Lecture Notes in Bioengineering

Vasilis Marmarelis
Georgios Mitsis *Editors*

Data-driven Modeling for Diabetes

Diagnosis and Treatment

Springer

Data-driven Modeling for Diabetes

Lecture Notes in Bioengineering

For further volumes:
http://www.springer.com/series/11564

Vasilis Marmarelis · Georgios Mitsis
Editors

Data-driven Modeling for Diabetes

Diagnosis and Treatment

Springer

Editors
Vasilis Marmarelis
Department of Biomedical Engineering
University of Southern California
Los Angeles, CA
USA

Georgios Mitsis
Department of Electrical and Computer Engineering
University of Cyprus
Nicosia
Cyprus

and

McGill University
Montreal, QC
Canada

ISSN 2195-271X
ISBN 978-3-642-54463-7
DOI 10.1007/978-3-642-54464-4
Springer Heidelberg New York Dordrecht London

ISSN 2195-2728 (electronic)
ISBN 978-3-642-54464-4 (eBook)

Library of Congress Control Number: 2014937279

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The motivation for this book is two-fold. On one hand, the book seeks to promote the notion that mathematical and computational *modeling* is critically important in advancing the scientific understanding of the physiological processes underpinning the regulation of blood glucose and, therefore, can be valuable in medical/clinical efforts to manage diabetes mellitus. On the other hand, the book advocates the view that these modeling efforts must rely primarily on *data-based* approaches as the scientific and methodological pathway leading to our ultimate objective in a most efficient and expeditious manner. Although both aforementioned reasons for the publication of this book are widely accepted by the peer community as being valid and meritorious, they have not been given—in the view of the co-editors—the requisite attention in the vast literature on glycemic regulation and the clinical management of diabetes mellitus. Through the publication of this book, the co-editors and the distinguished contributing co-authors aspire to make a contribution in addressing this need for rigorous, reliable and effective methodological tools that can be useful in achieving the goal of quantitative data-based modeling of the physiological process of glycemic control in a practical context—thus providing a valuable vehicle for the improved clinical management of diabetes mellitus.

These modeling tools should take into account the intrinsic characteristics of the human metabolic system, particularly in diabetic patients, which render the problem of achieving reliable and robust automatic glucose regulation very difficult. These include, but are not limited to, the *patient-specific* and *time-varying* nature of glucose metabolism in diabetic patients, the presence of large, possibly unpredictable and/or unobservable *disturbances*, and the *incomplete information* that is available to the modeler due to the complex interplay between a large number of physiological and behavioral factors that cannot be accounted for in a practical setting. Due to the aforementioned characteristics, choosing the proper model type, the estimation methodology and the control strategy are crucial for achieving the desired result.

The contents of the book can be divided into two parts conceptually. The first half (Chapters "Data-Driven and Minimal-Type Compartmental Insulin-Glucose Models: Theory and Applications" to "Pitfalls in Model Identification: Examples from Glucose-Insulin Modelling") focuses mostly on methodological considerations related to modeling approaches in diabetes, whereas the second half

(Chapters "Simulation Models for In-Silico Evaluation of Closed-Loop Insulin Delivery Systems in Type 1 Diabetes" to "Nonlinear Modeling of the Dynamic Effects of Free Fatty Acids on Insulin Sensitivity") focuses on the applications aspect of such approaches. Specifically, the first chapter (Mitsis and Marmarelis) demonstrates the relation between widely used compartmental differential equation models of glucose metabolism and input–output nonlinear Volterra-type models in an analytical manner. It also illustrates the feasibility of obtaining Volterra models from simulated data generated by these compartmental models as well as from experimental animal data.

The second chapter (by Ståhl, Johansson and Renard) concerns the important issue of whether a single optimal model may be identified for glucose prediction, given the highly complex nature of metabolism outlined above. In this context, the authors discuss the use of the model merging/switching approach to achieve robust and reliable prediction.

The third chapter (by Bequette) considers the crucial, potentially life-threatening occurrence of hypoglycemic events, particularly those occurring during the night. Various predictive algorithms that may be used to detect such events and consequently schedule the function of the insulin pump, as well as challenges linked to their implementation, are discussed.

The fourth chapter (by Daskalaki, Diem and Mougiakakou) addresses ways to handle the time-varying nature of glucose metabolism, which is one of the most important challenges in designing reliable therapeutic schemes. These time-varying characteristics may arise from intrinsic nonstationarities and/or the effect of exogenous unobserved factors and the authors consider the use of adaptive algorithms to perform on-line glucose regulation and detection of hypo/hyperglycemic events.

The fifth chapter (by Panunzi and DeGaetano) discusses the possible pitfalls that may arise when constructing parametric models of glucose-insulin interactions, i.e., differential equation models that assume a specific structure for the underlying system. They discuss the importance of examining the qualitative behavior of a model (e.g., whether this behavior is physiologically plausible) and investigate the presence of undesired noise in the input signal used to obtain parameter estimates—an issue that is particularly important in closed-loop systems, such as the glucose metabolism system.

The second, applications-oriented, half of the book commences with the sixth chapter (by Wilinska and Hovorka), which reviews the use of realistic virtual patient models to assess the performance of automated closed-loop glucose regulation systems. These models are beginning to constitute a crucial part of the design process in the quest for an artificial pancreas, as they may accelerate the design process and the transition to clinical trials, as demonstrated by their recent approval as a substitute for pre-clinical trial test beds by regulatory bodies (e.g., the Food and Drug Administration of the USA).

The seventh chapter (by Tura and Pacini) presents applications of the oral glucose tolerance test, as well as recently proposed mathematical models for the analysis of such data, in order to quantify sex-related differences in glucose

metabolism as well as glucose absorption in pregnant women with gestational diabetes. The results demonstrate that rich information may be extracted when models that are well suited to the experimental data at hand are designed.

The eighth chapter (by DeGaetano et al.) takes a longer-term perspective on the effects of diabetes, presenting a model (Diabetes Progression Model) that aims to quantify the evolution of the glucose–insulin system and particularly its compensation to progressively worsening insulin resistance. This perspective is equally important in the design of therapeutic interventions, such as metabolic surgery or beta cell protection, as it provides a way to quantify the long-term effects of these (or other) therapies.

The ninth chapter (by Cescon and Johansson) considers the use of widely used systems identification techniques, such as autoregressive and state-space models, in order to perform individualized short-term (up to 2 h) prediction of future glucose values in a group of patients with Type 1 diabetes. While the overall performance of these approaches was deemed reasonable, the prerequisites to achieve this performance were partially met, suggesting room for improvement, such as the design of optimized experimental protocols to perform system identification.

Finally, the tenth chapter (by Marmarelis, Shin and Mitsis) considers the application of a multivariate, data-driven approach to quantifying the dynamic effects of spontaneous fluctuations of plasma insulin and free fatty acids on glucose concentration in a fasting dog. The obtained dynamic, nonlinear, data-driven models are then linked to more traditional measures of glucose metabolism, such as insulin sensitivity and its interactions with free fatty acid levels, illustrating the importance of obtaining models that are amenable to physiological interpretation.

Concluding, we wish to extend our warm thanks to all contributors for submitting high quality work to this research volume, which we hope will prove useful in promoting further research on this important topic.

March 2014

Vasilis Marmarelis, Los Angeles, CA
Georgios Mitsis, Montreal, QC

Contents

Data-Driven and Minimal-Type Compartmental Insulin-Glucose Models: Theory and Applications

Georgios D. Mitsis and Vasilis Z. Marmarelis

Abstract This chapter initially presents the results of a computational study that compares simulated compartmental and Volterra models of the dynamic effects of insulin on blood glucose concentration in humans. In this context, we employ the general class of Volterra-type models that are estimated from input-output data, and the widely used "minimal model" as well as an augmented form of it, which incorporates the effect of insulin secretion by the pancreas. We demonstrate both the equivalence between the two approaches analytically and the feasibility of obtaining accurate Volterra models from insulin-glucose data generated from the compartmental models. We also present results from applying the proposed approach to quantifying the dynamic interactions between spontaneous insulin and glucose fluctuations in a fasting dog. The results corroborate the proposition that it may be feasible to obtain data-driven models in a more general and realistic operating context, without resorting to the restrictive prior assumptions and simplifications regarding model structure and/or experimental protocols (e.g. glucose tolerance tests) that are necessary for the compartmental models proposed previously. These prior assumptions may lead to results that are improperly constrained or biased by preconceived (and possibly erroneous) notions—a risk that is avoided when we let the data guide the inductive selection of the appropriate model.

G. D. Mitsis (✉)
Department of Electrical and Computer Engineering, University of Cyprus,
P.O. Box 20537 Kallipoleos 75 1678 Nicosia, Cyprus
e-mail: gmitsis@ucy.ac.cy

G. D. Mitsis
Department of Bioengineering, McGill University, P.O. Box 20537
817 Sherbrooke St. W., Macdonald Engineering Building, Montreal,
QC H3A 0C3, Canada

V. Z. Marmarelis
Department of Biomedical Engineering, University of Southern California,
1042 Downey Way, Denney Research Center 140,
Los Angeles 90089 CA, USA
e-mail: vzm@bmsr.usc.edu

V. Marmarelis and G. Mitsis (eds.), *Data-driven Modeling for Diabetes*,
Lecture Notes in Bioengineering, DOI: 10.1007/978-3-642-54464-4_1,

1 Introduction

Diabetes mellitus represents an alarming threat to public health with rising trends and severity in recent years worldwide and is characterized by multiple and often not readily observable clinical effects [15]. Hence, there is urgent need for improved diagnostic methods that provide more precise clinical assessments and sensitive detection of symptoms at earlier stages of the disease [33]. This critical task may be facilitated (or enabled) by the utilization of advanced mathematical models that reliably describe the dynamic interrelationships among key physiological variables implicated in the underlying physiology (i.e. blood glucose concentration and various hormones such as insulin, glucagon, epinephrine, norepinephrine, cortisol etc.) under a variety of metabolic and behavioral conditions (e.g. pre-/post-prandial, exercise/rest, stress/relaxation). Such models would not only provide a powerful diagnostic tool, but may also enable long-term glucose regulation in diabetics through closed-loop model-reference control using frequent insulin micro-infusions administered by implanted programmable micro-pumps. This may prevent the onset of the pathologies caused by elevated blood glucose over prolonged periods in diabetic patients [15].

The primary effect on blood glucose is exercised by insulin and most efforts to date have focused on the study of this causal relationship. Prolonged hyperglycemia is usually caused by defects in insulin secretion by the pancreatic beta cells or in the efficiency of insulin-facilitated glucose uptake by the cells. The exact quantitative nature of the dependence between blood glucose concentration and the action of the other hormones mentioned above, or factors such as diet, endocrine cycles, exercise, stress etc. remains largely unknown—primarily because of lack of appropriate data although the qualitative effect has been established. Thus, the aggregate effect of all these other factors for modeling purposes is viewed as random "disturbances", additive to the blood glucose level.

Starting from the initial work of Bolie [7] and Ackerman [1], most modeling studies of the causal relationship between insulin and glucose (as the "input" and "output" of a system representing this relationship) have relied on the concept of compartmental modeling [9]. In this context, the minimal model (MM) of glucose disappearance, combined with the intravenous glucose tolerance test, has been the most widely used method to study whole body glucose metabolism in vivo [3]. The MM postulates that insulin acts from a remote compartment and affects glucose utilization, in addition to the insulin-independent utilization that depends on the glucose level *per se*. These insulin-dependent and insulin-independent effects on glucose utilization/kinetics are combined in a single compartment. Certain parameters of the MM (i.e., insulin sensitivity SI and glucose effectiveness SG) have been shown to be of clinical importance and can be estimated from IVGTT data, using nonlinear least-squares methods [4, 34] or, more recently, Bayesian estimation techniques [22, 23].

The accuracy of the estimates obtained from the MM has been questioned because of the single-compartment assumption [10, 16, 34], and two-compartment

models for glucose kinetics have been proposed [8,11]. Moreover, more complex models that aim to capture the complexity of the underlying physiology under more general operating conditions have been also proposed [13, 14, 38]. However, models of this type cannot be identified from standard clinical tests and they require the design of specialized experimental protocols. For instance, the model proposed in [14] includes 12 differential equations and 35 parameters, and a specialized triple-tracer experiment was designed in order to identify its parameters in 204 healthy individuals and 58 pre-diabetic/diabetic patients [14]. Other modeling approaches that have been explored—in the context of glucose control—include artificial neural networks [43], probabilistic models [2] and linear/nonlinear impulse response and Volterra models [18, 27, 35]. In addition to modeling insulin-glucose interactions, attempts have been made to take into account the influence of additional relevant physiological signals, such as glucagon [13, 26] and free fatty acids [37].

Most of the aforementioned compartmental models rely on *a priori* assumptions and simplifications regarding the underlying physiological mechanisms and their primary aim is often to extract clinically important parameters in conjunction with specific experimental protocols (e.g., the IVGTT). Therefore, their ability to quantify glucose metabolism under actual, more general operating conditions remains limited. On the other hand, recent technological advances in the development of reliable continuous glucose sensors and insulin micro-pumps [6, 19] have provided time-series data that enable the application of data-driven modeling approaches. These approaches offer new opportunities towards the goal of obtaining reliable models of the insulin-glucose interrelationships in a more general context. Using spontaneous or externally infused insulin and glucose data, one can obtain data-driven models that are not constrained by *a priori* assumptions regarding their structure.

In addition to the effects of external insulin or glucose stimuli, the interactions between spontaneous fluctuations of plasma glucose and insulin are of great interest. The precise characteristics of pulsatile insulin secretion patterns influence blood glucose regulation and are altered in Type II diabetes [24, 36]. Furthermore, insulin secretion is regulated in vivo by plasma glucose oscillations over various time scales (from rapid to ultradian) and changes in this relation may be potentially useful as an early marker of Type II diabetes development [36, 40].

Therefore, the purpose of the present chapter is twofold: First, we examine the relation between existing compartmental (differential equation) and Volterra-type models, both analytically and computationally. The results demonstrate the feasibility of obtaining Volterra models of insulin-glucose dynamics that are equivalent to widely accepted compartmental models, using data-records that are practically obtainable. They also illustrate the physiological interpretation of nonlinear Volterra models by providing direct links to a well-known parametric model with parameters of clinical significance. Since the Volterra approach does not require prior assumptions about model structure, it can provide the effective means for obtaining accurate data-true, patient-specific and time-adaptive models in a clinical context. Second, we showcase the application of the proposed data-

driven approach to modeling the dynamic interrelationships between spontaneous variations of plasma insulin and glucose in fasting dogs in the closed-loop context of the problem. Specifically, we examine both causal directions of the loop—i.e. considering first the insulin as input and the glucose as output, and vice versa. The resulting Principal Dynamic Modes (PDM) models are equivalent to Volterra models and have a modular form that facilitates physiological interpretation. The analysis of the experimental data yields PDM models that are comprised of two parallel branches in each causal direction, describing the primary physiological mechanisms of slow and fast dynamic interactions between variations of plasma insulin and glucose for each causal direction. Spectral analysis of the resulting insulin and glucose residuals (representing internal secretions and systemic disturbances) indicate the presence of oscillatory spontaneous variations of insulin and glucose at preferred frequencies in agreement to previous reported observations. Overall, our results demonstrate the potential of the proposed black-box modeling approach to advance our quantitative understanding of this system.

2 Insulin-Glucose Models

The present study concerns compartmental and Volterra-type nonlinear dynamic models; among compartmental models, we select the minimal model of glucose disappearance (MM), as well as an augmented version of it (AMM), which incorporates an insulin secretion equation. The structure and parameter values of these models are taken from the literature [3, 4, 20, 25, 41, 45]. The equivalent Volterra models [30] are estimated using simulated input–output data from the compartmental models in Sect. 3.

2.1 The Minimal Model of Glucose Disappearance

The MM of glucose disappearance is described by the following two differential equations [3], which describe the nonlinear dynamics of the insulin-to-glucose relationship during an IVGTT:

$$\frac{dg(t)}{dt} = -p_1 g(t) - x(t)[g(t) + g_b] \tag{1}$$

$$\frac{dg(t)}{dt} = -p_2 x(t) + p_3 i(t) \tag{2}$$

where $g(t)$ is the deviation of glucose plasma concentration from its basal value g_b (in mg/dl), $x(t)$ is the internal variable of insulin action (in min^{-1}), $i(t)$ is the deviation of insulin plasma concentration from its basal value i_b (in μU/ml), p_1 and

p_2 are parameters describing the kinetics of glucose and insulin action respectively (in min^{-1}) and p_3 is a parameter (in $\text{min}^{-2}\text{ml}/\mu\text{U}$) that affects insulin sensitivity SI (see below). The initial conditions for the simulations are: $g(0) = 0$ and $x(0) = 0$ (i.e. we assume that we start at basal conditions—which is a reasonable assumption in the context of simulating the model for situations where the initial "transient" phase can be ignored). Note that the MM is nonlinear, due to the presence of the bilinear term between the internal variable $x(t)$ representing insulin action and the variable $g(t) + g_b$ representing the plasma glucose concentration in the first equation. This bilinear term describes the modulation of the effective kinetic constant of the glucose utilization by insulin action (i.e. insulin concentration increases cause faster disappearance of blood glucose).

The physiological interpretation of the MM parameters can be made in terms of insulin-dependent and insulin-independent processes that enhance glucose uptake and suppress net glucose output. The parameter p_1, termed "glucose effectiveness" SG, represents the insulin-independent effect, while the insulin-dependent effect is represented by the ratio p_3/p_2 (in $\text{min}^{-1}/\mu\text{Uml}^{-1}$) and is termed "insulin sensitivity" SI. The values of SG and SI are typically estimated from IVGTT data and the MM has proven to be successful in a clinical context, requiring a relatively simple test procedure [5]. Nonetheless, the accuracy and physiological interpretation of the MM parameter estimates has been questioned because of the use of a single compartment for glucose kinetics [10, 11].

The MM, as formulated in Eqs. (1) and (2), does not include an equation describing the secretion of insulin from pancreatic beta cells in response to an elevation in blood glucose concentration, i.e., it is an open-loop model, which may be used along with properly designed experimental protocols (IVGTT) for parameter estimation. However, the actual glucose metabolism process is a closed-loop system, except in conditions of severe Type I diabetes where the pancreatic beta cells are considered totally inactive. In order to account for this, an insulin secretion equation may be included, as described below (closed loop MM or AMM). Limitations of the MM (and the AMM) include the absence of an explicit glucogenic component reflecting production of new glucose by the liver in response to elevated plasma insulin and/or glucose (such as the model presented in [26]) and the associated glucagon secretion process (from the alpha cells of the pancreas) among others. The aggregate effect of these processes, as well as the effect of other factors (free fatty acids, epinephrine etc.), can be incorporated by "disturbance" terms that are added to the glucose rate and insulin action equations.

2.2 *Closed-Loop Compartmental Model: The Augmented Minimal Model*

The closed-loop nature of insulin-glucose interactions requires the incorporation of an additional equation describing the insulin secretion dynamics by the pancreatic beta cells. Of several equations that have been proposed [4, 39, 41, 42, 45], we

select one that utilizes a threshold function, similar to the one reported in [42]. The resulting closed-loop model becomes:

$$\frac{dg(t)}{dt} = -p_1 g(t) - x(t)[g(t) + g_b] \tag{3}$$

$$\frac{dx(t)}{dt} = -p_2 x(t) + p_3[i(t) + r(t)] \tag{4}$$

$$\frac{dr(t)}{dt} = -ar(t) + \beta T_h[g(t)] \tag{5}$$

where $r(t)$ is the secreted insulin by the pancreatic beta cells in response to an elevation in plasma glucose concentration and $i(t)$ correspond to insulin concentration changes due to externally administered insulin. The secretion is triggered by elevated plasma glucose concentrations according to the threshold function $T_h[g(t)]$ defined as:

$$T_h[g(t)] = \begin{cases} g(t) - \theta & g(t) \geq \theta \\ 0 & otherwise \end{cases} \tag{6}$$

where θ corresponds to the glucose concentration value above which insulin is secreted. The dynamics of this triggered secretion process and the kinetics of the secreted insulin are described (in first approximation) by the kinetic constant a (in min^{-1}) in Eq. (5). The parameter β (in $\mu\text{U min}^{-1}$/ml per mg/dl) determines the rate of insulin secretion (i.e. the strength of the feedback pathway). Note that some alternative similar models [4, 41] of insulin secretion include a time-varying term that multiples the last term of Eq. (5) with t. This is based on the hypothesis that the rate of insulin secretion in response to hyperglycemia increases linearly with time. However, this term may not admit a steady-state solution but instead result in unbounded state variable values for physiologically reasonable values of the model parameters. This is not plausible; therefore, it should be taken into account when constructing models that are intended to be physiologically realistic, as discussed in more detail in the chapter by Panunzi and de Gaetano in the present volume.

2.3 Volterra-Type Models

The Volterra-Wiener framework has been employed extensively for modeling nonlinear physiological systems [30]. In this context, the input-output dynamic relationship of a causal, nonlinear system of order Q and memory M is described by the Volterra functional expansion:

$$g(t) = \sum_{n=0}^{Q} \int_0^M \dots \int_O^M k_n(\tau_1, \dots, \tau_n) i(t - \tau_1) \dots i(t - \tau_n) d\tau_1 \dots d\tau_n \tag{7}$$

The Volterra model can be formulated in discrete-time as follows:

$$g(t) = \sum_{n=0}^{Q} \sum_{\tau_1=0}^{M} \dots \sum_{\tau_n=0}^{M} k_n(\tau_1, \dots, \tau_n) i(t-\tau_1) \dots i(t-\tau_n) \tag{8}$$

In both the above models (Eqs. 7 and 8), $i(t)$ and $g(t)$ are the input and output of the system at time t (deviations of plasma insulin and glucose concentrations from their basal values, respectively). The unknown quantities of the Volterra model that are estimated from the input-output data are the Volterra kernels $k_n(\tau_1,\dots, \tau_n)$. The first-order kernel ($n = 1$) is the linear component of the system dynamics, while the higher order kernels ($n > 1$) form a hierarchy of the nonlinear dynamics of the system. The highest order Q defines the nonlinear order of the system. Many physiological systems can be described adequately by Volterra models of relatively low order (second or third) [30]. The Volterra-Wiener approach is well-suited to the complexity of physiological systems since it yields data-true models, without requiring a priori assumptions about system structure.

Among various methods that have been developed for the estimation of the discrete-time Volterra kernels (Eq. 8), a Volterra-equivalent network in the form of the Laguerre-Volterra Network (LVN) is selected because it has been proven to be an efficient approach that yields accurate representations of high-order systems in the presence of noise using short input-output records [31]. The LVN model consists of an input layer of a Laguerre filter-bank and of a hidden layer with K hidden units with polynomial activation functions (Fig. 1) [31]. At each discrete time t, the input signal $i(t)$ (insulin) is convolved with the Laguerre filter-bank and weighted sums of the filter-bank outputs V_j (where $v_j = i * b_j$, * denotes convolution and b_j is the j-th order discrete-time Laguerre function) are transformed by the hidden units through polynomial transformations.

The model output $g(t)$ (glucose) is formed as the summation of the hidden unit outputs z_k and a constant corresponding to the glucose basal value g_b:

$$u_k(t) = \sum_{j=0}^{L-1} w_{k,j} v_j(t) \tag{9}$$

$$g(t) = \sum_{k=1}^{K} z_k(t) + g_b = \sum_{k=1}^{K} \sum_{n=1}^{Q} c_{n,k} u_k^n(t) + g_b \tag{10}$$

where L is the number of functions in the filter bank and $w_{k,j}$ and $c_{n,k}$ are the weighting and polynomial coefficients respectively. The insulin and glucose time-series are used to train the LVN model parameters ($w_{k,j}$, $c_{n,k}$ and the Laguerre parameter which determines the Laguerre functions dynamic properties) with a gradient-descent algorithm, as described in [31].

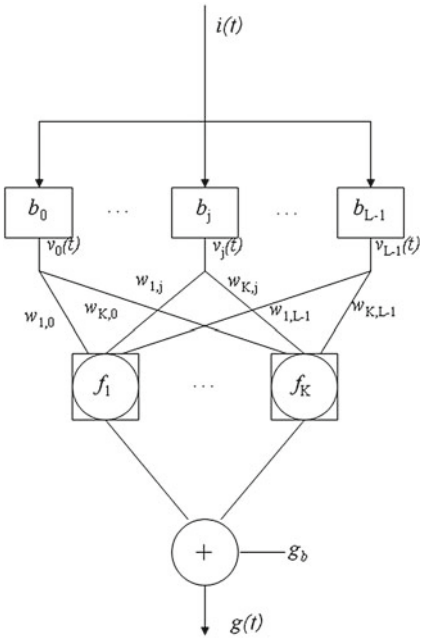


Fig. 1 The Laguerre-Volterra network. The system input $i(t)$ is convolved with a Laguerre filter bank with impulse responses b_j, the outputs of which ($v_j(n)$) are fed into a layer of K hidden units with polynomial activation functions f_K that produce the system output $g(t)$

The equivalent Volterra kernels are then obtained in terms of the LVN parameters as:

$$k_n(\tau_1,\ldots,\tau_n) = \sum_{k=1}^{K} c_{n,k} \sum_{j_1=0}^{L-1} \ldots \sum_{j_n=0}^{L-1} w_{k,j_1} \ldots w_{k,j_n} b_{j_1}(\tau_1) \ldots b_{j_n}(\tau_n) \qquad (11)$$

The structural parameters of the LVN model (L, K, Q) are selected on the basis of the normalized mean-square error (NMSE) of the output prediction achieved by the model, defined as the sum of squares of the model residuals divided by the sum of squares of the demeaned true output. The statistical significance of the NMSE reduction achieved for model structures of increased order/complexity is assessed by comparing the percentage NMSE reduction with the alpha-percentile value of a chi-square distribution with p degrees of freedom (p is the increase of the number of free parameters in the more complex model) at a significance level alpha, typically set at 0.05.

