**16.53** [10.04, 28.52]
+40	ReplayBG	**22.47** [8.85, 54.80]	**13.25** [5.29, 29.19]
	Net-Effect	**37.93** [16.47, 76.61]	**21.34** [12.00, 38.51]
+50	ReplayBG	**23.92** [9.76, 66.57]	**11.58** [5.71, 30.62]
	Net-Effect	**50.56** [24.76, 110.60]	**25.28** [13.88, 51.55]

TABLE IV
SCENARIO 4: MEDIAN [5TH–95TH PERCENTILE RANGE] OF $MARD_4$ AND $RMSE_4$ OBTAINED WITH REPLAYBG AND NET-EFFECT

Amount (g)	Method	$MARD_4$ (%)	$RMSE_4$ (mg/dl)
5	ReplayBG	**2.15** [1.43, 3.16]	**3.39** [2.34, 5.04]
	Net-Effect	**2.35** [1.46, 3.35]	**3.87** [2.54, 5.49]
10	ReplayBG	**3.36** [2.36, 5.29]	**5.86** [4.05, 8.84]
	Net-Effect	**4.43** [2.55, 6.32]	**7.43** [5.03, 10.86]
15	ReplayBG	**4.37** [3.22, 7.11]	**8.46** [5.74, 12.44]
	Net-Effect	**6.32** [3.71, 9.10]	**11.13** [7.50, 16.42]
20	ReplayBG	**5.26** [3.86, 8.24]	**10.42** [7.39, 15.82]
	Net-Effect	**8.09** [4.67, 11.68]	**14.72** [9.61, 21.60]
25	ReplayBG	**6.11** [4.45, 9.12]	**12.85** [9.16, 19.43]
	Net-Effect	**9.39** [5.45, 13.81]	**18.04** [11.87, 26.58]

ReplayBG. Notably, the error increases when the meal amount is totally removed (-100%) or doubled ($+100\%$), but the resulting profiles are still able to correctly reflect the overall dynamics.

Table II shows the obtained $MARD_2$ and $RMSE_2$ distributions in terms of median and 5th–95th percentile range. $MARD_2$ results show that, when the original meal carbohydrate amount is modulated by -60% up to $+100\%$, ReplayBG obtains reliable results. On the other hand, when a modulation of -100% or 80% is applied, the median MARD is above 10% being 13.40% and 10.60%, respectively. Comparing ReplayBG and Net-Effect in scenario 2, in terms of $MARD_2$ and $RMSE_2$ distributions, the same qualitative conclusions drawn for scenario 1 can also be made for scenario 2. Indeed, as shown by Fig. 3(b), ReplayBG allows to achieve statistically better accuracy than Net-Effect for all considered meal content modulations, except for the -100% modulation, estimating better IG profiles.

Fig. 2(c) shows the results obtained for scenario 3 in a representative subject for the ten different considered modulations of the original basal insulin infusion rate. In particular, the target IG data (continuous lines) are plotted against the IG profiles obtained in step 2 of ReplayBG (dashed lines) when the same input alterations are applied. Results shows that, when basal insulin is modulated from -50% to $+20\%$, the target IG profiles are really close to the respective traces obtained with ReplayBG. The error increases when the basal insulin infusion rate is increased by more than $+20\%$, where ReplayBG overestimates the impact of insulin on glucose levels.

The median and the 5th–95th percentile range results of $MARD_3$ and $RMSE_3$ are reported in Table III. ReplayBG is able to well-reproduce IG fluctuations when the original basal infusion rate is modulated from -40% to $+20\%$ being the obtained median $MARD_3$ below or really close to 10%. For all other considered modulations, $MARD_3$ is higher than the fixed threshold. ReplayBG vs. Net-Effect results are consistent with the results above: ReplayBG better estimates the target

IG traces than the Net-Effect for all considered modulations. Analyzing Fig. 3(c), however, the MARD results statistically lower using ReplayBG only when basal insulin is modulated by -10%, $+40\%$ and $+50\%$.

Fig. 2(d) compares, for a representative subject, the reference IG data obtained using T1DS (continuous lines) when the five different insulin boluses considered in scenario 4 are added to the original insulin input profile with the respective IG profiles obtained by step 2 of ReplayBG (dashed lines). ReplayBG obtains very accurate results and is able to almost perfectly approximate the target IG profile.

