

# Genetic Programming-based Induction of a Glucose-Dynamics Model for Telemedicine

Ivanoe De Falco<sup>a</sup>, Antonio Della Cioppa<sup>b</sup>, Tomas Koutny<sup>c</sup>, Michal Krcma<sup>d</sup>,  
Umberto Scafuri<sup>a</sup>, Ernesto Tarantino<sup>a,1,\*</sup>

<sup>a</sup>ICAR, National Research Council of Italy, Via P. Castellino 111, Naples, Italy

<sup>b</sup>NCLab, DIEM, University of Salerno, Via Giovanni Paolo II 132, Salerno, Italy

<sup>c</sup>NTIS – New Technologies for the Information Society, Faculty of Applied Sciences,  
University of West Bohemia, Univerzitni 8, Pilsen 306 14, Czech Republic

<sup>d</sup>Diabetology Center at the Pilsen Hospital University, Alej Svobody 80, 323 00 Pilsen,  
Czech Republic

---

## Abstract

This paper describes our preliminary steps towards the deployment of a brand-new original feature for a telemedicine portal aimed at helping people suffering from diabetes. In fact, people with diabetes necessitate careful handling of their disease to stay healthy. As such a disease is correlated to a malfunction of the pancreas that produces very little or no insulin, a way to enhance the quality of life of these subjects is to implement an artificial pancreas able to inject an insulin bolus when needed. The goal of this paper is to extrapolate a regression model, capable of estimating the blood glucose (BG) through interstitial glucose (IG) measurements, that represents a possible revolutionizing step in constructing the fundamental element of such an artificial pancreas. In particular, a new evolutionary-based strategy is illustrated to stem a mathematical relationship between BG and IG. To accomplish the task, an automatic evolutionary procedure is also devised to estimate the missing BG values within the investigated real-world database made up of both BG and IG measurements of people suffering from Type 1 diabetes. The discovered model is validated through a comparison with other models during the experimental phase on global and personalized data treatment. Moreover, investigation is performed about the accuracy of one

---

\*Corresponding author

*Email address:* `ernesto.tarantino@icar.cnr.it` (Ernesto Tarantino)

<sup>1</sup>Tel: +39 081 6139525 – Fax: +39 081 6139531

*Preprint submitted to Elsevier*

*December 21, 2017*

Published as:

De Falco, Ivanoe, Della Cioppa, Antonio, Koutny, Tomas, Krcma, Michal, Scafuri, Umberto, Tarantino, Ernesto (2018). Genetic Programming-based induction of a glucose-dynamics model for telemedicine. JOURNAL OF NETWORK AND COMPUTER APPLICATIONS, vol. 119, p. 1-13, ISSN: 1084-8045, doi: 10.1016/j.jnca.2018.06.007

single global relationship model for all of the subjects involved in the study, as opposed to that obtained through a personalized model found for each of them. This work will help us on adding the important feature of estimating BG in a web portal we are developing for telemedicine purposes. Once this research is clinically validated, this BG estimation feature will be added to the web portal for diabetic subjects to freely use it from their homes.

*Keywords:*

Blood glucose estimation, Interstitial glucose, Regression models, Evolutionary algorithms.

---

## 1. Introduction

Diabetes mellitus (DM) is a lifelong disease that causes a high level of blood glucose (BG). The amount of glucose in the blood is controlled by the insulin hormone produced by the pancreas. Insulin lowers BG by promoting glucose utilization by cells. Without insulin, glucose accumulates in various compartments. Subsequently, it binds to certain parts of multiple organs, thus damaging them and eventually leading to their failure.

Let us briefly describe diabetes pathogenesis with Type 1, Type 2 and gestational diabetes. With DM Type 1, there is an absolute lack of insulin as pancreas fails to produce it. With DM Type 2, patient exhibits insulin resistance as cells need increased amount of insulin to work properly. With gestational diabetes, insulin resistance is coupled with insufficient insulin secretion. The insulin secretion is too low to satisfy the increased requirements triggered by the insulin resistance.

Over time, having too much glucose in the blood can cause serious medical problems ranging from retinopathy, neuropathy, and nephropathy or even more grave side effects such as an increased risk of heart disease and stroke [1]. The diabetes has no cure and is spreading at alarming level across a large portion of the world population. Fortunately it is possible to improve the quality of life of the patients by controlling the medical risks associated to high BG levels through adequate treatments. These treatments last for many years and are very expensive for the society. Therefore, a methodology able to obtain as precise as possible BG-estimate to establish the right amount of insulin to inject assumes a very important role.

Implied monetary costs are not isolated from the problem of BG estimation only. We need to stress the patient education. Diabetic patient

must be educated about the disease, redesigned diet, new hygiene habits, drugs, devised therapy and available medical devices. Moreover, diabetic subjects may live far from doctors, or may have impairment in their movements due to problems to feet caused by this disease, or doctors may have not too much time to suitably follow all of their patients. All these issues call for remote management. Partially, it is possible to offload these issues using telemedicine. Specifically, it is possible to create a specialized web portal for diabetic patients. In fact, several such portals already exist. These portals can be classified into three, possibly overlapping, categories:

1. informative: informative portals provide general advices for diabetic patients about diet, hygiene habits, medical devices, etc., but they offer no glucose-signal processing.
2. glucometer-oriented: these portals can include the general advices along the possibility to upload and visualize glucose measurements.
3. advanced glucose-signal processing: on top of the functionality of the previous two portal categories, portals in this category are capable of calculating a new glucose signal from the measured glucose signal. For example, such a portal can calculate continuous BG from continuous IG measurements. Such a portal requires a model of glucose dynamics to operate, because its processing goes beyond a simple processing of single glucose signal to produce a statistics summary and to visualize the signal itself.

Informative portals are easy to find and several glucometer-oriented portals exist as well. They are usually associated with specific software that download data from the meters. Nevertheless, to the best of our knowledge, we are aware about only one portal that offers advanced glucose-signal processing [2]. This portal is an innovative tool and we are committed to its further development, for which we need to keep researching new ways to derive highly personalized models of glucose dynamics - which is the aim of this paper.

As face-to-face interaction between patient and physician becomes costly due to spent physician's time, telemedicine, a specialized web portal in our case, becomes an important tool along the medical devices as reported in Section 3.

Several estimation devices are present on the market that take BG measures in intervals that range from about 15 minutes to a couple of

hours with no measurements effected during the night. Since most of these devices are invasive, patients are reluctant to be subject to this continuous and invasive BG control because of the associated pain. Hence, it can become difficult to adequately take care of the sick persons. Instead there is the possibility of measuring more simply the IG, i.e., the glucose in the subcutaneous tissue, through the minimally invasive and easier-to-use Continuous Glucose Monitoring System (CGMS) devices [3]. With CGMS, a needle is inserted into subcutaneous tissue to measure electric current in the interstitial fluid of that tissue. A glucose-triggered reaction produces the current.

CGMS can be programmed to take measures with a prefixed frequency for a number of days, also during the night. Nevertheless, CGMS needs BG to calibrate - to transform the measured electric current to glucose level. Patient has to calibrate at least two times a day, when BG and IG are steady. While the patient collects only a few BGs, CGMS provides 288 measurements a day. Naturally, these measurements vary as a function of the patient.

Considered that BG and IG can differ considerably due to physiological reasons, the availability of the large amount of IG measures IS highly recommendable to capture as much of the BG-IG dynamics as possible. Then, it is possible to derive a precise estimation of BG by exploiting the availability of a large amount of IG measures for taking care of patients.

Although IG is not considered a perfect indicator for BG, nonetheless it is the only one to be available with continuous and non-invasive measurements. However, the complexity of the relationship between glucose dynamics in BG and IG is far too complex to be captured in the very simple calibration algorithms implemented in CGMS devices available in the market [4, 5]. In fact, CGMS is a low-power device that implies low computational capabilities and this invalidates the accuracy of the measurements.

In the literature different analytical models have been presented attempting to derive a mathematical relationship of IG and BG, as reported in Section 2. All these models represent a basic step to design and implement an *artificial pancreas*, i.e., an artificial device capable of automatically regulating insulin injections according to the needs so to assure an as-normal-as-possible life to the patients. This device must be able to carry out a glycemic control by estimating BG values through the analysis of the IG signal. Reliable prediction based on IG only is still out of the reach and present algorithms must rely on supplementary, imperfect information such as an assessment of carbohydrate intake by the patient.

