

## Article

# On the Use of Population Data for Training Seasonal Local Models-Based Glucose Predictors: An *In Silico* Study

Antonio Aslan <sup>1</sup>, José-Luis Díez <sup>1,2,\*</sup> , Alejandro José Laguna Sanz <sup>2</sup>  and Jorge Bondia <sup>1,2</sup> 

- <sup>1</sup> Instituto Universitario de Automática e Informática Industrial, Universitat Politècnica de València, Camino de Vera s/n, 46022 València, Spain; antonio@antonioaslan.com (A.A.); jbondia@isa.upv.es (J.B.)
- <sup>2</sup> Centro de Investigación Biomédica en Red de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM), Instituto de Salud Carlos III, 28029 Madrid, Spain; allasan@upvnet.upv.es
- \* Correspondence: jldiez@isa.upv.es

**Abstract:** Most advanced technologies for the treatment of type 1 diabetes, such as sensor-pump integrated systems or the artificial pancreas, require accurate glucose predictions on a given future time-horizon as a basis for decision-making support systems. Seasonal stochastic models are data-driven algebraic models that use recent history data and periodic trends to accurately estimate time series data, such as glucose concentration in diabetes. These models have been proven to be a good option to provide accurate blood glucose predictions under free-living conditions. These models can cope with patient variability under variable-length time-stamped daily events in supervision and control applications. However, the seasonal-models-based framework usually needs of several months of data per patient to be fed into the system to adequately train a personalized glucose predictor for each patient. In this work, an *in silico* analysis of the accuracy of prediction is presented, considering the effect of training a glucose predictor with data from a cohort of patients (population) instead of data from a single patient (individual). Feasibility of population data as an input to the model is asserted, and the effect of the dataset size in the determination of the minimum amount of data for a valid training of the models is studied. Results show that glucose predictors trained with population data can provide predictions of similar magnitude as those trained with individualized data. Overall median root mean squared error (RMSE) (including 25% and 75% percentiles) for the predictor trained with population data are {6.96[4.87, 8.67], 12.49[7.96, 14.23], 19.52[10.62, 23.37], 28.79[12.96, 34.57], 32.3[16.20, 41.59], 28.8[15.13, 37.18]} mg/dL, for prediction horizons (PH) of {15, 30, 60, 120, 180, 240} min, respectively, while the baseline of the individually trained RMSE results are {6.37[5.07, 6.70], 11.27[8.35, 12.65], 17.44[11.08, 20.93], 22.72[14.29, 28.19], 28.45[14.79, 34.38], 25.58[13.10, 36.60]} mg/dL, both training with 16 weeks of data. Results also show that the use of the population approach reduces the required training data by half, without losing any prediction capability.

**Keywords:** type 1 diabetes; glucose prediction; clustering; seasonal local models; individual data; population data



**Citation:** Aslan, A.; Díez, J.-L.; Laguna Sanz, A.J.; Bondia, J. On the Use of Population Data for Training Seasonal Local Models-Based Glucose Predictors: An *In Silico* Study. *Appl. Sci.* **2023**, *13*, 5348. <https://doi.org/10.3390/app13095348>

Academic Editor: Marco G. Alves

Received: 15 March 2023

Revised: 14 April 2023

Accepted: 20 April 2023

Published: 25 April 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Diabetes affected 537 million people with diabetes worldwide in 2021 [1], and this figure is estimated to rise to 643 million by 2030 and 783 million by 2045, which makes diabetes one of the major health challenges of the 21st century. Type 1 diabetes (T1D) is the most severe variant of this condition, and affects about 10% of people with diabetes. In people with T1D,  $\beta$ -cells in the pancreas are destroyed by an autoimmune response, impairing or completely suppressing insulin secretion, causing blood glucose (BG) concentrations to escape the safe range of 70–180 mg/dL. Since the discovery of insulin, T1D treatment usually consists in exogenous administration of insulin, or more recently, insulin analogs, in order to avoid long-term complications due to hyperglycemia (high concentration of BG). Unfortunately, insulin overdosing may reduce BG concentration to hypoglycemic

levels, which may lead to severe consequences such as coma and death when untreated. Ambulatory BG control is demanding for patients, who carry out many critical decisions and calculations per day, and are required to constantly monitor BG levels to avoid hyper- and hypoglycemia.

Nowadays, glucose concentration can be measured in real time by using continuous glucose monitoring (CGM) devices [2], which are increasingly common medical devices in the treatment and supervision on T1D. However, accurate glucose predictions over long time horizons are still needed to help in preventing hyperglycemic and, especially, hypoglycemic episodes [3,4]. Glucose prediction is important in sensor-pump integrated systems [5] or artificial pancreas (AP) systems [6–9].

There are a variety of alternatives for BG prediction in T1D in the literature, such as, among others, linear empirical dynamic models, multivariate nonlinear regression techniques, extended Kalman Filters, data mining or artificial intelligence approaches [10–12]. More recently, deep learning and meta-learning [13–15] approaches have yielded very accurate predictions when trained on population data, albeit those models need to be fed exogenous data that may not be reliably available or accurate. A promising approach is the use of seasonal stochastic local models [16,17], since these models provide reliable glucose predictions under high intra-patient variability for long prediction horizons. The rationale lies in the fact that observed behaviors in past similar scenarios can help to improve predictions at the present moment. Seasonal stochastic time series models such as SARIMA (seasonal autoregressive integrated moving average) or SARIMAX (SARIMA including exogenous variables) have exhibited good accuracy for postprandial BG prediction accuracy over long prediction horizons [16], including challenging scenarios with a variety of meals and exercise sessions [17]. Seasonal local modeling also proved successful under real-life scenarios [18], but at least four months of data from each patient were needed to properly train the personalized system. SARIMA models have recently been proven to provide valid short term predictions in real life experimental datasets, as shown in [19], holding their ground against neural networks when predicting glucose in two different experimental benchmark datasets.

Thus, it is desirable to reduce as much as possible the individual historical data needed to train these models in order to provide valid glucose predictions to the patient and clinical staff as soon as possible. The question that motivated this work is whether training from population data, composed of smaller datasets from many patients, provides accurate enough glucose predictions when used in an individual. Additionally, we investigated to what extent individual data needs can be reduced in the training process to maintain prediction accuracy.

The main purpose of this work is to train population seasonal data-based models and perform an *in silico* study of prediction capabilities to analyze if the use of population data can substitute individual patient data while maintaining accuracy. Additionally, the effect of the reduction of the volume of data needed (number of weeks per patient included in the population data) is studied and how it affects the feasible prediction horizon.

## 2. Material and Methods

The main contribution of this work is the use of global glucose predictors to estimate future glucose, which uses seasonal stochastic models trained with data from a population of patients, in contrast to those trained with data from the same individual. The methodology, apart from the data curation and partitioning, is very similar to the procedure described in [18] for seasonal stochastic local modeling. A summary of the method is reproduced next where additions and modifications to the original procedure have been included.