Table IV shows the overall metrics calculated on the 100 virtual subjects for the five different added insulin boluses considered in scenario 4. In this scenario, ReplayBG achieved good accuracy in terms of $MARD_4$ for considered situations ranging from 2.15% to 6.11%. Results in terms of RMSE are consistent. Again, ReplayBG achieves better performance compared to Net-Effect. It can be appreciated that ReplayBG is able to estimate the target IG time course more accurately than Net-Effect. In particular, the larger the considered carbohydrate intake amount, the larger the performance gap. This can be observed also in

Amount (U)	Method	MARD$_5$ (%)	RMSE$_5$ (mg/dl)
10/CF	ReplayBG	**1.44** [0.95, 1.90]	**1.99** [1.42, 2.63]
	Net-Effect	**1.60** [1.08 - 2.60]	**2.03** [1.50 - 3.44]
20/CF	ReplayBG	**2.41** [1.52, 3.31]	**2.91** [2.19, 4.26]
	Net-Effect	**3.41** [2.27 - 5.29]	**4.13** [2.97 - 7.08]
30/CF	ReplayBG	**3.37** [2.25, 4.81]	**3.86** [2.65, 6.34]
	Net-Effect	**5.56** [3.55 - 8.53]	**6.35** [4.56 - 10.01]
40/CF	ReplayBG	**4.56** [2.85, 6.46]	**5.12** [3.14, 8.23]
	Net-Effect	**7.88** [4.89 - 12.00]	**8.76** [6.05 - 13.40]
50/CF	ReplayBG	**5.93** [3.83, 8.35]	**6.00** [3.77, 10.51]
	Net-Effect	**10.48** [6.46 - 16.17]	**11.51** [7.71 - 17.30]

Fig. 3(d), where the MARD distributions obtained using ReplayBG and Net-Effect in scenario 4 are shown and compared. In particular, MARD significantly improves when a carbohydrate intake of 15 g or more is added to the original meal input. Of note, median MARD$_4$ is always below the 10% threshold using both methodologies ranging from 2.15% to 6.11% for ReplayBG and from 2.35% to 9.39% for Net-Effect.

Fig. 2(e) shows, for a representative subject, the simulated IG profiles obtained with T1DS (continuous lines) when the five different snacks considered in scenario 5 are added to the original meal input profile, and compares them with the respective IG traces resulting from step 2 of ReplayBG (dashed line). ReplayBG is able to replicate with high fidelity the target glucose dynamics and the fluctuations due to the added snack intakes for all the considered case.

Table V shows the overall metrics calculated on the 100 virtual subjects for the five different added snacks considered in scenario 5. In this case, ReplayBG is very accurate in reproducing the target IG traces obtaining a median MARD$_5$ that ranges from 1.44% to 5.93%, which is below the 10% threshold for all considered insulin bolus amounts, and a median RMSE$_5$ that ranges from 1.99 mg/dl to 6.00 mg/dl.

Comparing ReplayBG with Net-Effect, results are consistent with the other scenarios. ReplayBG performs better than Net-Effect by obtaining lower MARD and RMSE. Similarly to scenario 4, the larger the insulin bolus amount the larger the performance gap. As indicated in Fig. 2(e), this improvement is statistically significant for all considered insulin bolus amounts except for 10/CF. Of note, while ReplayBG MARD is below the 10% threshold for all the insulin bolus amounts, Net-Effect MARD exceed this limit when 50/CF U is considered.

IV. DEMONSTRATION OF REPLAYBG USE WITH REAL DATA

To demonstrate the use of the proposed methodology with real data, we considered two representative applications of ReplayBG using data of a representative patient with T1D. Specifically, the data of this patient were collected during a multicentre clinical trial conducted by the Juvenile Diabetes Research Foundation (JDRF) for the Artificial Pancreas Project Consortium [26], [27]. The study aimed to evaluate the efficacy of the platform "Diabetes Assistant" (DiAs) [28] on glucose control at home in overnight-only and 24/7 closed-loop control

(CLC) modes, compared to the baseline sensor augmented pump (SAP). The patients used the Accu-Check insulin pump (Roche Diagnostics, Mannheim, Germany), the Dexcom G4 Platinum CGM sensor (Dexcom, San Diego, CA, USA), and the DiAs smartphone-based medical platform. Participants underwent a 2-week period at home using the DiAs, the study pump, and the CGM in SAP mode, followed by 2 weeks of overnight-only CLC system use and 2 weeks of 24 h per day and 7 days per week (24/7) CLC system use. From the first two weeks of data (i.e., those collected in open loop mode), we extracted two six-hour-long portion of data containing a meal and the consequent postprandial period, but not containing any corrective actions, neither assumptions of rescue carbohydrates nor injection of corrective insulin boluses.

