The final publication is available at

https://doi.org/10.1002/cmm4.1064

Additional Information

A computational technique to predict the level of glucose of a diabetic patient with uncertainty in the short term

Clara Burgos, Carlos Cervigón, José-Ignacio Hidalgo, Rafael-J. Villanueva

Abstract

One of the most important problems in a diabetic patient are the sudden changes he/she can suffer in the levels of glucose. These changes, which are characteristic of the disease, may lead to acute situations, even to death, if extreme values are reached. Therefore, the control of the patient maintaining reasonable levels of glucose is the goal of the physicians. However, the prediction in advance of the levels of glucose becomes an issue of paramount importance as well as a very difficult task because of the mentioned sudden changes. If we were able to give reliable predictions in advance, it would facilitate the making of therapeutic decisions to control the disease and improve the patient's health. As a first approach in this sense, in this work we present a technique to estimate the glucose level of a second kind diabetic patient at different time instants, capturing the measurement errors produced by the device. To do so, we use a model relying on a system of difference equations and we calibrate it with the aim to capture the glucose level data of the patient and the uncertainty given by the measurement error. Then, we use the calibrated parameters to predict the levels of glucose over the next 15 minutes. To make this procedure dynamic, then, we reject some of the first data values and feed the model with new ones, then we calibrate the model again and predict over the next 15 minutes. Repeating this procedure, we are able to give good short term predictions and adapt the predictions when the sudden changes appear.

Keywords Diabetic Patient, Glucose Level, Mathematical Model, Uncertainty Quantification, Model Prediction.

1   Introduction

Diabetes Mellitus, usually referred as *Diabetes*, is a group of metabolic disorders related with the level of glucose in the blood. They are produced when the pancreas can not produce enough or effective insulin, consequently, the ingested glucose is not digested correctly and there is an increase in the level of glucose in the blood. People suffering this disorder are most likely to have cardiovascular

illnesses. Therefore, it is essential to control the glucose level in order to avoid serious health problems.

Depending on the amount and the efficacy of the insulin, there are two different kinds of diabetes: first and second kind. The former occurs when the pancreas can not produce enough insulin. It is usually diagnosed in people younger than 35 years old. The latter happens when the pancreas produces enough insulin but the cells are unable to absorb it and becomes ineffective. It usually appears in middle-age people. Here, we are going to focus on patients with diabetes of second kind.

There are in the market electronic devices that measure the glucose levels over the time. With these gadgets, it is possible to monitor the trend of the patient's glucose levels. Nevertheless, these devices return measurements with an intrinsic error of about 5% and it is important to take them into account if the goal is to provide a reliable control of the patient.

Monitoring the patient allows the physicians to give appropriate health recommendations as to increase or decrease the injected doses of insulin, to make changes in the diet or to do some exercise. These recommendations would be more helpful if we were able to predict the levels of glucose over appropriate periods of time. This would permit the physician to maintain the patient's glucose levels in healthy ranges.

However, the prediction in advance of the levels of glucose becomes a very difficult task because these levels use to change suddenly and largely in diabetic patients. Nevertheless, taking into account the paramount importance of this problem in the improvement and the benefit of the health of diabetic patients, in this paper, we are going to propose a first approach with the aim to give short time (15 minutes) predictions.

To address the problem, we start considering a discrete version of the model introduced in [3] that we are going to adapt for a particular patient. This model is developed with the objective of achieving control of postprandial glucose levels and it contains six unknown parameters related to the bioavailability of the meal, the insulin absorption and sensitivity, and the endogenous glucose production and effectiveness at zero insulin. Although [3] presents a minimal model, we can characterize it for each patient, and the parameters of the model can be associated to values used in the daily therapy. Also, this models has been used to study the evolution of the levels of glucose in particular patients [1].

Then, taking 30 minutes data of the patient's levels of glucose, we calibrate the model considering the data uncertainty. With the model parameters obtained in the calibration, we predict the levels of glucose over the next 15 minutes. After 15 minutes, we discard the first 15 minutes of glucose levels data and add the new 15 minutes provided by the device attached to the patient. Again, we calibrate the model and predict over the next 15 minutes. The procedure can be repeated several times to have updated predictions that can be adapted to the possible sudden changes in the levels of glucose.