### 2.1. Data Overview

CGM usage has experienced an extraordinary growth in recent years and its data is easier to access by researchers than before. Indeed, diabetes management software such as Tidepool or SocialDiabetes gather and store this data, facilitating data curation and access

by researchers [20,21]. Furthermore, full repositories of experimental CGM data are freely available to download e.g., the rePLACE-BG diabetes dataset [22]. In this work, we opted to use in silico data consisting of two datasets (one for training, one for validation) of ten virtual adults from the educational version of the UVA/Padova simulator [23], with an exercise model add-on [24], and additional sources of intra-patient variability. Both sets of data were generated following open loop therapy (basal insulin and ratios provided by the simulator were used). The main characteristics (events and variability) of the simulated scenario were chosen to represent as closely as possible the real life of people with T1D, and are summarized next:

- A nominal day with three meals of 40, 90 and 60 g of carbohydrates at 7:00, 14:00 and 21:00 h were simulated.
- Actual mealtime and carbohydrate intake varied around nominal values following a normal distribution with a standard deviation of 20 min for mealtime, and a 10% coefficient of variance for meal size.
- Meal absorption dynamics were changed at each meal by randomly selecting one of the meal model parameter sets available from the simulator [23].
- Meal absorption rate at each meal changed within a uniform distribution in  $\pm 30\%$ , and carbohydrate bioavailability in  $\pm 10\%$  around selected nominal values.
- Carbohydrate counting errors by the patient were considered to follow a uniform distribution between  $-30\%$  and  $+10\%$ , following results in [25] where a trend to meal underestimation is reported.
- Insulin absorption pharmacokinetics varied in  $\pm 30\%$ , according to the intra-patient variability reported in [26].
- Circadian variability of insulin sensitivity was considered with variations in  $\pm 30\%$  around nominal sensitivity, reproducing changes in basal insulin requirements in the adult population reported in [27].
- The 15–15 rule [28] was used to treat hypoglycemia (15 g of carbohydrates were administered when glucose went below 70 mg/dL, and repeated if after 15 min hypoglycemia was still present).

Training and validation datasets were generated using the same scenario. Four months of data were generated for the training dataset, representing the maximum data per patient available. The actual length of training data used was varied in the study performed according to the scenarios described in Section 2.8. A fixed length of two months was considered for the validation dataset.

In both datasets, besides CGM data, time stamps and labels for the following events were generated: *Breakfast*, *Lunch*, *Dinner*, *Night* and *Hypoglycemia treatment*. The *Night* event was considered to start 6 h after dinner. All rescue carbs administered in a hypoglycemia treatment were gathered in a single *Hypoglycemia treatment* event, starting at the time of the first rescue carb intake. This will be represented, for each patient  $p = 1, \dots, 10$ , by the set of time-ordered pairs

$$\mathcal{E}^p := \{(t_{e_i}, \mathcal{L}_{e_i}) \mid i = 1, \dots, n_e\}, \quad (1)$$

where  $t_{e_i}$  is the timestamp for the  $i$ -th event,  $t_{e_i} < t_{e_{i+1}}$ ,  $\mathcal{L}_{e_i}$  is the label of the  $i$ -th event,  $\mathcal{L}_{e_i} \in \{\text{Breakfast, Lunch, Dinner, Night, Hypoglycemia treatment}\}$ , and  $n_e$  is the number of events.

## 2.2. CGM Data Partitioning

Seasonality must be enforced before building seasonal stochastic local models from the training dataset. Provided that the starting time and duration of events in the CGM historical data is variable between event instances, the original CGM time series data is partitioned into a set of “event-to-event” time subseries from the reported timestamps and labels in the set  $\mathcal{E}^p$ . Although different options exist to build groups (from a single partition if event labels are neglected to multiple partitions of all event labels are considered), three

partitions are considered in this work (meals, night and hypoglycemia treatment). This partitioning is a compromise requiring only information about meal and hypoglycemia time instants (which can be extracted automatically from data) and performed well in previous studies [18]. Thus, consider for patient  $p$

$$\mathcal{E}^p = \mathcal{E}_{\mathcal{M}}^p \cup \mathcal{E}_{\mathcal{N}}^p \cup \mathcal{E}_{\mathcal{H}}^p, \tag{2}$$

where

$$\mathcal{E}_{\mathcal{M}}^p := \{(t_{e_i}, \mathcal{L}_{e_i}) \mid \mathcal{L}_{e_i} \in \{\text{Breakfast, Lunch, Dinner}\}\} \tag{3}$$

$$\mathcal{E}_{\mathcal{N}}^p := \{(t_{e_i}, \mathcal{L}_{e_i}) \mid \mathcal{L}_{e_i} \in \{\text{Night}\}\} \tag{4}$$

$$\mathcal{E}_{\mathcal{H}}^p := \{(t_{e_i}, \mathcal{L}_{e_i}) \mid \mathcal{L}_{e_i} \in \{\text{Hypoglycemia treatment}\}\}. \tag{5}$$

Given the index sets for patient  $p$

$$\mathcal{I}_{\mathcal{M}}^p := \{i \mid (t_{e_i}, \mathcal{L}_{e_i}) \in \mathcal{E}_{\mathcal{M}}^p\} \tag{6}$$

$$\mathcal{I}_{\mathcal{N}}^p := \{i \mid (t_{e_i}, \mathcal{L}_{e_i}) \in \mathcal{E}_{\mathcal{N}}^p\} \tag{7}$$

$$\mathcal{I}_{\mathcal{H}}^p := \{i \mid (t_{e_i}, \mathcal{L}_{e_i}) \in \mathcal{E}_{\mathcal{H}}^p\} \tag{8}$$

and denoting by  $[G]_{t_0}^{t_1}(t) := \{G(t) \mid t_0 \leq t \leq t_1\}$  the time subseries given by the segment of CGM data from  $t_0$  to  $t_1$  for patient  $p$ , the following CGM data partitions are built:

$$\Pi_{\mathcal{M}}^p := \{[G]_{t_{e_i}}^{t_{e_{i+1}}}(t) \mid i \in \mathcal{I}_{\mathcal{M}}^p\} \tag{9}$$

$$\Pi_{\mathcal{N}}^p := \{[G]_{t_{e_i}}^{t_{e_{i+1}}}(t) \mid i \in \mathcal{I}_{\mathcal{N}}^p\} \tag{10}$$

$$\Pi_{\mathcal{H}}^p := \{[G]_{t_{e_i}}^{t_{e_{i+1}}}(t) \mid i \in \mathcal{I}_{\mathcal{H}}^p\}. \tag{11}$$

In this work, population data is also needed. In this case, some partitions can be aggregated to create population partitions, such are those including the meals, the nights and the hypoglycemia treatments for the whole population.

$$\Pi_{\mathcal{M}} = \bigcup_{p=1}^{N_p} \Pi_{\mathcal{M}}^p \tag{12}$$

$$\Pi_{\mathcal{N}} = \bigcup_{p=1}^{N_p} \Pi_{\mathcal{N}}^p \tag{13}$$

$$\Pi_{\mathcal{H}} = \bigcup_{p=1}^{N_p} \Pi_{\mathcal{H}}^p \tag{14}$$

where  $N_p$  is the number of patients in the cohort; in our case,  $N_p = 10$ .