As additional selection criterion, we extracted only windows that do not present neither insulin bolus administrations, nor carbohydrate intakes, in the 4 hours preceding their starting point. As remarked in II-A2, this is crucial to avoid that an advice resulting from the application of our methodology would be biased by the starting point of the selected window, since, before that point, events that are close to it might affect the identification procedure. Specifically to this point, the aim being analyzing whether our data selection criteria would lead to a very minute subset of all data generated by a patient with T1D, we studied how many postprandial windows in the dataset at hand qualify to be used by ReplayBG. To do so, first we counted how many breakfasts, lunches, and dinners have been logged by the patients overall, then we computed, in terms of percentages, of many of these meals would have been potentially used to run assessments with ReplayBG. We found that patients logged a total of 135 breakfasts, 169 lunches, and 169 dinners (totaling 473 traces). Applying our selection criteria, we were able to extract 352 single meal traces which corresponds to roughly 74% worth of data. We acknowledge that we are losing some potential traces, however we also believe that this percentage is high enough to run ISCTs to preliminary validate new decision support methodologies using our tool on the dataset at hand. Furthermore, we also analyzed if the selection criteria leads to over representing specific time of day. We saw that the potential 352 traces were composed of 119 breakfasts, 112 lunches, and 121 dinners. As such, we can safely state that, at least for this dataset, we are not over/under representing specific time of the day.

The two selected CGM traces, reported in Figs. 4 and 5 with a red line, have been selected because the former presents an hyperglycemic event, suggesting that the insulin bolus has been probably underestimated, whereas the latter presents an hypoglycemic event, calling for an hypotreatment. These are two cases in which we would like to test if alternative treatments would have been able to improve glucose control. Here, we apply ReplayBG methodology to test the potential benefit that could come from the adoption of more advanced treatment solutions. As case study, to deal with the hyperglycemic event, we tested whether the insulin dose provided by the new formula for its calculation that we recently developed [29] could have been more effective than the one used by the patient. On the other hand, to deal with the hypoglycemic episode, we are
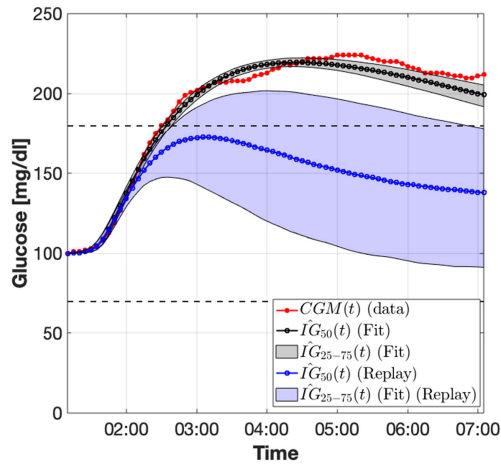
Fig. 4. Portion of real data extracted from [26], [27] characterized by a postprandial hyperglycemic event. In red, the original CGM data. In dotted black, the $\hat{IG}(t)_{50}$ profile obtained from ReplayBG using the original input data, i.e., the fitted median glucose profile; in shaded black the respective $\hat{IG}(t)_{25}$ and $\hat{IG}(t)_{75}$ profiles. In dotted blue, the $\hat{IG}(t)_{50}$ profile obtained from ReplayBG by substituting the original meal insulin bolus with the one computed by [29]; in shaded blue the respective $\hat{IG}(t)_{25}$ and $\hat{IG}(t)_{75}$ profiles.
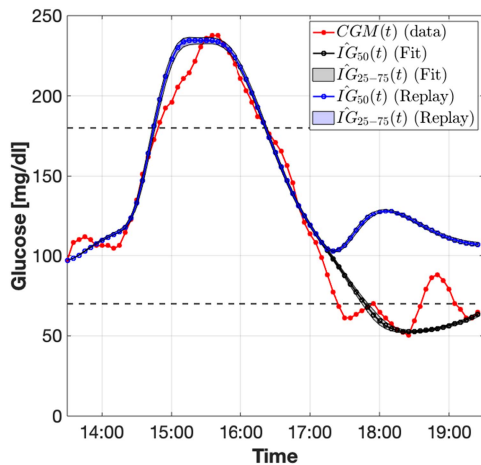