The paper is a revised and extended version of a conference paper [6]. In particular, a more thorough review of the research contribution in terms of related works and of a description of a web portal for telemedicine along with additional experimentations are included. This paper provides a twofold contribution. The first one is the introduction of an innovative evolutionary strategy to exert new BG values without performing further measurements. Such a strategy allows increasing the number of the BG values included in the database, which is a critical problem when searching for a prediction model. The second contribution consists in leveraging this modified database to symbolically derive a law able to estimate BG values by using IG measurements. These estimation problems are known as *regression problems*. Considered the stated complexity, we exploit the capability of the Genetic Programming (GP) [7] in tackling regression problems [8] to find a good and effective approximation of relationship between IG and BG values. The experimental trials are performed over a real-world database containing both BG and IG measures for several subjects suffering from Type 1 DM. The aim is to discover an explicit relationship, i.e., a mathematical expression, between BG and IG values that could be the core of the knowledge base of an artificial pancreas.

An important question our paper wishes to tackle is whether or not a global approach leading to one single general model for a given set of subjects can be competitive with respect to the results obtained by finding a specific model for each of the subjects involved in the study. This is significant, because of course the personalized approach has the cost of needing learning over each subject. Learning needs a preliminary phase of data gathering from the subject, followed by sending this data to GP experts who have to run the algorithm and find the model for that subject, then add it to the measurement device to be given for the subject under account to use.

The paper is organized as follows. In Section 2 a review of the relevant related work is given. As a telemedicine tool, a specialized and interactive web portal for diabetic patients is introduced in Section 3. Section 4 outlines the strategy employed, along with an innovative evolutionary procedure to enlarge the original database presenting too many missing BG values in Section 4.1, and the genetic-based regression model in Section 4.2. A discussion and a comparison with other models on the achieved results obtained on global data treatment is reported in Section 5. In the same section, the comparison between the results obtained with the global and the personalized approaches is performed, and a discussion is provided.

Conclusions and future work are exposed in Section 6.

## 2. Related research

Several analytical models have been introduced with the aim to find a mutual relationship between BG and IG values.

The first, and most widely used, model attempting to relate BG and IG was devised by Steil and Rebrin in 1999 [9]. It is represented by the following equation:

$$\frac{\tau}{g} \cdot \frac{di(t)}{dt} + \frac{1}{g} \cdot i(t) = b(t) \quad (1)$$

where  $b(t)$  and  $i(t)$  are the BG and the IG at time  $t$ , and the parameters  $g$  and  $\tau$  represent the steady-state gain and the IG equilibration time constant, respectively. An important task is the estimation of the best possible values for  $g$  and  $\tau$  so that the precision of the model can be improved.

Makroglou et al. (2006) [10] displayed an overview of several mathematical models appearing in the literature and aimed at describing the glucose-insulin regulatory system with reference to DM. The described models ranged from ordinary differential equations to partial differential equations, to delay differential equations to the integro-differential ones.

In 2007 Kildegaard et al. [11] suggested a physiological model, consisting of an insulin, a meal and a glucose metabolism model coupled with a Monte Carlo simulation to predict blood glucose values of people with diabetes. Simulations showed how the coefficient of variance for the different model compartments changes over time. For average people with diabetes the inaccuracies of blood glucose meters and carbohydrate estimates contribute to more than half of the variance. The results of the simulation demonstrated that the blood glucose prediction is severely affected by the inaccuracy in the input variables and by metabolic fluctuations so that an effective prediction model was very difficult to extract.

In 2007 Facchinetti et al. [12] investigated how well plasma glucose can be reconstructed from interstitial fluid measured by means of CGM sensors. Six diabetic volunteers were monitored for 2 days using a minimally invasive device that, on the basis of an initial calibration procedure, returns interstitial glucose concentration. Simultaneously, plasma glucose concentration was also measured every 15 minutes. A nonparametric regularization deconvolution method was used to reconstruct plasma values from IG. The results demonstrated that the quality of the reconstruction

was unsatisfactory. Only after a recalibration procedure the relative error in reconstructing plasma glucose was reduced significantly.

In 2010, Leal et al. [13] monitored 18 patients for three days, and used an autoregressive model to learn the relationship between the IG measured by CGMS and BG. This resulted in 98.5% of the points being in A and B zones of the Clarke error grid analysis [14].

In 2010 Pérez-Gandía [15] used an artificial neural network for the prediction of the blood glucose concentration. The network receives as inputs the CGM sensor measurements during the preceding 20 minutes and returns as output the prediction of glucose concentration at the selected prediction horizon time. The accuracy of the tool was estimated by employing the root mean square error and the prediction delay.

In 2011 Koutny [16] proposed a method to assess the BG-to-IG delay based on the hypothesis that the change in the blood glucose level includes information about the estimated rate with which the hypothalamus expects the blood glucose level to return to normal range.

In 2012 Koutny [17] proposed a model of glucose dynamics that allows IG prediction. By accounting the delay, it was possible to relate the present BG and IG to future IG through coefficients of the proposed model of glucose dynamics.

More recently, in 2014 Del Favero et al. [18] presented a new model that tries to improve that by Steil and Rebrin. Basically, they reckoned that the value of  $g$  is equal to 1, and added a way to calibrate the model so that a true IG can be restored.

In the same year Koutny [19] suggested to improve the devised model of glucose dynamics to reduce its calculation error, especially with rapid changes of BG and IG, e.g., due to short-action insulin. This model heavily relies on biological considerations, as for instance the importance of capillaries and the fact that capillaries in different compartments have different permeabilities. Through the proposed model tested on hyperglycemic-clamp data, he succeeded in achieving an improved model to compute the BG levels.

All the above analytic models present the problem of estimation of several parameters. This estimation is usually carried out by leveraging mathematical, biological or physiological considerations, or also by performing an a-posteriori manual tuning. Despite its importance, only in the last years some attempts have been accomplished to automatically extrapolate the parameter estimation by means of techniques able to deal with this optimization problem such as the evolutionary methods. Some of

these attempts are outlined in the following.

In 2016, Koutny [20] combined the analytic method proposed in 2014 with meta-Differential Evolution (DE) [21, 22]. Namely, starting from a continuously measured level of IG for human Type 1 diabetic patients, he computed a continuous BG level. Six different scenarios were employed to guarantee robust validation of the calculation, and a DE was used to evaluate the parameters for the model in all the scenarios. All the six scenarios, even the simplest ones, presented better performance than CGMS in estimating BG values.

In 2017 De Falco et al. [23, 6] proposed evolutionary-based tools to estimate the BG values exploiting the easily available IG values. Relationships under the form of explicit mathematical expressions were discovered. The experimentation was carried out on a real-world database containing subjects suffering from Type 1 diabetes. The comparison against state-of-the-art models stated the effectiveness of the proposed evolutionary approaches.

### **3. The web portal**

Telemedicine is a distant form of health care. Using information technologies, patient can receive the medical care without actually needing to visit a physician. This reduces a possibility of infection between the patient and medical staff or another, already infectious patient. Diabetic patients are in increased risk of developing further complications due to infection. Moreover, it may be difficult to reach the physician if the patient has already developed peripheral nerve dysfunction associated, e.g., with diabetic foot. In such a case, minor injuries may develop without being noticed, thus increasing the risk of infection.

For diabetic patients, we have developed a specialized web portal – diabetes.zcu.cz. CGMS allows its user to extract the glucose recordings into a single file. The portal lets the user upload this file. By removing any personal information, the portal anonymizes the file first, before committing it to any processing. Then, the portal applies an improved model of glucose dynamics to the IG glucose-signal contained in this file. Such a model reduces the CGMS error when estimating BG. Eventually, the user can analyze the improved BG signal using a time-plot, ambulatory glucose profile, Clarke and Parkes grids, and raw data. The portal is designed to address the needs

Home My account Diabetes About the project Patient Expert Media Contact

## Upload data

To process, calculate and visualize glucose levels, you have to upload at least a CGMS data. Such a file contains subcutaneous tissue glucose level and self-monitored blood glucose levels, which were used to calibrate CGMS sensor. Usually, it is a spreadsheet (e.g. MS Excel) file format such as xlsx or csv. Optionally, it is possible to upload more blood glucose levels along the CGMS data. This should be a spreadsheet file format as well.