This aggregates sequentially (the partition from one patient after the previous one) the partitions of all patients into a single partition. Therefore, when using the aggregated population partition, the identification procedure is patient-blind.

### 2.3. Data Regularization

Each CGM time subseries in the training dataset is comprised from a given event timestamp,  $t_{e_i}$ , to the next event timestamp,  $t_{e_{i+1}}$ . In order to enforce seasonality, the time subseries' lengths must be regularized. Time subseries of shorter length than the corresponding maximum length found in the partition where it is included are expanded with blank values ("not a number"—NaN) until reaching maximum length. As a result, time subseries length is regularized and seasonality can be enforced (see Section 2.6). Note that most of time subseries will have blanks in the final positions, which should be

adequately treated as missing data in the clustering and identification processes. Thus, regularized length will be different for each partition  $\Pi_{\mathcal{M}}$ ,  $\Pi_{\mathcal{N}}$ ,  $\Pi_{\mathcal{H}}$ ,  $\Pi_{\mathcal{M}}^p$ ,  $\Pi_{\mathcal{N}}^p$ , and  $\Pi_{\mathcal{H}}^p$ .

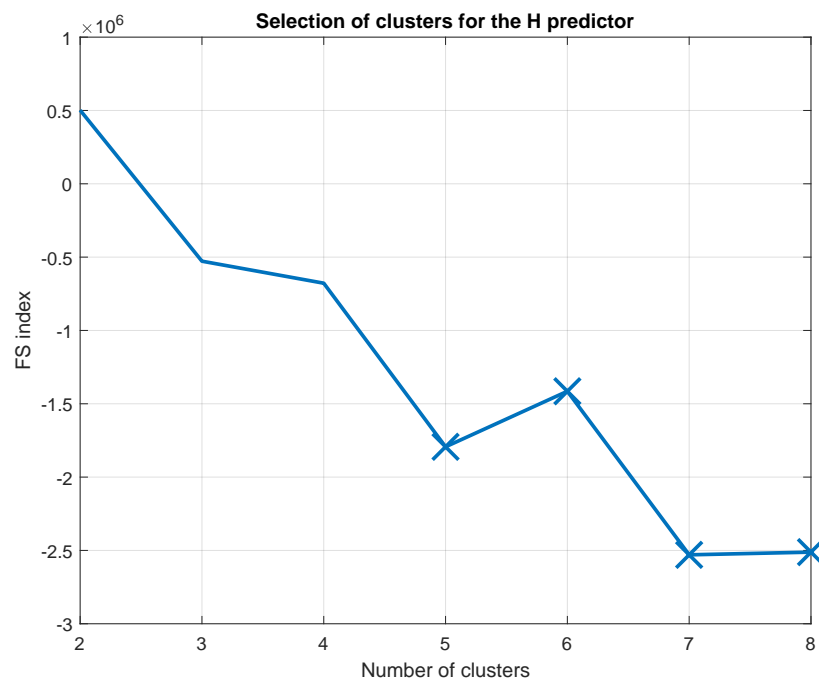
#### 2.4. Data Clustering

Regularized partitions in the training dataset must be clustered to gather similar glucose responses prior to local modelling, as required by seasonal techniques. Fuzzy C-Means (FCM) clustering [29] requires data to be complete and, therefore, it is not directly applicable to the available partitions with missing data resulting from the seasonality enforcing step. The partial distance strategy FCM (PDSFCM) algorithm is used instead, combined with the partial distance as defined in [30].

Most clustering algorithms, such as FCM and its modification PDSFCM, require knowing the expected number of clusters in the data before running the process. Fortunately, there exists a large variety of indices to help determine the optimal number of clusters. In this work, minimization of the Fukuyama–Sugeno (FS) index has been used [31] for this purpose, due to the good results obtained with it in previous works [18].

Additionally, provided that a local model will be identified later for each cluster, a minimum number of time series is required in a cluster for the identification process to be performed. A threshold of 10 time series has been found adequate in this work, then ensuring subsequent local models a minimum set for identification and validation.

Figure 1 shows an example where FS decreases as the number of clusters increases, with some local minimum in the evolution. If only FS is considered, five or seven clusters are the best candidates (local minimum with less clusters). However, when five or more clusters are considered, the number of time series drops below 10 in (at least) one cluster (marked with a cross in the figure), thus leading to possible problems in the subsequent identification step (due to the small number of data for training and validation). Therefore, four clusters are chosen in this case.



**Figure 1.** Example of number of clusters selection using the FS index in combination with a minimum number of time series in a cluster. The X indicates the existence of one or more clusters with less than 10 time series.

### 2.5. Data Preparation for Identification

This work will compare models obtained using individual patient data vs population data. Although the used data (amount, partitions, etc.) is the same, the data must be organised in a different way before feeding it to the identification algorithm. In both cases, 80% of data will be used for local models training and the remaining 20% for local models validation, but sorted in a different way.

In the case of data for individual models (one for each patient), partitions (9), (10), and (11) will be split into two parts: the first 80% of data will be used for training and last 20% for validation.

However, if the same splitting procedure is applied to the population partitions in (12), (13), and (14), an undesirable effect is detected. As an example, the top image in Figure 2 shows the application of the procedure to a cluster from the partition  $\Pi_{\mathcal{M}} = \bigcup_{p=1}^2 \Pi_{\mathcal{M}}^p$ . It can be seen that, provided that each cluster contains data from different patients and that data is stored in order in the original clustered partitions, the two parts for identification (training and validation) will contain events from completely different patients (patient 1 is used for training, and only part of patient 2 is used for validation). In the case of this work (10 patients), this procedure will lead to a scenario where patients 1 to 8 will be used to train the local models and the remaining patients will be used to validate the local model. In order to avoid this problem, it is extremely important to properly split the training and validation data from each patient. Therefore, the total number of patients present in each cluster will be identified and then the data from each one of them will be split using the 80/20 criteria, as presented in the bottom image in Figure 2 for the partition  $\Pi_{\mathcal{M}} = \bigcup_{p=1}^2 \Pi_{\mathcal{M}}^p$ . In this way, the right proportion of data from each patient is ensured in the training and validation split.