The LVN representation is equivalent to a variant of the general Wiener-Bose model termed the Principal Dynamic Mode (PDM) model. The PDM model consists of a set of parallel branches, each one of which is the cascade of a linear dynamic filter (PDM) followed by a static nonlinearity [29, 30]. Each of the K hidden units of the LVN corresponds to a separate branch and defines the respective PDM $p_k(t)$ and polynomial nonlinearity $f_k(.)$. This leads to model representations that allow physiological interpretation, since the resulting number of branches is typically low

in practice. According to the PDM model form, the insulin input signal is convolved with each of the PDMs $p_k(t)$, where $k = 1, \ldots, K$ and $p_k(t) = \sum_{j=0}^{L-1} w_{k,j} b_j(t)$, and the PDM outputs u_k are subsequently transformed by the respective polynomial nonlinearities $f_k(.)$ to produce the model-predicted blood glucose output (the asterisk denotes convolution) [30]:

$$\begin{aligned} g(t) &= g_b + f_1[u_1(t)] + \ldots + f_K[u_K(t)] \\ &= g_b + f_1[p_1(t) * i(t)] + \ldots + f_K[p_K(t) * i(t)] \end{aligned} \tag{12}$$

Therefore, once an LVN model is trained based on input-output data, the PDMs and their associated nonlinearities can be readily obtained using the final (trained) values of the LVN parameters, i.e., weights $w_{k,j}$, polynomial coefficients $c_{n,k}$ and Laguerre parameter α. For more details, the reader is referred to [32] (Chap. 2).

3 Comparison Between Compartmental and Volterra Models

3.1 Generalized Harmonic Balance Method

In order to examine the mathematical relationship between the aforementioned compartmental and Volterra models, we employ the generalized harmonic balance method to derive analytical relations between the two model forms, as outlined below for the second-order case of the nonparametric model [28]. This procedure can be extended to any order of interest.

By setting the input $i(t)$ equal to 0, e^{st} and $e^{s_1 t} + e^{s_2 t}$ in the general Volterra model of Eq. (7) successively, the output $g(t)$ becomes equal to k_0, $k_0 + e^{st}K_1(s) + e^{2st}K_2(s, s) + \ldots$ and $k_0 + e^{s_1 t}K_1(s_1) + e^{s_2 t}K_1(s_2) + e^{s_1 t + s_2 t}K_2(s_1, s_2) + \cdots$ where $K_1(s)$ and $K_2(s_1, s_2)$ are the Laplace transforms of $k_1(\tau)$ and $k_2(\tau_1, \tau_2)$ respectively. If we substitute these three input-output pairs into the differential equations of the compartmental models (Eqs. 1 and 2) for the open-loop model and (3–5) for the closed-loop model) and equate the coefficients of the resulting exponentials of the same kind, we can obtain analytical expressions for k_0, $K_1(s)$ and $K_2(s_1, s_2)$, in terms of the parameters of the respective compartmental model.

To define the computational equivalence between the two model forms, we simulate the compartmental models with broadband input (insulin) data and we then estimate the kernels of the equivalent Volterra model, from the simulated input- output data. The accuracy of the estimated first and second-order Volterra kernels is assessed by comparison with the exact kernels of the equivalent Volterra model that is derived in analytical form from the differential equations of the compartmental models. The accuracy and robustness of the kernel estimates is evaluated under measurement noise conditions, in order to assess the performance of the Volterra approach.

3.2 Analytical Expressions of the Volterra Kernels of the Compartmental Model: Open-Loop Case

The bilinear term between insulin action and glucose concentration in Eq. (1) of the MM gives rise to an equivalent Volterra model of infinite order. However, for parameter values within the physiological range, a second-order Volterra model offers an adequate approximation for all practical purposes. Considering the insulin and glucose deviations from the respective basal values $i(t)$ and $g(t)$ as the input and the output respectively, we can derive analytically the Volterra kernels of the open-loop MM by applying the procedure outlined in Sect. 3.1 to the integro-differential equation:

$$\dot{g}(t) + p_1 g(t) + p_3 \int_0^\infty exp(-p_2\tau) i(t-\tau) d\tau = -g_b p_3 \int_0^\infty exp(-p_2\tau) i(t-\tau) d\tau \tag{13}$$

The above equation is derived from the MM by substituting the convolutional solution of Eq. (2):

$$x(t) = p_3 \int_0^\infty exp(-p_2\tau) i(t-\tau) d\tau \tag{14}$$

into Eq. (1). Upon application of this method, we derive the following analytical expressions in the Laplace domain for the first- and second-order Volterra kernels of the MM ($k_0 = 0$):

$$K_1(s) = p_3 g_b \frac{1}{(s+p_1)(s+p_2)} \tag{15}$$

$$K_2(s_1, s_2) = \frac{p_3^2 g_b}{2} \frac{1}{(s_1+p_1)(s_1+p_2)} \frac{1}{(s_2+p_1)(s_2+p_2)} \left[1 + \frac{p_2}{s_1+s_2+p_1}\right]. \tag{16}$$

The MM has, in principle, Volterra kernels of any order. However, it can be shown that the magnitude of the n-th order kernel is proportional to the n-th power of p_3 and, subsequently, an adequate Volterra model may only include the first two kernels (since the value of p_3 is on the order of 10^{-5}–10^{-4}). The resulting expressions for the first and second order kernels in the time domain are given in Eqs. (17) and (18) respectively:

$$k_1(\tau) = -g_b \frac{p_3}{p_2 - p_1} [\exp(-p_2\tau_1) - \exp(-p_2\tau)] \tag{17}$$

$$k_2(\tau_1,\tau_2) = \frac{p_3^2 g_b}{2(p_2-p_1)^2}\Bigg[[\exp(-p_1\tau_1)-\exp(-p_2\tau_1)][\exp(-p_1\tau_2)-\exp(-p_2\tau_2)] + p_2\Bigg[\Bigg[\frac{1}{p_1}\exp[-p_1(\tau_1+\tau_2)].[\exp(p_1\min(\tau_1,\tau_2)-1)]\Bigg] - \frac{1}{p_2}\exp[(-p_1\tau_1-p_2\tau_2)+\exp(-p_1\tau_2-p_2\tau_1)](\exp[p_2\min(\tau_1,\tau_2)]-1) + \frac{\exp(-p_2(\tau_1-\tau_2))}{2p_2-p_1}(\exp[(2p_2-p_1)\min(\tau_1,\tau_2)]-1)\Bigg]\Bigg] \tag{18}$$

These first and second-order Volterra kernels are plotted in Fig. 2 (top panel) for typical MM parameter values within the physiological range [25, 34]: g_b= 80 mg/dl, p_1= S_G = 0.02 min^{-1}, p_2 = 0.028 min^{-1} and p_3= 10^{-4} min^{-2}ml/μU, which yield S_1= 0.0036 min^{-1}/μU ml^{-1}. Since the specific parameter values define the MM description of insulin-glucose dynamics, they also define the form of the equivalent Volterra kernels. The form of the first-order kernel in Fig. 2 (top left panel) indicates that an 10 μU/ml insulin concentration increase will cause a first-order drop in plasma glucose concentration that will reach a minimum of about −1.2 mg/dl about 36 min later, rising after that to half the drop in about 1 h and relaxing back to the basal value about 4 h after the minimum. The positive values of the second-order Volterra kernel indicate that the actual glucose drop caused by the insulin infusion will be slightly less than the first-order prediction (sublinear response). For instance, an insulin concentration increase of 100 μU/ml will not cause a maximum glucose drop of 12 mg/dl (as predicted by its equivalent first-order kernel) but a drop of about 10.5 mg/dl due to the antagonistic second-order kernel contribution.

Changes in these parameter values affect the form and the values of the kernels in the precise manner described by Eqs. (17) and (18). The effects of changes in the two MM parameters p_1 and p_2 on the equivalent first-order kernel are illustrated in Fig. 2 (bottom panels) for a range of physiological values (p_1 between 0.01 and 0.04 min^{-1} and p_2 between 0.02 and 0.05 min^{-1} [34], keeping $p_3 = 10^{-4}$ min^{-2} ml/μU constant). Note that changes in p_3 simply scale the first-order kernel according to Eq. (17) and do not alter its form (proportional dependence)—nor do they alter the form of the second-order kernel (they scale it quadratically). A direct sense of the effects of parameter changes is obtained by the waveforms of Fig. 2: for instance, as $p_1(S_G)$ increases, the maximum drop of the first-order kernel becomes smaller and its dynamics (i.e. the drop to the minimum and the return to basal value) become faster. Similar effects are observed when p_2 increases (or S_I decreases).

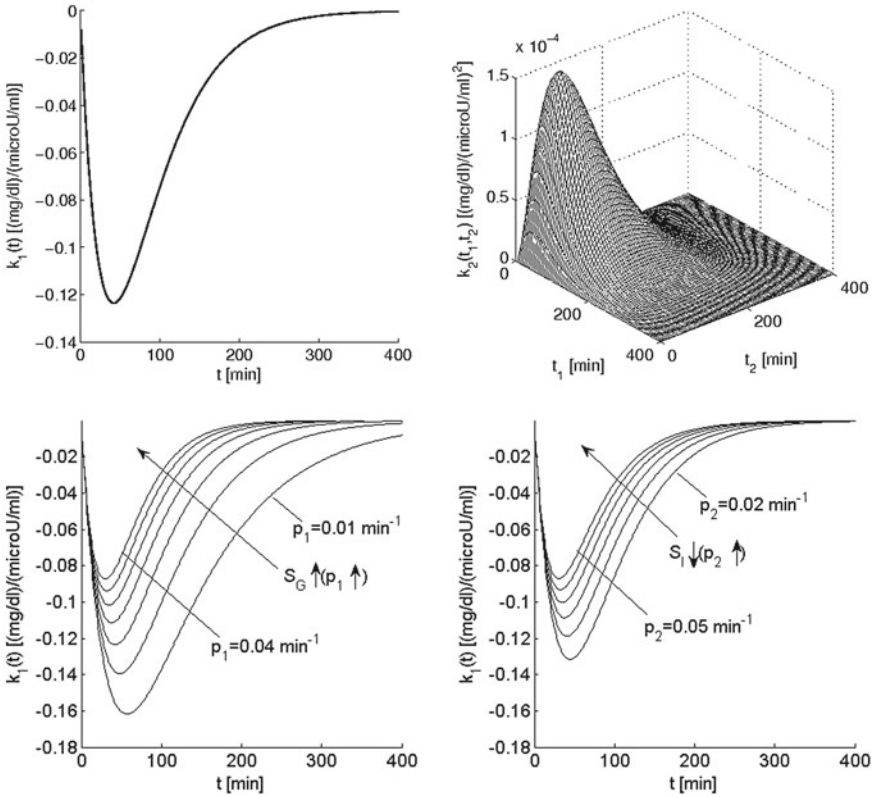


Fig. 2 *Top panel* The first-order (*left*) and second-order (*right*) Volterra kernels of the minimal model for typical values of its parameters within the physiological range ($S_G = 0.02\ \text{min}^{-1}$ and $S_I = 0.0036\ \text{min}^{-1}/\mu\text{U ml}^{-1}$). *Bottom panel* Effect of the two key parameters p_1 and p_2 of the open-loop MM on the form of the equivalent first-order kernel. Note that the glucose effectiveness S_G is equal to p_1 and the insulin sensitivity S_I is inversely proportional to p_2 (and proportional to p_3). These plots offer a visual understanding of the effects of changes in these parameters (p_1 between 0.01 and 0.04 min^{-1}, p_2 between 0.02 and 0.05 min^{-1}) on the first-order insulin-glucose dynamics (see text)

3.3 Analytical Expressions of the Volterra Kernels of the Compartmental Model: Closed-Loop Case

To derive the analytical expressions of the kernels in the closed-loop case, we approximate the threshold function of Eq. (6) with a polynomial as indicated below, assuming that θ is equal to zero (i.e. insulin secretion is triggered when the glucose concentration rises above its basal value):

$$\beta T_h[g(t)] \approx \beta_1 g(t) + \beta_1 g^2(t) + \ldots \tag{19}$$

where $g(t)$ is the deviation of glucose plasma concentration from its basal value. Equation (19) provides an accurate representation of (6) within a desired dynamic range for glucose values $g(t)$ (Weierstrass theorem), whereby the coefficients β_i can be estimated in a mean-square sense. However, note that the subsequent analysis is valid for any values of these coefficients. Equation (5) can be rewritten as:

$$\frac{dr(t)}{dt} = -ar(t) + \beta_1 g(t) + \beta_2 g^2(t) + \ldots \tag{20}$$

The solution of Eq. (20) is given by:

$$r(t) = \beta_1 f(t) * g(t) + \beta_2 f(t) * g^2(t) + \ldots \tag{21}$$

where the asterisk denotes convolution and $f(t) = e^{-at}u(t)$. Also, from Eq. (4) we have

$$\frac{dx(t)}{dt} = -p_3 h(t) * [i(t) + r(t)] \tag{22}$$

where $h(t) = e^{-p_2 t}u(t)$. Then, Eq. (3) becomes:

$$\frac{dg(t)}{dt} + p_1 g(t) = -p_3 g(t)\big[h(t) * i(t) + \beta_1 h(t) * f(t) * g(t) + \beta_2 h(t) * f(t) * g^2(t) + \ldots\big]. \tag{23}$$

The above equation can be used to obtain the equivalent Volterra kernels of the closed-loop model, following the procedure outlined before for the open-loop model. The resulting expressions for the first-order and the second-order kernels in the Laplace domain are given by Eqs. (24) and (25) respectively ($k_0 = 0$):

$$K_1(s) = -p_3 g_b \frac{H(s)}{(s + p_1 + p_3 g_b \beta_1 F(s) H(s))} \tag{24}$$

$$\begin{aligned} K_2(s_1, s_2) = & -p_3[(\beta_1 + \beta_2)] g_b \frac{H(s_1 + s_2) F(s_1 + s_2) K_1(s_1) K_1(s_2)}{s_1 + s_2 + p_1 + p_3 g_b \beta_1 H(s_1 + s_2) F(s_1 + s_2)} \\ & + \frac{1}{2} \frac{H(s_1) K_1(s_2) H(s_2) K_1(s_1)}{s_1 + s_2 + p_1 + p_3 g_b \beta_1 H(s_1 + s_2) F(s_1 + s_2)} \end{aligned} \tag{25}$$

where $F(s)$, $H(s)$ are the Laplace transforms of $f(t)$, $h(t)$ respectively, i.e. $F(s) = \frac{1}{s+a}$, $H(s) = \frac{1}{s+p_2}$.

The above relations were inverted numerically to yield the time-domain expressions for the first-order kernel, which are shown in Fig. 3 for the following parameter values: a varying between 0.1 and 0.3 min^{-1} with β remaining constant at 0.05 $\mu\text{U-min}^{-2}$/ml per mg/dl (left panel) and β varying between 0.0001 and 0.1

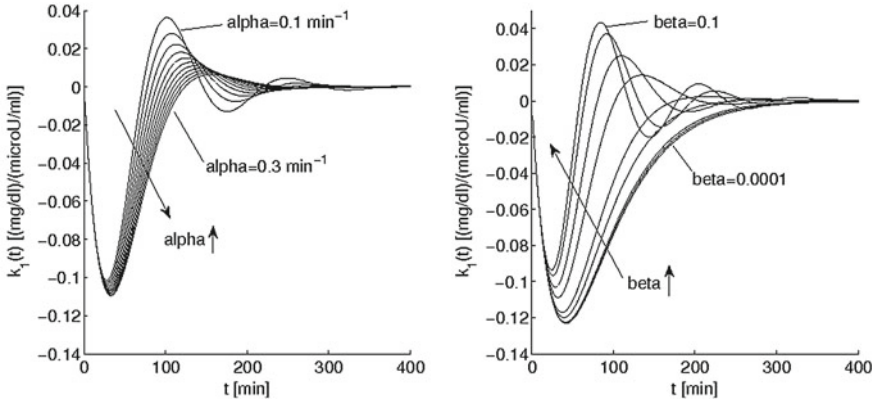


Fig. 3 The first-order kernels of the AMM for a varying between 0.1 and 0.3 min^{-1} with constant $\beta = 0.05$ (*left panel*) and for β varying between 0.0001 and 0.1 μU-min^{-2}/ml per mg/dl with constant $a = 0.13$ min^{-1} (*right panel*)

μU-min^{-2}/ml per mg/dl with a remaining constant at 0.13 min^{-1} (right panel). The nominal value of a (0.13 min^{-1}) was taken from [45], while the value of β was set at 0.05 μU-min^{-2}/ml, since the value reported in [45] (0.0054) resulted in negligible effects of endogenous insulin secretion for the stimuli used in this study. The decrease of a (slower insulin secretion dynamics) and increase of β (stronger feedback) affect the AMM first-order kernel waveform similarly—i.e., they result in faster dynamics with a small decrease of the negative peak value and the appearance of an overshoot which is characteristic of closed-loop systems.

3.4 Simulation Results: Open-Loop Model

In order to demonstrate the feasibility of estimating the Volterra kernels of the open-loop MM directly from input-output measurements, we simulate it by numerical integration of Eqs. (1) and (2) for the following values of MM parameters: $p_1 = 0.020$ min^{-1}, $p_2 = 0.028$ min^{-1}, $p_3 = 10^{-4}$ min^{-2} ml/μU, $g_b = 80$ mg/dl that are around the middle of the physiological ranges reported in the literature [4, 34]. The input signal for this simulation is a zero-mean Gaussian white noise (GWN) sequence of insulin time-series (i.e. independent samples every 5 min), with a standard deviation of 4 μU/ml, which may be viewed as spontaneous fluctuations around its basal value or arising from step-wise continuous infusions of insulin at random levels, changed every 5 min, superimposed on a constant (positive) baseline infusion. Due to the low-pass dynamic characteristics of the model, one sample every 5 min is sufficient for representing the input-output data. An input-output record of 144 sample points (i.e., 12 h long) is used to perform the training of the LVN and the estimation of the kernels of the equivalent Volterra model.

Table 1 Output prediction NMSEs for various LVN model structures and values of p_3, GWN input (open-loop case)

L	$p_3 = 5 \cdot 10^{-5}$		$p_3 = 5 \cdot 10^{-4}$		$p_3 = 5 \cdot 10^{-4}$	
	$\text{min}^{-2}\text{ml}/\mu\text{U}$		$\text{min}^{-2}\text{ml}/\mu\text{U}$		$\text{min}^{-2}\text{ml}/\mu\text{U}$	
	Linear NMSE	Nonlinear NMSE	Linear NMSE	Nonlinear NMSE	Linear NMSE	Nonlinear NMSE
2	13.55	8.97	16.24	15.46	22.42	4.89
3	0.39	0.32	0.68	0.30	23.23	1.13
4	4.62	4.85	3.33	3.63	23.88	3.73
5	0.17	0.14	0.40	0.09	21.35	0.61
6	0.22	0.31	0.39	0.17	21.82	0.61

The value of p_3 determines the relative contribution of the nonlinear terms: note that for $p_3 = 5 \cdot 10^{-5}$ min^{-2} ml/μU the NMSE reduction achieved by nonlinear models is marginal, while for $p_3 = 5 \cdot 10^{-4}$ min^{-2} ml/μU it is over 20 %. Using $L > 5$ does not improve model performance further

In order to illustrate model structure selection, we show the obtained NMSEs for various values of L, as well as for linear ($Q = 1$) and nonlinear ($Q = 2$) models for three different values of p_3, which determines the strength of the MM nonlinearity, in Table 1. For $p_3 = 5 \cdot 10^{-5}$ $\text{min}^{-2}\text{ml}/\mu\text{U}$ the model is weakly nonlinear, whereas for $p_3 = 5 \cdot 10^{-4}$ $\text{min}^{-2}\text{ml}/\mu\text{U}$ the NMSE reduction achieved for $Q = 2$ is over 20 %. The contribution of the n-th order Volterra term is proportional to the n-th power of the product of parameter p_3 with the power level of the input (i.e., this contribution increases for larger insulin variations); however, for the range of values examined, a second-order model is found to be sufficient. Also, using $L > 5$ reduces the NMSE minimally in all cases. Therefore, we select a second-order LVN with one hidden unit and five Laguerre functions (i.e., $L = 5$, $K = 1$, $Q = 2$) for the estimation of the equivalent Volterra model, with the resulting output prediction NMSE being 0.09 % ($p_3 = 10^{-4}$ $\text{min}^{-2}\text{ml}/\mu\text{U}$). The estimated kernels of first (Fig. 4—dotted) and second order for the noise free case are almost identical to the true kernels given by Eqs. (17) and (18) (Fig. 2—top panel).

In order to examine the effect of measurement noise on the kernel estimates, we repeat the kernel estimation with the aforementioned input-output data after the addition of 20 independent white-noise signals with maximum amplitude equal to approximately 20 % of the basal glucose value (i.e., error range of ±16 mg/dl) to the output [21]. This corresponds to an SNR of around 6.5 dB relative to the demeaned glucose deviations output. The resulting kernel estimates are also shown in Fig. 4 (top panels) and demonstrate the robustness of this modeling approach in the presence of measurement noise. The corresponding linear and nonlinear NMSEs are equal to 24.0 ± 2.7 and 23.6 ± 2.7 % respectively (mean ± standard deviation), i.e., the output additive noise is not accounted by the model. Also in Fig. 4 (bottom panels), we present the kernel estimates obtained with an insulin input of the same length (144 points) composed of a random sequence of impulses (representing insulin concentration increases that could be due to insulin infusions), with a mean frequency of 1 impulse every 2 h and a normally distributed random amplitude with

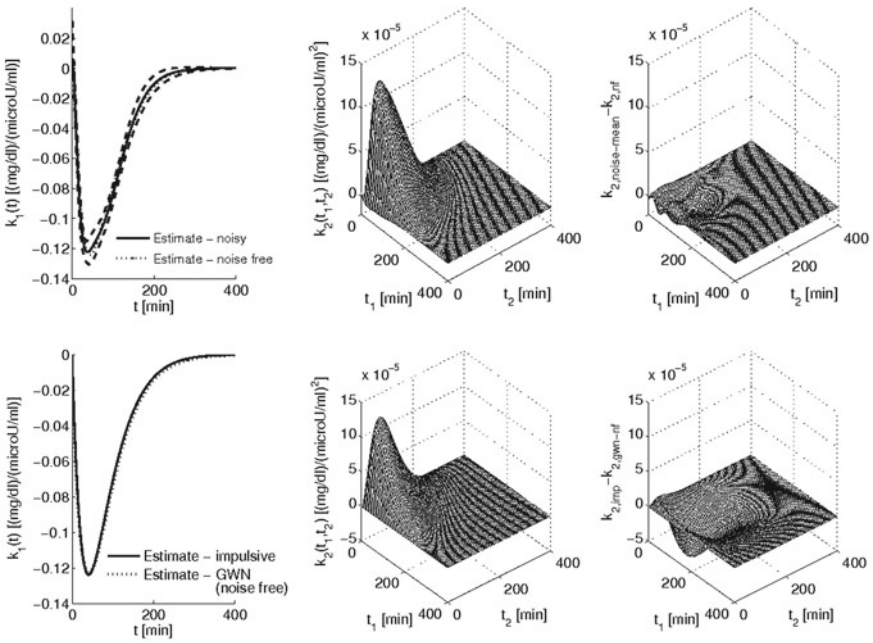


Fig. 4 *Top panel* The estimated first and second order Volterra kernels of the MM using a GWN input of 144 points (12 h) when 20 different realizations of independent GWN signals are added to the output for an SNR of 6.5 dB. The obtained first-order (*left panel—solid* mean value, *dashed* ± one standard deviation, *dotted* noise-free estimate) and second-order kernel estimates (*middle panel*—mean value, *right panel* standard deviation) are not affected significantly relative to their exact counterparts (Fig. 1—top panel), demonstrating the robustness of this approach. *Bottom panel* The estimated first and second order Volterra kernels of the MM for an insulin input composed of 8 insulin infusions over 12 h. The timing and amplitude of each infusion are random (see text). Note the similarity of these estimates to the estimates obtained from GWN inputs

standard deviation 20 μU/ml. The resulting kernel estimates are almost identical to their GWN-input counterparts, demonstrating the feasibility of estimating accurate Volterra models using sparser, infusion-like stimuli.