A similar technique, where we feedback the model with new data discarding the old data, has been applied with success previously [2] to predict the number of downloads of an app. Here, we are going to adapt some ideas from [2] and

introduce some new algorithms in order to provide reliable predictions.

This paper is organized as follows. In Section 2 we introduce the model, a discrete version of the one presented in [3]. In Section 3, data and their uncertainty treatment to include the measurement errors are shown. In Section 4 the model calibration and the model prediction procedure is described. Results are depicted in Section 5. Finally, Section 6 is devoted to conclusions.

2   Model description and data uncertainty treatment

In that section we are going to present the model we are going to work with, in order to describe the glucose behavior over the time of a second kind diabetic patient. It is an adaptation of the one presented in [3] described in [1] and it is based on a system of five difference equations

$$U_{t+1} = U_t + V_t, \tag{1}$$

$$V_{t+1} = V_t - 2a_g V_t - a_g^2 U_t + K_g a_g^2 Ch_t, \tag{2}$$

$$G_{t+1} = G_t - X_t G_t - S_{g_0} G_t + U_{endo} + C\frac{U_t}{M}, \tag{3}$$

$$X_{t+1} = X_t - a_x X_t + a_x X_t^1, \tag{4}$$

$$X_{t+1}^1 = X_t^1 - a_x X_t^1 + K_x a_x \frac{I_t}{M}, \tag{5}$$

where U_t represents the gut glucose absorption at time t, V_t is the variation rate of the gut glucose absorption at time t, G_t stands for the level of Glucose at time t, X_t is the insulin action and X_t^1 represents the intermediate insulin action at time t.

The parameters of the model are related with the daily clinics of the patient, is to say, they are related with the biology of the patient. Ch_t and I_t are the level of ingested carbohydrates and insulin, respectively, C is the constant $50/9$, M is the weight of the patient. Also a_g is the inverse of the meal time constant, K_g is the unitless bioavalibility of the meal of interest, S_{g_0} is the glucose effectiveness at zero insulin, U_{endo} is the insulin independent endogenous glucose production, a_x is the inverse of the insulin absorption/action time constant and K_x is the insulin sensitivity [1].

In Table 1 we have collected the units of the model variables, parameters and constants.

3   Data and uncertainty treatment

Once we have introduced the model, we consider the levels of glucose of a specific patient given by $\mathbf{g}_t$ for the time instants $t = 1, 2, \ldots n$. These glucose levels have been measured with an electronic device with 5% of measurement error

Variables			Constants		Parameters	
U_t	g/min	Ch_t	g/min		a_g	$1/min$
V_t	g/min^2	I_t	mU/min		K_g	unitless
G_t	mg/dl	C	$mg \times Kg/(dl \times g)$		S_{g_0}	$1/min$
X_t	$1/min$	M	Kg		U_{endo}	mg/dl
X_t^1	$1/min$				a_x	$1/min$
					K_x	kg/mU

Table 1: Units of the model variables, constants and parameters.

described in the technical specifications. As we mentioned before, we want that our method also captures the measurement error in order to avoid undesired effects produced by non considering the behavior of the electronic device. To this end we are going to assume that each glucose level is represented by a random variable instead a single value.

The measurement error of a device is usually modeled as a Gaussian Random Variable $N(\mu, \sigma)$ with mean μ and the standard deviation σ. Therefore, we assume that each glucose level is given by a Gaussian Random Variable with mean the glucose value $\mathbf{g}_t$ and standard deviation $0.05 \times \mathbf{g}_t$, that is, $N(\mathbf{g}_t, 0.05\mathbf{g}_t)$, $t = 1, 2, \ldots n$.

Then, we compute the percentiles 2.5 and 97.5 of $N(\mathbf{g}_t, 0.05 \times \mathbf{g}_t)$, $t = 1, 2, \ldots n$, given in this case by $LP_t = 0.902\mathbf{g}_t$ and $UP_t = 1.098\mathbf{g}_t$, respectively. Thus, we have the 95% confidence intervals of the data (CI95%), for the particular patient and these CI95% can be seen graphically in the Figure 1. These CI95% $= [0.902\mathbf{g}_t, 1.098\mathbf{g}_t]$, $t = 1, 2, \ldots n$, determine the data uncertainty and will allow us to seek the sets of parameters that best capture these data uncertainty.