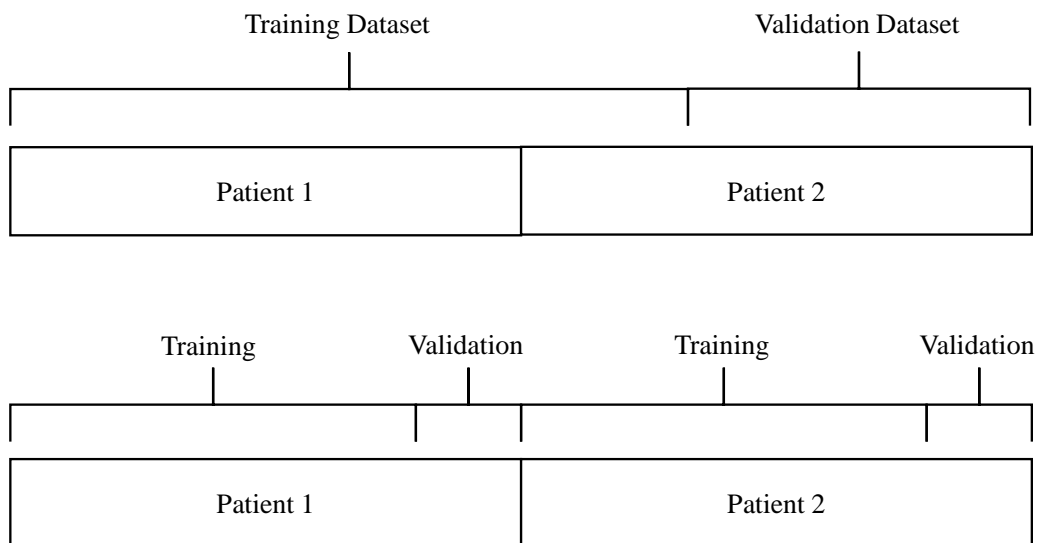


Figure 2. Training and validation split procedures for population data.

### 2.6. Local SARIMA Modeling

After data in a partition are clustered, local seasonal models are trained for each cluster. If the clustering process is successful, all the time series in a cluster will be similar. Then, a SARIMA model can be identified for each cluster  $i$ , yielding a set of local glucose predictors (one per cluster). Note that the cost index during identification will neglect blank data in the residuals computation. Box–Jenkins methodology [32] is used in this identification process. Exogenous inputs could also be considered with SARIMAX models [17], although only glucose and mealtimes will be considered available in this work.



### 2.7. Models Integration in a Glucose Predictor

Once local models are properly trained for each cluster, a glucose predictor is built by the integration of such local models during real-time operation. Although a detailed description of this process can be found in [18], the glucose predictor working process is quickly described next: When the glucose predictor receives a new real time glucose value  $G(i)$ , a time series is built with this value and a given window of previous glucose measurements from the time the last event occurred. This time series is then compared with the same time window from the cluster prototypes of each local model, and a membership value  $\mu_i$  is calculated, representing the similarity between the current time series and each prototype  $i$  for that time window. Then, local models are evaluated providing a local prediction, based on the new glucose measurement and past historical data in the corresponding cluster, depending on the seasonal order of the local models. In the end, all those predictions are integrated using the weighted sum of each response by their corresponding membership  $\mu_i$ . Therefore, the final (global) prediction  $\hat{G}(i)$  will depend mostly on the local model's response that is more similar to the current patient's situation, but all the other responses will be taken into account in some degree.

### 2.8. Glucose Predictors Validation

As described in Section 2.1, the training dataset comprised sixteen-weeks in silico data for a ten patient cohort. For each patient  $p$ , training data was partitioned (following the descriptions in Section 2.2 for partition, and Section 2.3 for regularization) in  $\Pi_{\mathcal{M}}^p$ ,  $\Pi_{\mathcal{N}}^p$ , and  $\Pi_{\mathcal{H}}^p$  sets. Clustering (see Section 2.4) and local model training and integration (Sections 2.6 and 2.7) was then applied to each partition, yielding a set of individual glucose predictors (one per patient  $p$  and event group  $\mathcal{M}$ ,  $\mathcal{N}$ , and  $\mathcal{H}$ ):  $\hat{G}_{\mathcal{M}}^p(i)$ ,  $\hat{G}_{\mathcal{N}}^p(i)$ , and  $\hat{G}_{\mathcal{H}}^p(i)$ . The same procedure has been applied to the population data in the partitions  $\Pi_{\mathcal{M}}$ ,  $\Pi_{\mathcal{N}}$ , and  $\Pi_{\mathcal{H}}$ , then resulting in a single population glucose predictor per event ( $\mathcal{M}$ ,  $\mathcal{N}$ , and  $\mathcal{H}$ ), but valid for all patients:  $\hat{G}_{\mathcal{M}}(i)$ ,  $\hat{G}_{\mathcal{N}}(i)$ , and  $\hat{G}_{\mathcal{H}}(i)$ .

It must be recalled that in both cases, individual and population training, 80% of the available data for each cluster was used for local models identification and the remaining 20% was kept for validation time series following the procedure described in Section 2.5. These partial results on local modeling are kept internal and not shown in this work, since it is not deemed relevant for the conclusions.

Individual (patient-personalized) glucose predictors performed very well in previous works [18], and this performance will be compared to the population predictor in this work. Additionally, the population predictor performance degradation will be studied as the number of data is reduced (the number of weeks of available data per patient is reduced sequentially from 16 to 2 weeks) for prediction horizons in the set  $PH \in \{15, 30, 60, 120, 180, 240\}$  min.

Both glucose predictors (individual and population) performance have been evaluated using an independent 2-month validation dataset (see Section 2.1). Each glucose predictor has been matched, at each event, according to its type (meal, night, hypoglycemia treatment). For all time series in the validation data, predictions have been computed as described in Section 2.7. In order to facilitate results comparison with previous studies in glucose prediction in the literature, the standard metric RMSE (mg/dL) has been used here to measure forecasting accuracy:

$$RMSE := \sqrt{\frac{1}{n} \sum_{i=1}^n e^2(i)} \quad (15)$$

where  $n$  is the number of observations,  $e(i) = G(i) - \hat{G}(i)$  is the forecasting error (residual),  $G(i)$  denote the  $i$ th glucose observation, and  $\hat{G}(i)$  is the forecast of  $G(i)$ . The residuals of the glucose trajectory are built from predictions at time  $t_p$  for a prediction horizon  $PH$ , i.e.  $G(t_p + PH | t_p)$ . It evaluates how successfully the actual glucose trajectory can be predicted  $PH$  time samples ahead. In this case, for each time series in the validation

data, and denoting  $l_i$  as the length of the time series,  $l_i - PH$  observations are available for the computation of residuals. Metrics were averaged for all those time series in the validation data.

### 3. Results and Discussion

The main contribution of this work is using CGM data from a given population to train prediction models for each individual, instead of using data from longer periods of time provided by the same individual. In this section, the prediction accuracy metrics for our approach are presented.

#### 3.1. Individual vs. Population Predictor

For the population predictor, the number of clusters (and, therefore, the number of local models needed) for each partition predictor are reported in Table 1, where the first row displays the results using the whole 16-week dataset. The optimal number of clusters in the day predictor is independent of the volume of data used for training, always requiring eight clusters for optimum clustering performance. For the individual predictor case, a detailed list of the number of clusters of each partition and each patient is detailed in [18].

**Table 1.** Number of clusters in each predictor in function of the training dataset size.

Dataset Size (weeks)	Number of Clusters		
	Day Predictor	Night Predictor	Hypoglycemic Predictor
16	8	5	8
12	8	5	6
8	8	4	6
6	8	4	6
4	8	3	5
2	8	3	5

The glucose predictor ensembled for the day partition is consistent in finding eight clusters of data independently of the amount of data used to train it, suggesting a homogeneous distribution of easy to identify behaviors across the dataset. The night predictor ranges from three to five clusters as the volume of data increases. The three base clusters may represent common nighttime patterns, which appear consistently even in small datasets. The fourth and fifth clusters, however, are only identified when the data volume grows, suggesting that they are sparsely represented in the data, and several repetitions of those behaviors are required to appear in the data in order to assign them a cluster. A similar conclusion can be drawn from the number of clusters in the hypoglycemic predictor.