3.5 Simulation Results: Closed-Loop Model

The closed-loop AMM was simulated by numerical integration of Eqs. (3)–(5), using the same GWN input used for the open-loop MM. The parameter values used for this model were $p_1 = 0.020\ \text{min}^{-1}$, $p_2 = 0.028\ \text{min}^{-1}$, $p_3 = 10^{-4}\ \text{min}^{-2}\text{ml}/\mu\text{U}$ and parameter values of $a = 0.13\ \text{min}^{-1}$, $\beta = 0.05\ \mu\text{U-min}^{-2}/\text{ml}$ per mg/dl, $\theta = 80$ mg/dl for the additional insulin-secretion equation. Representative time-series data of the resulting insulin input, insulin secretion, insulin action and glucose, used for training the equivalent LVN model, are shown in Fig. 5, where the effect

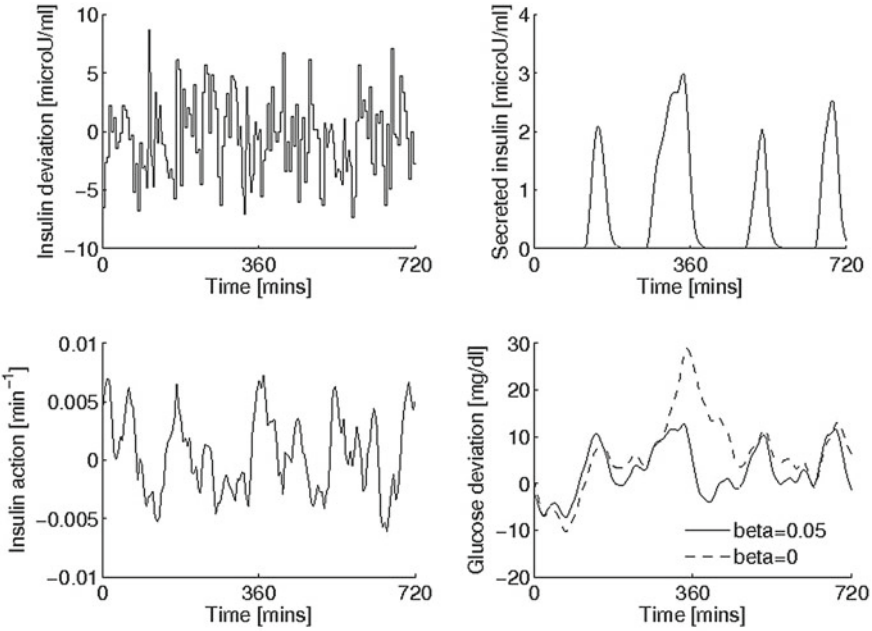


Fig. 5 Representative realization of the closed-loop AMM time-series data for a GWN insulin input used for LVN training (length: 12 h). The insulin time series represent deviations from its basal value. The effect of the secretion equation is seen by comparing the two output waveforms of glucose deviations shown in the bottom right panel (*dashed* open-loop, *solid* closed-loop for $\beta = 0.05$)

of insulin secretion, relative to the open-loop case, can be seen in the bottom right panel (solid: closed-loop output, dashed: open-loop output).

An LVN with $L = 5$, $K = 2$ and $Q = 3$ was employed in this case—i.e., a more complex structure of higher order is required relative to the open-loop case. In the noise-free case, the obtained nonlinear model reduces the prediction NMSE considerably, from 12.41 %—yielded by the linear model—to 2.18 % (Fig. 6, top left panel). As before, we repeat the kernel estimation after adding 20 independent white noise sample signals (with the same variance as above) to the output. Note that the resulting SNR is now around 4.5 dB, i.e. lower than the open-loop case, since the noise-free output (glucose deviations) has a smaller mean-square value in the closed-loop case, due to the effect of the endogenous insulin secretion. Therefore, the corresponding NMSEs are larger—i.e. 48.2 % for the linear model and 34.2 $\pm$ 4.0 % for the nonlinear model—and correspond, for the nonlinear model, to the noise present in the signal. This demonstrates the predictive capability of the obtained models in the presence of considerable output-additive noise that emulates the observed errors in the measurements of current continuous glucose monitors [21]. The kernel estimates for both cases are shown in Fig. 6, illustrating the robustness of this approach. The first and second order kernels of the closed-loop AMM exhibit biphasic characteristics (i.e., regions of positive and

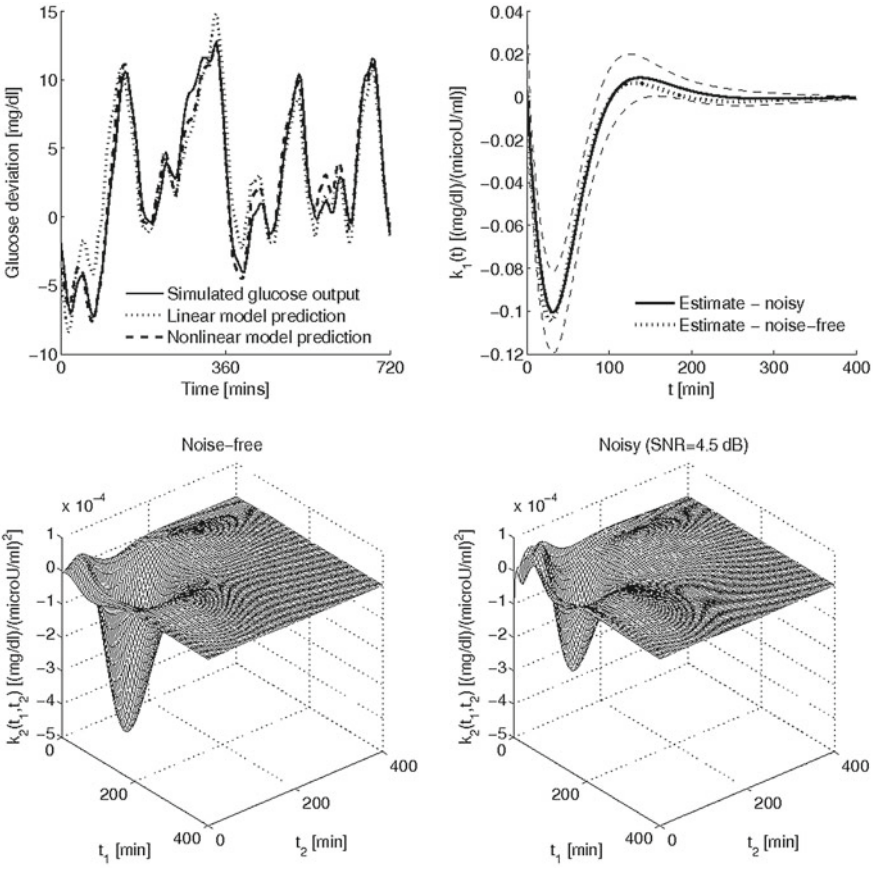


Fig. 6 Representative model predictions (noise-free output, *top left*) and estimated first and second order Volterra kernels of the closed-loop AMM for a GWN input of 144 points (12 h) for noise-free output (*top right—dotted* and *bottom left*) and when 20 different realizations of independent GWN measurement noise are added to the output for an SNR of 4.5 dB (*top right—solid black* mean, *dashed black* ± one standard deviation and *bottom right*—mean). Nonlinear models achieve better predictions (over 10 % NMSE reduction). The obtained kernel estimates are not affected significantly relative to their noise-free counterparts despite the low SNR

negative response to a positive change in the input, and vice versa). The first-order kernel contribution to the output remains dominant over the second-order kernel contribution for impulsive inputs up to about 100 μU/ml.

3.6 Simulation Results: Principal Dynamic Mode Models

The obtained equivalent PDM models for both the open-loop and closed-loop models are shown in Fig. 7. In the open-loop case (top panel), since we have used $K = 1$ in the LVN model, the equivalent PDM model has one branch, with the

PDM dynamics exhibiting similar characteristics to the open-loop first-order kernel (Fig. 7) and the static nonlinearity being close to linear, due to the relatively low value of p_3 used in this particular case. In the closed-loop case (bottom panel), we have used $K = 2$; therefore, the equivalent PDM model has two branches. The lower PDM exhibits a clear biphasic response characteristic (corresponding to a glucose decrease and increase respectively, in response to an insulin increase) that is not present in the open-loop model. The upper PDM branch exhibits slower dynamics (peak latency of about 80 min) than the open-loop PDM (peak latency at 40 min) and a strictly negative nonlinearity (i.e., always leading to a reduction of glucose), while the nonlinearity of the open-loop model has both positive and negative response regions. The PDM of the lower branch exhibits faster dynamics (shorter latency of the first peak of about 30 min) and has a nonlinearity that resembles a sigmoidal (soft saturating) characteristic.

3.7 Simulation Results: Sorensen Model

Finally, we present results from fitting the MM and LVN models to simulated data obtained from the model proposed by Sorensen [38], which has been used as a comprehensive representation of the metabolic system in several studies (e.g., [25, 35]) for insulin input signals considered above (i.e., random insulin variations around a putative basal value). Note that we do not make claims about the universal validity of this particular model, but we use it as a third-party metabolic simulator for comparative purposes. We considered two distinct cases of Sorensen model parameters: one that corresponds to a healthy subject and another that corresponds to a Type-1 diabetic subject, following the procedure described in [26]. Briefly, Type 1 diabetes is characterized by complete failure of the pancreatic beta cells, decreased (40–50 % of normal) insulin stimulated hepatic and periphery glucose uptake and impaired, glucose-induced, endogenous glucagon production (assumed 50 % of normal in this study). Therefore, the corresponding parameters in Sorensen's model were changed accordingly [26].

The MM parameters were obtained by using a nonlinear optimization method (Levenberg-Marquardt method) in order to fit p_1, p_2 and p_3 to the Sorensen model generated data. We considered 10 different realizations of the insulin input signal (of the same length considered above) and provide the results in Fig. 8. The results show that the output prediction performance of the LVN model is superior in both cases, particularly for the Type-1 diabetic case. Specifically, the average NMSEs for the LVN approach were equal to 3.62 ± 1.92 % and 4.95 ± 4.90 % for the healthy and Type-I diabetic cases respectively, while for the MM approach the corresponding NMSEs were equal to 11.11 ± 7.52 % and 25.37 ± 10.73 %, i.e., considerably higher. We note that an LVN model structure with $L = 5$, $K = 1$ and $Q = 2$ was deemed appropriate in this case.

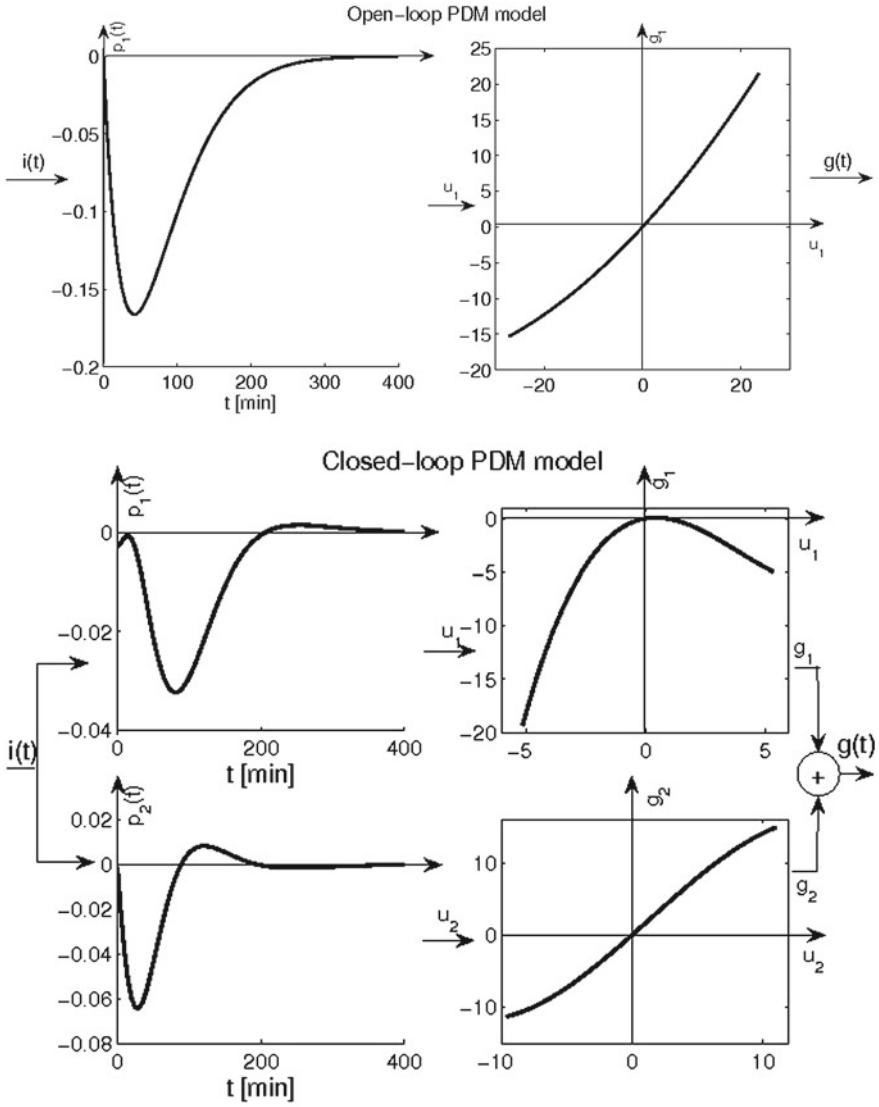


Fig. 7 The obtained PDM model for the open- and closed-loop models, which consist of one and two branches (*top* and *bottom panels* respectively). The open-loop single PDM (*top left panel*) exhibits a glucoleptic characteristic (reduces the glucose output) for positive insulin inputs in a mildly sublinear manner. The closed-loop upper PDM branch exhibits a glucoleptic characteristic for positive or negative insulin inputs in a mildly supralinear manner, unlike the single PDM branch of the open-loop MM. Note that the latency of the peak response (about 80 min) is much longer for this closed-loop PDM than for the open-loop PDM (about 40 min), and the slope of its output nonlinearity is different for positive/negative input (about 4 to 1). The lower PDM is biphasic with the first glucoleptic peak having a latency comparable to the open-loop PDM (about 30 min) and the second glucogenic peak being much smaller (about 15 %) and having a latency of about 120 min. The nonlinearity of the lower PDM branch retains the biphasic response characteristic (increase of insulin leads to glucose decrease and vice versa) and is mildly sublinear (resembling a soft saturating characteristic)

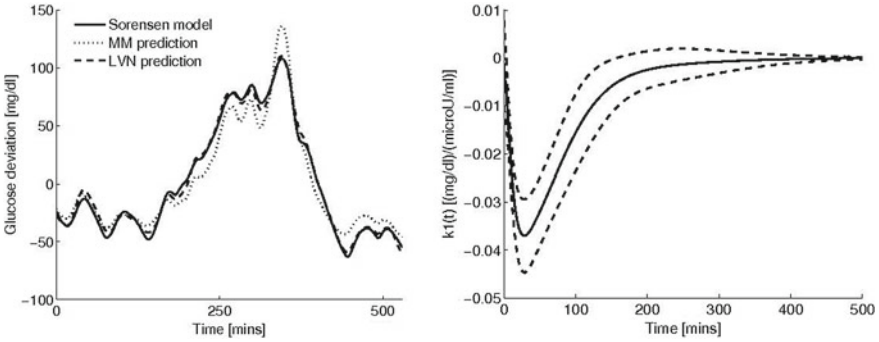


Fig. 8 The predictions of the MM and LVN models for a representative Sorensen-model simulated data set (healthy subject; *left panel*) and the average first-order kernel estimate of the LVN model for 10 different insulin input realizations (*right panel—solid black* mean *dashed line* ± one standard deviation)

4 Insulin-Glucose Dynamics in a Fasting Dog

4.1 Experimental Data Collection and Analysis

Plasma glucose and insulin were measured in a healthy male mongrel dog every 3 min over a 10-h period (200 samples). The data were collected under conditions of spontaneous activity (there were no insulin or glucose injections). The animal was judged to be in good health by visual observation, weight stability, body temperature, and hematocrit. The University of Southern California Institutional Animal Care Committee approved all surgical and experimental procedures. The dog was fed standard can food and had free access to standard dry chow and tap water. A chronic catheter (Tygon, ID = 0.13 cm) was implanted one week prior to the experiment. The catheter was placed into the femoral vein and advanced into the inferior vena cava and tunneled subcutaneously to the neck and exteriorized. The experiments were performed in the morning after a 12 h fast and the dog was not fed during the experiment. One hour prior to the beginning of blood sampling, an infusion of 500 ml 0.9 % saline was given to avoid dehydration of the animal due to the extended fast. Arterial blood was sampled at 3-minute intervals for 10 h from 6:00 a.m. till 4:00 p.m. Heart rate and blood pressure were monitored every 30 min over the entire experimental period. Samples were collected in tubes containing EDTA and 0.275 mg/ml lipoprotein lipase inhibitor paraoxon to avoid in vitro lipolysis. Samples were immediately centrifuged, and the plasma separated and stored at −20 °C. Plasma glucose was analyzed by the glucose oxidase method on an automated analyzer (model 23A, Yellow Springs Instruments). Insulin was measured in singlicate by an enzyme-linked immunospecific assay (ELISA, CV = 2 %) originally developed for human plasma by Novo-Nordisk and adapted for dog plasma. The collected insulin/glucose time-series data are

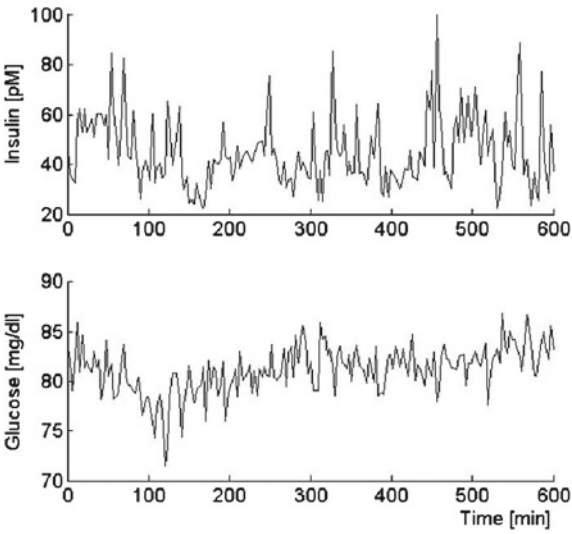


Fig. 9 Experimental time-series data of plasma insulin (*top panel*) and glucose (*bottom panel*) collected from a fasting dog over 10 h (sampling interval is 3 min)

shown in Fig. 9. All signals were demeaned prior to processing and the means were considered to be the respective basal values.

The time-series data were analyzed using the LVN methodology outlined above (Sect. 2.3) in both causal directions (i.e. when the spontaneous plasma insulin variations are viewed as the input and the plasma glucose variations are viewed as the output, and when the roles of input and output are reversed). The LVN model contains the Laguerre parameter α (between 0 and 1), which determines the rate of relaxation of the Laguerre functions. We utilized an LVN model with two filter-banks with distinct α parameter values [31], in order to capture the identified fast and slow dynamics between insulin and glucose [36] efficiently. The resulting Volterra-equivalent model has predictive capability for all possible inputs within the dynamic range of the data used for its estimation. The structural parameters of the LVN model were selected through the search procedure described above (Sect. 2.3). After training the LVN model with the insulin-glucose data, equivalent PDM models (Eq. 12) were obtained in both causal directions.

4.2 Insulin-to-Glucose Branch

The mean and standard deviation of the plasma insulin and glucose data were equal to 44.9 ± 13.8 pM and 81.1 ± 2.4 mg/d1 (mean ± standard deviation) respectively. Figure 10 shows the estimated modular PDM model for the insulin-to-glucose relationship. The structural parameters L, K and Q of the model were selected as follows: two Laguerre functions for each filter bank ($L_1 = L_2 = 2$), two parallel PDM branches ($K = 2$) and cubic nonlinearity ($Q = 3$). The abscissa of the

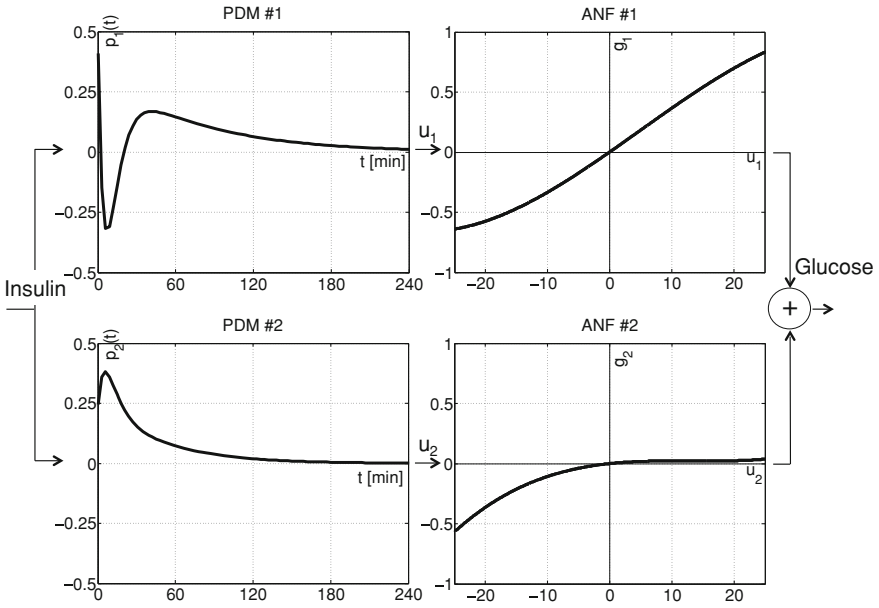


Fig. 10 The insulin-to-glucose model is composed of two branches, each of which is the cascade of a linear filter (PDM) followed by a static nonlinearity (ANF). The upper branch consists of an early glucoleptic response to insulin increases, followed by a subsequent glucogenic response, while the lower branch produces non-zero responses to negative insulin deflections only, due to the halfwave rectifying characteristic of the corresponding nonlinearity. The input signal (plasma insulin) and the output signal (plasma glucose) represent deviations from the respective basal values that are computed as the mean values of the recorded data and subtracted prior to processing

k-th static nonlinearity f_k is the output u_k of the k-th PDM filter p_k and the ordinate is the respective additive component $g_k(t)$ of the model output. The input and output signals (plasma insulin and glucose respectively) of this model represent deviations from the respective mean values that were subtracted prior to model estimation.

The obtained PDM model suggests the correspondence of the upper branch to the primary effect of insulin on plasma glucose. Specifically, the PDM of this branch exhibits an early glucoleptic phase, i.e. it causes reduction of glucose levels in response to an increase of insulin that is due to insulin-facilitated uptake of glucose by tissue cells and glycogen synthesis in the liver, over the first 20–25 min (with a trough at 10 min). It also exhibits a subsequent glucogenic phase that extends over about 150 min, which describes mainly the processes of production and release in the blood of new glucose by the liver and other organs (glucogenesis) in response to insulin increases. In general, the biphasic form of this PDM suggests there is counter-regulation of glucose changes in response to insulin deflections due to the closed-loop nature of insulin-glucose interactions. The

respective static nonlinearity (ANF#1) is odd-symmetric and it consists of a linear part, as well as a soft saturating characteristic. The form of the second branch suggests that it produces a glucose reduction in response to negative deflections of insulin only, as its respective nonlinearity exhibits a negative half-wave rectifying characteristic. The latter assures the reduction of plasma glucose in response to a negative deflection of plasma insulin from the basal value, since this generates a negative value in the output u_2 of PDM#2, but no major change in glucose for a positive deflection of plasma insulin from the basal value. Combined with the characteristics of the upper branch, the characteristics of the lower branch suggest that the response of the two model branches to insulin concentration reductions are antagonistic and may serve to regulate the overall glucose response.

In order to illustrate the net effect of both branches to a simple insulin input, we show the total model-predicted glucose response to an impulse of insulin with amplitude of 10 pM, which corresponds to around 1 standard deviation of the actual experimental insulin data, in Fig. 11. The overall response exhibits an early reduction of blood glucose followed by an overshoot and settles to zero after around 200 mins and it resembles in waveform the upper-branch PDM as expected, since a positive insulin input will elicit no contribution from PDM#2 (Fig. 11).

The normalized mean-square error (NMSE) of the output prediction of this third-order PDM model was 60.2 %, while the linear model prediction NMSE was 89.9 %—a fact that indicates the presence of strong nonlinearities in the causal relation of insulin-to-glucose variations. The glucose variations not explained by the model (i.e., not caused by prior insulin variations) are the residuals of the model prediction, which are viewed as "glucose disturbances" caused by numerous internal and external factors (hormonal, neural, metabolic) that are not insulin-dependent. Moreover, the effect of measurement errors may be considerable; for instance, the reader is referred to [12] for a quantitative description of the performance of the glucose analyzer used in the present experiment. The relatively high NMSE value of the overall output prediction indicates that the contributions of these insulin-independent unobserved factors, as well as the effects of measurement errors are significant.