4 Procedure design

Once we have presented the model and the data available, we are going to describe a technique to determine the sets of parameters that capture the data together with their uncertainty. Remember that our initial purpose is to model the glucose behavior to detect, in advance, if there is going to be an increasing or a decreasing of glucose level in order to administrate the adequate dose of insulin or to give proper health recommendations.

To do that, we consider the following procedure,

Step 1: we take the know glucose levels of a given patient during 30 minutes, $\mathbf{g}_{t_1}, \ldots, \mathbf{g}_{t_{30}}$;

Step 2: we perform several model calibrations, considering as a fitness function the one that returns 0 if the model output of levels of the glucose are inside the CI95% of the data $[0.902\mathbf{g}_{t_i}, 1.098\mathbf{g}_{t_i}]$, $i = 1, \ldots, 30$;
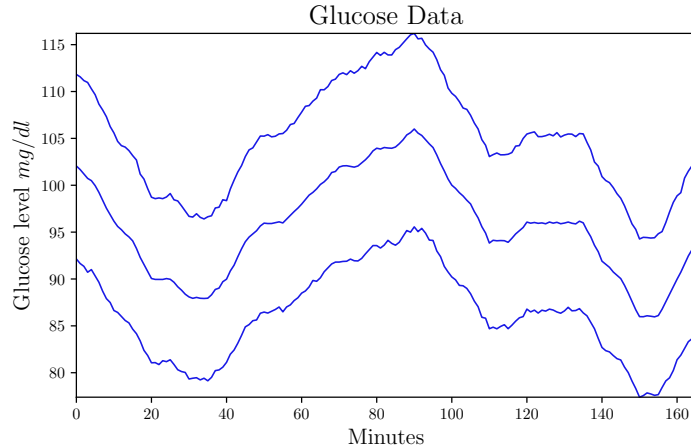
4

Figure 1: Glucose levels of the selected patient (in the middle) with the 95% confidence intervals given by the measurement error of the device (data uncertainty).

Step 3: among all the model evaluations performed during the calibrations, we select the model parameters such that their CI95% model output in each time instant t_i, $i = 1, \ldots, 30$, be as close as possible as $[0.902\mathbf{g}_{t_i}, 1.098\mathbf{g}_{t_i}]$, $i = 1, \ldots, 30$ (capture the data uncertainty of the levels of glucose);

Step 4: then, we substitute the selected model parameters into the model and obtain the model output over the next 15 minutes, that is, for the time instants G_{t_i}, $i = 31, \ldots, 45$;

Step 5: for every time instant over the next 15 minute, we calculate the percentiles 2.5 and 97.5 of the model output to obtain the prediction with uncertainty, that is, the CI95% of model output levels of glucose over the next 15 minutes.

The above procedure can be repeated every 15 minutes if, in the data in the Step 1, we reject the first 15 minutes of glucose levels and add the new 15 minutes. The data feedback makes that the predictions keep the accuracy.

Even though this kind of feedback techniques have been successfully used in other papers [2] to predict the number of downloads in apps over longer periods, here we decided to take short times (30 minutes) of data to predict over shorter times (15 minutes) due to the variability of the levels of glucose in diabetic patients.

5

Once the 30 minutes data $\mathbf{g}_{t_1}, \ldots, \mathbf{g}_{t_{30}}$ have been set (Step 1) jointly their uncertainty given by the intervals $[LP_t, UP_t] = [0.902\mathbf{g}_t, 1.098\mathbf{g}_t]$, $t = 1, 2, \ldots, 30$, we have to define the fitness function to perform the calibrations.

Thus, given a set of parameters by $Par = (a_g, K_g, S_{g_0}, U_{\text{endo}}, a_x, K_x)$, and denoting as G_t^{Par} the level of glucose returned by the model in the time instant t for the model parameters Par, we define the fitness function as

$$FF(Par) = \sum_{i=1}^{30} \tau(G_{t_i}^{Par}), \tag{6}$$

where

$$\tau(G_t^{Par}) = \begin{cases} 0, & \text{if } LP_t \leq G_t^{Par} \leq UP_t, \\ \min\left\{G_t^{Par} - LP_t, UP_t - G_t^{Par}\right\} & \text{otherwise.} \end{cases} \tag{7}$$