As expected from the population approach, the number of clusters needed when clustering the whole dataset is larger than those for any single individual (see [18]), which range from five to eight for the Day Predictor, from three to five in the Night Predictor, and between one and four for the hypoglycemia predictor. The case of the hypoglycemia predictor is particularly interesting, since using data from a population yields from five to eight clusters, more than that of any individual. This leads to the conclusion that hypoglycemia predictors are more readily built with fewer sets of data than those based on data from the individual, due to hypoglycaemic patterns appearing more often among many individuals rather than repeatedly in the same individual.

The accuracy performance of both predictors (trained with individual/population data) are reported in Table 2 for the initial training with 16-week data. The results are reported as mean (standard deviation), and median [25th–75th percentiles]. For each  $PH$ , results for each partition predictor (meal, night, hypoglycemia treatment) as well as overall metrics are presented. In both cases, results show the averaged behaviour over



the 10 patient cohort. The Wilcoxon rank sum test ( $p < 0.001$ ) was used to test statistically significance, marked with a red dot (•).

**Table 2.** Accuracy metrics of the individual and population glucose predictors for the 16-weeks validation dataset.

	Number of Predictions	RMSE Individual (mg/dL)	RMSE Population (mg/dL)
<b>PH = 15 min</b>			
Meals	133,540	6.58 (0.96) 6.54 [6.10, 6.79]	7.58 (1.20) 6.98 [6.84, 7.99]•
Night	26,426	4.90 (0.87) 4.71 [4.29, 5.07]	4.86 (0.87) 4.57 [4.39, 4.87]
Hypo	4308	8.08 (2.97) 6.57 [6.43, 10.80]	8.67 (1.41) 8.79 [7.63, 9.93]
Overall	164,274	6.52 (2.24) 6.37 [5.07, 6.70]	7.04 (1.99) 6.96 [4.87, 8.67]•
<b>PH = 30 min</b>			
Meals	128,144	12.08 (1.81) 11.71 [11.16, 12.33]	13.91 (2.31) 12.68 [12.40, 15.19]•
Night	24,659	7.93 (1.83) 7.42 [6.69, 8.35]	7.86 (1.82) 7.19 [6.66, 7.96]
Hypo	3746	13.95 (4.67) 12.66 [11.30, 14.53]	14.13 (2.80) 14.08 [12.51, 15.91]
Overall	156,549	11.32 (3.92) 11.27 [8.35, 12.65]	11.97 (3.72) 12.49 [7.96, 14.23]•
<b>PH = 60 min</b>			
Meals	117,356	19.49 (3.52) 18.58 [17.43, 20.09]	22.75 (4.28) 20.29 [19.89, 25.56]•
Night	21,142	10.46 (2.95) 9.67 [8.23, 11.08]	10.44 (3.02) 9.45 [8.51, 10.62]
Hypo	2709	23.60 (9.77) 20.94 [17.57, 23.39]	24.10 (7.41) 21.35 [18.81, 30.87]
Overall	141,207	17.85 (8.21) 17.44 [11.08, 20.93]	19.10 (8.04) 19.52 [10.62, 23.37]
<b>PH = 120 min</b>			
Meals	95,780	26.80 (6.03) 24.77 [22.77, 28.19]	30.75 (5.86) 29.63 [26.87, 31.95]•
Night	14,149	12.01 (3.62) 11.24 [8.71, 14.29]	12.59 (4.50) 11.58 [9.61, 12.96]
Hypo	1143	36.64 (21.12) 29.97 [20.72, 42.83]	40.45 (13.64) 36.30 [30.22, 42.10]
Overall	111,072	25.15 (16.12) 22.72 [14.29, 28.19]	27.93 (14.58) 28.79 [12.96, 34.57]•

Table 2. Cont.

	Number of Predictions	RMSE Individual (mg/dL)	RMSE Population (mg/dL)
<b>PH = 180 min</b>			
Meals	74,297	32.14(6.80) 30.31 [27.52, 34.38]	34.92 (6.70) 33.21 [31.40, 35.45]•
Night	7182	13.25 (4.41) 13.32 [9.33, 14.79]	14.89 (5.96) 13.38 [11.63, 16.20]
Hypo	374	41.50 (25.48) 35.28 [29.39, 41.73]	48.39 (15.44) 45.89 [39.24, 59.79]
Overall	81,853	28.96 (19.10) 28.45 [14.79, 34.38]	32.73 (17.17) 32.30 [16.20, 41.59]•
<b>PH = 240 min</b>			
Meals	53,118	34.53 (9.29) 33.93 [27.40, 37.77]	33.28 (7.43) 31.02 [28.80, 35.09]
Night	1384	12.87 (6.05) 10.38 [8.30, 16.19]	15.10 (7.23) 13.72 [9.35, 17.36]
Hypo	117	41.58 (14.18) 43.77 [30.66, 50.34]	55.26 (20.57) 63.35 [47.86, 65.74]
Overall	54,619	27.28 (15.18) 25.58 [13.10, 36.60]	30.40 (18.45) 28.80 [15.13, 37.18]

Results show a small increase in the mean error in the population approach when compared to the individual one in the case of small *PH*s. However, in the worst case scenario, an increase of 4.86 mg/dL is detected when considering the Meals and Night glucose predictors, but this value can not be considered relevant in a medical context. On the other hand, this loss of prediction capability becomes clearer when considering longer *PH* hypoglycemia predictors, where the RMSE increases to 6.89 mg/dL (*PH* = 180 min) and 13.68 mg/dL (*PH* = 240 min). It must be noted that this difference has not been found statistically significant.

However, considering that the population predictors were trained with data from a population and not from data of the very same patient, the prediction errors are remarkably low, especially when considered in a clinical environment. The overall predictor being able to accurately predict glucose 4h (*PH* = 240 min) into the future with a RMSE of 30.40 mg/dL is very promising, especially when the individual predictor only yields errors of 27.28 mg/dL, a difference of 3.12 mg/dL, which is not statistically significant. In general, we observe RMSE numbers of the same order of magnitude for both the individual and population approaches, which, given that population data can be readily used as a “seed” for predictions out-of-the-box without the need of individualization to the patient, suggests that the population approach to clustering and training local models is the way forward.