4.3 Glucose-to-Insulin Branch

We next examine the reverse causal pathway, i.e. how variations of plasma glucose (input) affect the variations of plasma insulin (output), using the same experimental data. The obtained PDM model of the glucose-to-insulin causal relationship is shown in Fig. 12 and has two PDM branches. The PDM of the upper branch is almost monophasic (with the exception of the zero-lag value that seems to counterbalance the zero-lag value of the lower-branch PDM) and causes a co-directional change of insulin for a change of glucose, consistent with our understanding of the function of pancreatic beta cells. The PDM-ANF cascade of

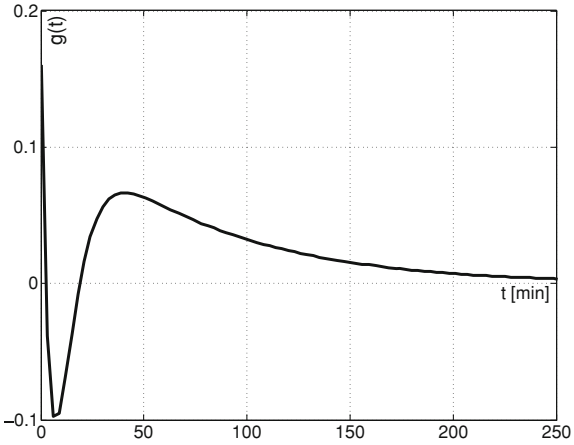


Fig. 11 Prediction of the insulin-to-glucose model (Fig. 10) to an impulse insulin input with amplitude 10 pM (around 1 standard deviation of the experimental data). The output waveform $g(t)$ resembles the form of the upper branch PDM of the model

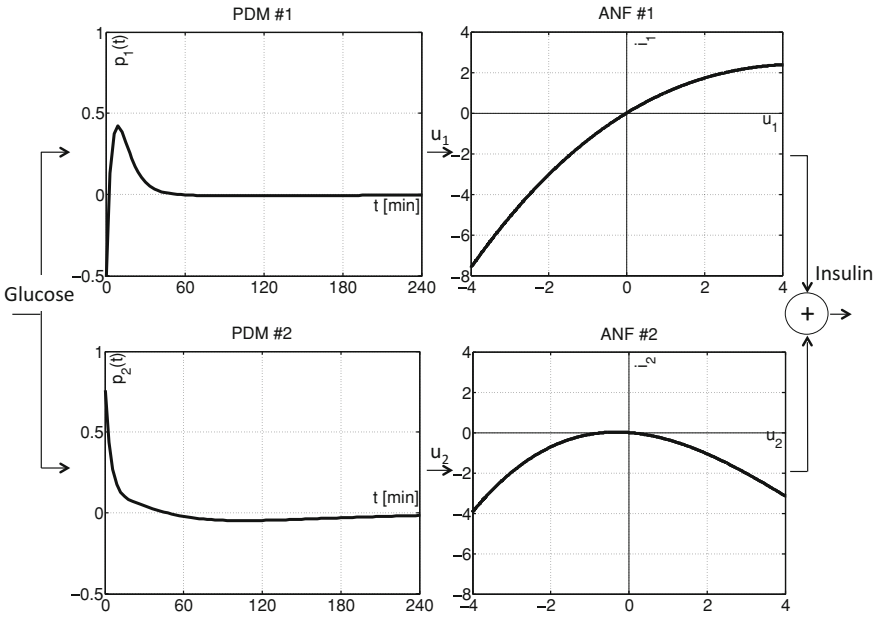


Fig. 12 The glucose-to-insulin model is composed of two PDM branches, the first of which causes a co-directional change of insulin for a change of glucose, while the second is strictly insulinoleptic, i.e. it causes reduction of insulin to either positive or negative change of glucose

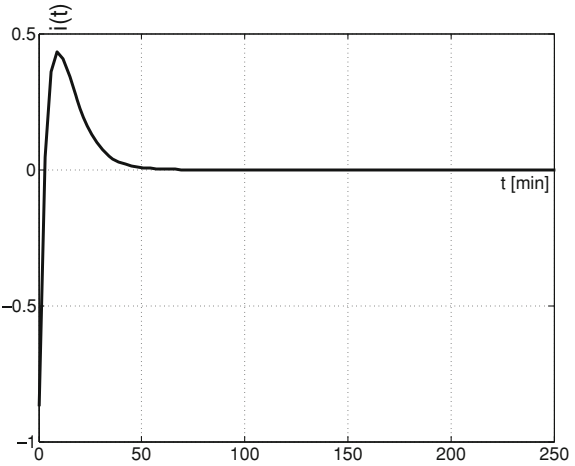


Fig. 13 Prediction of the glucose-to-insulin model (Fig. 12) to a glucose impulse insulin input with amplitude 1 mg/dl (around 0.5 standard deviations of the experimental data)

the lower branch is strictly insulinoleptic (i.e. it causes reduction of insulin to either positive or negative change of glucose) and operates in two time-scales: one fast over 10–15 min and the other slow over around 2 h.

Figure 13 shows the model-predicted insulin response to a positive impulse of glucose with amplitude of 1 mg/dl (around half the standard deviation of the actual experimental data). The model prediction resembles in waveform the upper-branch PDM of Fig. 12, since an increase of glucose is expected to elicit an increase of insulin by stimulating pancreatic secretion. In general, the second branch of both the above models (Figs. 11 and 13) appears to have a counter-regulatory role, balancing the effect of the respective first branch of each model. The model prediction achieved by the PDM model of Fig. 12 is equal to 69.9 %, while the corresponding linear NMSE is equal to 97.6 %. This implies a predominantly nonlinear relationship between glucose and insulin. The large nonlinear NMSE value suggests that spontaneous plasma glucose variations account for a fraction of spontaneous insulin variations, and that interferences from other physiological variables dominate in the formation of the insulin signal, along with the aforementioned possible effects of measurement errors.

5 Discussion

In the present chapter, we have rigorously examined the relation between nonlinear compartmental and Volterra models of glucose metabolism. We have also applied the proposed data-driven approach to investigating the spontaneous dynamic interactions between insulin and glucose in a fasting dog. Two widely used compartmental models, the minimal model (MM) of glucose disappearance and its closed-loop extension (AMM), which includes the effects of insulin

secretion, were formulated in the Volterra-Wiener framework and equivalent descriptions, in the form of Volterra models, were derived analytically. The effect of parametric model parameters of clinical importance on these descriptors (Volterra kernels) was examined. Using simulated data generated from the aforementioned compartmental models using both random-like and impulsive insulin stimuli, as well as the experimental dog data, we have demonstrated the feasibility of obtaining Volterra models that describe these data accurately. We have also shown that these estimates are not affected significantly by output-additive noise corresponding to measurement noise. The results provide evidence that Volterra models, free of a priori assumptions, may be estimated reliably from patient-specific data. These models may provide quantitative descriptions that reflect the underlying physiological mechanisms under general operating conditions and may prove useful in diagnostic or therapeutic [27] applications. We should note that for model-based glucose control applications, additional factors, such as the delay between plasma glucose and the sensor signal, should be taken into account.

5.1 Comparison Between Compartmental and Volterra Models

The parametric models examined herein are nonlinear due to the presence of a bilinear term in Eqs. (1) and (3), which modulates the effective time constant of glucose disappearance and depends on the action of plasma insulin (in the case of MM) and both plasma and endogenous secreted insulin (in the case of AMM) respectively. An additional nonlinearity is found in the endogenous insulin secretion Eq. (5) of the AMM in the form of a nonlinear threshold operator. The range of values for the MM and AMM parameters is taken from the literature [3, 4, 20, 25, 41, 45]. The value of p_3 was selected towards the upper limit of previously reported values in order to increase the contribution of the bilinear term, while the parameter β in Eq. (5), which determines feedback strength was selected to be larger than the value reported in [45] since, for the stimuli examined in the present paper, the effect of endogenous insulin was almost negligible for this latter value (it corresponds to low tolerance, obese patients [45]). Note that in the more general case, the value of β could be viewed as being dependent on g, in order to account for the effect of blood glucose concentration on insulin secretion. The value of the threshold β in the endogenous insulin secretion Eq. (5) was selected equal to zero in order to simplify the analytical derivations. This threshold can be generally set to a larger value, particularly when glucose disturbance terms that are non-insulin dependent, are included. However, in the context of the simulations presented herein, this value yielded reasonable patterns for the insulin secretion profile (Fig. 5).

Two types of inputs (variations of insulin concentration) were used in this computational study for the simulation of the parametric models: GWN fluctuations around a putative basal value (corresponding to the GWN mean) and random sequences of sparse insulin increases (about one every 2 h on the average), which may result from insulin/glucose infusions. It was shown that reliable and robust nonparametric models can be obtained with both types of stimuli in the presence of measurement noise. The GWN insulin fluctuations may also be viewed as internal spontaneous fluctuations and, therefore, the applicability of this approach can be extended to the case of spontaneous glucose/insulin measurements. The use of random sequences of larger sparse impulsive insulin increases, although unconventional, was shown to be effective in terms of model estimation and may offer clinical advantages as it is likely to mitigate the risk of induced hypoglycemia—an issue that must be examined carefully in future studies.

The Volterra approach does not require specific prior postulates of compartmental model structures (e.g. it is not committed to any particular number of compartments) and allows estimation of the model (i.e. the Volterra kernels) directly from arbitrary input-output data. Therefore, it offers the advantage of yielding models that are "true-to-the-data" and valid under all input conditions within the range of the experimental data. Therefore, this fundamentally different approach provides significant benefits relative to existing approaches in terms of modeling flexibility and accuracy.

The robustness of the Volterra modeling approach (i.e. the effect of output-additive noise on the obtained kernel estimates) was studied by selecting as noise sample signals from a GWN process with variance consistent with what is known about glucose measurement errors (i.e. a standard deviation equal to 14–20 % of the glucose basal value [21]). However, we must make the distinction between noise (which is primarily related to measurement errors) and systemic disturbance (which is related to systemic perturbations that are not explicitly accounted for in the model). The systemic disturbance signal may include the effect of meals [17], the effect of circadian and ultradian endocrine cycles [44] and the effect of randomly occurring events of accelerated metabolism (due to exercise or physical exertion) as well as neuro-hormonal excretions (due to stress or mental exertion). The amplitudes and the relative phases of these disturbance components will generally vary among subjects and over time. Since the selection of such disturbance components is rather complex, the study of their effect on the robustness of the model estimation is deferred to future studies.

The MM approach is based on the notion that estimates of the three model parameters (p_1, p_2 and p_3), obtained through a glucose tolerance test, provide the necessary clinical information for diagnostic purposes in the form of the equivalent indices of glucose effectiveness (S_G) and insulin sensitivity (S_I). Although this proposition has merit and has proven to be useful so far, it is widely recognized that it has serious limitations [10, 11]. To overcome some of these potential limitations, our approach advances the notion that a Volterra-type model (in the form of kernels or the PDM model) provides the requisite clinical information in a more complete manner (i.e., no model constraints). In order to compare the

relative utility of the Volterra approach with the conventional MM approach in a clinical context, we must define clinically relevant attributes for the two approaches that are directly comparable. For instance, if we are interested in deriving quantitative descriptions/measures of how insulin affects the plasma glucose concentration in specific subjects (i.e. based on collected data), we may use certain features of the estimated first-order kernels, such as the integrated area, peak value and initial slope, which determine the linear component of the overall effect of an insulin injection, its maximum instantaneous effect and how fast this effect occurs respectively, instead of the estimated MM parameters.

In this context, the combined effect of errors in the estimates of the three parameters of the MM (p_1, p_2, p_3) may be compared to estimation errors in the integrated area of the first-order kernel, which is equal to the ratio S_I/S_G (i.e. $p_3/(p_1p_2)$), as a measure of how much a unitary insulin impulse will affect the plasma glucose concentration. Also, since $S_G = p_1$ is the inverse of the long time-constant of the kernel (providing a measure of the extent of the kernel), it follows that "insulin sensitivity" S_I is akin to the average kernel value. Thus, one may suggest that the clinical index of "insulin sensitivity" may be defined alternatively by the average kernel value and "glucose effectiveness" by the extent of the kernel in the data-driven modeling context. It also stands to reason that the peak value of this kernel is likely to have some clinical significance, since it quantifies the maximum effect of an insulin injection on blood glucose in a given subject. Finally, the slope of the first-order kernel at the origin (a measure of how rapidly glucose drops in response to an insulin infusion) is equal to $-(g_b p_3)$. Since the basal glucose value is known, a quick estimate of p_3 can be obtained from the slope of the first-order kernel. In the above context, PDM models (Figs. 7, 8, 9, 10, 12) may prove very beneficial, since they facilitate meaningful physiological interpretations relative to the general Volterra formulation. Therefore, certain characteristics of the PDM branches (e.g., the dynamics of the linear filters and the characteristics of the nonlinearities) may also be associated to clinical indices that describe insulin action and its efficiency in specific subjects.

Existing insulin-glucose models can be divided in two broad categories; the first includes models that describe these dynamics in a simplified manner by using a limited number of parameters of diagnostic importance, in order to make these parameters identifiable from relatively simple experimental protocols (IVGTT, OGTT, EHC) and the second category includes more complicated models that capture the complexity of the underlying physiology under more general operating conditions. Choosing the most appropriate model type depends strongly on the particular application. For instance, minimal-type models have met with success for diagnostic purposes; however, they do not describe the metabolic system realistically. On the other hand, more complicated models achieve this but they are typically not identifiable from experimental data that are available in practice, requiring considerably more complicated protocols (such as multiple tracer experiments).

It is worth noting that a recent large-scale model by Dalla Man et al. [14] has been approved by the Food and Drug Administration Agency of the USA and the

Juvenile Diabetes Research Foundation as a substitute for clinical trials for the preclinical testing of glucose control algorithms, illustrating the importance of mathematical modeling. In our view, data-driven models, such as the ones presented hereby, present an attractive alternative that lies between the two aforementioned main model categories. Data-driven models are identifiable in practical settings in a patient-specific and adaptive manner, and provide more flexibility than minimal-type models. However, they are not based on physiological considerations; therefore this flexibility should be exploited carefully and their complexity should be selected judiciously in order to assure that they are accurately describing the underlying phenomena without overfitting to particular data sets.

5.2 Insulin-Glucose Dynamics in a Fasting Dog

There are four salient issues that deserve further elaboration and discussion: (1) the validity and utility of the obtained models; (2) the notion of "indirect effects" captured by the model of a closed-loop system; (3) the physiological interpretation of the obtained PDMs and (4) the pivotal role of internal "disturbance" signals in closed-loop systems. These issues are intertwined and their successful resolution will determine the potential utility of the advocated approach.

The validity of the obtained model is based on its predictive capability and the consistency of the modeling results across different experiments. Random-like spontaneous variations enable us to perform broadband analysis of the input-output data and obtain reliable models that are capable of predicting the output for arbitrary inputs within the dynamic range of the available measurements, removing the restrictive effect of specialized test inputs (such as impulses or sinusoids). The significant reduction in the prediction NMSE achieved by nonlinear models (especially for the glucose-to-insulin branch) suggests the presence of strong dynamic nonlinearities. This reduction is unlikely to be due to overfitting, which is an issue that may arise when using nonlinear models to model short experimental data records. The number of free parameters for the LVN approach depends linearly on system order [31], and is much lower compared to alternative approaches, such as standard function expansions or orthogonal based approaches, for which the same dependence is exponential [30]. The free parameters of the linear and nonlinear models examined herein were equal to 25 and 29 respectively, i.e. third-order models required the estimation of 4 additional parameters only. The availability of data from one dog, as well as the relatively limited dynamic range of the experimental data limits our ability to generalize our findings; however, the unique nature of the employed data yields novel insights regarding the dynamic interactions between spontaneous insulin and glucose fluctuations.

The notion of indirect effects is fundamental in the context of closed-loop physiological systems, which are observed under conditions of spontaneous activity. Multiple physiological processes participate in the transition of the system to the state of dynamic equilibrium (homeostasis). These processes are typically

nonlinear, dynamic and interacting in a complex manner. Therefore, a change in one of the implicated variables will generally cause some direct effects (immediate consequences) and indirect effects (subsequent consequences of a chain of events). The indirect effects are harder to notice/delineate experimentally and even harder to analyze, since the number of overlapping effects from interacting factors increases with system complexity. The advocated approach offers the practical means for obtaining a unique glance of these indirect effects in a quantitative and reliable manner, as illustrated in the cases of the glucogenic and insulinogenic branches of the insulin- to-glucose and glucose-to-insulin models.

The physiological interpretation of the obtained PDM models attains critical importance for the advancement of scientific understanding as well as their potential use for improved clinical diagnosis. The ability to detect subtle early defects in the complex cascade of metabolic events involved in insulin and glucose regulation may prove crucial for the improved diagnosis of diseases such as Type II diabetes [36]. The presented models quantify the relation between naturally occurring insulin/glucose variations and may be used to identify changes in the glucose effectiveness and insulin sensitivity of a subject—thus leading to improved diagnostic methods. They may also disclose physiological deficiencies (e.g. an undeveloped glucoleptic or insulinogenic component) in a quantitative manner. The potential clinical utility can be expanded if additional physiological variables are incorporated (e.g. free fatty acids). The dynamic characteristics of the PDM model, which exhibit multiple time constants (Figs. 10 and 12), are in agreement with the previously reported entrainment between insulin and glucose patterns over different time scales (rapid to ultradian) [36, 40]. In these studies, experimental impulsive or step stimuli were induced, whereas in the present study we studied natural variations, an approach that removes the need for such interventions.

The role of internal "disturbance" signals is pivotal for closed-loop physiological systems seeking homeostasis. Although homeostasis is typically mediated through multiple nested closed-loops, let us consider the case of two physiological variables operating in a closed-loop configuration (e.g. insulin and glucose) as shown schematically in Fig. 14. The two measured variables (spontaneous variations of plasma insulin $I(t)$ and plasma glucose $G(t)$, are causally linked in a mutual interdependence described by the general system operators **A** and **B**. Operator **A** transforms the insulin signal $I(t)$ into the causally linked (i.e. insulin-dependent) glucose component $G_c(t)$ in a dynamic and nonlinear fashion. Likewise, operator **B** transforms the glucose signal $G(t)$ into the causally linked (i.e. glucose-dependent) insulin component $I_c(t)$. Each of these nonlinear dynamic operators can be described by a Volterra-equivalent model (such as the PDM models of Figs. 10 and 12) that is estimated directly from spontaneous input-output data. The model residuals of the two operators (Figs. 11 and 13) are the "disturbance" signals, $G_d(t)$ for glucose and $I_d(t)$ for insulin, that represent the aggregate effects of multiple systemic factors (e.g. glucagon, cortisol, epinephrine etc.) and external factors (e.g. meals, exercise, stress, mental exertion etc.). These disturbance signals are added to the causally linked

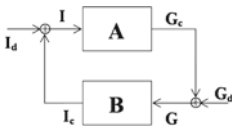


Fig. 14 The closed-loop configuration representing the plasma insulin-glucose dynamic interactions (glucose-to-insulin and insulin-to-glucose models, **B** and **A** respectively). The "disturbance" signals I_d and G_d are the residuals of the obtained causal models and represent the aggregate effects of multiple internal (e.g. glucagon, epinephrine, norepinephrine, cortisol, growth hormone, somatostatin etc.) and external factors (e.g. meals, exercise, stress, mental exertion etc.)

components, $G_c(t)$ and $I_c(t)$ (Figs. 11 and 13, left panels), to form the observed variables $G(t)$ and $I(t)$, respectively.

Regarding the overall functional characteristics of the closed-loop configuration of Fig. 14, we examine the closed-loop relation in operator notation:

$$G(t) = \mathbf{A}[\mathbf{B}[\mathrm{G(t)}] + \mathrm{I_d(t)}] + \mathrm{G_d(t)} \tag{26}$$

where $\mathbf{B}[G(t)]$ denotes that the operator **B** (glucose-to-insulin PDM model) acts on the glucose signal $G(t)$. Equation (26) is equivalent to a nonlinear auto-regressive model with exogenous input (NARX) with stochastic terms due to the interaction of $I_d(t)$ with **A**.

6 Conclusion

Mathematical modeling has played a central role in the context of diabetes diagnosis and treatment. The results of the present chapter demonstrate the relative advantages and disadvantages of the Volterra modeling methodology versus the compartmental approach for specific minimal-type parametric models (MM and AMM). The Volterra approach is inductive (data-driven) and yields models with minimum prior assumptions. The compartmental approach is deductive (hypothesis-based) and yields models with the desired level of complexity that are directly interpretable but not necessarily inclusive of all functional characteristics of the system. The recent availability of continuous measurements of glucose (through continuous glucose sensors) and the feasibility of frequent infusions of insulin (through implantable insulin micro-pumps) make possible the realistic application of data-driven modeling approaches in a subject-specific and adaptive context, which does not require the prior postulates of compartmental models. The potential benefits include: (i) the inherent completeness of the obtained models, in the sense that they will include all functional characteristics of the system contained within the data (ii) the robustness of their estimation in a practical context (iii) their subject-specific customization and (iv) their time-dependent adaptability when the system characteristics are changing slowly over time, allowing for

effective tracking of these changes in each subject. This could be achieved by piecewise stationary models or by employing recursive estimation schemes, which may be readily implemented for the LVN models employed hereby.

This is corroborated by the presented models of the inter-relationships between spontaneous plasma insulin and glucose variations in a fasting dog, which offer quantitative means for advancing our scientific understanding and potentially improving clinical diagnosis and treatment of diabetes. This will depend on further validating our findings and on assigning correct physiological meaning to the various components of the obtained models. It is worth noting that data-driven methods typically require richer stimulus patterns than simple compartmental models. For instance, spontaneous physiological variability, such as the dog insulin and glucose measurements examined above, exhibits broadband characteristics that provide rich information for system operation over a wide dynamic range. However, measuring plasma insulin with a high temporal resolution in humans can not be performed easily and/or with a low cost with current technology, so developing such technologies deserves attention in future related research. Concluding, in our view the aforementioned results illustrate that the relation between data-driven and compartmental models deserves further attention in the future in order to more effectively combine and exploit the advantages of each approach in various clinical applications including diabetes diagnosis and glucose regulation.

Acknowledgements This work has been funded by the University of Cyprus internal research grant "Nonlinear, data-driven modeling and model-based control of blood glucose". The authors would like to thank Prof. Richard N. Bergman and Dr. Katrin Huecking for providing the experimental data. This work was supported in part by the NIH/NIBIB Center Grant No. P41-EB001978 to the Biomedical Simulations Resource at the University of Southern California.