That is, the function τ is zero if the glucose level returned by the model lies inside the 95% confidence interval, see Figure 2(a). Otherwise, the minimum of the distance between the glucose level and each percentile, see Figure 2(b).
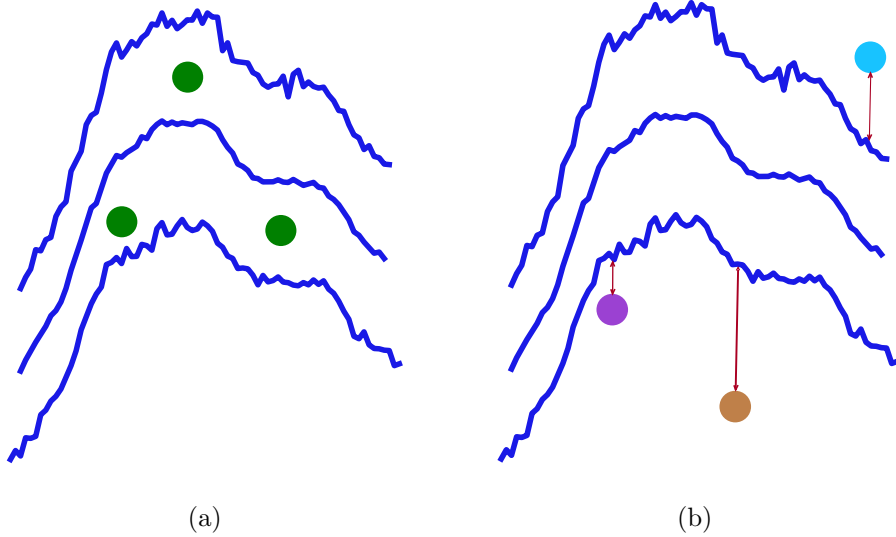

(a)   (b)

Figure 2: (a) FF is zero if the glucose level, green dots, lie inside the CI95%, blue lines. (b) FF is the minimum distance between the glucose level, colored dots, and the CI95%, blue lines, if they are not inside CI95%.

With the fitness function FF, we calibrate the model using the rPSO algorithm [4] (Step 2), several times. All the evaluations performed during the calibrations, are stored as follows: in the file called *Parameters.txt* we store, in rows, the model parameters Par; in the file called *model_output.txt* we store

the model output corresponding to parameters Par in the time instants t_i, $i = 1, \ldots, 30$, in the same order as the model parameters are in the file *Parameters.txt*; in the same way, we store in the file *fitnesses.txt*, the fitnesses $FF(Par)$.

We now turn to Step 3. In order to select properly the model outputs that best capture the data uncertainty, we need to design a specific algorithm. This algorithm is based on the general PSO algorithm and an *ad hoc* new fitness function. First, we are going to call $\Theta(i)$ the vector made up of the 30 model outputs stored in the row i of the file *model_output.txt*, for $i = 1, \ldots, M$, where M is the number of evaluations performed during the calibration and stored in the file *model_output.txt*. Thus, we define the following fitness function F:

INPUT: k, a value between 2 and M. $I = \{i_1, \ldots, i_k\}$, a index subset of k elements of $\{1, \ldots, M\}$.

Step 1. Select the model output vectors $\Theta(i_1), \ldots, \Theta(i_k)$ and calculate the mean, percentile 2.5 and percentile 97.5 for each time instant.

Step 2. Calculate the norm of the difference between the mean, percentile 2.5 and percentile 97.5 of the k model output vectors obtained in Step 2 and the data, and sum them up.

Then, the selection algorithm to choose the k model outputs the best capture the data uncertainty, is given by:

1. Initialization.

   - Initialize P index subsets $S_1, \ldots, S_P$ with k elements of the set $\{1, \ldots, M\}$ (particles) chosen randomly without repetition. Evaluate the fitness of all the particles $F(S_1), \ldots, F(S_P)$.
   - Define the individual best fitness as $S_i^{best} = S_i$, $i = 1, \ldots, P$ and the global best fitness S_{global}^{best} as the S_i^{best} which fitness is the minimum.

2. For $i = 1, \ldots, P$, we extract without repetition k indexes from the union of the current particle S_i, its individual best S_i^{best} and the global best S_{global}^{best}, and we denote it as the new S_i. Evaluate the fitness of all the new particles $F(S_1), \ldots, F(S_P)$.