Files were uploaded

CGMS data

Blood-glucose levels

Comment

Diabetes type  E08  E09  Type 1  Type 2  Gestational  Other

Target folder

Previously uploaded data

There is a list recently uploaded data files. To see all previously uploaded data, please navigate to:

21. 04. 2017 16:24:00 - 23. 04. 2017 20:50:00

- Blood glucose level count: 39
- Interstitial fluid glucose level count: 629
- Comment:
- Calculation in progress...

24. 04. 2017 00:04:00 - 25. 04. 2017 19:23:00

- Blood glucose level count: 19
- Interstitial fluid glucose level count: 517
- Comment:
- Calculation in progress...

© since 2015 · Medical Informatics · DCSE · FAV · UWB top ▲

Figure 1: Uploading data to the portal.

of diabetic patients, physicians, and researchers. Researchers are allowed to modify model parameters.

Using the portal is straightforward. A user just uploads CGMS-exported file and wait for the result. Once displayed, the user can download any resulting figure. For example, the patient can go through the time plot to recall all events which could affect the BG to learn what could be improved, without imposing any stress on the patient by letting her/him take as much time as needed. Then, the patient can discuss these findings with a physician.

The portal is free to access and the available options differ depending on chosen role of the user. Users with an advanced role such as researchers, should register first to access full functionality of the portal. Users with other roles can use the portal without being registered.

Figure 1 depicts a form, where the user can upload CGMS data at the portal. This figure depicts the choices for a registered user, which include a track of previously uploaded CGMS data. Figure 2 depicts the results window, once BG calculation has finished. Detailed description of the portal is available in study [2].

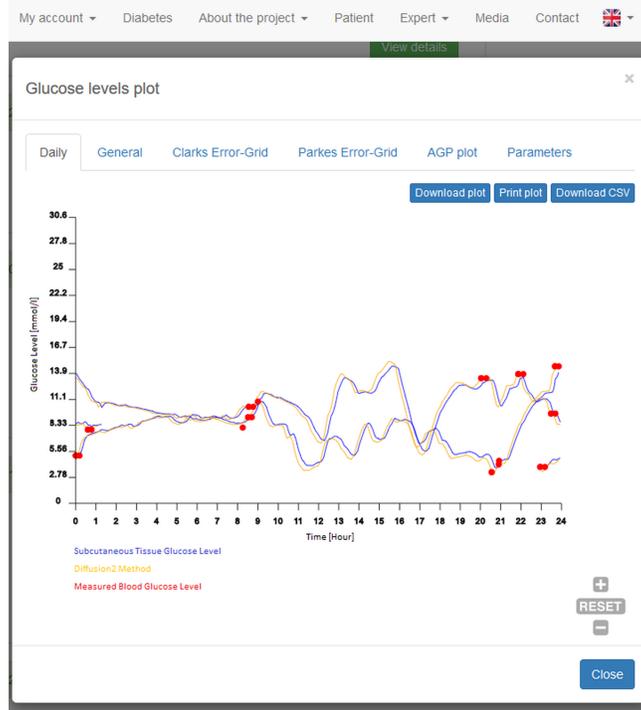


Figure 2: Calculated results at the portal.

#### 4. The evolutionary strategy

The evolutionary methodology proposed within this paper can be summarized in two basic steps: i) the leverage of an evolutionary optimizer for the evaluation of the model parameters to define an automatic procedure that allows enriching the original database with estimated BG values; ii) the extraction of an explicit symbolic regression model by exploiting an appropriate evolutionary technique, i.e., a GP algorithm.

##### 4.1. Database enlargement

One of the major problems in discovering a mutual and efficient mathematical relationship between BG and IG values is represented by the scarce availability of the first values with respect to the latter ones. To overcome this problem we devise a new procedure able to estimate a number of missing BG values within the database used to solve the related regression problem.

The Steil-Rebrin model [9] has been employed to estimate the missing BG values in the original database. To enhance the model precision, it is

necessary to evaluate the model parameters of Eq. (1), i.e., the steady-state gain  $g$  and the equilibration constant  $\tau$ . Following the approach of Koutny in [20], this estimation is performed by using an evolutionary optimization algorithm, namely the DE. Such an algorithm works on a set, known as *population*, of potential candidate solutions of the problem under investigation. Each solution, called *individual*, is encoded and evaluated through a fitness function that measures the current solution quality. The encoding and the fitness are set according to the problem under investigation. In our case each solution is represented by the model parameters while the fitness function calculates the Root Mean Square Error (*RMSE*) between the estimated and the measured values.

The algorithm finds a suboptimal solution to a problem by evolving the population of the potential candidate solutions. Specifically, a starting population of randomly chosen parameter values evolves for successive generations creating new candidate solutions by combining existing ones through specific evolutionary operators, i.e., recombination and mutation. The details related to these operators can be found in [21, 22]. These operators depend on two parameters called scale factor and crossover ratio. Moreover, several different DE strategies, still described in [21, 22], can be adopted. The evolutionary strategy is applied iteratively until either a fitness with desired quality is achieved or a fixed number of generations is performed. During the successive generations of the evolution the fitness function computes the *RMSE* between the BG estimated through the Steil-Rebrin model endowed with the current parameters and the available measured BG values of the original database.

By exploiting the parameter values found at the end of the evolution, it is possible to estimate the missing counterpart of BG values for all the IG values. In this way a complete correspondence between all the IG and the BG values is made available. Naturally, the enrichment procedure is not applied for the truly measured BG values contained in the original database, which are left unchanged. The so-modified database is then employed for solving the symbolic regression problem as outlined in Sect. 4.2.1.

By taking into account that the number of estimated BG values in this enriched database is much higher than those truly measured, it will be necessary to add a correction factor to avoid a bias in the model extraction. The detail of this correction factor will be reported in the experimental section.

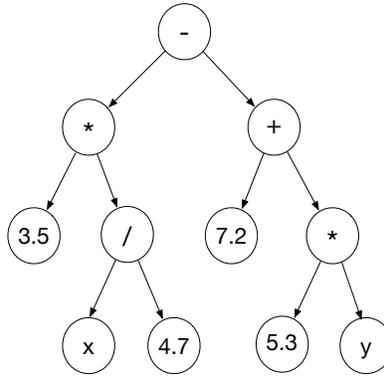


Figure 3: An exemplary tree-structured solution in GP.

#### 4.2. The general GP skeleton

Genetic Programming (GP) [7] is a heuristic methodology well suited for optimization purposes [24, 25], and has its roots in the implementation in a computer of mechanisms borrowed from the natural evolution process that happens in populations.

GP operates on a set, referred to as *population*, of potential solutions of the problem under exam. Each solution, called *individual*, is a program encoded as a tree structure. The inner nodes in any individual indicate elementary functions, while the leaf nodes comprehend terminals, i.e., either variables of the problem or constant values.

The corresponding program is attained by visiting in pre-order the starting tree as outlined in the example of Fig. 3 in which the tree represents the in-order expression  $3.5 * x / 4.7 - (7.2 + 5.3 * y)$ .

As in each evolutionary algorithm, also in GP the quality of each individual in solving the given problem is evaluated quantitatively by means of a *fitness function*  $\Phi$  that should be tailored to the specific problem. The population of solutions evolves iteratively from one generation to the next one by applying evolutionary operators with the aim to improve the fitness values. This evolution ends when a fixed maximum number of generations  $g_{max}$  is reached. A general pseudo-code describing GP is reported in the following:

- randomly generate an initial population with  $P$  individuals;
- assess the quality of each individual through the appropriate fitness function;

- at each generation create a new population by repeating the steps below:
  - randomly select an evolutionary operator; i.e., crossover, mutation, and copy;
  - choose in the current population as many individuals as needed by the operator picked;
  - let the selected individuals undergo the operator so as to generate a new individual;
  - include this individual into the new population being created;
  - evaluate the quality of the new individual through the fitness function;
- keep on creating a new population at each generation until reaching a predefined  $g_{max}$ .

The quality of the solutions attained by GP, expressed in terms of fitness function values, typically enhances as the number of generations increases. For the sake of brevity, a detailed description of the three operators is not reported here. The interested reader can refer to [7].

#### 4.2.1. GP for regression

Solving a regression problem through an exhaustive enumeration of all the possible models for the relationship between the independent and the dependent variables would be practically unfeasible due to the huge amount of time required, hence GP becomes of high interest for our purposes. As a heuristic technique, in general GP provides no guarantee about the attainment of the globally best model for a problem to be faced, yet usually a satisfactory, suboptimal, solution is found in a relatively low amount of time.