References

1. Ackerman E, Gatewood LC, Rosevear JW, Molnar GD (1965) Model studies of blood-glucose regulation. Bull Math Biophys 27(Suppl):21–37
2. Andreassen S, Benn JJ, Hovorka R, Olesen KG, Carson ER (1994) A probabilistic approach to glucose prediction and insulin dose adjustment: description of metabolic model and pilot evaluation study. Comput Methods Programs Biomed 41:153–165
3. Bergman RN, Ider YZ, Bowden CR, Cobelli C (1979) Quantitative estimation of insulin sensitivity. Am J Physiol 236:E667–E677
4. Bergman RN, Phillips SM, Cobelli C (1981) Physiologic evaluation of factors controlling glucose tolerance in man: measurement of insulin sensitivity and beta-cell glucose sensitivity from the response to intravenous glucose. J Clin Invest 68:1456–1467
5. Bergman RN, Lovejoy JC (1997) The minimal model approach and determinants of glucose tolerance, vol 7. Louisiana State University Press, Baton Rouge
6. Bode BW, Sabbah HT, Gross TM, Fredrickson LP, Davidson PC (2002) Diabetes management in the new millennium using insulin pump therapy. Diab Metab Res Rev 18(Suppl. 1):S14–S20

7. Bolie VW (1961) Coefficients of normal blood glucose regulation. J Appl Physiol 16:783–788
8. Callegari T, Caumo A, Cobelli C (2003) Bayesian two-compartment and classic single-compartment minimal models: comparison on insulin modified IVGTT and effect of experiment reduction. IEEE Trans Biomed Eng 50:1301–1309
9. Carson ER, Cobelli C, Finkelstein L (1983) The mathematical modeling of endocrine-metabolic systems. Model formulation, identification and validation. Wiley, New York
10. Caumo A, Vicini P, Cobelli C (1996) Is the minimal model too minimal? Diabetologia 39:997–1000
11. Caumo A, Vicini P, Zachwieja JJ, Avogaro A, Yarasheski K, Bier DM, Cobelli C (1999) Undermodeling affects minimal model indexes: insights from a two-compartment model. Am J Physiol 276:E1171–E1193
12. Chua KS, Tan IK (1978) Plasma glucose measurement with the Yellow Springs Glucose Analyzer. Clin Chem 24(1):150–152
13. Cobelli C, Mari A (1983) Validation of mathematical models of complex endocrine-metabolic systems. A case study on a model of glucose regulation. Med Biol Eng Comput 21:390–399
14. Man CD, Rizza RA, Cobelli C (2007) Meal simulation model of the glucose-insulin system. IEEE Trans Biomed Eng 54(10):1740–1749
15. The Diabetes Control and Complications Trial Research (1993) The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. New England J Med 329:977–986
16. Finegood DT, Tzur D (1996) Reduced glucose effectiveness associated with reduced insulin release: an artifact of the minimal-model method. Am J Physiol Endocrinol Metab 271(3):E485–E495
17. Fisher ME (1991) A semiclosed-loop algorithm for the control of blood glucose levels in diabetics. IEEE Trans Biomed Eng 38:57–61
18. Florian JA, Parker RS (2005) Empirical modeling for glucose control in diabetes and critical care. Eur J Control 11:605–616
19. Freckmann G, Kalatz B, Pfeiffer B, Hoss U, Haug C (2001) Recent advances in continuous glucose monitoring. Exp Clin Endocrinol Diab 109(Suppl 2):S347–S357
20. Furler SM, Kraegen EW, Smallwood RH, Chisolm DJ (1985) Blood glucose control by intermittent loop closure in the basal model: computer simulation studies with a diabetic model. Diab Care 8:553–561
21. Ginsberg J (2007) The current environment of CGM technologies. J. Diab Sci Technol 1:111–127
22. Godsland IF, Agbaje OF, Hovorka R (2006) Evaluation of nonlinear regression approaches to estimation of insulin sensitivity by the minimal model with reference to Bayesian hierarchical analysis. Am J Physiol Endocrinol Metab 291:E167–E174
23. Krudys KM, Kahn SE, Vicini P (2006) Population approaches to estimate minimal model indexes of insulin sensitivity and glucose effectiveness using full and reduced sampling schedules. Am J Physiol Endocrinol Metab 291:E716–E723
24. Lefebvre PJ, Paolisso G, Sheen AJ, Henquin JC (1987) Pulsatility of insulin and glucagon release: physiological significance and pharmacological implications. Diabetologia 30:443–452
25. Lynch SM, Bequette BW (2002) Model predictive control of blood glucose in Type 1 diabetics using subcutaneous glucose measurements. In: Proceedings of American control conference, Anchorage, AK, pp 4039–4043
26. Markakis MG, Mitsis GD, Marmarelis VZ (2008) Computational study of an augmented minimal model for glycaemia control. In: Proceedings of the 30th annual IEEE-EMBS conference, Vancouver, BC, Canada, pp 5445–5448
27. Markakis, M.G., Mitsis, G.D., Papavassilopoulos, G.P., Marmarelis, V.Z.: Model Predictive Control of Blood Glucose in Type 1 Diabetics: the Principal Dynamic Modes Approach. Proc. 30th Annual IEEE-EMBS Conf., Vancouver, BC, Canada, 5466-5469 2008

28. Marmarelis VZ (1991) Wiener analysis of nonlinear feedback in sensory systems. Ann Biomed Eng 19:345–382
29. Marmarelis VZ (1997) Modeling methodology for nonlinear physiological systems. Ann Biomed Eng 25:239–251
30. Marmarelis VZ (2004) Nonlinear dynamic modeling of physiological systems. IEEE-Wiley, Piscataway
31. Mitsis GD, Marmarelis VZ (2002) Modeling of nonlinear physiological systems with fast and slow dynamics. I. Methodology. Ann Biomed Eng 30:272–281
32. Mitsis GD (2002) Nonlinear physiological system modeling with Laguerre-Volterra networks: methods and applications. Ph.D. thesis, Department of Biomedical Engineering, University of Southern California
33. Muniyappa R, Lee S, Chen H, Quon MJ (2008) Current approaches for assessing insulin sensitivity and resistance in vivo: advantages, limitations, and appropriate usage. Am J Physiol Endocrinol Metab 294(1):E15–E26
34. Ni TC, Ader M, Bergman RN (1997) Reassessment of glucose effectiveness and insulin sensitivity from minimal model analysis: a theoretical evaluation of the single-compartment glucose distribution assumption. Diabetes 46:1813–1821
35. Parker RS, Doyle FJ, 3rd Peppas NA (1999) A model-based algorithm for blood glucose control in type I diabetic patients. IEEE Trans Biomed Eng 46:148–157
36. Porksen N (2002) The in vivo regulation of pulsatile insulin secretion. Diabetologia 45:3–20
37. Roy A, Parker RS (2006) Dynamic modeling of free fatty acid, glucose, and insulin: an extended "minimal model". Diab Technol Ther 8:617–626
38. Sorensen J (1985) A physiologic model of glucose metabolism in man and its use to design and assess insulin therapies for diabetes. Ph.D. thesis, Department of Chemical Engineering, Massachussetts Institute of Technology, Cambridge, MA
39. Steil GM, Rebrin K, Janowski R, Darwin C, Saad MF (2003) Modeling beta-cell insulin secretion-implications for closed-loop glucose homeostasis. Diab Technol Ther 5:953–964
40. Sturis J, Van Cauter EV, Blackman JD, Polonsky KS (1991) Entrainment of pulsatile insulin secretion by oscillatory glucose infusion. J Clin Invest 87:439–445
41. Toffolo G, Bergman RN, Finegood DT, Bowden CR, Cobelli C (1980) Quantitative estimation of beta cell sensitivity to glucose in the intact organism: a minimal model of insulin kinetics in the dog. Diabetes 29:979–990
42. Toffolo G, Campioni M, Basu R, Rizza RA, Cobelli C (2006) A minimal model of insulin secretion and kinetics to assess hepatic insulin extraction. Am J Physiol Endocrinol Metab 290:E169–E176
43. Tresp V, Briegel T, Moody J (1999) Neural-network models for the blood glucose metabolism of a diabetic. IEEE Trans Neur Netw 10:1204–1213
44. Van Cauter EV, Shapiro ET, Tillil H, Polonsky KS (1992) Circadian modulation of glucose and insulin responses to meals-relationship to cortisol rhythm. Am J Physiol 262:R467–R475
45. Van Herpe T, Pluymers B, Espinoza M, Van den Berghe G, De Moor B (2006) A minimal model for glycemia control in critically ill patients. In: Proceedings of the 28th IEEE EMBS annual international conference, New York, NY

Ensemble Glucose Prediction in Insulin-Dependent Diabetes

Fredrik Ståhl, Rolf Johansson and Eric Renard

Prediction of glucose changes in type 1 *Diabetes Mellitus* has received a considerable amount of scientific and commercial interest over the last decade. In large, the driving force behind this surge in research can be explained by the recent advances in sensor technology [101], and the thereto attached promises and hopes of closed, or semi-closed, loop control of diabetic glucose dynamics. Predicting models play a key role in many of these concepts—providing the essential simulation tool in MPC-oriented closed loop arrangements of an artificial pancreas [20], or as a component in a decision support system—providing predictions directly to the user [82].

However, insulin-dependent diabetic glucose dynamics are known to be subject to time-shifting dynamics. Considering this, as well as the vast number of models developed in the literature, it is unclear if a single model can be determined to be optimal under every possible situation. This raises the question whether it is more useful to use one of the models solely, or if it is possible to gain additional prediction accuracy by combining their outcomes. Accuracy may be gained from merging, due to mismodeling or to changing dynamics in the underlying data creating process, where a single model capturing the system behavior may be infeasible, e.g., for practical identification concerns. Thus, by an ensemble approach, robustness and performance may be improved.

In this chapter, a novel merging approach—combining elements from both switching and averaging techniques, forming a 'soft' switcher in a Bayesian framework—is presented for the glucose prediction application.

F. Ståhl (✉) · R. Johansson
Department of Automatic Control, Lund University, Lund, Sweden
e-mail: fredrik.stahl@control.lth.se

E. Renard
Department of Endocrinology, Diabetes and Nutrition,
Montpellier University Hospital, Montpellier, France

V. Marmarelis and G. Mitsis (eds.), *Data-driven Modeling for Diabetes*,
Lecture Notes in Bioengineering, DOI: 10.1007/978-3-642-54464-4_2,

1 Related Research

In this section some related research to glucose prediction and model merging are presented.

1.1 Models for Glucose Prediction

Models of glucose dynamics for predictive purposes can mainly be divided into two categories; physiologically-oriented models and data-driven black-box approaches. The latter sometimes incorporate physiological sub models of insulin and glucose infusion following insulin administration and meal intake, but the main part of the dynamics stem from the statistically derived relationships.

The development of physiological diabetic glucose modeling started with the simple linear models of [2, 12], aiming at describing the relationship between glucose and insulin utilization. Following these efforts, the slightly more complex, and well-established, minimal model [10] was suggested as a means to estimate insulin sensitivity from an intravenous glucose tolerance test (IVGTT). Detailed models of the glucose metabolism; separating insulin and non-insulin dependent glucose utilization, incorporating models of hepatic balance, renal clearance, glucose rate of appearance following meal intake, insulin pharmacokinetics, and in some cases pancreatic insulin synthesis and release, have surged since then.

The transport of rapid-acting insulin from the subcutaneous injection site to the blood stream has been described in quite a few models of insulin pharmacokinetics. Most of these are linear compartment models, and reviews can be found in [72, 104]. This phenomenon has generally been considered independent to the metabolic interaction, and thus separated as a stand-alone model. In [104], 11 different models (10 compartment models and the model from [9] were fitted to empirical meal test data from seven type 1 patients using rapid-acting bolus insulin. A third-order compartment model, with local degradation of insulin at the injection site (modeled as a Michaelis–Menten relationship), turned out to be the best choice, according to the Akaike criterion [53], and this may serve as a typical example of how the insulin kinetics have been be modeled.

The corresponding flux of glucose from the intestines following a meal intake, has been modeled with different approaches. There is evidence that gastric emptying, to some extent, is dependent on current glucose level, see, e.g., [94], but this relationship has not been incorporated in any model so far. Thus, the digestive process is also considered as a stand-alone model, without dependencies to the glucose metabolism. Two models have been widely used; the models by [24, 62]. In [62], the model consists of single compartment with fixed limited gastric emptying rate constant, and with a duration dependent on the meal size. Earlier work on models of glucose rate of appearance during an OGTT [22] and mixed meal test [21] formed the basis for the model in [24]. Here, a third-order nonlinear

compartment model was used, and also in this case, the gastric emptying rate was limited dependent upon the amount of ingested carbohydrates.

Turning to general models of glucose metabolism, a sparse fourth-order linear model, with physiological interpretation of the state variables, was suggested in [92], with six tunable parameters. The original model was validated on data from intravenous experiments involving diabetic dogs. Thereafter, the model has been both reduced, and extended to include exercise load, and to also consider oral hyperglycaemic agents. The model order is still four, but the number of tunable parameters has been reduced to five, and incorporated into a decision support system (DSS) called KADIS [93].

In [62], a simulation model based on the insulin kinetics from [9], and including hepatic balance (described by a look-up table), peripheral and insulin-independent glucose utilization (Michaelis–Menten like relationship), renal clearance and the meal digestion model from the same paper (described above), was presented. Overall, the model contains only two tunable parameters, the rest are considered patient invariant. Later, the freely downloadable educational simulation software AIDA [61] was developed using this model. The system was validated on a set of 24 subjects with parameter convergence achieved in 80 % of the cases [60].

Another simulation model, that has been turned into an advisory system, is the DIAS model [48]. Especially noteworthy of this model is the nonlinear model of the hepatic balance [6], fitted to tracer literature data, and the model extension to include the delayed hypoglycemic effect of alcohol intake [81]. The model was incorporated into a prototype eHealth tool called DiasNet [52], with a central server-based web service, which also communicates over the cellular network with the user's mobile application implemented on a smartphone. The system has been tested in a small field trial, but was mainly evaluated on overall data acquisition, transmission and application usability aspects, and not on results concerning model performance.

A large model with 19 tunable parameters was proposed in the Sorensen thesis [95], a model often used as a verification tool to assess different control approaches, e.g., [34]. The web-based educational simulation model GlucoSim [3] has been developed based on another thesis [84]. Generally, these models are difficult to fit to an individual person, and may lack structural identifiability. This makes them unsuitable for predictive purposes, but synthetic subjects may be created for simulation studies.

Currently, the most influential simulation model is the University of Virginia and Padova University (UVa/Padova) model described in [23, 24], which has been accepted by the Food and Drug Administration of the U.S. (FDA) to be used as a substitute for animal trials in preclinical trials of closed-loop development [57]. To this purpose, 300 artificial subjects have been derived from estimated parameters from population studies, and used in, e.g., [59]. This model is based upon the classical minimal model [10], and the glucose rate-of-appearance model in [21]. The population data for estimating the 300 artificial subjects were derived using the triple-tracer protocol described in [8].

In [89], the minimal model was augmented with additional states to include the dynamical interaction between free fatty acids and the insulin and glucose

compartments. The model parameters were partly fixed, and partly identified using experimental data, and showed reasonable resemblance to data. In [90], the model was used, together with the gastric emptying function taken from [62], to fit the model against data from one mixed meal consumed by normal subjects, with good correspondence.

The limitation of the classical minimal model to provide consistent estimates of insulin sensitivity, when different insulin concentrations arise during an IVGTT, was addressed in [83]. Modifications to the model was suggested to incorporate the saturation effect of insulin on insulin-dependent glucose utilization [69, 88], as well as a saturation effect on insulin transport from the plasma to the interstitial compartment. Generally, the saturation effect is not pronounced at insulin infusion levels of most insulin-dependent diabetic patients. However, the critically ill may often experience reduced insulin sensitivity, and are treated with intensive insulin treatment with abnormal insulin levels to maintain normoglycemia, thereby reducing mortality and morbidity outcome [102]. Thus, for the purpose of improved glycemic control of the critically ill in Intensive Care Units (ICU), this model was picked up in [64]. Thereafter, the table-based protocol SPRINT, which acts as a decision support in the manual infusion control for the ICU personal, was derived [18]. This approach has been successfully validated in a large study covering 371 subjects, achieving a very tight glucose control [17].

Another extension of the minimal model was proposed in [28], by incorporating effects of physical exercise by adding parameters, which increase insulin sensitivity, insulin-independent glucose utilization and insulin clearance during exercise, to the model. The model has not been evaluated empirically. Also the UVa/Padova model has been extended to cover physical activity in [67], based on the model in [15]. The model links elevated heart rate to increased insulin sensitivity and insulin-independent glucose utilization. In [15], the model was fitted to data from a hyper-insulinemic clamp test, including a 15-min exercise period (50 % VO_{2max}), for 21 type 1 subjects, with a weighted mean square estimation error of 7.7 mg/dl (unclear how the weights were chosen).

Yet another ambitious extension with 19 parameters, whereof 10 are subject to identification, and including modeling of the circadian rhythm was given in [37]. In [38], the model was validated by simulation comparisons on two data set of six and nine type 1 patients with excellent results (RMSE about 1 mmol/L), however, apparently without cross-validation.

Before leaving the minimal model, the work in [54] needs to be commented. Here, the minimal model, extended with a simple pharmacokinetic compartment model for the insulin kinetics and a compartment meal model of the same type as in [105], was tested on closed-loop data from a trial involving 10 type 1 subjects. Intraday variations of the model parameters related to the insulin sensitivity, hepatic balance and insulin-independent glucose utilization was allowed over three different sections of the day. Also in this case, the model was validated without crossvalidation, but with an impressive average simulation prediction error (RMSE about 16 mg/dl).

A simpler model, with only five tunable parameters, is the Hovorka model [50], later extended and altered for the critically ill in [51]. The former model has been validated for predictive capacity on 15 subjects with a RMSE of 3.6 mg/dl for a prediction horizon of 15 min. Parameter estimates were retrieved recursively from a sliding data window using a Bayesian approach. This model is also used extensively for MPC-oriented closed-loop validation in a simulation environment, including a cohort of 18 virtual patients [103]. 8 out of the 18 parameter sets have been derived from experimental data, and the rest from so-called informed prior distributions. The model has also been used, e.g., in the evaluation of PID control in [39], which also make use of the Sorensen [95] and the minimal model [10].

Data-driven models have been investigated on CGM time-series alone, or by considering inputs as well. The meal sub models of [24, 62] are furthermore often used as input generating components in data-driven models to approximate the glucose flux input from the gut following a meal intake. Here, the focus has been prediction for the purpose of early hypoglycemic detection, e.g., to be used for alarm triggering in CGM devices, or temporary insulin pump shut-off, as well as establishing models suitable for model-based control.

Time-series analysis by Auto-regressive (AR) models started with [14], who evaluated the basic underlying assumptions concerning stationarity and auto-covariance that AR modeling is based upon, concluding that diabetic data generally is non-stationary, but highly auto-correlated, thus recommending the models to be recurrently re-estimated. Following this, AR and ARMA models were developed in [97, 99] using glucose data from a recently diagnosed type 1 diabetic. In [96], first-order recursive AR models were investigated for 28 subjects using a low-pass filtered CGM signal from the GlucoDay CGM system. The results indicate that hypoglycemia can be detected by the model 25 min before the CGM signal passes the same threshold. Another example of recursive AR and ARMA models of third order, incorporating a change detection feature for more rapid parameter re-estimation when large changes in the dynamics are detected, is found in [35]. The models were evaluated for 30 healthy, 7 glucose-intolerant and 25 type II diabetic subjects, with less than 4 % mean Relative Average Deviation (RAD) and almost no values in D or E zones of the Clarke Error Grid [19] for the 30-min predictions in comparison to the CGM Medtronic Gold reference [68]. Contrary to the above, the authors of [42] claim that a generic patient- and time-invariant AR model of order 30 can be identified from any patient and used for glucose prediction for any other patient. Very promising results were achieved in [41], where the model was evaluated for three different datasets, each utilizing a different CGM device, and the patient cohorts included both type I and type II diabetes. The prediction error was on average, in terms of RMSE, less than 3.6 mg/dl for a 30-min prediction, with negligible delay, and with 99 % of the paired prediction-reference points in the A and B zones of the p-CGA. However, these results were achieved by filtering the CGM signal in both training and test data using a non-causal filter, removing the high frequency components. In [65] the causality aspect of the input filtering was addressed. The AR model, here reduced to order 8 after model complexity considerations, was reformulated as a

linear model with a Kalman filter, and the filter parameters were adjusted to account for the filtering of the CGM signal. For evaluation purposes, the reference was however still filtered in the same non-causal way as before. Using this approach on the same data set as in [41], yielded more moderate results with an average prediction error of 16 mg/dl, and a 9-min lag for the 20-min prediction.

Algorithms specifically developed for hypoglycemic detection have also been proposed. In [76], a Kalman filter approach was suggested, estimating the states corresponding to the interstitial glucose level, and the first and second derivative thereof, i.e., rate of glucose change and acceleration. In [75], this method was evaluated for 13 hypoglycemic clamp data sets. Using a hypoglycemic threshold of 70 mg/dl, the sensitivity and specificity were 90 and 79 %, respectively, with unknown alarm time. Combining three different methods for hypoglycemic detection with the ARMA model of [35], data from insulin-induced hypoglycemic tests for 54 type 1 subjects were evaluated in [33]. With a hypoglycemic threshold of 60 mg/dl, sensitivity of 89, 88, and 89 % and specificity of 67, 74, and 78 % were reported for each method, respectively. Mean values for time to detection were 30, 26, and 28 min.

A short-coming of the AR models and the algorithms above is the lack of input-output relationship, excluding them from being used in a model-based control framework. A natural extension to the AR concept is to include external inputs, transforming the model to an ARX model. This type of model has been considered in, e.g., [40], where both batch-wise and recursively identified patient-specific ARX models have been analysed for nine patients with a mean 30-min prediction error RMSE of 26 mg/dl. In [16] both ARX, ARMAX and state-space models were investigated using different identification methods for 30-, 60-, 90- and 120-min prediction for nine Montpellier patients from a trial in the DIAdvisor project [30]. The best performance was achieved with the ARX and the ARMAX models. The ARX model gave a standard deviation of the prediction error of 17, 34, 46 and 56 mg/dl on average for the 30-, 60-, 90- and 120-min prediction, respectively. The corresponding results for the ARMAX model were 16, 30, 39 and 44 mg/dl.

Another type of transfer function model, cast in the continuous domain, was approached in [78], where it was evaluated for nine type I subjects on separated meal and insulin intakes. Model parameters were determined both heuristically and by least-squares estimation. The carbohydrate and insulin impacts of the model, i.e., the steady-state rise and drop of glucose following these intakes, were further compared to the corresponding practically used estimates of these factors. No independent prediction validation was given. This model was later evaluated in a control framework in [79], where two data sets were created by the Hovorka (4 subjects) and Padova (10 subjects) simulation models. Here, the model could approximate the simulated data very well, with a 3-h look-ahead prediction error of 26 mg/dl reported. A very similar model structure was used in [55], the difference being a time delay changed into a time lag. In this chapter, breakfast glucose excursion prediction was addressed for 10 patient datasets collected in the DIAdvisor project [30]. For each patient, model parameters were determined by

constrained least squares for two breakfast meals and cross-validated on a third breakfast, with an average fit value of 42 %.

Neural network (NN) models have been shown to be a competitive approach in [26], where a recurrent NN model was compared against an AR and an ARX model on a 30 patient dataset, retrieved from the Padova simulation model. Here, the NN clearly outperformed the competing models with an average RMSE of 4.9 mg/dl versus 29 mg/dl (AR) and 26 mg/dl (ARX) for the 45-min prediction. Apart from meal and insulin information, emotional factors, hypoglycemic/hyperglycemic symptoms and lifestyle/ activities, were collected in an electric diary and used as inputs in the NN model of [77]. Training was performed on a dataset from 17 patients, and performance was evaluated on 10 patient data sets not included in the training set, with a RMSE of 44 mg/dl for the 45-min prediction.

A fully connected three-layer (5, 10, 1 neuron per layer) NN, with sigmoidal transfer functions in the first two layers and a linear for the output block was used in [80]. No insulin nor meal information were used, but the concurrent and previous CGM values, up to 20 min back, acted as inputs. The model was evaluated on two datasets with different CGM devices (Abbott Freestyle and MedTronic Guardian). Three subject data sets were used for training for each patient group and were thereafter excluded from the validation data. For the six Guardian patients and the three Abbott Freestyle patients the performance was 10, 18 and 27 mg/dl for the 15, 30 and 45-min prediction, with a delay of around 4, 9, and 14 min for upward trends, and 5, 15, and 26 min for downward trends. In [106], the linear predictor from [96] worked in a cascade-like configuration with a NN model, which also used both CGM and glucose flux from the meal model of [24] as inputs. Training and validation was done using 15 patient records from the 7-day free-living conditions set of the DIAdvisor DAQ trial [30]. The NN was trained and validated on 25 time series, each one of 3 days, selected so as to ensure a wide variety of glycemic dynamics. Nine daily profiles, containing several hypo- and hyperglycemic events, were used to test the NN with an average of 14 mg/dl and a 14 min delay for the 30-min prediction. For an assessment on 20 simulated subjects using the UVa/Padova model, the corresponding metrics were 9.4 mg/dl and 5 min. Both insulin and carbohydrate digestion were considered by incorporating input-generating sub models in the support vector machine of [45]. Additionally, exercise-induced glucose and insulin absorption variations were also considered as inputs by processing a metabolic equivalent (MET) estimate, derived from a SenseWear body monitoring system (BodyMedia Inc.) used in the study, in a model by [91]. The NN was trained individually for seven type 1 patients with RMSE of 9.5, 16, 25 and 36 mg/dl for the 15, 30, 60 and 120-min prediction.

Examples of other machine learning approaches that have been considered, include, e.g., support vector regression [44] and random forests [46]. Both techniques were evaluated on the same dataset of 27 type 1 patient records from free-living conditions collected within the METABO project [43]. The recorded insulin injections as well as the meal intakes were fed into compartment models to provide estimated profiles of plasma insulin and glucose rate of appearance. Furthermore, physical activity, estimated from a body monitoring system, and the time of the

day were also added as input variables. The predictive performance of each method was assessed for a 15-, 30-, 60- and 120-min ahead prediction horizon with impressive results. The reported RMSE of the support vector regression for these predictions horizons was 5.2, 6.0, 7.1 and 7.6 mg/dl, whereas the random forest method managed slightly worse; 6.6, 8.2, 9.3 and 10.8 mg/dl.

Further reviews can be found in, e.g., [7, 45, 66].

1.2 Model Merging

Merging models for the purpose of prediction has been developed in different research communities. In the meteorological and econometric communities regression-oriented ensemble prediction has been a vivid research area since the late 1960s, see, e.g., [31, 85].

Also in the machine learning community, the question of how different predictors or classifiers can be used together for increased performance has been investigated, and different algorithms have been developed, such as the bagging, boosting [13] and weighted majority [63] algorithms, and online versions of these [56, 74].

In most approaches the merged prediction $\hat{y}_k^e$ at time k is formed by a linear weighted average of the individual predictors $\hat{\mathbf{y}}_k$.

$$\hat{y}_k^e = \mathbf{w}_k^{\mathbf{T}} \hat{\mathbf{y}}_k \tag{1}$$

It is also common to restrict the weights $\mathbf{w}_k$ to [0,1]. The possible reasons for this are several, where the interpretation of the weights as probabilities, or rather Bayesian beliefs, is the dominating. Such restrictions are however not always applicable, e.g. in the related optimal portfolio selection problem, where negative weight (short selling) can reduce the portfolio risk [32].