3. Update the individual best fitness $S_i^{best}, i = 1, \ldots, P$ and the global best fitness S_{global}^{best}. Go to Step 2.

In the above algorithm, we consider 10% of randomness (with a probability of 10%, we reject the particle and generate another one randomly) and 10% of mutation when updating new particles (in this case, mutation consists of changing some of the indexes in the current particle by other randomly chosen indexes, avoiding repetitions).

The S_{global}^{best} returned by the last iteration of the algorithm will correspond to the k indexes $\{j_1, \ldots, j_k\}$ which 95% confidence interval of the model outputs $\Theta(j_1), \ldots, \Theta(j_k)$ capture the best the data uncertainty.

The model parameters of the rows $j_1, \ldots, j_k$ in the file *Parameters.txt* are the model parameters which model output is $\Theta(j_1), \ldots, \Theta(j_k)$. Now, in Step 4, we take the selected model parameters, we substitute into the model and obtain the model output in the time instants t_i, for $i = 31, 32, \ldots, 45$, that is, the 15 minutes prediction. Finally, in Step 5 we calculate the mean and the CI95% of the prediction model outputs given in Step 4. These mean and the CI95% constitutes the model prediction with uncertainty over the next 15 minutes.

As we mentioned before, after 15 minutes, we can reject the first 15 minutes of data and add the new 15 minutes of patient's levels of glucose, and repeat the Steps 2 to Step 5 in order to provide a new model prediction with uncertainty over the next 15 minutes. And so on.

5   Results

For this experiment, we consider 160 data of levels of glucose of a patient, one per minute (2 hours and 40 minutes), shown in Figure 1. Thus, we take the first 30 values and their uncertainty (Step 1). We perform 5 calibrations using rPSO with 60 particles, 300 iterations and fitness function FF given by (6) (Step 2). We store all the 90000 evaluations. In order to decrease the number of eligible evaluations, we take, among them, those with fitness less than 70, reducing the number until about 40000. Then, let us apply the selections algorithm with 30 particles, 1000 iterations and $k = 100$ the number of outputs selected (Step 3). Now, the $k = 100$ model parameters related to the outputs selected, are substituted into the model and we obtain the model output for the next 15 minutes, that is, from minute 31 to 45 (Step 4). Hence, calculating the CI95% of the model outputs from minute 31 to 45, we have the 15 minutes prediction with uncertainty (Step 5) as shown in the first graph of Figure 3.

Now, rejecting the first 15 minutes of data, adding 15 new minutes more of data and repeating the above procedure, we obtain the calibration and prediction shown in the second graph of Figure 3.

In the Figures 3, 4 and 5 we can see the results of the calibration and prediction with uncertainty feeding the model with new data every 15 minutes.

In every graphs, we show the previous calibrations with uncertainty, in intervals of 15 minutes, between the green vertical lines, the new one between the rightest green vertical line and the red vertical line, being the part of the graph on the right of the red line the prediction with uncertainty.

It can be seen that the calibration and the predictions are accurate and, when there is a sudden change in the levels of glucose (for instance, in the last 2 graphs of Figure 5), the prediction adapts to change with a minimum delay.

6   Conclusions

Predicting the glucose level for a diabetic patient is not a easy task since there are sudden changes in the glucose levels consequence of the insulin sensitivity

or the endogenous production of glucose. The insulin sensitivity control the sudden changes of glucose and it is difficult to estimate. If it does not work correctly there may be disorders in the glucose level and consequently, the diabetic patients have more risk to suffer cardiovascular illnesses.

In this paper, we use an adapted model based upon a system of difference equations to fit the data of the glucose level of a patient and capture the uncertainty of these data due to the measurement error of the device. The obtained parameters in the model calibration procedure and their values, may help us to understand the condition of a patient because the model parameters are related to the daily therapy. Also, the obtained parameters allow us to predict the evolution of the glucose levels over the next 15 minutes. Feeding the model with new glucose level data and rejecting the old ones and repeating the process, we have a computational procedure that permits to provide, every 15 minutes, updated predictions with the ability to adapt to the possible sudden changes in the levels of glucose of the patient.

In future works, we would like to improve our prediction for linger times with good accuracy, adding to the prediction known previous behavior of the patient's glucose level in similar situations to the ones we want to predict.