Since we wish to tackle here a regression task, the population will be constituted by a set of regression models. Each such model is encoded as a ‘formula’, represented as a tree whose nodes can comprehend either functions or terminals. The complete set of the functions employed within this paper is summarized in Table 1.

The terminal set consists of the set of the independent variables of the problem, plus the *Const* symbol that denotes a constant value. This latter

Table 1: The set of the elementary functions along with the corresponding symbols.

Symbol	Description
+	Addition
-	Subtraction
*	Multiplication
/	Protected division (is 1 if the denominator is 0)
sqrt	Square root
pow	Power

is always used in relation to a problem variable, and its value is randomly selected in a range suited to the specific variable involved.

Aiming at obtaining a (sub)-optimal regression model, a division of the database items into either two or three sets takes typically place for learning purposes. Due to the limited amount of measured items in the available database, we have opted for a two-set division, namely *train* and *test* sets. Learning is carried out over the items of the train set with the goal to achieve a model useful to approximate the dependent variable values as a function of the independent variable values. The quality of the best model (in terms of best fitness value over the train set) provided by GP at the end of the execution is evaluated over the test set, whose items had been never previously displayed to the GP algorithm.

To numerically evaluate the quality of each regression model  $S$  obtained during the GP execution, we have employed the *RMSE* as fitness function  $\Phi$ , i.e:

$$\Phi(S) = \sqrt{\frac{\sum_{i=1}^n (y_{comp}(i) - y(i))^2}{n}} \quad (2)$$

where  $y_{comp}(i)$  represents the output value for the  $i$ -th item of the database by the model  $S$  under examination, whereas  $y(i)$  is the value of the dependent variable for the same  $i$ -th item. With this choice the regression problem becomes a minimization problem.

## 5. Experimental findings

### 5.1. The database

From the Diabetology Center at the Pilsen University Hospital, we received anonymized datasets of Type-1 diabetic patients. We transformed the datasets into a database. The database comprises 5 different patients, their IDs being 1, 2, 4, 5, and 6, respectively. Each patient comprises several

time segments. A time segment is a period for which the patient wore a CGMS. There are 9, 30, 31, 38 and 38 time segments per patient, respectively. In total, there are 146 time segments, which contain 342 BGs and 36256 IGs. The IG values are taken with an interval  $\Delta t = 5$  minutes.

We have divided our experiments into two parts. In the former, we have taken into account the global database as described above, and have aimed at finding one general model that could work well for all of the five subjects considered. These experiments are described in Sec. 5.2. In the latter part, instead, we have used a personalized approach, so that the goal has been to find the best model for each of the five subjects, as reported in Sec. 5.3. Moreover, in Sec. 5.4 we have compared the results obtained in both cases, with the aim to assess whether or not a global approach leading to one single general model can be competitive with the results attained over each subject by a specifically achieved model.

### *5.2. The results for the global database*

The strategy described in 4.1 to estimate the missing BG values is applied to the global database. This procedure requires the estimation of the parameters of the Steil-Rebrin model. The estimation is carried out by means of the DE algorithm. The strategy used is the DE/rand/1/bin [21]. The DE parameters values have been determined by a preliminary tuning as follows: the population size equal to 50, the maximum number of generations equal to 200, the scale factor and the crossover ratio equal to 0.85 and 1, respectively.

The outcome of the DE algorithm has provided the following values for the Steil-Rebrin parameters:  $g = 0.98$  and  $\tau = 0.02$ . These values have been used to enlarge the dataset with the missing BG values. As already said, the truly measured BG are left unchanged.

The enriched dataset has been used to evaluate whether or not a general behavior, able to suitably describe all of the involved subjects, can be determined. If it is possible, this will result in a unique model without the necessity of personalizing it as a function of the subject. The existence of a single model would simplify the use of an artificial pancreas for more patients.

A tool executing GP and running in the MATLAB environment, i.e., GPTIPS [26], has been employed to perform all the experiments reported throughout this section. After a preliminary tuning phase, the values of the GP parameters have been set as follows: the population size  $P = 500$ , the

maximum generation number  $g_{max} = 500$ , the selection  $tourn\_size = 30$ , the mutation probability  $p\_mutate = 0.1$ , the crossover probability  $p\_cross = 0.9$  and copy probability  $p\_copy = 0.02$ . To permit a simple explication of the results, the maximum tree depth and the maximum number of the tree nodes have been set equal to 4 and 8, respectively.

We have assigned the first 70% of the database items to the train set (25,370 items) and the remaining 30% of items to the test set. Then the tool has been executed for 10 times. The reason is that GP is a nondeterministic algorithm whose results depend on a random integer value that should be assigned as seed to a random number generator. Different seeds can yield different results. Hence, the multiple execution over the database. The run reaching the lowest  $RMSE$  value over the test has been taken into account because its associated final model has the highest ability to correctly compute unseen data.

Moreover, given a time  $t$  at which the dependent variable BG has to be computed, the GP considers for the independent variable IG a time interval of 30 minutes before and 30 minutes after around  $t$ . CGMS provides a discrete signal as IG is reported with a fixed delay. Therefore, we restrict GP to avoid calculation of IG derivative although SR model does so. We consider IG derivative as possibly problematic due to the IG-signal discretization.

Considered that the IG values are taken with a  $\Delta t = 5$  minutes, the values considered with respect to IG at time  $t$ , i.e.,  $i(t)$ , range from  $i(t - \Delta t)$  to  $i(t - 6\Delta t)$  for the past, and from  $i(t + \Delta t)$  to  $i(t + 6\Delta t)$  for the future. Consequently there is a total of 13 independent variables, yielding for GPTIPS 13 possible terminals, plus the const node. In addition to these 13 variables, the dataset contains also the estimated BG values  $b_s(t)$  and the measured BG values  $b_m(t)$ .

In Eq. (2)  $y(i)$  represents the measured BG value in the  $i$ -th item of the set, and  $y_{calc}(i)$  is its estimate through the use of the IG values.

As explained in Section 4.1, to avoid bias in the model extraction a fitness function with a correction factor  $p_s$  has been devised. To this aim a global fitness function  $RMSE_{ALL}$  arranged as the sum of two sub-fitness functions appropriately weighted by an appropriate correction factor  $p_s$  is conceived. This function can be expressed in formula as follows:

$$RMSE_{ALL} = (1 - p_s) \cdot RMSE_s + p_s \cdot RMSE_m \quad (3)$$

where  $RMSE_s$  is the error evaluated on the estimated BG values while  $RMSE_m$  is the error calculated on the measured BG values.

The correction factor is given by  $p_s = n_s/n_t$  where  $n_s$  is the number of the estimated values and  $n_t$  is the total number of values in the dataset. The choice for the weights of  $RMSE_s$  and  $RMSE_m$  in Eq. 3 is due to the fact that the number of measured values is about 1% with respect to the number of estimated values.

Throughout our experiments, we compare our GP-based method endowed with the new strategy to estimate the missing BG values against a GP approach which makes use of the measured IG values as if they were the exact measured values of the BG. This method is denoted with IGBG. A further comparison is carried out with a state-of-the art model, i.e., Steil-Rebrin [9] with the parameters optimized through DE. Hereinafter this model is named  $S-R_{opt}$ .

The best model found by GP for the global database is:

$$b(t) = 1.16 \cdot i(t + \Delta t) - 3.69 \cdot 10^{-2} \cdot i(t - 6\Delta t) - 0.141 \cdot i(t - \Delta t) + 10.6 \cdot \frac{i(t) + i(t - \Delta t)}{i(t + 3\Delta t)} - 20.6$$

where  $i$  is the measured IG and  $b$  is the computed BG. The achieved model uses five out of the thirteen inputs, namely the IG values at times  $t$ ,  $t - \Delta t$ ,  $t - 6\Delta t$ ,  $t + \Delta t$  and  $t + 3\Delta t$ .

From a quantitative viewpoint, Table 2 shows the lowest  $RMSE_m$  of the different methods over train and test sets.

Table 2:  $RMSE_m$  values attained by the different methods.