A special case, considering distinct switches between different linear system dynamics, has been studied mainly in the control community. The data stream and the underlying dynamic system are modelled by pure switching between different filters derived from these models, i.e., the weights w_k can only take value 1 or 0. A lot of attention has been given to reconstructing the switching sequence, see, e.g., [47, 73]. From a prediction viewpoint, the current dynamic mode is of primary interest, and it may suffice to reconstruct the dynamic mode for a limited section of the most recent time points in a receding horizon fashion [4].

Combinations of specifically adaptive filters has also stirred some interest in the signal processing community. Typically, filters with different update pace are merged, to benefit from each filter's specific change responsiveness, respectively steady state behaviour [5].

Finally, in fuzzy modeling, soft switching between multiple models is offered using fuzzy membership rules in the Takagi–Sugeno systems [100].

Merging of predictions in the glucose prediction context has previously been investigated in terms of hypo- or hyperglycemic warning systems. In [25], the

glucose prediction from a so-called output corrected ARX predictor (see the reference for method details) was linearly combined with the prediction from an adaptive recurrent neural network model. The balancing factor for the linear combination was determined offline by optimizing a trade-off between hypo- and hyperglycemic sensitivity, effective prediction horizon and the false alarm rate. This factor was determined individually for each patient and the balance may be different for hypo- and hyperglycemia. A different mechanism was used in [27]. Here, five different predictors were running simultaneously, and the hypoglycemic alarm was based upon a voting scheme between the individual predictors. If a number of the five predictors exceeded the predefined hypoglycemic threshold value an alarm was raised. Both studies indicated an improvement in alarm sensitivity compared to the individual predictors.

2 Problem Formulation

As seen from the review above, many different approaches to glucose modeling and predicting have been established. These methods may each be more suitable to specific conditions for the glucose dynamics, and improvements in robustness and prediction performance may be achieved by combining their outcomes, as indicated from the studies from the hypo-/hyperglycemic alarm systems. Such a situation is depicted in Fig. 1, where two prediction models try to capture the true glucose level. In different situations, each predictor is clearly outperforming the other and is capable of providing good estimates of the true glucose level. However, as the conditions change the performance deteriorates, and instead the other predictor is more suitable to rely upon. Given this informal background a more formal problem formulation is now outlined.

A non-stationary data stream $z_k : \{y_k, u_k\}$ arrives with a fixed sample rate, set to 1 for notational convenience, at time $t_k \in \{1, 2, \ldots\}$. The data stream contains a variable of primary interest called $y_k \in \mathbb{R}$ and additional variables u_k. The data stream can be divided into different periods T_{S_i} of similar dynamics $S_i \in S = [1, \ldots, n]$, and where $s_k \in S$ indicates the current dynamic mode at time t_k. The system changes between these different modes according to some unknown dynamics.

Given m number of expert q-steps-ahead predictions, $\hat{y}^j_{k+q|k}, j \in \{1, \ldots, m\}$ of the variable of interest at time t_k, each utilizing different methods, and/or different training sets; how is an optimal q-steps-ahead prediction $\hat{y}^e_{k+q|k}$ of the primary variable, using a predefined norm and under time-varying conditions, determined?

3 Sliding Window Bayesian Model Averaging

Apart from conceptual differences between the different approaches to ensemble prediction, the most important difference is how the weights are determined. Numerous different methods exist, ranging from heuristic algorithms [5, 100] to

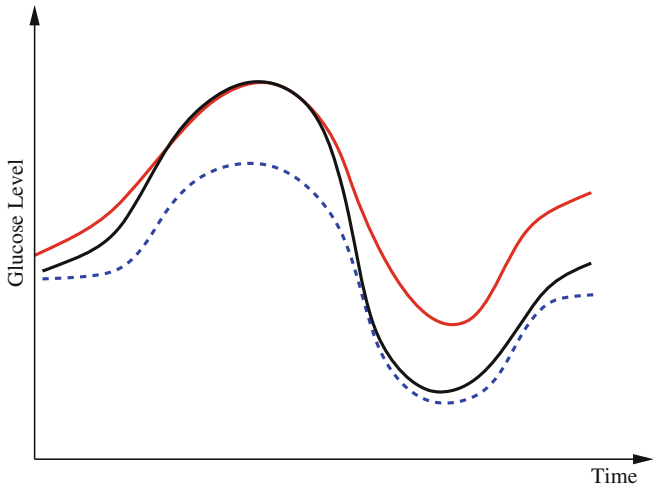


Fig. 1 Example of when merging between different predictors may be beneficial. Initially the model corresponding to the *red dash-dotted* prediction resembles the true reference (*black solid curve*) best, but as the conditions change the prediction given by the other prediction model (*blue dashed curve*) gradually takes the lead

theory based approaches, e.g., [49]. Specifically, in a Bayesian Model Averaging framework [49], which will be adopted in this chapter, the weights are interpreted as partial beliefs in each predictor M_i, and the merging is formulated as:

$$p\left(y_{k+q}|D_k\right) = \sum_i p\left(y_{k+q}|M_i, D_k\right) p(M_i|D_k) \tag{2}$$

where $p\left(y_{k+q}|D_k\right)$ is the conditional probability of y at time t_{k+q} given the data, $D_k : \{z_{1:k}\}$ received up until time k, and if only point-estimates are available, one can, e.g., use:

$$\hat{y}^e_{k+q|k} = \mathbb{E}\left(y_{k+q}|D_k\right) \tag{3}$$

$$= \sum_i \mathbb{E}(M_i|D_k)\mathbb{E}\left(y_{k+q}|M_i,\ D_k\right) \tag{4}$$

$$= \mathbf{w}_k^T \hat{\mathbf{y}}_k \tag{5}$$

$$\mathbf{w}_k^{(i)} = \mathbb{E}(M_i|D_k) \tag{6}$$

$$p\left(\mathbf{w}_k^{(i)}\right) = p(M_i|D_k) \tag{7}$$

where $\hat{y}^e_{k+q}$ is the combined prediction of y_{k+q} using information available at time k, and $\mathbf{w}_k^{(i)}$ indicates position i in the weight vector. The conditional probability of

predictor M_i can be further expanded by introducing the latent variable $\theta_k \in \Theta = [1, \ldots, p]$.

$$p(M_i|D_k) = \sum_j p(M_i|\theta_k = j,\ D_k)p(\theta_k = j|D_k) \tag{8}$$

or in matrix notation

$$\mathbf{p}(\mathbf{w}_k) = [\mathbf{p}(\mathbf{w}_k|\theta_k = 1), \ldots, \mathbf{p}(\mathbf{w}_k|\theta_k = p)]\ [p(\theta_k = 1|D_k), \ldots, p(\theta_k = p|D_k)]^T \tag{9}$$

Here, Θ represents a *predictor mode* in a similar sense to the dynamic mode S, and likewise θ_k represents the prediction mode at time k. $\mathbf{p}(\mathbf{w}_k|\theta_k = j)$ is a column vector of the joint prior distribution of the conditional weights of each predictor model given the predictor mode $\theta_k = j$. Generally, there is a one-to-one relationship between the predictor modes and the corresponding dynamic modes, i.e., $p = n$.

Data for estimating the distribution for $\mathbf{p}(\mathbf{w}_k|\theta_k = i)$ is given based upon using a constrained optimization on the training data. In cases of labelled training data sets, the following applies:

$$\begin{aligned} \left\{\mathbf{w}_{k|\theta_k=i}\right\}T_{S_i} \ &= \ \arg\min \sum_{m=k-N/2}^{k+N/2} \mathscr{L}\left(y(t_m),\ \mathbf{w}_{k|\theta_k=i}^T \mathbf{\hat{y}}_i\right), \quad k \in T_{S_i} \\ &\text{s.t.} \sum_j \mathbf{w}_{k|\theta_k=i}^{(j)} = 1 \end{aligned} \tag{10}$$

where T_{S_i} represents the time points corresponding to dynamic mode S_i, the tunable parameter N determines the size of the evaluation window and $\mathscr{L}(y, \hat{y})$ is a cost function. From these data sets, the prior distributions can be estimated by the Parzen window method [11], giving mean $\mathbf{w}_{0|\theta_k=i}$ and covariance matrix $\mathbf{R}_{\theta_k=i}$. An alternative to the Parzen approximation is of course to estimate a more parsimoniously parametrized probability density function (pdf) (e.g., Gaussian) for the extracted data points. For unlabelled training data, with time points T, the corresponding datasets $\{\mathbf{w}_k|\theta_k = i\}_T$ are found by cluster analysis, e.g., using the k-means algorithm or a Gaussian Mixture Model (GMM) [11]. A conceptual visualisation is given in Fig. 2. Now, in each time step k, the $\mathbf{w}_k|\theta_{k-1}$ is determined from the sliding window optimization below, using the current active mode θ_{k-1}. For reasons soon explained, only $\mathbf{w}_k|\theta_{k-1}$ is thus calculated:

$$\begin{aligned} \mathbf{w}_{k|\theta_{k-1}} = \arg\min & \sum_{j=k-N}^{k-1} \mu^{k-j} \mathscr{L}\left(y_j,\ \mathbf{w}_{k|\theta_{k-1}}^T \mathbf{\hat{y}}_j\right) \\ & + \left(\mathbf{w}_{k|\theta_{k-1}} - \mathbf{w}_{0|\theta_{k-1}}\right) \Lambda_{\theta_{k-1}} \left(\mathbf{w}_{k|\theta_{k-1}} - \mathbf{w}_{0|\theta_{k-1}}\right)^T \\ & \text{s.t.} \sum_j \mathbf{w}_{k|\theta_{k-1}}^{(j)} = 1 \end{aligned} \tag{11}$$

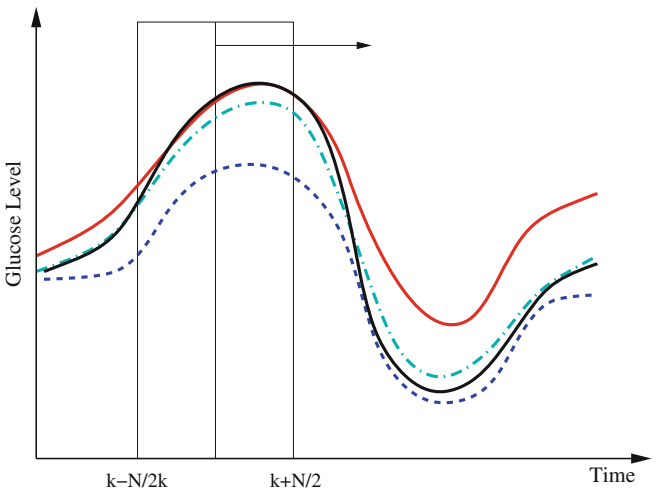


Fig. 2 Principle of finding the predictor modes for unlabelled data over the training data set time period T. For every time point $t_k \in T$, the optimal $\mathbf{w}_k$ is determined by Eq. (10), where the optimal prediction $\mathbf{w}_k\hat{\mathbf{y}}$ (*light green dash-dotted curve*) formed from the individual predictions $\hat{\mathbf{y}}$ (the *blue dashed* and the *red solid curves*) is evaluating against the reference (*black solid curve*) using the cost function $\mathscr{L}$ over a sliding data window between $t = k - N/2$ and $t = k + N/2$. The aggregated set $\{\mathbf{w}_k\}_T$ is thereafter subjected to clustering to find the different mode centers $\mathbf{w}_{0|\theta=i},\ i = [1, \ldots, p]$

Here, μ_j is a forgetting factor, and $\Lambda_{\theta_k=i}$ is a regularization matrix. To infer the posterior $\mathbf{p}(\theta_k|D_k)$ in (9), it would normally be natural to set this probability function equal to the corresponding posterior pdf for the dynamic mode $\mathbf{p}(S|D_k)$. However, problems arise if $\mathbf{p}(S|D_k)$ is not directly possible to estimate from the dataset D_k. This is circumvented by using the information provided by the $\mathbf{p}\left(\mathbf{w}_{k|\theta_k}\right)$ estimated from the data retrieved from Eq. (10) above. The $\mathbf{p}\left(\mathbf{w}_{k|\theta_k}\right)$ prior density functions can be seen as defining the region of validity for each predictor mode. If the $\mathbf{w}_{k|\theta_{k-1}}$ estimate leaves the current active mode region θ_{k-1} (in a sense that $\mathbf{p}\left(\mathbf{w}_{k|\theta_{k-1}}\right)$ is very low), it can thus be seen as an indication of that a mode switch has taken place. A logical test is used to determine if a mode switch has occurred. The predictor mode is switched to mode $\theta_k = i$, if:

$$p(\theta_k = i|\mathbf{w}_k,\ D_k) > \lambda \tag{12}$$

where

$$p(\theta_k = i|\mathbf{w}_k,\ D_k) = \frac{p(\mathbf{w}_k|\theta_k = i,\ D_k)p(\theta_k = i|D_k)}{\sum_j p(\mathbf{w}_k|\theta_k = j,\ D_k)p(\theta_k = j|D_k)} \tag{13}$$

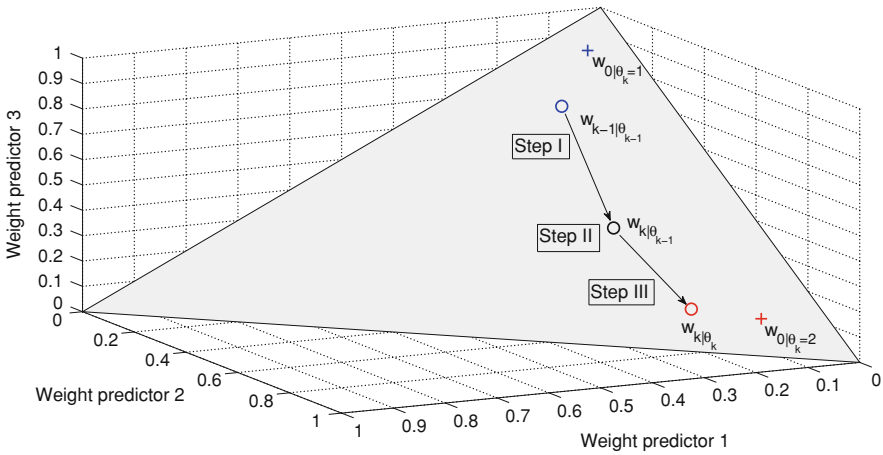


Fig. 3 Predictor mode switching for an example with three individual predictor models. *Step I* At time instance t_k the new $w_{k|\theta_{k-1}}$ is determined from Eq. (11) In this case, the data forces the optimal weight away from the active prediction mode center. *Step II* The likelihood values $p(w_k|\theta_k = i),\ i = [1, \ldots, p]$ are calculated and if the condition according to Eq. (12) is fulfilled, a predictor mode switch occurs. *Step III* Based on the new predictor mode, Eq. (11) is rerun and the weight vector now gravitates towards the new mode center

A λ somewhat larger than 0.5 gives a hysteresis effect to avoid chattering between modes. Unless otherwise estimated from data, the conditional probability of each prediction mode $p(\theta_k = i|D_k)$ is set equal for all possible modes, and thus cancels in (13). The logical test is evaluated using the priors received from the pdf estimate and the $\mathbf{w}_{\mathbf{k}|\theta_\mathbf{k}}$ received from (11). If a mode switch is considered to have occurred (11) is rerun using the new predictor mode.

Now, since only one prediction mode θ_k is active; (9) reduces to $\mathbf{p}(\mathbf{w}_k) = \mathbf{p}\left(\mathbf{w}_{k|\theta_k}\right)$. The predictor mode switching concept is visualised in Fig. 3.

3.1 Parameter Choice

The length N of the evaluation period is, together with the forgetting factor μ, a crucial parameter determining how fast the ensemble prediction reacts to sudden changes in dynamics. A small forgetting factor will put much emphasis on recent data, making it more agile to sudden changes. However, the drawback is of course that the noise sensitivity increases.

$\Lambda_{\theta_k=i}$ should also be chosen, such that a sound balance between flexibility and robustness is found, i.e., a too small $\|\Lambda_{\theta_k=i}\|_2$ may result in over-switching, whereas a too large $\|\Lambda_{\theta_k=i}\|_2$ will give a stiff and inflexible predictor. Furthermore, $\Lambda_{\theta_k=i}$ should force the weights to move within the perimeter defined by $p(\mathbf{w}|\theta_k = i)$. This is approximately accomplished by setting $\Lambda_{\theta_k=i}$ equal to the

inverse of the covariance matrix $\mathbf{R}_{\theta_k=i}$, thus representing the pdf as a Gaussian distribution in the regularization.

Optimal values for N and μ can be found by evaluating different choices for some test data. However, from our experience we have seen that $N = 10$–20 and $\mu = 0.8$ are suitable choices for this application.

3.2 Nominal Mode

Apart from the estimated prediction mode centres, an additional predictor mode can be added, corresponding to a heuristic fall-back mode. In the case of sensor failure, or other situations where loss of confidence in the estimated predictor modes arises, each predictor may seem equally valid. In this case, a fall-back mode to resort to may be the equal weighting. This is also a natural start for the algorithm. For these reasons, a nominal mode $\theta_k = 0 : p(\mathbf{w_k}|\theta_\mathbf{k} = \mathbf{0}) \in N(\mathbf{1/m}, \mathbf{I})$ is added to the set of predictor modes.

Summary of algorithm

1. Estimate m numbers of predictors according to best practice.
2. Run the predictors and the constrained estimation (10) on labelled training data and retrieve the sequence of $\left\{\mathbf{w}_{k|\Theta=i}\right\}_{T_{S_i}}$, $\forall i \in \{1, \ldots, n\}$.
3. Classify different predictor modes, and determine density functions $\mathbf{p}\left(\mathbf{w}_{k|\Theta=i}\right)$ for each mode $\Theta = i$ from the training results by supervised learning. If possible; estimate $p(\Theta = i|D)$.
4. Initialize mode to the nominal mode.
5. For each time step; calculate $\mathbf{w_k}$ according to (11).
6. Test if switching should take place by evaluating (12) and (13), and switch predictor mode if necessary and recalculate new $\mathbf{w}_k$ according to (11).
7. Go to 5.

The ensemble engine outlined above will hereafter be referred to as Sliding Window Bayesian Model Averaging (SW-BMA) Predictor.

4 Choice of Cost Function $\mathscr{L}$

The cost function should be chosen with the specific application in mind. A natural choice for interpolation is the 2-norm, but in certain situations asymmetric cost functions are more appropriate. For the glucose prediction application, a suitable

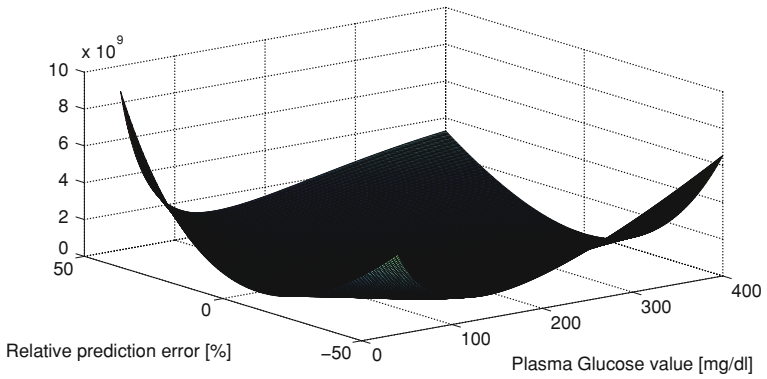


Fig. 4 Cost function of relative prediction error

candidate for determining appropriate weights should take into account that the consequences of acting on too high glucose predictions in the lower blood glucose (G) region (<90 mg/dl) could possibly be life threatening. The margins to low blood glucose levels, that may result in coma and death, are small, and blood glucose levels may fall rapidly. Hence, much emphasis should be put on securing small positive predictive errors and sufficient time margins for alarms to be raised in due time in this region. In the normoglycemic region (here defined as 90–200 mg/dl), the predictive quality is of less importance. This is the glucose range that healthy subjects normally experience, and thus can be considered, from a clinical viewpoint in regards to possible complications, a safe region. However, due to the possibility of rapid fluctuation of the glucose into unsafe regions, some considerations of predictive quality should be maintained.

Based on the cost function in [58], the selected function incorporates these features; asymmetrically increasing cost of the prediction error depending on the absolute glucose value and the sign of the prediction error.

In Fig. 4 the cost function can be seen, plotted against relative prediction error and absolute blood glucose value.

4.1 Correspondence to the Clarke Grid Error Plot

A de facto accepted standardized metric of measuring the performance of CGM signals in relation to reference measurements, and often used to evaluate glucose predictors, is the Clarke Grid Plot [19]. This metric meets the minimum criteria raised earlier. However, other aspects makes it less suitable; no distinction between prediction errors within error zones is made, switches in evaluation score are instantaneous, etc.

In Fig. 5, the isometric contours of the chosen function for different prediction errors at different G values has been plotted together with the Clarke Grid Plot.

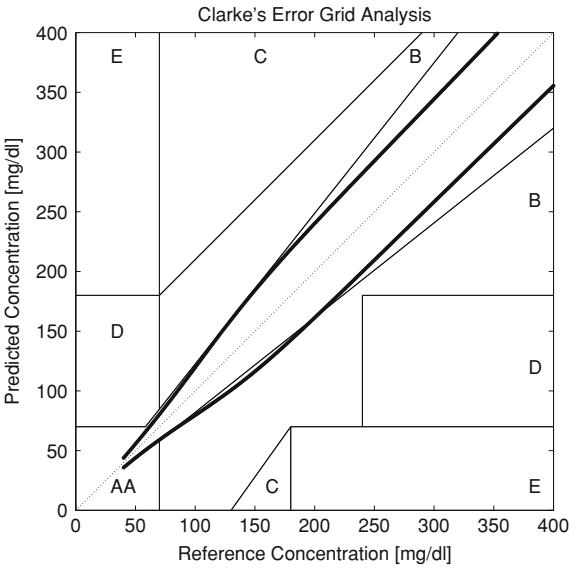


Fig. 5 Isometric cost in comparison to the Clarke Grid

The boundaries of the A/B/C/D/E areas of the Clarke Grid can be regarded as lines of isometric cost according to the Clarke metric. In the figure, the isometric value of the cost function has been chosen to correspond to the lower edge, defined by the intersection of the A and B Clarke areas at 70 mg/dl. Thus, the area enveloped by the isometric border can be regarded as the corresponding A area of this cost function.

Apparently, much tougher demands are imposed both in the lower and upper glucose regions in comparison to the Clarke Plot.

5 Example I: The UVa/Padova Simulation Model

5.1 Data

Data were generated using the nonlinear metabolic simulation model, jointly developed by the University of Padova, Italy and University of Virginia, U.S. (UVa) and described in [24], with parameter values obtained from the authors. The model consists of three parts that can be separated from each other. Two sub models are related to the influx of insulin following an insulin injection and the rate of appearance of glucose from the gastro-intestinal tract following meal intake, respectively.

The transport of rapid-acting insulin from the subcutaneous injection site to the blood stream was based on the compartment model in [23, 24], as follows.

$$\dot{I}_{sc1}(t) = -(k_{a1} + k_d) \cdot I_{sc1}(t) + D(t) \tag{14}$$

$$\dot{I}_{sc2}(t) = k_d \cdot I_{sc1}(t) - k_{a2} \cdot I_{sc2}(t) \tag{15}$$

$$\dot{I}_p(t) = k_{a1} \cdot I_{sc1}(t) + k_{a2} \cdot I_{sc2}(t) - (m_2 + m_4) \cdot I_p(t) + m_1 \cdot I_l(t) \tag{16}$$

$$\dot{I}_l(t) = m_2 \cdot I_p(t) - (m_1 + m_3) \cdot I_l(t) \tag{17}$$

$$m_2 = 0.6 \frac{C_L}{H_{Eb} \cdot V_i \cdot M_{BW}} \tag{18}$$

$$m_3 = \frac{H_{Eb} \cdot m_1}{1 - H_{Eb}} \tag{19}$$

$$m_4 = 0.4 \frac{C_L}{V_i \cdot M_{BW}} \tag{20}$$

Following the notation in [23, 24], I_{sc1} is the amount of non-monomeric insulin in the subcutaneous space, I_{sc2} is the amount of monomeric insulin in the subcutaneous space, k_d is the rate constant of insulin dissociation, k_{a1} and k_{a2} are the rate constants of non-monomeric and monomeric insulin absorption, respectively, $D(t)$ is the insulin infusion rate, I_p is the level of plasma insulin, I_l the level of insulin in the liver, m_3 is the rate of hepatic clearance, and m_1, m_2, m_4 are rate parameters. The parameters m_2, m_3, m_4 are determined based on steady-state assumptions—relating them to the constants in Table 1 and the body weight M_{BW}.