Fig. 5. Portion of real data extracted from [26], [27] characterized by a postprandial hypoglycemic event. In red, the original CGM data. In dotted black, the $\hat{IG}(t)_{50}$ profile obtained from ReplayBG using the original input data, i.e., the fitted median glucose profile; in shaded black the respective $\hat{IG}(t)_{25}$ and $\hat{IG}(t)_{75}$ profiles. In dotted blue, the $\hat{IG}(t)_{50}$ profile obtained from ReplayBG by adding rescue carbohydrate intakes as suggested by [30]; in shaded blue the respective $\hat{IG}(t)_{25}$ and $\hat{IG}(t)_{75}$ profiles.

interested in understanding if the algorithm designed by Camerlingo et al. [30] could have avoided the event by generating a preventive hypotreatment.

Figs. 4 and 5 show the simulation results obtained from the application of [29] and [30], respectively, on the two considered portions of data. By comparing the black line, which represents the median $IG(t)$ profile, i.e., $\hat{IG}(t)_{50}$ obtained from the simulation step of ReplayBG, and CGM data (red), we can first observe that, in both cases, our methodology is able to describe the glucose-insulin system and and fit the CGM traces.

Focusing on the portion of data with hyperglycemia (Fig. 4), replaying the same scenario by substituting the original bolus with the one calculated by the algorithm of Noaro et al. [29], ReplayBG suggests that it would have been possible to prevent, on average, hyperglycemia without inducing any hypoglycemia. In particular, the algorithm increases the original insulin dose from 0.10 U to 1.88 U. The same qualitative results are observed in Fig. 5, when replaying the same scenario with the addition of the preventive hypotreatment suggested by the methodology of Camerlingo et al. [30], ReplayBG allows to retrospectively prove that, on average, it would have been possible to prevent the adverse event thanks to the assumption of one rescue carbohydrate intake of 20 g (assumed 45 minutes in advance with respect to the start of hypoglycemia, at 17:00).

It is interesting to analyse also the confidence intervals $\hat{IG}_{25-75}(t)$ observed in the two scenarios (reported as blue shaded areas). It is clear that in this case the simulated glucose profile obtained in the first scenario is much more uncertain than the one achieved using the algorithm of Camerlingo et al. Such a difference is a direct implication of the uncertainty around the estimated model parameters obtained from step 1 of ReplayBG. In fact, as suggested by the intuition, the wider the posterior distributions of model parameters, the more the 1000 realizations of model parameters, $\hat{\boldsymbol{\theta}}_r, r = 1, \ldots, 1000$, are far apart, the more each of the resulting 1000 simulated glucose profiles is different, leading to a wider confidence interval. On the contrary, as in the case of the second case of study, the narrower the model parameter posterior, the closer the values of $\hat{\boldsymbol{\theta}}_r, r = 1, \ldots, 1000$, the less the resulting 1000 simulated $\hat{IG}(t)$ are different, leading to a narrower $\hat{IG}_{25-75}(t)$.

V. DISCUSSION AND CONCLUSIONS

In this work, we proposed and evaluated ReplayBG, a simulation methodology for the in silico assessment of new or alternative treatments for people with T1D. By leveraging already collected data of insulin and carbohydrate intake and CGM, and a nonlinear physiological model of glucose-insulin dynamics identified at the individual level through a Bayesian parameter estimation procedure, ReplayBG provides an indication of what would have been possible to obtain, in terms of (interstitial) glucose time-course, using different carbohydrates intakes and/or insulin regimens. In the evaluation reported in the present article, ReplayBG reconstructed with high accuracy the IG concentration profile resulting from insulin bolus modulations (scenario 1), and snack/bolus additions (scenarios 4 and 5), for all conditions considered (with extreme cases included). ReplayBG also proved to be a reliable simulation methodology to test the modulation of the amount of meal intake (scenario 2) from -60% to $+100\%$, and the modulation of the basal insulin infusion rate (scenario 3) from -40% to $+20\%$. On the other hand, for greater alterations in scenarios 2 and 3, the performance of ReplayBG degraded resulting in a median MARD above 10%. However, the situations in which ReplayBG does not perform satisfactorily, i.e., large alterations of the basal insulin infusion rate and complete removal of meal carbohydrate amounts, are

extreme cases that are unlikely to be applied or tested in common real-life applications.