GP	GP	IGBG	IGBG	$S-R_{opt}$	$S-R_{opt}$
Train	Test	Train	Test	Train	Test
9.65	10.15	17.51	18.27	14.71	13.82

Our GP-based method has the best performance with respect to the other models, and obtains an  $RMSE_m$  value equal to 9.65 over the train set, and to 10.15 over the test set. This is a very good result, because this means that the difference between a BG value estimated by this model and the real BG value is about 10 mg/dl in a typical range between 50 and 400. This relatively small  $RMSE_m$  assures that the use of the above model can be of great help in assessing whether the subject is in a normal state, or is undergoing a hyperglycemic or a hypoglycemic episode. Therefore, it is to underline that our GP-based model is superior with respect to the state-of-the art  $S-R_{opt}$  methods.

Figure 4 shows how well the BG values estimated by the found model fit the BG measured data over both train and test sets. As a general comment, fitting is very satisfactory: also over test set data, never shown to the system during learning, BG values, especially in peaks, are very often caught in terms of both time and magnitude.

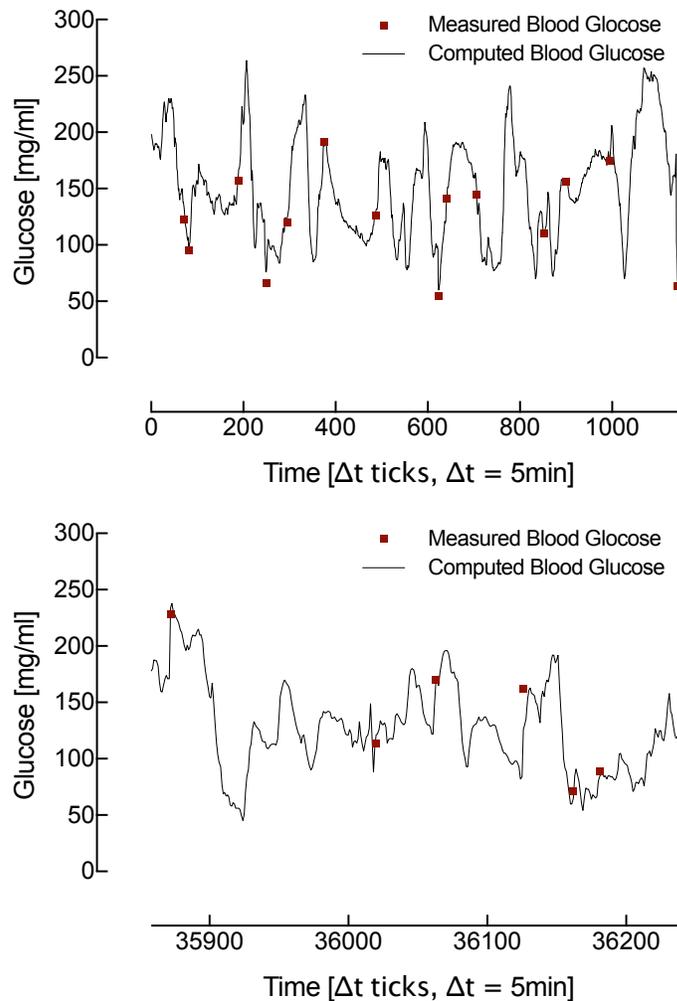


Figure 4: The estimate of the values through GP. Top: first 2 segments on train set. Bottom: last 4 segments on test set.

The frequency of relative errors for the three investigated methods is reported in Table 3. As it can be appraised from the table, for our model

87.74% of the items has a relative error lower than 10%, which is an excellent result. Moreover, more than 95% of the items has an error lower than 20%, which is also good with respect to ISO 15197:2003 and 15197:2013 accuracy standards for blood glucose meters. The table also evidences that these results are much better than those obtained when IG is used as the real value of BG, and better also than  $S-R_{opt}$  model for the items with relative errors lower than 10% and 20%.  $S-R_{opt}$  model has slightly better values than GP for a relative error lower than 5% while the results are the same for the other relative errors. This means that the proposed GP-based approach can actually help in better estimating BG.

Table 3: The frequency of relative errors: cumulative probability of lower than or equal to relative error.

Relative error	GP	IGBG	$S-R_{opt}$
$\leq 5\%$	59.43%	41.51%	60.38%
$\leq 10\%$	87.74%	72.64%	83.02%
$\leq 15\%$	90.57%	84.91%	90.57%
$\leq 20\%$	95.28%	90.57%	92.45%
$\leq 25\%$	98.11%	93.40%	98.11%
$\leq 30\%$	98.11%	96.23%	98.11%
$\leq 35\%$	98.11%	96.23%	98.11%
$\leq 40\%$	100.00%	98.11%	100.00%
$\leq 45\%$	100.00%	98.11%	100.00%
$\leq 50\%$	100.00%	100.00%	100.00%

Figure 5 shows the results for GP in terms of the Clarke error grid analysis [14], that is widely used in the studies about diabetes. Basically, the 2D-estimate space is subdivided into five zones labeled from A to E. The higher the number of points falling in zones A and B, the better the estimate. Points falling in the other zones represent different types of situations that should be avoided. As it can be seen, almost all of the points lie in the zones A and B, and this holds true not only for the train set but for the test set too, which is a hint of a very good estimate. Numerically, the percentage of points falling in the different zones is reported in Table 4 for both train and test sets.

Figure 6 reports the same diagram on all the dataset.

For GP the zones A and B contain more than 99% of the points in test set (ISO 15197:2013 requirement), which is very good and is superior to IGBG, and slightly outperforms also the  $S-R_{opt}$  method given the results in zone A. Also the absence of points related to situations C and E is an excellent outcome of the proposed model. As a comparison, GP has a lower number

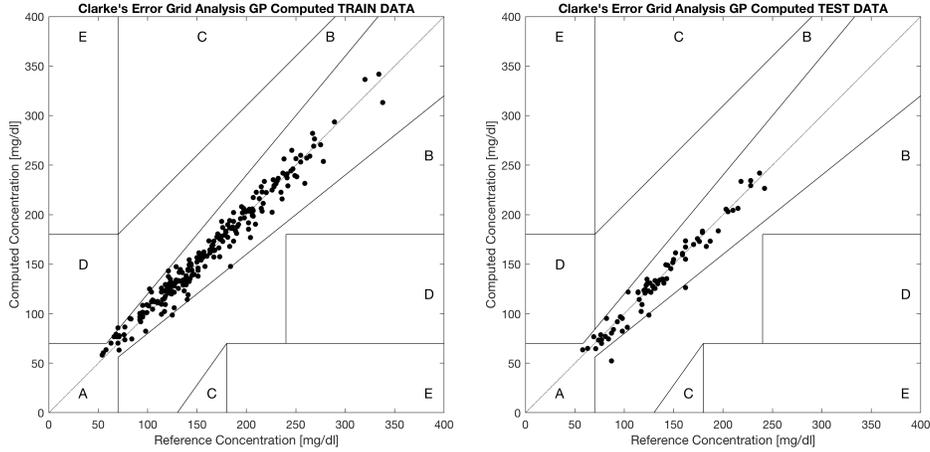


Figure 5: Clarke grid analysis. Left: train set. Right: test set.

Table 4: The percentage of points falling in the different zones of the Clarke error grid.

Zone	GP Train	GP Test	IGBG Train	IGBC Test	$S-R_{opt}$ Train	$S-R_{opt}$ Test
A	98.73%	95.28%	94.49%	90.57%	97.46%	93.40%
B	0.85%	4.72%	4.24%	8.49%	1.69%	6.60 %
A + B	99.58%	100%	98.73%	99.06%	99.15%	100%
C	0.00 %	0.00%	0.00%	0.0%	0%	0%
D	0.42%	0%	1.27%	0.94%	0.85%	0%
E	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

of points lying in the zone D than IGBG and the  $S-R_{opt}$  models. This is a very important result because zone D points out a possible risky inadequacy for hypoglycemia or hyperglycemia identification.

Summarizing, the findings obtained seem to imply that a unique model extracted by using GP can fit all of the subjects involved in this study, and the artificial pancreas for all of them could be based on evolutionary-devised models.

### 5.3. The results for the personalized databases

In this case the procedure described in the previous subsection has been performed separately on each of the five subjects considered in this study. For the sake of conciseness, we do not report here the five explicit models obtained, rather we concentrate our attention on the numerical results obtained.