Acknowledgements

References

[1] Luis Acedo, Marta Botella, Juan-Carlos Cortés, José-Ignacio Hidalgo, Esther Maqueda, and Rafael-Jacinto Villanueva. Swarm hybrid optimization for a piecewise model fitting applied to a glucose model. *Journal of Systems and Information Technology*, Accepted, 2018.

[2] Juan Alegre-Sanahuja, Juan-Carlos Cortés, Rafael-Jacinto Villanueva, and Francisco-José Santonja. Predicting mobile apps spread: An epidemiological random network modeling approach. *SIMULATION*, 94(2):123–130, jun 2017.

[3] Alain Bock, Grégory François, and Denis Gillet. A therapy parameter-based model for predicting blood glucose concentrations in patients with type 1 diabetes. *Computer Methods and Programs in Biomedicine*, 118(2):107–123, feb 2015.

[4] Namrata Khemka and Christian Jacob. Exploratory toolkit for evolutionary and swarm-based optimization. *The Mathematica Journal*, 11(3):376–391, feb 2010.
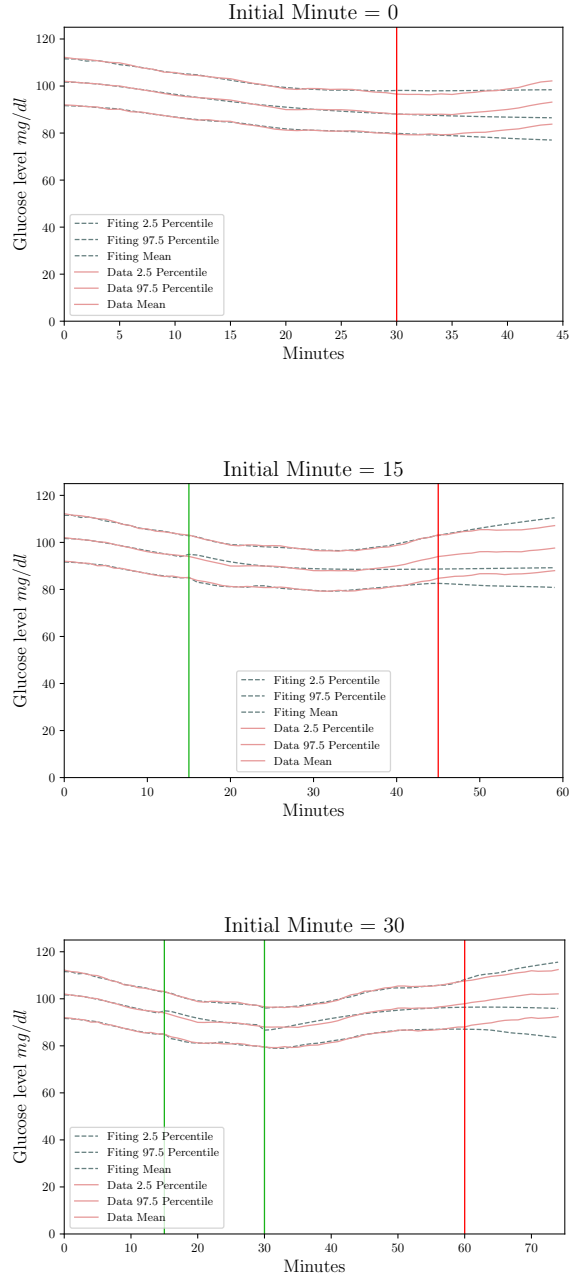
Figure 3: Calibration and prediction with uncertainty using 30 minutes of glucose data starting in the minutes 0, 15 and 30. Calibration is performed from 0 or the rightest green vertical line until the red vertical line, where the prediction starts. The calibration and the prediction are very accurate.
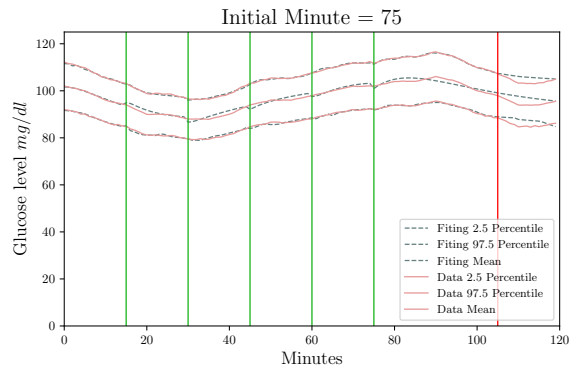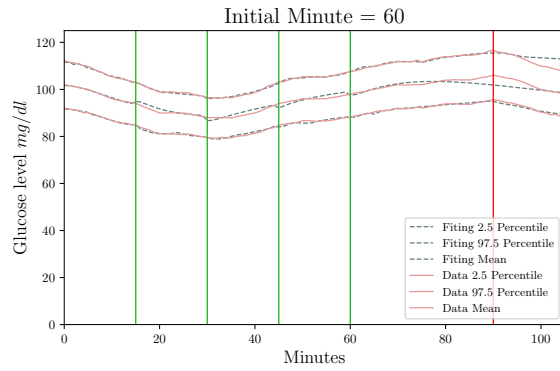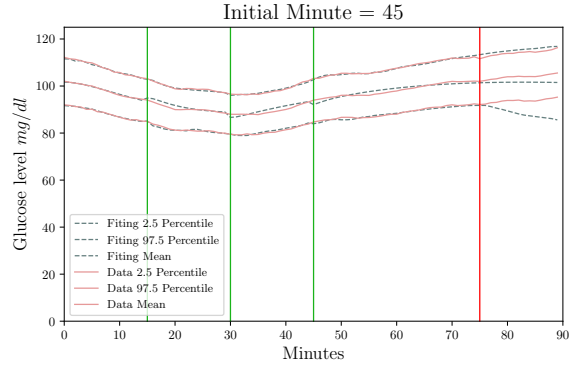
10