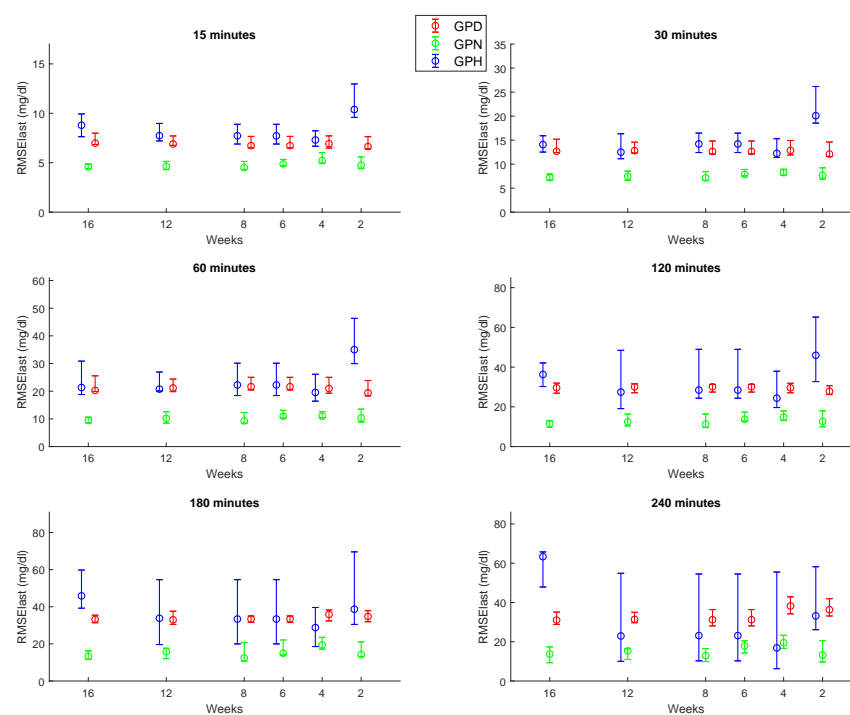
### 3.2. Effect of the Number of Weeks of Data on the Population Predictor

So far, it has been observed that using population data to train local models does not jeopardize prediction accuracy when applied to an individual, compared to training with individual data and assuming the same individual dataset size. Now, the effect of the data size (in number of weeks of data used to train the local models) on the predictions  $\hat{G}_{\mathcal{M}}(i)$ ,  $\hat{G}_{\mathcal{N}}(i)$ , and  $\hat{G}_{\mathcal{H}}(i)$  is analysed. The initial hypothesis was that, as the size of the training dataset is reduced, so is the predictive power of the ensemble of models that comprise the global predictors. To confirm this hypothesis, we evaluate the performance of the predictors

using population sets of  $\{2, 4, 6, 8, 12, 16\}$  weeks of data in  $\Pi_M, \Pi_N, \Pi_H$ , and prediction horizons ranging  $PH \in \{15, 30, 60, 120, 180, 240\}$  min.

The number of clusters for each partition predictor was calculated for each set of identification data (from 16 to 2 weeks) using the FS index (see Section 2.4, and results are reported in Table 1). As can be seen in Table 1, the ideal number of clusters is lower in scenarios for smaller numbers of weeks, which is expected since the volume of data is lower. This behaviour affects particularly to the number of clusters in the Night and Hypoglycemia predictor, since the number of these events decrease drastically in smaller datasets.

The same validation procedure followed in Section 3.1 was used here to compare the predictive power of the different glucose predictors when trained with different population data volumes. Figure 3 shows the distribution of the prediction error (RMSE median and quartiles) vs the different training dataset sizes for all the prediction horizons.



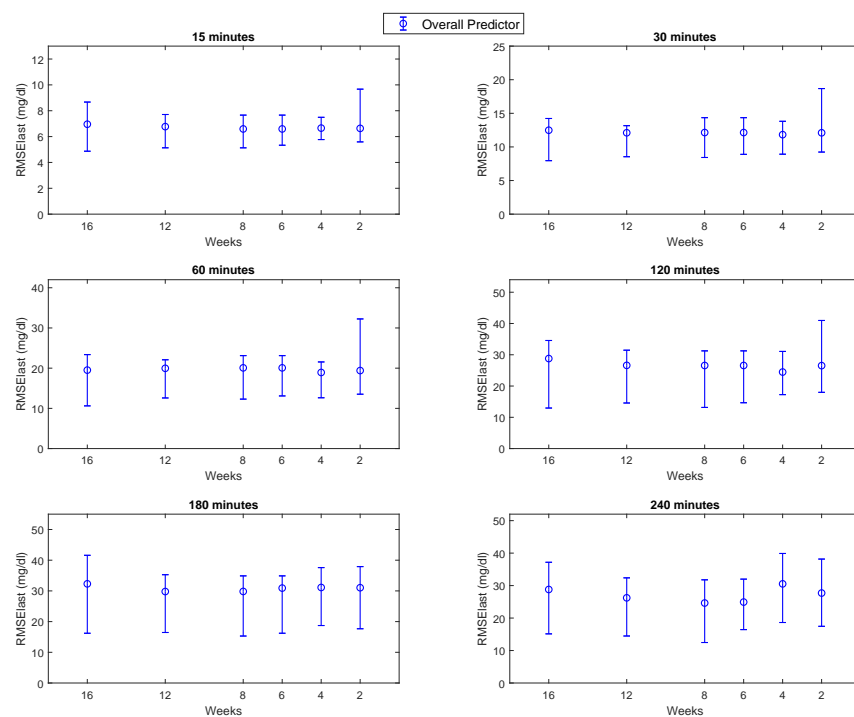
**Figure 3.** RMSE Median and Quartiles metrics for the different Glucose Predictor for all prediction horizons for all the training dataset sizes. Hypoglycemia predictor in blue, Night predictor in green, and Meal predictor in red.

For all considered PHs, the hypoglycemia glucose predictor is clearly the most erratic (larger error bars) and on average, the one that shows higher median errors. It can also be observed that for low PHs ( $PH \in \{15, 30, 60\}$ ) no significant difference is observable in the RMSE, except for an abnormally high error when only 2 weeks of data were used for training. For higher PHs ( $PH \in \{120, 180, 240\}$ ), the RMSE remains almost constant when training with 4, 6, 8 or 12 weeks of data, but is worse for 2 or 16 weeks, suggesting a “sweet spot” of training data volume between 4 and 12 weeks of data. The most plausible explanation for this phenomenon is a tendency of over-fitting anomalous or highly variable behaviors in the scenario with 16 weeks of data, and a lack of relevant dynamics in the 2 weeks of data.

The day and night time predictors show a very consistent predictive power, with little to no difference at low PHs when using different dataset volumes for training, and a slight trend of decreasing RMSE as the number of weeks increases for longer PHs.