Table 5 shows the  $RMSE_m$  values obtained over each of the five subjects by the same methods investigated above. In this table, the first column

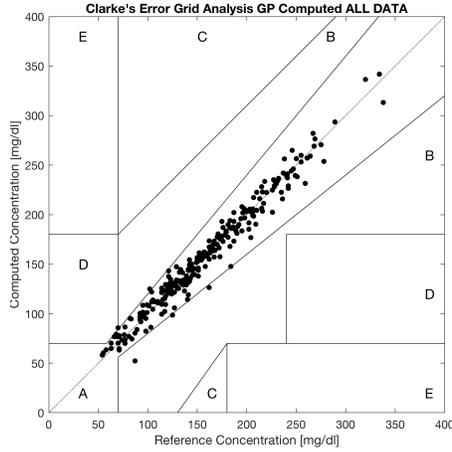


Figure 6: Clarke grid analysis on the whole dataset.

indicates subjects' IDs. Also, the four bottom rows outline the results for each method averaged over the five subjects in terms of the average value, the related standard deviation, and the best and the worst from among the five values obtained for the subject under account.

Table 5:  $RMSE_m$  values achieved by the different methods.

Subject ID	GP Train	GP Test	IGBG Train	IGBG Test	$S-R_{opt}$ Train	$S-R_{opt}$ Test
1	4.10	6.01	12.17	5.56	1.37	0.57
2	9.07	14.57	15.04	19.23	6.18	6.31
4	9.26	14.54	23.46	19.51	5.79	5.00
5	7.61	14.79	13.75	22.97	6.25	6.06
6	7.03	14.03	15.81	23.50	11.83	7.03
Average	7.41	12.79	16.05	18.15	6.28	4.99
St. dev.	2.09	3.80	4.37	7.30	3.71	2.58
Best	4.10	6.01	12.17	5.56	1.37	0.57
Worst	9.26	14.79	23.46	23.50	11.83	7.03

Considering  $RMSE_m$ ,  $S-R_{opt}$  is the best method over both train and test sets. GP is the runner-up, whereas IGBG is very far from both. It is interesting to note that GP and  $S-R_{opt}$  are, on average, almost equivalent over train sets and GP has smaller standard deviation, while  $S-R_{opt}$  produced smaller  $RMSE_m$  per individual subject.

Moreover, as concerns the numerical results related to Clarke error grid analysis, Table 6 reports for each of the five subjects the percentages of points falling in the different zones as obtained by the three models investigated,

over both the train set and the test set. In the bottom of the table, the averaged results obtained over the five subjects are also shown.

Table 6: The percentage of points falling in the different zones of the Clarke error grid for the personalized case.

Zone	GP Train	GP Test	IGBG Train	IGBC Test	$S-R_{opt}$ Train	$S-R_{opt}$ Test
subject 1						
A	100.00%	100.00%	92.00%	100.00%	96.00%	100.00%
B	0.00%	0.00%	4.00%	0.00%	4.00%	0.00%
A + B	100.00%	100.00%	96.00%	100.00%	100.00%	100.00%
C	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
D	0.00%	0.00%	4.00%	0.00%	0.00%	0.00%
E	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
subject 2						
A	98.15%	95.45%	96.30%	95.45%	98.15%	95.45%
B	1.85%	4.55%	3.70%	4.55%	1.85%	4.55%
A + B	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
C	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
D	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
E	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
subject 4						
A	100.00%	95.45%	96.36%	90.91%	96.36%	90.91%
B	0.00%	4.55%	1.82%	9.09%	1.82%	9.09%
A + B	100.00%	100.00%	98.18%	100.00%	98.18%	100.00%
C	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
D	0.00%	0.00%	1.82%	0.00%	1.82%	0.00%
E	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
subject 5						
A	100%	87.50%	92.16%	87.50%	92.16%	91.67%
B	0.00%	12.50%	5.88%	12.50%	5.88%	8.33%
A + B	100.00%	100.00%	98.04%	100.00%	98.04%	100.00%
C	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
D	0.00%	0.00%	1.96%	0.00%	1.96%	0.00%
E	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
subject 6						
A	100.00%	84.21%	93.18%	84.21%	100.00%	84.21%
B	0.00%	15.79%	6.82%	15.79%	0.00%	15.79%
A + B	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
C	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
D	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
E	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Averaged values						
A	99.63%	92.52%	94.00%	91.61%	96.53%	92.45%
B	0.37%	7.48%	4.44%	8.39%	2.71%	7.55%
A + B	100.00%	100.00%	98.44%	100.00%	99.24%	100.00%
C	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
D	0.00%	0.00%	1.56%	0.00%	0.76%	0.00%
E	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

By looking at the averaged values, it can be noticed that all of the three methods allow obtaining 100.00% of correctly placed points (A+B zones) in the test sets of all the five subjects. As concerns the train sets, instead, GP

obtains perfect placement for all the points, followed by  $S-R_{opt}$  and by IGBC. If the perfect case of zone A is considered only, GP has the best performance over both train and test sets,  $S-R_{opt}$  being the runner-up, whereas IGBC is quite far from the other two.

If a closer look is taken at the results over the single subjects, it can be seen that for subject 1 the three methods are equivalent over the test set, whereas GP and  $S-R_{opt}$  are perfect over the train set too and IGBC places 4% of the points in the dangerous zone D. For subject 2, the three methods are equivalent over both sets, although IGBC has poorer performance over the train set if zone A only is taken into account. Subject 4 shows that GP is the best over both sets, especially if we consider the zone A only, while IGBC and  $S-R_{opt}$  obtain exactly the same results. For subject 5, instead, all the three methods being equal over the test set in terms of A+B percentage,  $S-R_{opt}$  is better if only zone A is looked at; over the train set, instead, GP is perfect and outperforms the other methods. Finally, subject 6 shows that the three methods are about equivalent over the test set, IGBC being worse over the train set especially if zone A only is taken into account.

Given the apparently contrasting conclusions that could be drawn by looking at the results in Table 5 and at those in Table 6, we need to further investigate the actual differences between GP and  $S-R_{opt}$ . With this aim, Figures 7 and 8 show the results in terms of the Clarke error grid analysis obtained for each of the five subjects over the test by using GP (on the left) and  $S-R_{opt}$  (on the right).

In general, it can be seen that the points are closer to the main diagonal line for  $S-R_{opt}$  than they are for GP, which well reflects the fact that  $RMSE_m$  values are lower for the former method. On the other hand, in terms of numbers of points in zone A we can see that GP is in many cases able to place the same number of points in zone A as  $S-R_{opt}$  can, or even more. As an example of this latter, we can take a look at subject 4: GP has only one point in zone B, whereas  $S-R_{opt}$  has two, although the value for  $RMSE_m$  is much lower for the latter method, due to the points being closer to the diagonal. As a consequence, the two conclusions that can be separately drawn from the two tables are not actually contrasting each other, rather each of them focuses on a different aspect of the problem.

Summarizing, on average, predictions are numerically better when  $S-R_{opt}$  method is used, whereas the number of calculated points in zone A is equal in both cases, or is slightly better when GP method is considered.

#### 5.4. Discussion

As we have seen in the two previous subsections, as long as a global approach is taken over a database constituted by data coming from more subjects, GP method is the most suitable one in terms of both lower  $RMSE_m$  and better Clarke Error Grid analysis results. When personalized management of the data is to be carried out, instead,  $S-R_{opt}$  is the most precise method.

It is interesting to understand the reasons for this different behavior. The main difference between the two different cases lies in the number of items onto which carry out the supervised learning phase. This number is, of course, higher for the global approach. This could mean that, in order to obtain a reliable model, GP needs a larger number of items than  $S-R_{opt}$  does. Thus, we believe the reason for this resides in the internal features of GP that allow it to abstract a model starting from raw data.  $S-R_{opt}$ , on its turn, is a well-known model with a reliable general structure being successfully applied many times in the field, and only the values of its parameters should be optimized, so it does not need too many items to find satisfactory values for its parameters through DE.

In addition, GP could not find the SR model because the set of elementary functions does not contain derivative due to the discrete nature of the IG signal - see Section 5.2.

Another difference between the two approaches when dealing with personalized databases lies in the type of data they work with.  $S-R_{opt}$  deals only with measured glucose values with the aim to derive the best parameters for the Steil-Rebrin model for a specific patient. On the other hand, GP deals with a mixture of measured glucose values and glucose values derived by  $S-R_{opt}$ . As a consequence, according to Eq. (3) the GP is focused on searching for a personalized model on the basis of the underlying optimized Steil-Rebrin model by  $S-R_{opt}$  endowed with measured data. Therefore, it is not surprising to expect a larger  $RMSE_m$  value for the GP with respect to  $S-R_{opt}$ .