Figure 4: Calibration and prediction with uncertainty using 30 minutes of glucose data starting in the minutes 45, 60 and 75. Calibration is performed with the glucose values between the rightest green vertical line and the red vertical line, where the prediction starts.
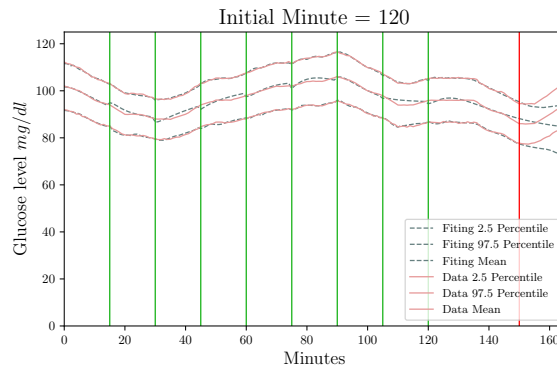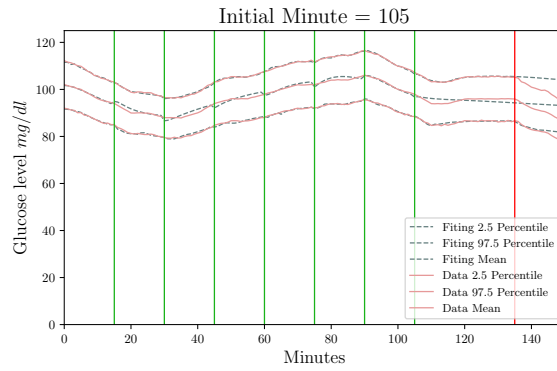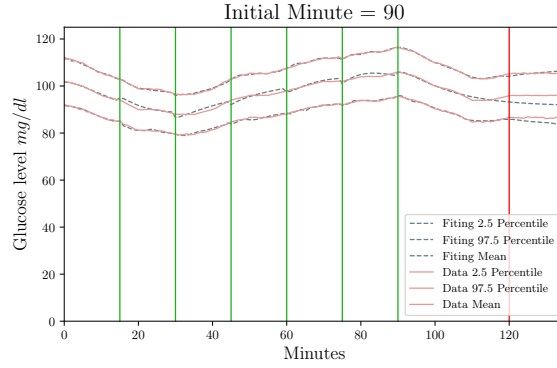
Figure 5: Calibration and prediction with uncertainty using 30 minutes of glucose data starting in the minutes 90, 105 and 120. Calibration is performed between the rightest green vertical line and the red vertical line. Then, the prediction with uncertainty stars.

12