ReplayBG has also been compared with the state-of-the-art approach proposed by Hughes et al. [10], which, as ReplayBG, uses data already collected in people with T1D to predict the glucose time-course resulting from the adoption of alternative carbohydrate/insulin therapies. Results demonstrated that ReplayBG is able to better estimate the target IG trace than Net-Effect in all five considered scenarios, making it a viable and robust tool to run such analyses. Anyway, an important remark must be made. From the computational burden point-of-view, the Net-Effect is faster than ReplayBG, since it employs a linear model and a simpler identification procedure. For this reason, depending on the specific setup, in which the potential researcher needs to adopt a methodology like ReplayBG or Net-Effect, the user has to pay attention to this aspect, privileging Net-Effect when the computational resources are limited. Furthermore, we would like to acknowledge that it is possible that, in other more difficult and challenging scenarios (here not evaluated), as those encountered in real data, Net-Effect could achieve comparable or better performance than ReplayBG. Following this direction, and for the sake of fairness, our research group will focus on this peculiar aspect during future dedicated and diligent work.

A limitation of this work could relate to the validation procedure. Indeed, we acknowledge that it would have been preferable to compare the glucose predictions made by ReplayBG with real data rather than vs the simulated outcome of a more complex, but reliable simulator, i.e., T1DS. However, as discussed in Section I, the assessment solution here proposed is the only one that allows obtaining the glycemic ground truth resulting from an alteration of the original treatment regime, since it enables one to run multiple ISCTs maintaining the same surrounding conditions, which is impossible to achieve in the "real" setup.

Furthermore, the validation procedure may suffer from the so-called "model-matching" bias since the model we used to generate the synthetic data [12], thus the ground truth, is, in some aspects, similar to the model to be identified. However, two main points must be taken into account. First, the model used in ReplayBG is composed of three main subsystems, i.e., (1), (2), and (4), and each of which differs from the one used in the UVa/Padova simulator [12] in a certain way. Starting from the closest, the subcutaneous insulin absorption subsystem neglects the direct flux between the non-monomeric insulin state compartment (I_{sc1}) and the plasma insulin compartment (I_p) modeled in [12]. The oral glucose absorption subsystem describes glucose transit through the stomach and upper small intestine as a simple linear chain of three compartments (Q_{sto1}, Q_{sto2}, and Q_{gut}) unlike the model of [12], which has a nonlinear structure to enable the description of both the biphasic nature of gastric emptying and the nonlinear dependence on gastric emptying of liquids on the size of the meal, its energy density, and the amount of nutrient in the stomach. Finally, the core glucose-insulin kinetics subsystem is a minimal model that does not account for the many different factors that, on the other hand, are modeled in [12]. In turn, these simplifications plus other unmodeled phenomena (e.g., glucagon subsystem) make the model of ReplayBG still distinct, by many aspects, from

the model used to generate the target ground-truth, mitigating the model-matching bias. Second, as already stressed, when validating methodologies such as ReplayBG, where it is necessary to evaluate if the approach under assessment can correctly reproduce the impact of certain input alterations (carbohydrate and insulin) on a given output (in our case, glucose) of a baseline scenario maintaining the same surrounding conditions, this bias is not totally avoidable, since data must be generated via a simulation tool. In fact, this is a common problem faced by other literature techniques as Patek et al. [8] and Hughes et al. [10], where they used the same validation procedure deployed in this work and the same maximal physiological model to generate the synthetic data. It is also important to note that the potential user of ReplayBG must be careful when choosing the portion of data to replay. As other similar methodologies for the scope, e.g., [8] and [10], ReplayBG formulates glucose predictions whose reliability depends on the quality of the data, e.g., reported meal/insulin inputs, whose correctness is key to model identification. In fact, incorrect model inputs, such as incorrectly entered meal carbohydrate amounts, inevitably affect first the model parameter estimates and, secondly, the glucose output predictions when different altered inputs are used to "replay" the scenario at hand. To mitigate this issue, future work will explore the possibility of adding, as a preliminary step to model identification, the application of ad-hoc dedicated modules to detect and correct potential data errors and faults, e.g., the approaches proposed by Meneghetti et al. [31], [32]. Future work will also deal with the possibility of releasing the assumption that all model states start from steady state. This will be possible by modifying the identification scheme of ReplayBG to also include the estimation of the corresponding initial conditions.