The initial stages of glucose metabolism, describing the digestive process and the flux of glucose from the intestines, have been modeled as follows:

$$q_{sto}(t) = q_{sto1}(t) + q_{sto2}(t) \tag{21}$$

$$\dot{q}_{sto1}(t) = -k_{gri} \cdot q_{sto1}(t) + C(t) \tag{22}$$

$$\dot{q}_{sto2}(t) = k_{gri} \cdot q_{sto1}(t) - k_{empt} \cdot q_{sto1}(t) \cdot q_{sto2}(t) \tag{23}$$

$$\dot{q}_{gut}(t) = -k_{abs} \cdot q_{gut}(t) + k_{empt} \cdot q_{sto}(t) \cdot q_{sto2} \tag{24}$$

$$R_a(t) = \frac{f \cdot k_{abs} \cdot q_{gut}(t)}{M_{BW}} \tag{25}$$

where, again following the notation in [21], q_{sto} is the amount of glucose in the stomach (q_{sto1} solid, and q_{sto2} liquid phase), q_{gut} is the glucose mass in the intestine, k_{gri} the rate of grinding, k_{empt} is the rate constant of gastric emptying, k_{abs} is the rate constant of intestinal absorption, f is the fraction of intestinal absorption

Table 1 Generic parameter values used for the GSM and ISM

Parameter	Value	Unit	Parameter	Value	Unit
$\mathbf{k_{gri}}$	0.0558	[min^{-1}]	$\mathbf{k_{a1}}$	0.004	[min^{-1}]
$\mathbf{k_{max}}$	0.0558	[min^{-1}]	$\mathbf{k_{a2}}$	0.0182	[min^{-1}]
$\mathbf{k_{min}}$	0.008	[min^{-1}]	$\mathbf{k_{d}}$	0.0164	[min^{-1}]
$\mathbf{k_{abs}}$	0.0568	[min^{-1}]	$\mathbf{k_{d}}$	0.0164	[min^{-1}]
b	0.82	[-]	$\mathbf{m_1}$	0.1766	[min^{-1}]
d	0.01	[-]	$\mathbf{V_i}$	0.05	[L/kg]
f	0.9	[-]	$\mathbf{C_L}$	1.1069	[L/min]

Table 2 Meal amount and timing randomization

Meal	Time (30 min)	Amount carbohydrates (g)
Breakfast	08:00	45 (5)
Lunch	12:30	70 (10)
Dinner	19:00	80 (10)

Standard deviation in parenthesis

which actually appears in the blood stream, $C(t)$ is the amount of ingested carbohydrates and $R_a(t)$ is the appearance rate of glucose in the blood. k_{empt} is a nonlinear function of q_{sto} and $C(t)$:

$$k_{empt}(q_{sto}) = k_{min} + k \cdot \{\tanh[\alpha(q_{sto} - b \cdot G(t))] + \tag{26}$$

$$- \tanh[\beta(q_{sto} - d \cdot G(t))] + 2\} \tag{27}$$

With $k = (k_{max} - k_{min})/2, \quad \alpha = 5/2D(1b), \quad \beta = 5/2Dc$ with parameters k_{max}, k_{min}, b, and d

Both models were evaluated using generic population parameter values according to Table 1.

The final part of the total model is concerned with the interaction of glucose and insulin in the blood stream, organs and tissue, including renal extraction, endogenous glucose production and insulin and non-insulin dependent glucose utilization. The model equations are partly nonlinear and are found in [24].

Using a parameter set corresponding to a subject with type 1 diabetes (retrieved from the authors of [24]), 20 datasets, each 8 days long, were generated. The timing and size of meals were randomized for each dataset, according to Table 2. The amount of insulin administered for each meal was based on a fixed carbohydrate-to-insulin ratio, perturbed by normally distributed noise, with a 20 % standard deviation.

Process noise was added by perturbing some crucial model parameters p_i in each simulation step; $p_i(t) = (1 + \delta(t))p_i^0$, where p_i^0 represent nominal value and $\delta(t) \in N(0,0.2)$. The affected parameters were (again following the notation in [24]))

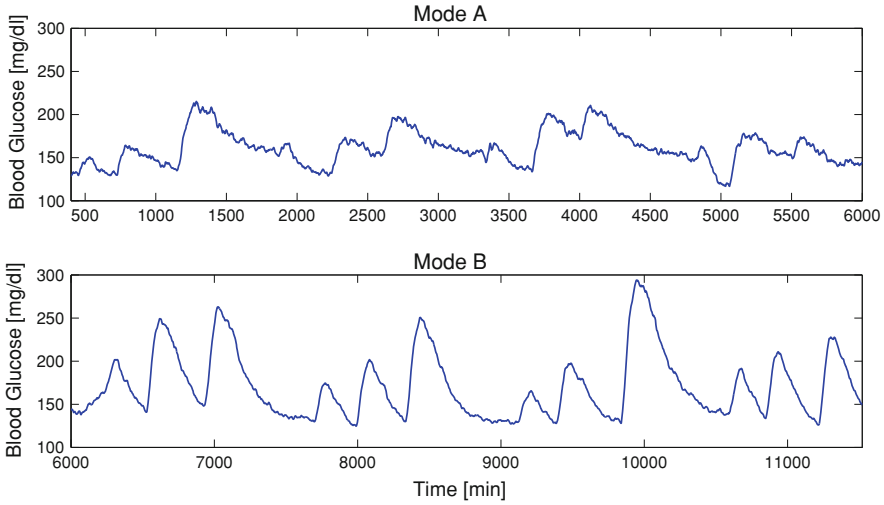


Fig. 6 The training data set. The *upper plot* represents 4 days of dynamic mode A data and the *lower plot* the corresponding last 4 days of dynamic mode B, where four model parameters have been modified. Example I: UVa/Padova Model

k_1, k_2, p_{2u}, k_i, m_1, m_{30}, m_2, k_{sc}, and represents natural intrapersonal variability in the underlying physiological processes.

Two dynamic modes A and B were simulated by, after 4 days, changing four model parameters (following the notation in [24]) k_1, k_i, k_{p3} and p_{2u}, related to the endogenous glucose production and insulin and glucose utilization. This represents an example of shift in the underlying patient dynamics, which may occur due to, e.g., sudden changes in physical or mental stress levels.

A section of 4 days, including the period when the dynamic change took place, of a data set can be seen in Fig. 6. One of the 20 datasets was used for training and the others were considered test data.

5.2 Predictors

For prediction modeling purposes, the system was considered to consist of three main parts in a similar sense as the simulation model was constructed. The absorption models of glucose and insulin were adopted and considered known. The outputs $I_p(t_k)$ and $R_a(t_k)$ from these models were fed into a linear state-space model of the Glucose-Insulin Interaction (GIIM), generating the final output—the blood

glucose $G(k)$ at time $t_k \in (5, 10, \ldots)$ min. Short-term predictions, p steps ahead, were evaluated using the Kalman filter:

$$\hat{x}(k+1) = A\hat{x}(k) + Bu(k) + K(y(k) - C\hat{x}(k)) \tag{28}$$

$$\hat{x}(k+p) = A\hat{x}(k+p-1) + Bu(k+p-1) \tag{29}$$

$$\hat{G}(k+p) = C\hat{x}(k+p) \tag{30}$$

where meal and insulin announcements were assumed at least T_{PH} minutes ahead, implying that $u(k+1)$ was known for all $0 < l < p$.

Three models were identified using the N4SID algorithm of the Matlab System Identification Toolbox. Model order (2–4) was determined by the Akaike criterion [53]. The first model *I* was estimated using data from dynamic mode A in the training data, and the second *II* from the mode *B* data, and the final model *III* from the entire training data set. Thus, model *I* and *II* are each specialized, whereas *III* is an average of the two dynamic modes. The models were evaluated for a prediction horizon of 60 min.

5.3 Results

5.3.1 Training the Mode Switcher

The three predictors were used to create three sets of 60 min ahead predictions for the training data. Using (10) with $N = 10$, the weights $\mathbf{w}_k$ were determined. The mode centers were found by k-means clustering, and the corresponding probability distribution for each mode, projected onto the (w_1, w_2)-plane, was thereafter estimated by Parzen window technique [11]. The densities are well concentrated to the corners [1,0,0] and [0,0,1], with means $\mathbf{w_{0|1}} = [0.96, 0.03, 0.01]$ and $\mathbf{w_{0|2}} = [0.03, 0.96, 0.01]$ defining the expected weights for each predictor mode. The nominal mode probability density function was set to $N\left(\frac{\mathbf{1}}{\mathbf{3}}\frac{\mathbf{1}}{\mathbf{3}}\frac{\mathbf{1}}{\mathbf{3}}, \mathbf{0.1I}\right)$. In Fig. 7 all density functions, including the nominal mode, projected onto the (w_1, w_2)-plane, can be seen together.

5.3.2 Ensemble Prediction Versus Individual Predictions

Using the estimated probability density functions and the expected weights $\mathbf{w}$ of the identified predictor modes, the ensemble machine was run on the test data. An example of the distribution of the weights for the two dynamic modes *A* and *B* can be seen in Fig. 8.

An example of how switching between the different modes occurs over the test period can be found in Fig 9.

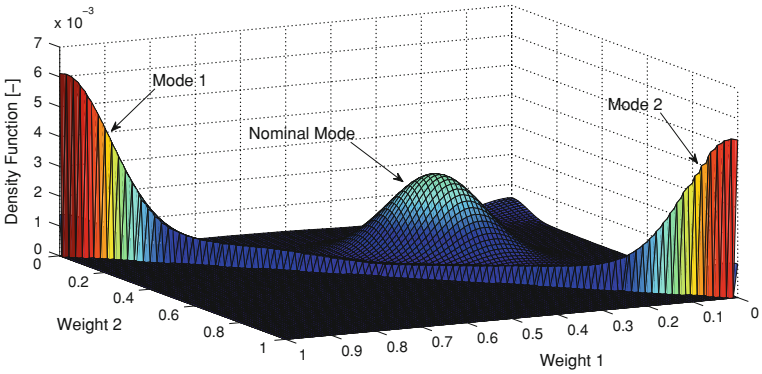


Fig. 7 Estimated probability density functions for the weights in the training data, including nominal mode. Example I: UVa/Padova model

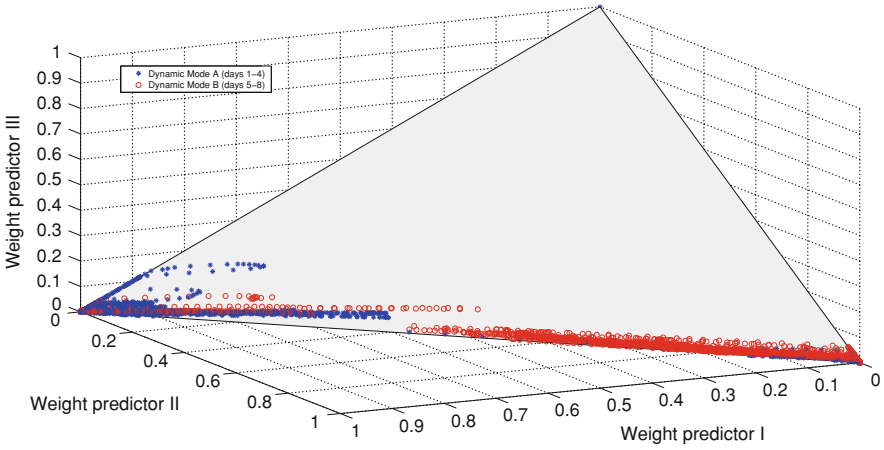


Fig. 8 Example of the distribution of weights in the test data using the estimated pdf:s and expected weights. Example I: UVa/Padova model

For evaluation purposes, all predictors were run individually. In Table 3, a comparative summary of the predictive performance of the different approaches over the test batches, in terms of mean Root Mean Square Error (RMSE), is given. It was also noted that the merged prediction did not introduce any extra prediction delay in comparison to the best individual prediction (not shown).

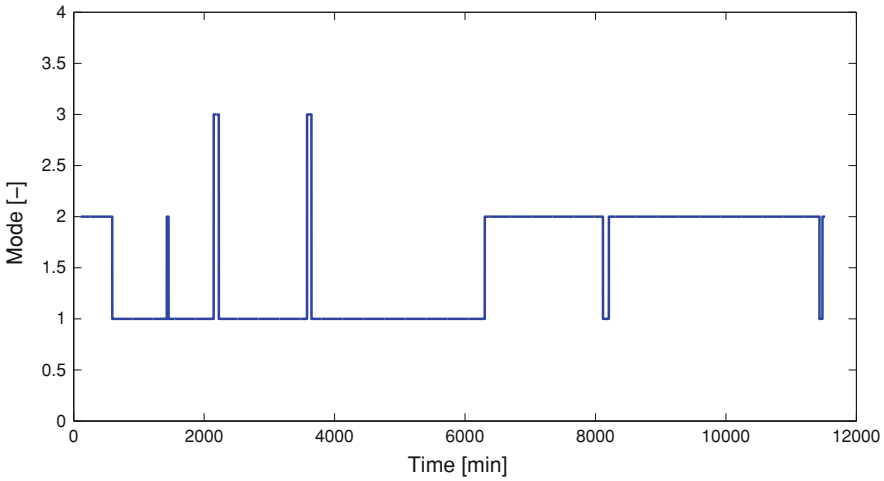


Fig. 9 Example of switching between different predictor modes in the test data. The transition from dynamic mode *B* to mode *A* takes place at 6000 min (c:a 4 days). Mode 3 represents the nominal mode. The late switch to predictor mode 2 in comparison to when the dynamic mode switch takes place is due to that the excitation for the first hours of the fifth day is low until the breakfast meal takes place, i.e., there is little incitement to switch predictor mode before that point. Example I: UVa/Padova model

Table 3 Performance evaluation by RMSE for the 60 min predictors using different approaches

Predictor type	RMSE [mg/dl]		
	Section A	Section B	A + B
Predictor I	8.0	16.1	12.6
Predictor II	15.3	7.2	12.1
Predictor III	9.8	9.9	9.9
Merged prediction	8.4	7.6	8.1

6 Example II: The DIAdvisor Data

6.1 Data

Data from the clinical part of the DAQ trial and the DIAdvisor I B and C trials, conducted within the DIAdvisor project [30], were used. A number of patients participated in all three trials. Based on data completeness, six of these were selected for this study with population characteristics according to Table 4. All selected data were collected at the Montpellier Hospital, and each trial ran over three days. The patients received standardized meals where the amount of carbohydrates included in each meal was about 40 (45 in DAQ), 70 and 70 g,

Table 4 Population Statistics of data

Parameter	Value
Gender	3 Men /3 Women
Therapy	3 Pump /3 Multi-dose injection
Age	54 [32–68]
HbA1c	7.9 [5.7–9.1]
BMI	25.8 [23.7–29.4]

Mean values and [min-max]

respectively. Additional snacks, in some cases related to counter-act hypoglycemia, were also digested. No specific intervention on the usual diabetes treatment was undertaken during the studies, since a truthful picture of normal blood glucose fluctuation and insulin-glucose interaction was pursued. Meal and insulin administration were noted in a logbook, glucose was monitored by the Abbott Freestyle [1] (DAQ) and the Dexcom Seven Plus [29] (DIAdvisor I) CGM systems, and frequent blood glucose measurements (>37 samples a day) were collected for calibration and as reference measurements. The CGM data were used for model identification, whereas the spline-interpolated frequent blood glucose reference measurements were used for validation purposes.

The first trial data (DAQ) were used to train the individual predictor models. The second and third trial data (DIAdvisor I.B and C) were used to train and cross-validate the SW-BMA, i.e., the SW-BMA was trained on B data and validated on C data, and vice versa.

6.2 Predictors

Three different predictors of different structure were developed within the DIAdvisor project, and used in this study; a state-space-based model (SS) [98], a recursive ARX model [36] and a kernel-based predictor [70]. For all three models, the CGM signal $G_{CGM}(t)$ was considered a proxy for the blood glucose $G(t)$, i.e., the lag between the interstitial glucose and the blood glucose, described in e.g. [87], was ignored.

The state-space model and the ARX model used the modeling approach depicted in Fig. 10, with insulin and glucose sub models according to Eqs. (14)–(27), and without interstitial and sensor dynamics modeling (M_2). The state-space model modeled the glucose-insulin interaction, and the glucose prediction, according to Eqs. (28)–(30). The ARX predictor was recursively updated at each time step with an adaptive update gain dependent upon the glucose level according to [36].

The kernel-based predictor did not directly utilize the insulin or meal data channels. Instead, the linear trend and offset parameters given by linear regression of recent CGM data were used as meta features to switch between different

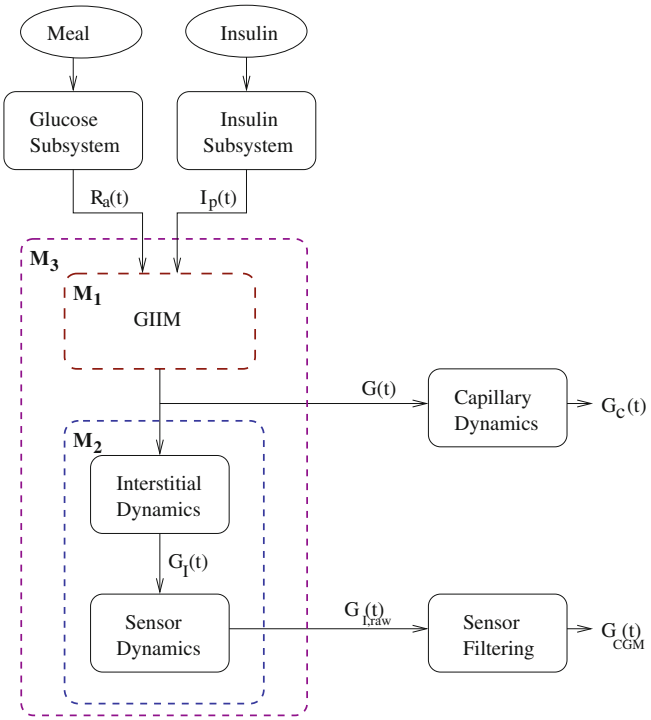


Fig. 10 Overview of the modeling approach. Notation: Plasma Insulin $I_p(t)$, Rate of Glucose Appearance following a meal $R_a(t)$, Blood glucose $G(t)$, Capillary glucose $G_C(t)$, Interstitial Glucose $G_I(t)$, CGM raw current signal $G_{Iraw}(t)$ and CGM signal $G_{cgm}(t)$. M_1 represent the model describing the glucose-insulin interaction in the blood and inner organs (GIIM), the M_2 model represents the diffusion-like relationship between blood and interstitial glucose and the CGM sensor dynamics, and M_3 is the joint model of M_1 and M_2

predefined kernel-based prediction functions, see [71] for a full explanation. Furthermore, this predictor was only trained on one patient data set and was thus considered patient invariant.

6.3 Evaluation Criteria

The prediction results were compared to the interpolated blood glucose G in terms of Clarke Grid Analysis [19] and the complementary Root Mean Square Error (RMSE).

6.4 Results

6.5 Training the Mode Switcher

6.5.1 Cluster Analysis: Finding the Modes

The three predictors were used to create 40 min ahead predictions for both training data sets $D_{T_{B(C)}}$. Using (10) with $N = 20$, the weights $\{\mathbf{w}_k\}_{T_{B(C)}}$ were obtained; example depicted in the (w_1, w_2) plane in Fig. 11. The weights received from the training are easily visually recognized as belonging to different groups (true for all patients, not shown). Attempts were made to find clusters using a Gaussian Mixture Model (GMM) by the EM algorithm, but without viable outcome. This is not totally surprising, considering, e.g., the constraints $0 \geq w_i \geq 1$ and $\Sigma w = 1$. A more suitable distribution, often used as a prior for the weights in a GMM, is the Dirichlet distribution, but instead the simpler k-means algorithm was applied using four clusters (number of clusters given by visual inspection of the distribution of $\{\mathbf{w}_k\}_{T_{B(C)}}$, providing the cluster centers $\mathbf{w}_{0|\Theta_i}$.

The corresponding probability distribution for each mode $p(\mathbf{w}|\Theta_i)$, projected onto the (w_1, w_2)-plane, was estimated by Parzen window technique, and an example can be seen in Fig. 12. Gaussian distributions were fitted to give the covariance matrices $\mathbf{R}_{\Theta_i}$ used in (11).

6.5.2 Feature Selection

The posterior mode probability $p(\theta_k|D_k)$ is likely not dependent on the entire data D_k, but only a few relevant data features, possible to extract from D_k. Features related to the performance of a glucose predictor may include meal information, insulin administration, level of activity, measures of the glucose dynamics, etc. By plotting the training CGM data, colored according to the best mode at the prediction horizon retrieved by the training, interesting correlations become apparent (Fig. 13). The binary features in Table 5 were selected.

When extracting the features, meal timing and content were considered to be known 30 min before the meal.

From the training data, the posterior mode probabilities $p(\theta_k = i|f_j)$, given each feature f_j, were determined by the ratio of active time for each mode over the time periods when each feature was present. Additionally, the overall prior $p(\theta_k = i)$ was determined by the total ratio of active time per cluster over the entire test period.

The different features are overlapping, and the combinations thereof could be regarded as features by themselves. However, the data support for each such new feature would be small and could potentially disrupt, rather than improve, the switching performance. To resolve this issue, the features were not combined

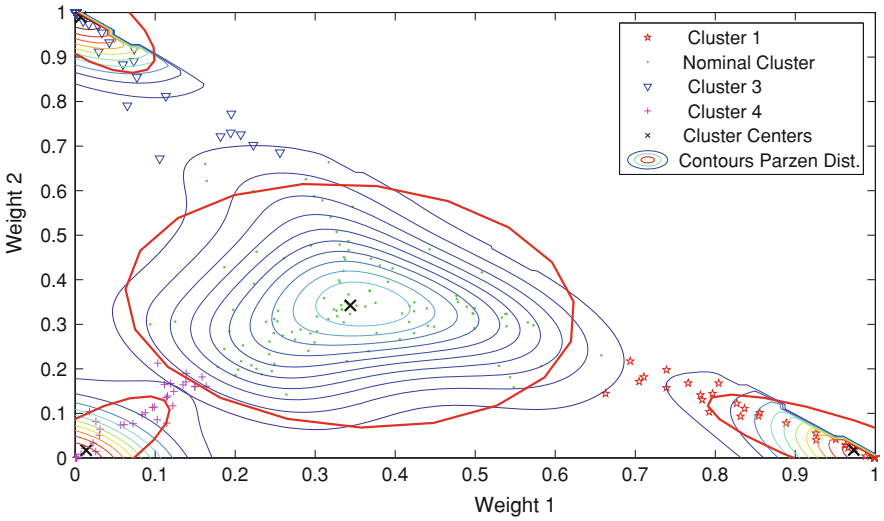


Fig. 11 Example of distribution of weights in the training data by (10) and clusters given by the k-means algorithm. The *red ellipses* represent the fitted Gaussian covariances of each cluster (patient 0103, Trial B). Example II: DIAdvisor Data

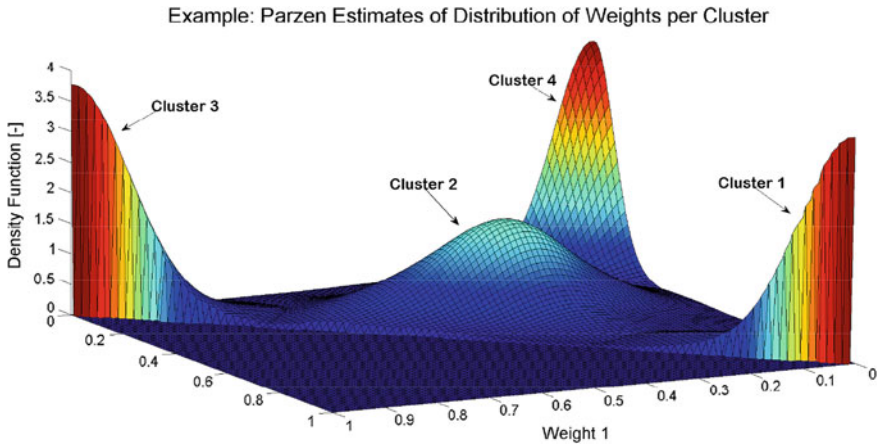


Fig. 12 Example of estimated probability density functions for the different predictor mode clusters in the training data (patient 0103, Trial B). Example II: DIAdvisor Data

(apart from concurrent rising glucose and meal intake, which formed a new feature), and each feature was given different priority—only allowing only the feature of highest priority, f_k^* to be present at each time step t_k. The priority rank was chosen to allow the more specific features to take precedence over the more general features. At each cycle, $p(\theta_k = i|D_k) = p\left(\theta_k = i|f_k^*\right)$ was determined, and if no feature was active, $p(\theta_k = i|D_k)$ was approximated by the $p(\theta_k = i)$ estimate.