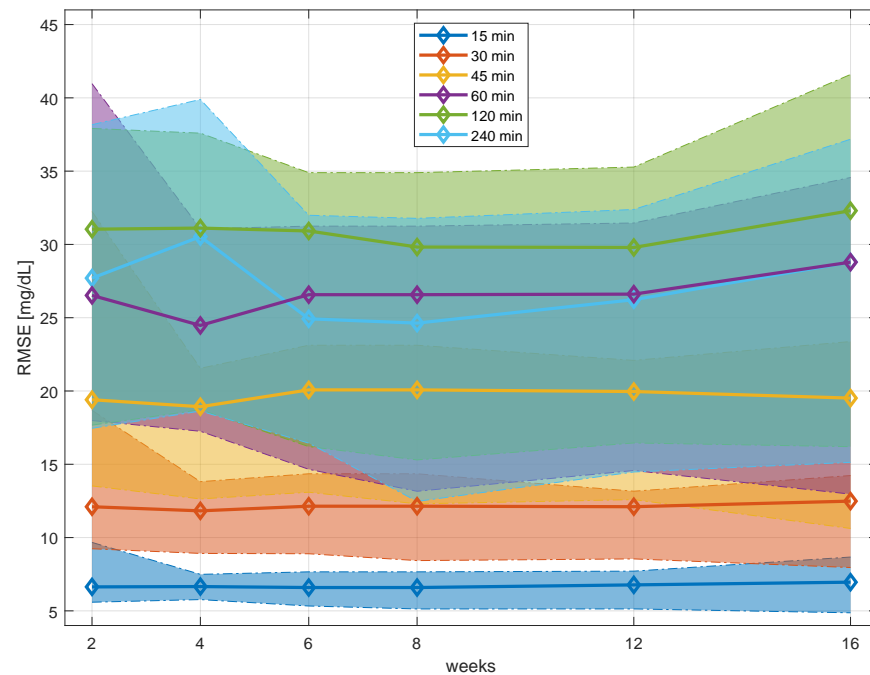
This is concordant with the initial hypothesis of the effect of the dataset volume on the predictive power.

Similar trends are observed if the Overall prediction error is analysed, as seen in Figure 4. For the shorter PHs ( $PH \in \{15, 30, 60, 120\}$ ), the error bars of the two-week dataset are significantly larger than all the other (larger) datasets used for training, suggesting that a critical volume of training data for obtaining consistently good predictions is more than two weeks. For longer PHs, lower RMSE values tend to be achieved when larger datasets are used for training (6, 8 or 12 weeks of data vs. 2 or 4), but this trend is inverted when the dataset volume is increased to 16 weeks. This suggests that using 16 weeks of data from a whole cohort of patients to train models might lead to subpar predictions, since more dynamics are modeled that may not be used to provide accurate predictions of an individual patient.



**Figure 4.** RMSE Median and Quartiles metrics for the Overall for all prediction horizons for all the training dataset sizes.

This is also observed in Figure 5, where the median and interquartile ranges for all the Overall predictors are illustrated overlapping. Median RMSE remains mostly unchanged for low PHs, and it reaches its minimum values for 4, 6 and 8 weeks for longer PHs. Interquartile range drastically increases for the 2 and 16 weeks datasets. When comparing the RMSE between PHs, it is curious to observe how better predictions are achieved for the 240 minutes prediction vs the 120 minutes one.



**Figure 5.** Median and inter-quartile range trends in RMSE for different PHs.

#### 4. Conclusions

In previous works, seasonal local modeling proved successful in glucose prediction under realistic scenarios for long prediction horizons. However, a minimum of 4 months of data for each patient were needed to properly train the system personalized to the user. The predictive power of these previously tested individualized glucose predictors has been compared to those obtained by a single model for all patients trained using population data. The results show that there is no loss of predictive power when population and individual approaches are compared. The population approach was easier and faster to implement and train, as only three different glucose predictors were trained for the whole cohort of patients, instead of the thirty glucose predictors that the individual approach requires. Additionally, the use of population data avoids the necessity of using specific data from a patient and transfers knowledge from a generic population to a specific patient by learning common behaviours (local models) useful for all. While the prediction accuracy metrics reported are promising, in order to extrapolate these results to a general population, these models will have to be assessed in a randomized experimental setting.

In summary, it was observed that no significant differences were found between using 16, 12 and 8 weeks for training the population glucose predictors when using them to forecast individual data. Therefore, using an ensemble of data from diverse patients reduces the amount of data required to train models by half, without losing any predictive power.

**Author Contributions:** Conceptualization, J.-L.D. and J.B.; methodology, J.-L.D., A.J.L.S. and J.B.; software, A.A., J.-L.D. and J.B.; validation, A.A., J.-L.D., A.J.L.S. and J.B. formal analysis, J.-L.D., A.J.L.S. and J.B.; investigation, J.-L.D., A.J.L.S. and J.B.; resources, J.-L.D. and J.B.; data curation, A.A., J.-L.D., A.J.L.S. and J.B.; writing—original draft preparation, A.A. and J.-L.D.; writing—review and editing, J.-L.D., A.J.L.S. and J.B.; visualization, A.A., J.-L.D., A.J.L.S. and J.B.; supervision, J.-L.D. and J.B.; project administration, J.-L.D. and J.B.; funding acquisition, J.-L.D. and J.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by grant PID2019-107722RB-C21 funded by MCIN/AEI/10.13039/501100011033, by grant CIPROM/2021/012 funded by Conselleria de Innovación, Universidades, Ciencia y Sociedad Digital from Generalitat Valenciana, by CIBER -Consortio Centro de Investigación Biomédica en Red- group number CB17/08/00004, Instituto de Salud Carlos III, Ministerio de Ciencia e Innovación and by European Union–European Regional Development Fund.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The simulated datasets used in this study are available on request from the corresponding author.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