It is worth mentioning that the proposed ReplayBG methodology can be potentially employed for a multitude of other purposes, such as refining and tuning state-of-the-art decision support algorithms and leveraging it as the methodological core of new educational tools for T1D. To expand the domain of applicability of ReplayBG, future work will concern the refinement of the methodology to better describe glucose dynamics by expanding the physiological model of ReplayBG to include a subsystem describing physical exercise [33], [34]. Furthermore, the methodology will be validated in more challenging multiple meal/multiple day scenarios. Indeed, the ReplayBG methodology, as it is presented and validated in this manuscript, limits its domain of applicability to single-meal scenarios, where it is possible to assume that the underneath system is stationary. In the direction of releasing this assumption, current undergoing work in our laboratory is focusing on expanding the ReplayBG methodology to multiple-meal scenarios by introducing dedicated model parameters able to capture, for example, the intra-/inter-day insulin sensitivity variability one can observe in real subjects [35], multiple mixed meals with different content, stress, and natural physiological fluctuations that will impact glucose dynamics in ways that are not included in the proposed model. Finally, as previously done in [29], ReplayBG will be deployed within the common pipeline used to validate newly developed algorithms for T1D treatment. In these terms, we believe that ReplayBG could be a useful tool to help researchers

in the field to test their methodologies, for example, new insulin bolus calculators, hypotreatment strategies, and, in general, decision support systems, for people with T1D. For this reason, ReplayBG is freely available to the community as open source software on GitHub (https://github.com/gcappon/replay-bg). This will hopefully allow leveraging the potential of the developer community to maintain and improve ReplayBG over time. Of note, it is worth underlining that the proposed approach provides just an indication of the performance of the treatment under assessment and it is not intended to substitute at any level its validation via formal clinical trials.

REFERENCES

[1] American Diabetes Association, "Diagnosis and classification of diabetes mellitus," in *Diabetes Care*, vol. 33, no. Suppl1, pp. S62–S69, 2010.

[2] American Diabetes Association, "Standards of medical care in diabetes," in *Diabetes Care*, vol. 44, no. Suppll, pp. S1–S232, 2021.

[3] F. Pappalardo et al., "In *silico* clinical trials: Concepts and early adoptions," *Brief Bioinform.*, vol. 20, no. 5, pp. 1699–1708, 2019.

[4] R. Hovorka et al., "Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes," *Physiol. Meas.*, vol. 25, no. 4, pp. 905–920, 2004.

[5] M. E. Wilinska et al., "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.

[6] S. S. Kanderian et al., "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, Dec. 2013.

[7] C. Dalla Man et al., "Meal simulation model of the glucose-insulin system," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 10, pp. 1740–1749, 2007.

[8] S. D. Patek et al., "Empirical representation of blood glucose variability in a compartmental model," in *Prediction Methods for Blood Glucose Concentration.* Berlin, Germany: Springer, 2016, pp. 133–157.

[9] M. Vettoretti et al., "Predicting insulin treatment scenarios with the net effect method: Domain of validity," *Diabetes Technol. Ther.*, vol. 18, no. 11, pp. 694–704, 2016.

[10] J. Hughes et al., "Replay simulations with personalized metabolic model for treatment design and evaluation in type 1 diabetes," *J. Diabetes Sci. Technol.*, vol. 15, no. 6, pp. 1326–1336, 2020.

[11] G. Cappon et al., "A bayesian framework to identify type 1 diabetes physiological models using easily accessible patient data," in *Proc. IEEE 41st Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2019, pp. 6914–6917.

[12] R. Visentin et al., "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.

[13] M. Schiavon, C. D. 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, Sep. 2018.

[14] C. Dalla Man et al., "A system model of oral glucose absorption: Validation on gold standard data," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 12, pp. 2472–2478, 2006.