These features have important consequences on the medical protocol to be followed for the subjects taking place in a medical trial. In fact, if the goal is to obtain one general model that can be applied to a wide set of subjects, then, based on the results in Sect. 5.2, GP is the most viable option. Therefore, in the design of a general-purpose artificial pancreas to be given to many subjects, GP is the best choice. This choice only requires a preliminary phase in which the data from many subjects is gathered, the use of GP on it, and one multi-subject model being added to the real-time

glucose monitoring system. A copy of this system will be given to any subject to monitor.

When, instead, a one-subject artificial pancreas is to be considered for each patient to be monitored, then the use of  $S-R_{opt}$  seems more fruitful in terms of better calculation of the BG values, as reported in Sec. 5.3. It should be underlined that the personalized approach has an additional cost consisting in the fact that learning should be carried out over each specific subject. This means for each subject the need of a preliminary phase of data gathering, followed by sending this data to the experts who have to run the  $S-R_{opt}$  method and find the values of the parameters of the model for that given subject. The model endowed with its parameter values is then to be added to the monitoring system that is finally given to the subject under account for she/he to use it.

## 6. CONCLUSIONS

The main problem for a regression model in finding a relationship between variables is the absence of a sufficient number of values of the variables to be correlated. This is the usual situation in the research area of diabetes where the easily available number of IG values contrast with the low number of corresponding number of measured BG values. Within this paper, to surmount the problem we have envisaged an evolutionary procedure to enrich the database made up of many missing BG values by leveraging the Steil-Rebrin model and the DE technique to estimate these missing BG values. Thereafter, a GP algorithm has been employed to extract an explicit relationship between BG and IG values under the form of a mathematical expression.

This model has been compared both against a GP approach which makes use of the measured IG values as if they were the exact measured values of the BG and against the state-of-the-art Steil-Rebrin model with optimized parameters. This comparison has taken place on global and on personalized bases.

For the global situation in which one model is needed to fit all subjects, the results have shown GP superiority in terms of lower  $RMSE$ , and of better fitting in the Clarke error grid. The findings obtained seem to imply that a unique model can fit all of the subjects involved in this study, and the artificial pancreas for all of them could be based on evolutionary-devised model.

The comparison of the approach in which one single global relationship model is found for all of the subjects involved in the study, and that in which a personalized model is obtained for each of them has shown that the error made by  $S-R_{opt}$  increases with an increased number of reference BG, i.e., the global scenario. The reason can be that in the personalized scenario  $S-R_{opt}$  is i) required to match a lower number of different glucose patterns and ii) easier to overdetermine than GP due to the small number (2) of parameters.

Based on the results, we reason that GP may provide more interesting insights into the BG patterns by considering multiple time segments, possibly coming from different sources.

Future work implies the use of evolutionary-devised model in a clinical trial to estimate the BG values of the involved subjects, so as to further test its effectiveness. In the positive case, this model could be added to an under-development artificial pancreas device for a real experimentation.

Moreover, we plan to enrich the specialized web portal for diabetic patients with the outcomes presented in this paper, once the research is clinically validated. Namely, the web portal will be endowed with this BG estimation feature to permit diabetic patients a free remote use.

## Acknowledgements

This publication was partially supported by the project LO1506 of the Czech Ministry of Education, Youth and Sports.

## References

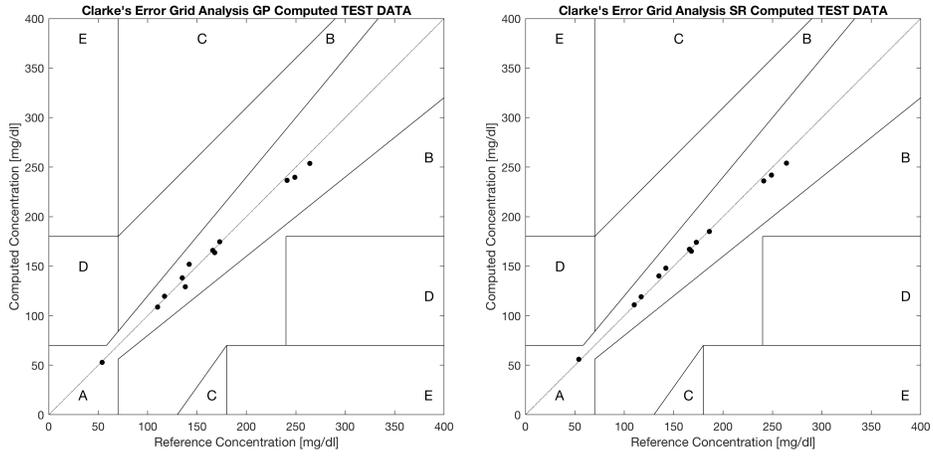
- [1] World Health Organization, Diabetes fact sheet n. 312 (October 2013).
- [2] T. Koutny, M. Krcma, J. Kohout, P. Jezek, J. Varnuskova, P. Vcelak, J. Strnadek, On-line blood glucose level calculation, *Procedia CompuT. Sci.* 98 (2016) 228–235.
- [3] S. K. Vashist, Continuous glucose monitoring systems: a review, *Diagnostics* 3 (4) (2013) 385–412.
- [4] Y. H. Baek, H. Y. Jin, K. A. Lee, S. M. Kang, W. J. Kim, M. G. Kim, J. H. Park, S. W. Chae, H. S. Baek, T. S. Park, The correlation and accuracy of glucose levels between interstitial fluid and venous plasma by continuous glucose monitoring system, *Korean Diabetes J.* 34 (2010) 350–358.

- [5] P. Rossetti, J. Bondia, J. Vehí, C. Fanelli, Estimating plasma glucose from interstitial glucose: the issue of calibration algorithms in commercial continuous glucose monitoring devices, *Sensors* 10 (2010) 10936–10952.
- [6] I. De Falco, A. Della Cioppa, T. Koutny, M. Krcma, U. Scafuri, E. Tarantino, An evolutionary approach for estimating the blood glucose by exploiting interstitial glucose measurements, in: *Proceedings of the International Workshop on Artificial Intelligence for Health*, Funchal, Madeira, Portugal, January 19-21, 2018. To be published.
- [7] J. Koza, *Genetic programming: on the programming of computers by means of natural selection*, Vol. 1, MIT Press, 1992.
- [8] A. Borrelli, I. De Falco, A. Della Cioppa, M. Nicodemi, G. Trautteur, Performance of genetic programming to extract the trend in noisy data series, *Physica A* 370 (1) (2006) 104–108.
- [9] K. Rebrin, G. Steil, W. van Antwerp, J. Mastrototaro, Subcutaneous glucose predicts plasma glucose independent of insulin: implications for continuous monitoring, *Am. J. Physiol.* 277 277 (1999) E561–E571.
- [10] A. Makroglou, J. Li, Y. Kuang, Mathematical models and software tools for the glucose-insulin regulatory system and diabetes: an overview, *Appl. Numer. Math.* 56 (2006) 559–573.
- [11] J. Kildegaard, J. Randløv, O. H. J. U. Poulsen, The impact of non-model-related variability on blood glucose prediction, *Diabetes Technol. The.* 9 (4) (2007) 363–371.
- [12] A. Facchinetti, G. Sparacino, C. Cobelli, Sensors & algorithms for continuous glucose monitoring reconstruction of glucose in plasma from interstitial fluid continuous glucose monitoring data role of sensor calibration, *J. Diabetes Sci. Technol.* 1 (2007) 617–623.
- [13] Y. Leal, W. Garcia-Gabin, J. Bondia, E. Esteve, W. Ricart, J. M. Fernandez-Real, J. Vehí, Real-time glucose estimation algorithm for continuous glucose monitoring using autoregressive models, *J. Diabetes Sci. Tech.* 4 (2) (2010) 391–403.