- International Diabetes Federation. *IDF Diabetes Atlas*, 10th ed.; International Diabetes Federation: Brussels, Belgium, 2021.
- Hovorka, R. Continuous glucose monitoring and closed-loop systems. *Diabet. Med.* **2006**, *23*, 1–12. [[CrossRef](#)] [[PubMed](#)]
- Palerm, C.C.; Willis, J.P.; Desemone, J.; Bequette, B.W. Hypoglycemia prediction and detection using optimal estimation. *Diabetes Technol. Ther.* **2005**, *7*, 3–14. [[CrossRef](#)] [[PubMed](#)]
- Li, L.; Xie, X.; Yang, J. A predictive model incorporating the change detection and Winsorization methods for alerting hypoglycemia and hyperglycemia. *Med. Biol. Eng. Comput.* **2021**, *59*, 2311–2324. [[CrossRef](#)]
- Forlenza, G.P.; Li, Z.; Buckingham, B.A.; Pinsky, J.E.; Cengiz, E.; Wadwa, R.P.; Ekhlaspour, L.; Church, M.M.; Weinzimer, S.A.; Jost, E.; et al. Predictive low-glucose suspend reduces hypoglycemia in adults, adolescents, and children with type 1 diabetes in an at-home randomized crossover study: Results of the PROLOG trial. *Diabetes Care* **2018**, *41*, 2155–2161. [[CrossRef](#)]
- Haidar, A. The artificial pancreas: How closed-loop control is revolutionizing diabetes. *IEEE Control Syst. Mag.* **2016**, *36*, 28–47.
- Kovatchev, B.; Tamborlane, W.V.; Cefalu, W.T.; Cobelli, C. The artificial pancreas in 2016: A digital treatment ecosystem for diabetes. *Diabetes Care* **2016**, *39*, 1123–1126. [[CrossRef](#)] [[PubMed](#)]
- Trevitt, S.; Simpson, S.; Wood, A. Artificial pancreas device systems for the closed-loop control of type 1 diabetes: What systems are in development? *J. Diabetes Sci. Technol.* **2016**, *10*, 714–723. [[CrossRef](#)]
- Bergental, R.M.; Garg, S.; Weinzimer, S.A.; Buckingham, B.A.; Bode, B.W.; Tamborlane, W.V.; Kaufman, F.R. Safety of a hybrid closed-loop insulin delivery system in patients with type 1 diabetes. *JAMA* **2016**, *316*, 1407–1408. [[CrossRef](#)]
- Georga, E.I.; Príncipe, J.C.K.; Fotiadis, D.I. Short-term prediction of glucose in type 1 diabetes using kernel adaptive filter. *Med. Biol. Eng. Comput.* **2019**, *57*, 27–46. [[CrossRef](#)]
- Hovorka, R.; Canonico, V.; Chassin, L.J.; Haueter, U.; Massi-Benedetti, M.; Federici, M.O.; Pieber, T.R.; Schaller, H.C.; Schaupp, L.; Vering, T.; et al. Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes. *Physiol. Meas.* **2004**, *25*, 905. [[CrossRef](#)]
- Oviedo, S.; Vehí, J.; Calm, R.; Armengol, J. A review of personalized blood glucose prediction strategies for T1DM patients. *Int. J. Numer. Methods Biomed. Eng.* **2017**, *33*, e2833. [[CrossRef](#)]
- Li, K.; Daniels, J.; Liu, C.; Herrero-Vinas, P.; Georgiou, P. Convolutional Recurrent Neural Networks for Glucose Prediction. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 603–613. [[CrossRef](#)]
- Zhu, T.; Li, K.; Chen, J.; Herrero, P.; Georgiou, P. Dilated Recurrent Neural Networks for Glucose Forecasting in Type 1 Diabetes. *J. Healthc. Inform. Res.* **2020**, *4*, 308–324. [[CrossRef](#)] [[PubMed](#)]
- Langarica, S.; Rodriguez-Fernandez, M.; Núñez, F.; Doyle, F. A meta-learning approach to personalized blood glucose prediction in type 1 diabetes. *Control Eng. Pract.* **2023**, *135*, 105498. [[CrossRef](#)]
- Montaser, E.; Díez, J.L.; Bondia, J. Stochastic Seasonal Models for Glucose Prediction in the Artificial Pancreas. *J. Diabetes Sci. Technol.* **2017**, *11*, 1124–1131. [[CrossRef](#)] [[PubMed](#)]
- Montaser, E.; Díez, J.L.; Rossetti, P.; Rashid, M.; Cinar, A.; Bondia, J. Seasonal Local Models for Glucose Prediction in Type 1 Diabetes. *J. Biomed. Health Inform.* **2020**, *24*, 2064–2072. [[CrossRef](#)] [[PubMed](#)]
- Montaser, E.; Díez, J.-L.; Bondia, J. Glucose Prediction under Variable-Length Time-Stamped Daily Events: A Seasonal Stochastic Local Modeling Framework. *Sensors* **2021**, *21*, 3188. [[CrossRef](#)]
- Prendin, F.; Díez, J.L.; del Favero, S.; Sparacino, G.; Facchinetti, A.; Bondia, J. Assessment of Seasonal Stochastic Local Models for Glucose Prediction without Meal Size Information under Free-Living Conditions. *Sensors* **2022**, *22*, 8682. [[CrossRef](#)]
- León-Vargas, F.; Martin, C.; Garcia-Jaramillo, M.; Aldea, A.; Leal, Y.; Herrero, P.; Reyes, A.; Henao, D.; Gomez, A. Is a cloud-based platform useful for diabetes management in Colombia? The Tidepool experience. *Comput. Methods Programs Biomed.* **2021**, *208*, 106205. [[CrossRef](#)]
- Vehí, J.; Regincós Isern, J.; Parcerisas, A.; Calm, R.; Contreras, I. Impact of Use Frequency of a Mobile Diabetes Management App on Blood Glucose Control: Evaluation Study. *JMIR Mhealth Uhealth* **2019**, *7*, e11933. [[CrossRef](#)]



22. Aleppo, G.; Ruedy, K.; Riddlesworth, T.; Kruger, D.; Peters, A.; Hirsch, I.; Bergenstal, R.; Toschi, E.; Ahmann, A.; Shah, V.; et al. REPLACE-BG: A Randomized Trial Comparing Continuous Glucose Monitoring With and Without Routine Blood Glucose Monitoring in Adults With Well-Controlled Type 1 Diabetes. *Diabetes Care* **2017**, *40*, 538–545. [[CrossRef](#)] [[PubMed](#)]
23. Dalla Man, C.; Micheletto, F.; Lv, D.; Breton, M.; Kovatchev, B.; Cobelli, C. The UVA/PADOVA type 1 diabetes simulator: New features. *J. Diabetes Sci. Technol.* **2014**, *8*, 26–34.
24. Schiavon, M.; Man, C.D.; Kudva, Y.C.; Basu, A.; Cobelli, C. In silico optimization of basal insulin infusion rate during exercise: Implication for artificial pancreas. *J. Diabetes Sci. Technol.* **2013**, *7*, 1461–1469. [[CrossRef](#)] [[PubMed](#)]
25. Brazeau, A.S.; Mircescu, H.; Desjardins, K.; Leroux, C.; Strychar, I.; Ekoé, J.M.; Rabasa-Lhoret, R. Carbohydrate counting accuracy and blood glucose variability in adults with type 1 diabetes. *Diabetes Res. Clin. Pract.* **2013**, *99*, 19–23. [[CrossRef](#)]
26. Haidar, A.; Elleri, D.; Kumareswaran, K.; Leelarathna, L.; Allen, J.M.; Caldwell, K.; Murphy, H.R.; Wilinska, M.E.; Acerini, C.L.; Evans, M.L.; et al. Pharmacokinetics of insulin aspart in pump-treated subjects with type 1 diabetes: Reproducibility and effect of age, weight, and duration of diabetes. *Diabetes Care* **2013**, *36*, e173–e174. [[CrossRef](#)]
27. Scheiner, G.; Boyer, B. Characteristics of basal insulin requirements by age and gender in Type-1 diabetes patients using insulin pump therapy. *Diabetes Res. Clin. Pract.* **2005**, *69*, 14–21. [[CrossRef](#)]
28. American Diabetes Association. Standards of Medical Care in Diabetes. *Diabetes Care*. **2014**, *1*, S14–S80.
29. Bezdek, J.C. *Pattern Recognition with Fuzzy Objective Function Algorithms*; Advanced Applications in Pattern Recognition Series; Springer: New York, NY, USA, 1981.
30. Dixon, J.K. Pattern recognition with partly missing data. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 617–621. [[CrossRef](#)]
31. Wang, W.; Zhang, Y. On fuzzy cluster validity indices. *Fuzzy Sets Syst.* **2007**, *158*, 2095–2117. [[CrossRef](#)]
32. Box, G.E.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*; John Wiley & Sons: Hoboken, NJ, USA, 2015.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.