[15] R. N. Bergman, C. R. Bowden, and C. Cobelli, "Quantitative estimation of insulin sensitivity," *Amer. J. Physiol.*, vol. 236, no. 6, 1979, Art. no. e667.

[16] C. D. Man, A. Caumo, and C. Cobelli, "The oral glucose minimal model: Estimation of insulin sensitivity from a meal test," *IEEE Trans. Biomed. Eng.*, vol. 184, no. 1, pp. 419–429, May 2002.

[17] C. D. Man et al., "The UVA/padova type 1 diabetes simulator: New features," *J. Diabetes Sci. Technol.*, vol. 8, no. 1, pp. 26–34, 2017.

[18] A. Facchinetti et al., "Reconstruction of glucose in plasma from interstitial fluid continuous glucose monitoring data: Role of sensor calibration," *J. Diabetes Sci. Technol.*, vol. 1, no. 5, pp. 617–623, 2007.

[19] K. Rebrin et al., "Subcutaneous glucose predicts plasma glucose independent of insulin: Implications for continuous monitoring," *Amer. J. Physiol.*, vol. 277, no. 3, pp. E561–E571, 1999.

[20] G. Pillonetto et al., "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.

[21] C. Dalla Man et al., "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.

[22] W. R. Gilks et al., *Markov Chain Monte Carlo in Practice*, 1st ed. Boca Raton, FL, USA: Chapman and Hall, 1996.

[23] S. Demarta and A. McNeil, "The *t* copula and related copulas," *Int. Statist. Rev.*, vol. 73, no. 1, pp. 111–129, 2007.

[24] S. Schmidt and K. Nørgaard, "Bolus calculators," *J. Diabetes Sci. Technol.*, vol. 8, no. 5, pp. 1035–1041, 2014.

[25] W. L. Clarke et al., "Evaluating clinical accuracy of systems for self-monitoring of blood glucose," *Diabetes Technol. Ther.*, vol. 10, no. 5, pp. 622–628, 1987.

[26] B. Kovatchev et al., "Feasibility of long-term closed-loop control: A multicenter 6-month trial of 24/7 automated insulin delivery," *Diabetes Technol. Ther.*, vol. 19, no. 1, pp. 18–24, 2017.

[27] S. M. Anderson et al., "Multinational home use of closed-loop control is safe and effective," *Diabetes Care*, vol. 39, no. 7, pp. 1143–1150, 2016.

[28] P. Keith-Hynes et al., "The diabetes assistant: A smartphone-based system for real-time control of blood glucose," in *Electronics*, vol. 3, no. 4, pp. 609–623, 2014.

[29] G. Noaro et al., "Machine-learning based model to improve insulin bolus calculation in type 1 diabetes therapy," *Diabetes Technol. Ther.*, vol. 68, no. 1, pp. 247–255, 2021.

[30] N. Camerlingo et al., "A real-time continuous glucose monitoring-based algorithm to trigger hypotreatments to prevent/mitigate hypoglycemic events," *Diabetes Technol. Ther.*, vol. 21, no. 11, pp. 644–655, 2019.

[31] L. Meneghetti et al., "Machine learning-based anomaly detection algorithms to alert patients using sensor augmented pump of infusion site failures," *J. Diabetes Sci. Technol.*, vol. 16, no. 3, pp. 641–648, 2022.

[32] L. Meneghetti, A. Facchinetti, and S. D. Favero, "Model-based detection and classification of insulin pump faults and missed meal announcements in artificial pancreas systems for type 1 diabetes therapy," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 1, pp. 170–180, Jan. 2021.

[33] D. Romeres et al., "691-P: Exercise effect on endogenous glucose production in type 1 diabetes: A modeling analysis," *Diabetes*, vol. 69, no. S1, p. 691-P, 2020.

[34] M. Rashid et al., "Simulation software for assessment of nonlinear and adaptive multivariable control algorithms: Glucose - Insulin dynamics in type 1 diabetes," *Comput. Chem. Eng.*, vol. 130, 2019, Art. no. E106565.

[35] R. Visentin et al., "Circadian variability of insulin sensitivity: Physiological input for in silico artificial pancreas," *Diabetes Technol. Ther.*, vol. 17, no. 1, pp. 1–7, 2015.