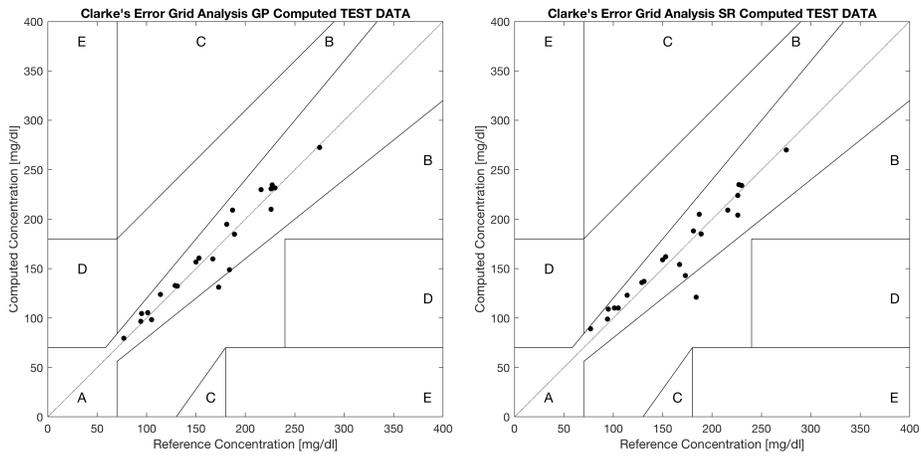
- [14] W. L. Clarke, D. Cox, L. Gonder-Frederick, W. Carter, S. Pohl, Evaluating clinical accuracy of systems for self-monitoring of blood glucose, *Diabetes Care* 10 (1987) 622–628.
- [15] C. Pérez-Gandía, A. Facchinetti, G. Sparacino, C. Cobelli, E. Gómez, M. Rigla, A. de Lieiva, M. E. Hernando, Artificial neural network algorithm for online glucose prediction from continuous glucose monitoring, *Diabetes Technol. The.* 12 (1) (2010) 81–88.
- [16] T. Koutny, Estimating reaction delay for glucose level prediction, *Med. Hypotheses* 77 (6) (2011) 1034–1037.
- [17] T. Koutny, Prediction of interstitial glucose level, *IEEE Trans. Inf. Technol. Biomed.* 16 (2012) 136–142.
- [18] S. Del Favero, A. Facchinetti, G. Sparacino, C. Cobelli, Improving accuracy and precision of glucose sensor profiles: retrospective fitting by constrained deconvolution, *IEEE Trans. Biomed. Eng.* 61 (4) (2014) 1044–1053.
- [19] T. Koutny, Blood glucose level reconstruction as a function of transcapillary glucose transport, *Comput. Biol. Med.* 53 (2014) 171–178.
- [20] T. Koutny, Using meta-differential evolution to enhance a calculation of a continuous blood glucose level, *Comput. Meth. Prog. Bio.* 133 (2016) 45–54.
- [21] K. Price, R. Storn, Differential evolution, *Dr. Dobb’s J.* 22 (4) (1997) 18–24.
- [22] R. Storn, K. Price, Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces, *J. Global Optim.* 11 (4) (1997) 341–359.
- [23] I. De Falco, A. Della Cioppa, U. Scafuri, E. Tarantino, Accurate estimate of blood glucose through interstitial glucose by genetic programming, in: *Proceedings of the International Symposium on Computers and Communications*, IEEE Press, 2017, pp. 284–289.
- [24] I. De Falco, A. Della Cioppa, E. Tarantino, F. Fontanella, A novel grammar-based genetic programming approach to clustering, in:

Proceedings of the ACM Symposium on Applied Computing, Vol. 2, ACM Computing, 2005, pp. 928–932.

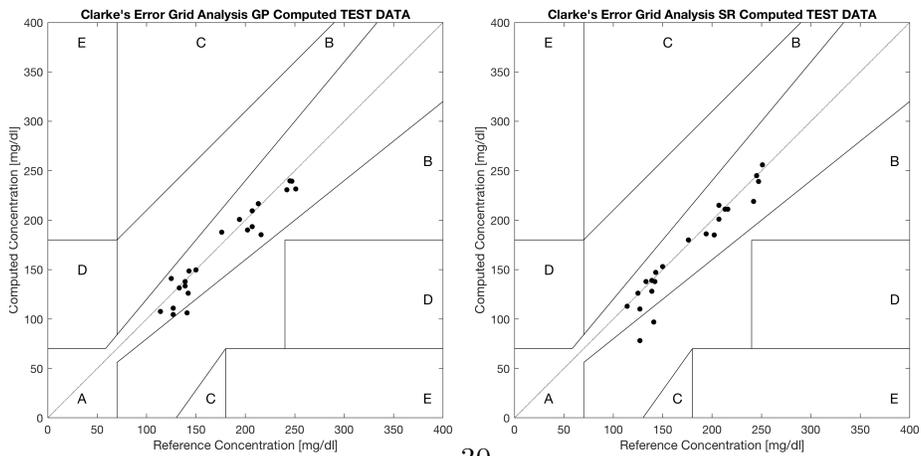
- [25] I. De Falco, A. Della Cioppa, E. Tarantino, F. Fontanella, An innovative approach to genetic programming-based clustering, Vol. 34 of Advances in Soft Computing, Springer, 2006, pp. 55–64.
- [26] D. Searson, GPTIPS: Genetic programming and symbolic regression for MATLAB. [http://gptips.sourceforge.net.](http://gptips.sourceforge.net/), 2009. (2009).



(a) Subject 1.

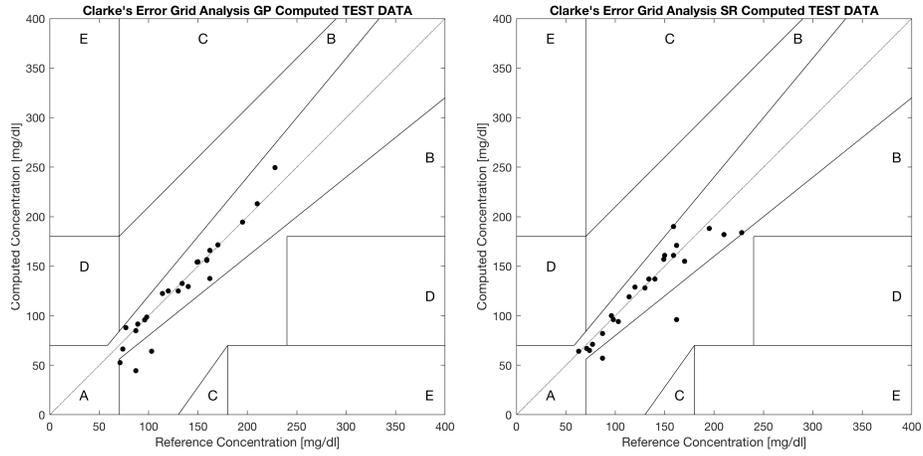


(b) Subject 2.

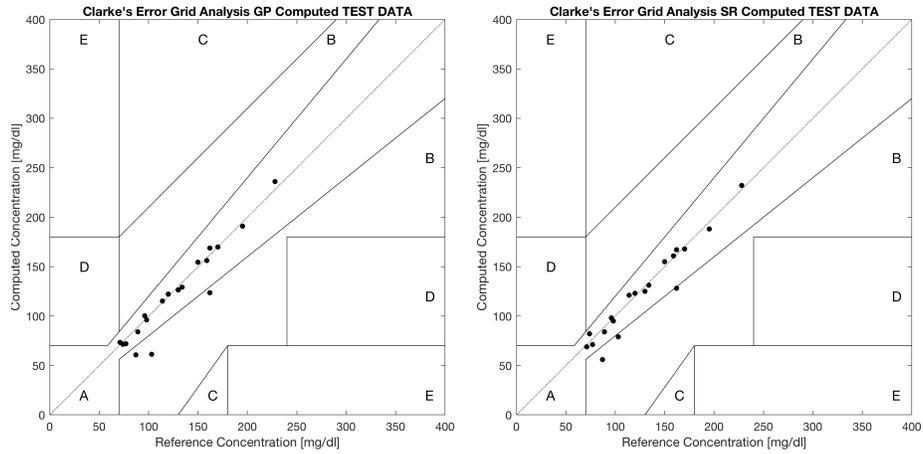


30  
(c) Subject 4.

Figure 7: Clarke error grid analysis over the first three subjects in order of their IDs (test sets only). Left: GP. Right:  $S-R_{opt}$ .



(a) Subject 5.



(b) Subject 6.

Figure 8: Clarke error grid analysis over the last two subjects in order of their IDs (test sets only). Left: GP. Right:  $S-R_{opt}$ .