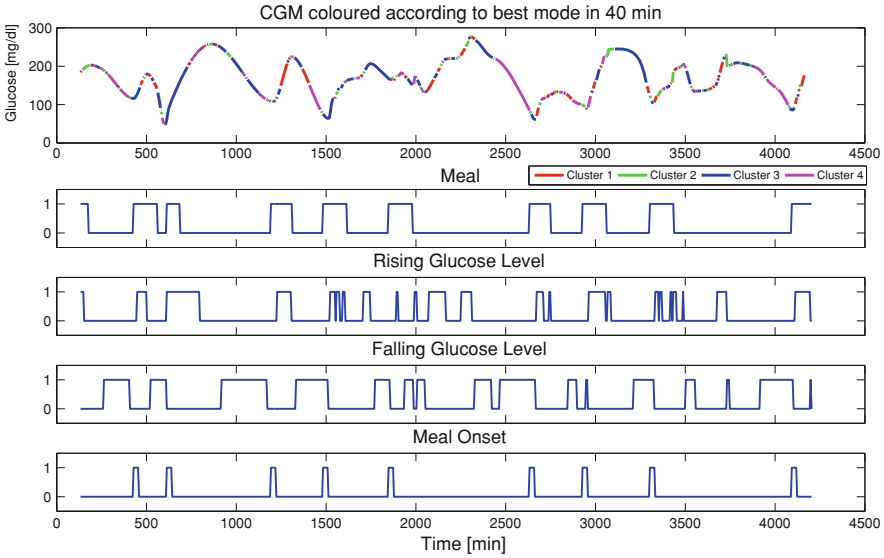


Fig. 13 Example of CGM coloured according to best predictor mode 40 min ahead, together with active features at the moment the prediction was made (patient 0103, Trial B). Example II: DIAdvisor Data

Table 5 Selected features

Feature	Threshold	Priority
Meal	max $(Ra_k, \ldots, Ra_{k+30}) > \varepsilon$	1
Rising G	mean $(\Delta G_{k-10}, \ldots, \Delta G_k) > 30$ mg/(dl·h)	2
Falling G	mean $(\Delta G_{k-10}, \ldots, \Delta G_k) < -18$ mg/(dl·h)	3
Meal and rising G	See above	4
Meal onset	max $Ra(k-30, \ldots, k) < \varepsilon$ and max $Ra(k, \ldots, k+30) > \varepsilon$	5

ε corresponds to the maximum amplitude of glucose rate-of- appearance, R_a after digesting 10 g CHO, and $\Delta G = G_k - G_{k\text{-}5}$

6.6 Prediction Performance on Test Data

Using the estimated mode clusters $\{w_{0li}, R_{0li}\}$, $i = [1, \ldots, M]$, and the estimated posteriors $p(\Theta_i|f^*)$ from Trial B (C), the ensemble machine was run on the Trial C (B) data. The parameter μ was set to 0.8 and N to 20 min. An example of the distribution of the weights $\mathbf{w}_k$ for the three predictors can be seen in Fig. 14.

Table 6 summarizes a comparison of predictive performance over the different patient test data sets for the RMSE evaluation criteria, and in Table 7 the evaluation in terms of Clarke Grid Analysis is given. The optimal switching approach, here defined as using the non-causal fitting by Eq. (10), is used as a measure of optimal performance of a linear combination of the different predictors.

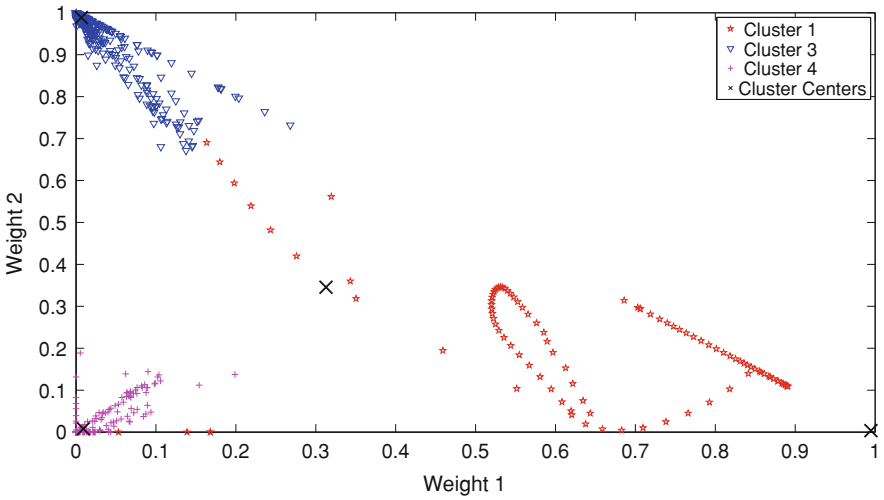


Fig. 14 Example of the distribution of weights in the test data using the estimated clusters and feature correlations (patient 0108, Trial B). Example II: DIAdvisor Data

Table 6 Performance evaluation for the 40 min SW-BMA prediction compared to the optimal switching and the individual predictors

Merging strategy	Median RMSE/RMSE$_{best}$ [min-max]	
	Trial B	Trial C
SW-BMA	1.03 [0.75–1.04]	1.03 [0.94–1.05]
Optimal switching	0.97 [0.54–1.0]	0.94 [0.73–1.0]
2nd best individual pred.	1.16 [1.09–1.27]	1.21 [1.04–1.37]
Worst individual pred.	1.44 [1.25–1.73]	1.45 [1.18–1.83]

The metric is the Root Mean Square Error (RMSE), normalized against the best individual predictor $M_1 - M_3$ for each patient

Table 7 Performance evaluation for the 40 min SW-BMA prediction compared to the optimal switching and the best individual predictor by the amount of data (%) in the acceptable A/B zones versus the dangerous D and E zones

Merging strategy	Trial B			Trial C		
	A/B	D	E	A/B	D	E
SW-BMA	95.5	2.2	0	95.3	3.0	0.1
Optimal switching	96.2	1.7	0	96.9	1.3	0
Best individual pred.	94.8	2.6	0	95.0	3.4	0

7 Discussion

Example I outlined how the technique may be applied to the specific example of diabetes glucose prediction under sudden changes in the underlying physiological dynamics. In this example, the merged prediction turned out to be the best choice. In Example II, applying the algorithm to real-world data, the SW-BMA has, for most patients, the same RMSE and Clarke Grid performance as the best individual predictor. In one case, the merged prediction clearly outperformed also the best predictor ($\mathrm{RMSE}/\mathrm{RMSE}_{\mathrm{best}} = 0.75$). However, comparison to the optimal switcher indicates that there is still further room for improvement. To fill this gap, timely switching is most important. The prediction models in Example II were not specifically designed for specialisation, but are diversified in terms of modeling and parameter identification methods in relation to each other. The state-space model is patient-specific, with fixed parameter values after training—making it agile to interpersonal differences but more sensitive to time-variability. The model is invariant to the absolute glucose level. The ARX model, on the other hand, is recursively updated to capture time-variability, but the approach may be vulnerable to fluctuating system excitation conditions. Both models utilize the insulin and meal data inputs. The kernel-based predictor is generic over the patient cohort, and considers the dynamics to be related to the glucose level rather than directly to the inputs' effects. Overall, the three models thereby complement each other in these aspects. The posterior mode probabilities, conditioned on each selected feature, show that some specialisation exists. For example, when feature 5 (meal onset) was active, cluster 3, dominated by the SS predictor, was clearly favoured an average (61 %). Exploiting these correlations may enhance timely switching, and further specialisation and diversification amongst the prediction models can thus be expected to further improve the added value of prediction merging.

The evaluation indicates that the proposed algorithm is robust to sudden changes and in reducing the impact of modeling errors. Apart from that, in many applications, transition between different dynamic modes is a gradual process rather than an abrupt switch, making the pure switching assumption inappropriate. The proposed algorithm can handle such smooth transitions by slowly sliding along a trajectory in the weight plane of the different predictors, perhaps with a weaker Λ if such properties are expected. Furthermore, any type of predictor may be used, not restricting the user to a priori assumptions of the underlying process structure.

In Takagi–Sugeno (TS) system, a technique that also gives soft switching, the underlying assumption is that the switching dynamics can be observed directly from the data. This assumption has been relaxed for the proposed algorithm, extending the applicability beyond that of TS systems.

In [86], another interesting approach to online Bayesian Model Averaging is suggested for changing dynamics. In this approach, the assumed transition dynamics between the different modes are based on a Markov chain. However, in our approach no such assumptions on the underlying switching dynamics are

postulated. Instead, switching is based on recent performance in regards to the applicable norm, and possibly on estimated correlations between predictor modes and features of the data stream $P(\theta_k = i|D_k)$, see Eq. (13).

8 Conclusions

A novel merging mechanism for multiple glucose predictors has been proposed for time-varying and uncertain conditions. The approach was evaluated on both artificial and real-world data sets, incorporating modeling errors in the individual predictors and time-shifting dynamics.

The results show that the merged prediction has a predictive performance in comparison with the best individual predictor in each case, and indicates that the concept may prove useful when dealing with several individual (glucose) predictors of uncertain reliability—reducing the risk associated with definite a priori model selection, or as a means to improve predictive quality if the predictions are diverse enough.

Further research will be undertaken to investigate how interesting features correlated to expected predictor mode changes should be extracted, and in regards to the possibility of making the algorithm unsupervised.

Acknowledgments This work has been financially supported by the European FP7 IP IST-216592 DIAdvisor project. Furthermore, the first two authors are members of the LCCC Linnaeus Center and the eLLIIT Excellence Center at Lund University.

References

1. Abbott Freestyle Navigator (2012) http://www.abbottdiabetescare.co.uk/your-products/freestyle-navigator
2. Ackerman E, Gatewood LC, Rosevear JW, Molnar GD (1965) Model studies of blood-glucose regulation. Bull Math Biophys 27(Special Issue):21–37
3. Agar B, Eren M, Cinar A (2005) Glucosim: educational software for virtual experiments with patients with type 1 diabetes. In: Proceedings of 2005 annual international conference of the IEEE engineering in medicine and biology (EMBC2005), pp 845–848
4. Alessandri A, Baglietto M, Battistelli G (2005) Receding-horizon estimation for switching discrete-time linear systems. IEEE Trans Autom Control 50(11):1736–1748. doi:10.1109/TAC.2005.858684
5. Arenas-Garcia J, Martinez-Ramon M, Navia-Vazquez A, Figueiras-Vidal AR (2006) Plant identification via adaptive combination of transversal filters. Signal Process 86(9):2430–2438. doi:10.1016/j.sigpro.2005.11.008. Special section: Signal processing in UWB communications
6. Arleth T, Andreasson S, Federici MO, Benedetti MM (2000) A model of the endogenous glucose balance incorporating the characteristics of glucose transporters. Comp Meth Prog Biomed 62:219–234

7. Balakrishnan NP, Rangaiah GP, Samavedham L (2011) Review and analysis of blood glucose (BG) models for type 1 diabetic patients. Ind Eng Chem Res 50(21):12041–12066. doi:10.1021/ie2004779
8. Basu R, Di Camillo B, Toffolo G, Basu A, Shah P, Vella A, Rizza R, Cobelli C (2003) Use of a novel triple-tracer approach to assess postprandial glucose metabolism. Am J Physiol 284:E55–E69
9. Berger M, Rodbard D (1989) Computer simulation of plasma insulin and glucose dynamics after subcutaneous insulin injection. Diabetes Care 12(10):725–736
10. Bergman RN, Cobelli C (1980) Minimal modeling, partition analysis, and the estimation of insulin sensitivity. Fed Proc 39(1):110–115
11. Bishop CM (2006) Pattern recognition and machine learning. Springer, Secaucus
12. Bolie VW (1961) Coefficients of normal blood glucose regulation. J Appl Phys 16(5):783–788
13. Breiman L (1996) Bagging predictors. Mach Learn 24(2):123–140
14. Bremer T, Gough DA (1999) Is blood glucose predictable from previous values? A solicitation for data. Diabetes 48:445–451
15. Breton MD (2008) Physical activity—the major unaccounted impediment to closed loop control. J Diab Sci Technol (Online) 2(1):169–174
16. Cescon M (2011) Linear modeling and prediction in diabetes physiology. Licentiate Thesis TFRT-3250. Department of Automatic Control, Lund University, Sweden
17. Chase JG, Shaw G, Le Compte A, Lonergan T, Willacy M, Wong XW, Lin J, Lotz T, Lee D, Hann C (2008) Implementation and evaluation of the SPRINT protocol for tight glycaemic control in critically ill patients: a clinical practice change. Crit Care 12(2):R49. doi:10.1186/cc6868
18. Chase JG, Shaw GM, Lotz T, LeCompte A, Wong J, Lin J, Lonergan T, Willacy M, Hann CE (2007) Model-based insulin and nutrition administration for tight glycaemic control in critical care. Curr Drug Deliv 4(4):283–296
19. Clarke WL, Cox D, Gonder-Frederick LA, Carter W, Pohl SL (1987) Evaluating clinical accuracy of systems for self-monitoring of blood glucose. Diabetes Care 10:622–628
20. Cobelli C, Renard E, Kovatchev B (2011) Artificial pancreas: past, present, future. Diabetes 60(11):2672–2682. doi:10.2337/db11-0654
21. Dalla Man C, Camilleri M, Cobelli C (2006) A system model of oral glucose absorption: validation on gold standard data. IEEE Trans Biomed Eng 53(12):2472–2478
22. Dalla Man C, Caumo A, Cobelli C (2002) The oral glucose minimal model: estimation of insulin sensitivity from a meal test. IEEE Trans Biomed Eng 49(5):419–429
23. Dalla Man C, Raimondo DM, Rizza RA, Cobelli C (2007) GIM, simulation software of meal glucose insulin model. J Diabetes Sci Technol 1(3):1–8
24. Dalla-Man C, Rizza RA, Cobelli C (2007) Meal simulation model of the glucose-insulin system. IEEE Trans Biomed Eng 54(10):1740–1749
25. Daskalaki E, Norgaard K, Zueger T, Prountzou A, Diem P, Mougiakakou S (2013) An early warning system for hypoglycemic/hyperglycemic events based on fusion of adaptive prediction models. J Diabetes Sci Technol 7(3):689–698
26. Daskalaki E, Prountzou A, Diem P, Mougiakakou SG (2012) Real-time adaptive models for the personalized prediction of glycemic profile in type 1 diabetes patients. Diabetes Technol Ther 14(2):168–174
27. Dassau E, Cameron F, Bequette BW, Zisser H, Jovanovič L, Chase HP, Wilson DM, Buckingham BA, Doyle FJ (2010) Real-time hypoglycemia prediction suite using continuous glucose monitoring. Diabetes Care 33(6):1249–1254. doi:10.2337/dc09-1487
28. Derouich M, Boutayeb A (2002) The effect of physical exercise on the dynamics of glucose and insulin. J Biomech 35:911–917
29. Dexcom Seven Plus (2012) http://www.dexcom.com/seven-plus
30. DIAdvisor (2012) http://www.diadvisor.eu
31. Elliott G, Granger CW, Timmermann A (eds) (2006) Handbook of economic forecasting, Chap. 10. Forecast combinations. Elsevier, Amsterdam

32. Elton EJ, Gruber MJ, Padberg MW (1976) Simple criteria for optimal portfolio selection. J. Financ 31(5):1341–1357
33. Eren-Oruklu M, Cinar A, Quinn L (2010) Hypoglycemia prediction with subject-specific recursive time-series models. J Diabetes Sci Technol 4(1):25–33
34. Eren-Oruklu M, Cinar A, Quinn L, Smith D (2008) Adaptive control strategy for regulation of blood glucose levels in patients with type 1 diabetes. J Proc Cont 19(8):1333–1346. doi:10.1016/j.jprocont.2009.04.004
35. Eren-Oruklu M, Cinar A, Quinn L, Smith D (2009) Estimation of future glucose concentrations with subject-specific recursive linear models. Diabetes Technol Ther 11(4):243–253. doi:10.1089/dia.2008.0065
36. Estrada G, Kirchsteiger H, del Re L, Renard E (2010) Innovative approach for online prediction of blood glucose profile in type 1 diabetes patients. In: American control conference (ACC2010), pp 2015–2020
37. Fabietti PG, Canonico V, Federici MO, Benedetti MM, Sarti E (2006) Control oriented model of insulin and glucose dynamics in type 1 diabetics. Med Bio Eng Comp 44(1–2):69–78. doi:10.1007/s11517-005-0012-2
38. Fabietti PG, Canonico V, Orsini-Federici M, Sarti E, Massi-Benedetti M (2007) Clinical validation of a new control-oriented model of insulin and glucose dynamics in subjects with type 1 diabetes. Diabetes Technol Ther 9(4):327–338. doi:10.1089/dia.2006.0030
39. Farmer TG, Edgar TF, Peppas NA (2009) Effectiveness of intravenous infusion algorithms for glucose control in diabetic patients using different simulation models. Ind Eng Chem Res 48(9):4402–4414. doi:10.1021/ie800871t
40. Finan DA, Doyle FJ, Palerm CC, Bevier WC, Zisser HC, Jovanovic L, Seborg DE (2009) Experimental evaluation of a recursive model identification technique for type 1 diabetes. J Diabetes Sci Technol 3(5):1192–1202
41. Gani A, Gribok AV, Lu Y, Ward WK, Vigersky RA, Reifman J (2010) Universal glucose models for predicting subcutaneous glucose concentration in humans. Trans Info Tech Biomed 14(1):157–165. doi:10.1109/TITB.2009.2034141
42. Gani A, Gribok AV, Rajaraman S, Ward WK, Reifman J (2009) Predicting subcutaneous glucose concentration in humans : data-driven glucose modeling. IEEE Trans Biomed Eng 56(2):246–254
43. Georga E, Protopappas V, Guillen A, Fico G, Ardigo D, Arredondo MT, Exar-chos TP, Polyzos D, Fotiadis DI (2009) Data mining for blood glucose prediction and knowledge discovery in diabetic patients: the METABO diabetes modeling and management system. Conference proceedings: annual international conference of the IEEE engineering in medicine and biology society. IEEE engineering in medicine and biology society. Conference 2009, pp 5633–5636. doi:10.1109/IEMBS.2009.5333635. http://www.ncbi.nlm.nih.gov/pubmed/19964403
44. Georga EI, Protopappas VC, Ardigò D, Polyzos D, Fotiadis DI (2013) A glucose model based on support vector regression for the prediction of hypoglycemic events under free-living conditions. Diabetes Technol Ther 15(8):634–643. doi:10.1089/dia.2012.0285. http://www.ncbi.nlm.nih.gov/pubmed/23848178
45. Georga EI, Protopappas VC, Fotiadis DI (2011) Glucose prediction in type 1 and type 2 diabetic patients using data driven techniques. In: Funatsu PK (ed) Knowledge-oriented applications in data mining, Chap. 17. InTech, Rijeka
46. Georga EI, Protopappas VC, Polyzos D, Fotiadis DI (2012) A predictive model of subcutaneous glucose concentration in type 1 diabetes based on random forests. In: 2012 annual international conference of the IEEE engineering in medicine and biology society (EMBC2012), pp 2889–2892
47. Gustafsson F (2000) Adaptive filtering and change detection. Wiley, Hoboken
48. Hejlesen OK, Andreassen S, Hovorka R, Cavan D.A (1997) DIAS—the diabetes advisory system: an outline of the system and the evaluation results obtained so far. Comput Meth Prog Biomed 54(1–2):49–58

49. Hoeting JA, Madigan D, Raftery AE, Volinsky CT (1999) Bayesian model averaging: a tutorial. Stat Sci 14(4):382–417
50. Hovorka R, Canonico V, Chassin LJ, Haueter U, Massi-Benedetti M, Federici MO, Pieber TR, Schaller HC, Schaupp L, Vering T, Wilinska ME (2004) Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes. Physiol Meas 25(4):905–920. doi:10.1088/0967-3334/25/4/010
51. Hovorka R, Chassin LJ, Ellmerer M, Plank J, Wilinska ME (2008) A simulation model of glucose regulation in the critically ill. Physiol Meas 29(8):959–978. doi:10.1088/0967-3334/29/8/008
52. Jensen K, Pedersen C, Larsen L (2007) Diasnet mobile: a personalized mobile diabetes management and advisory service. In: 2nd workshop on personalization for e-health, vol 1
53. Johansson R (2009) System modeling & identification. KFS AB, Lund
54. Kanderian SS, Weinzimer S, Voskanyan G, Steil GM (2009) Identification of intraday metabolic profiles during closed-loop glucose control in individuals with type 1 diabetes. J Diab Sci Technol 3(5):1047–1057
55. Kirchsteiger H, Estrada GC, Pölzer S, Renard E, Re L (2011) Estimating interval process models for type 1 diabetes for robust control design. In: IFAC world congress 2011, pp 11761–11766
56. Kolter JZ, Maloof MA (2003) Dynamic weighted majority: a new ensemble method for tracking concept drift. IEEE international conference on data mining, pp 123–130. doi:http://doi.ieeecomputersociety.org/10.1109/ICDM.2003.1250911
57. Kovatchev B, Breton C, Dalla-Man C, Cobelli C (2008) In silico model and computer simulation environment approximating the human glucose/insulin utilization. Technical Report. Food and Drug Administration Master File MAF 1521
58. Kovatchev B, Straume M, Cox D, Farhy L (2000) Risk analysis of blood glucose data: a quantitative approach to optimizing the control of insulin dependent diabetes. J Theor Med 3:1–10
59. Lee H, Buckingham BA, Wilson DM, Bequette BW (2009) A closed-loop artificial pancreas using model predictive control and a sliding meal size estimator. J Diabetes Sci Technol 3(5):1082–1090
60. Lehmann E, Hermanyi I, Deutsch T (1994) Retrospective validation of a physiological model of glucose-insulin interaction in type 1 diabetes mellitus. Med Eng Phys 16(4):351–352. doi:10.1016/1350-4533(94)90064-7
61. Lehmann ED (1994) AIDA: an interactive diabetes advisor. Comput Methods Programs Biomed 2607(93):183–203
62. Lehmann ED, Deutsch T (1992) A physiological model of glucose-insulin interaction in type 1 diabetes mellitus. J Biomed Eng 14:235–242
63. Littlestone N, Warmuth MK (1994) The weighted majority algorithm. Inf Comput 108(2):212–261
64. Lonergan T, Compte AL, Willacy M, Chase JG, Shaw GM, Hann CE, Lotz T, Lin J, Wong XW (2006) A pilot study of the SPRINT protocol for tight glycemic control in critically Ill patients. Diab Technol Ther 8(4):449–462. doi:10.1089/dia.2006.8.449
65. Lu Y, Rajaraman S, Ward WK, Vigersky RA, Reifman J (2011) Predicting human subcutaneous glucose concentration in real time: a universal data-driven approach. In: Proceedings of 2011 annual international conference of the IEEE engineering in medical and biology society (EMBC2011), pp 7945–7948. doi:10.1109/IEMBS.2011.6091959
66. Makroglou A, Li J, Kuang Y (2006) Mathematical models and software tools for the glucose-insulin regulatory system and diabetes: an overview. Appl Num Math 56:559–573
67. Man CD, Breton MD, Cobelli C (2009) Physical activity into the meal glucose-insulin model of type 1 diabetes: in silico studies. J Diab Sci Technol 3(1):56–67
68. MedTronic (2012) http://www.medtronic-diabetes.se/
69. Natali A, Gastaldelli A, Camastra S, Sironi AM, Toschi E, Masoni A, Ferrannini E, Mari A (2000) Dose-response characteristics of insulin action on glucose metabolism: a nonsteady-state approach. Am J Physiol Endocrinol Metab 278(5):E794–E801

70. Naumova V, Pereverzyev S, Sampath S (2011) A meta-learning approach to the regularized learning—case study: blood glucose prediction. Technical Report. Johann Radon Institute for Computational and Applied Mathematics (RICAM), Linz, Austria
71. Naumova V, Pereverzyev SV, Sivananthan S (2012) A meta-learning approach to the regularized learning-Case study: blood glucose prediction. Neural networks: the official journal of the International Neural Network Society 33:181–193. doi:10.1016/j.neunet.2012.05.004
72. Nucci G, Cobelli C (2000) Models of subcutaneous insulin kinetics. A critical review. Comput Methods Programs Biomed 62:249–257
73. Ohlsson H, Ljung L, Boyd S (2010) Segmentation of ARX-models using sum-of-norms regularization. Automatica 46(6):1107–1111
74. Oza N (2005) Online bagging and boosting. In: 2005 IEEE international conference on systems, man and cybernetics, vol 3, pp 2340–2345
75. Palerm CC, Bequette BW (2007) Hypoglycemia detection and prediction using continuous glucose monitoring-a study on hypoglycemic clamp data. J Diabetes Sci Technol 1(5):624–629
76. Palerm CC, Willis JP, Desemone J, Bequette BW (2005) Hypoglycemia prediction and detection using optimal estimation. Diabetes Technol Ther 7(1):3–14
77. Pappada SM, Cameron BD, Rosman PM, Bourey RE, Papadimos TJ, Olorunto W, Borst MJ (2011) Neural network-based real-time prediction of glucose in patients with insulin-dependent diabetes. Diabetes Technol Ther 13(2):135–141
78. Percival M, Bevier W, Wang Y (2010) Modeling the effects of subcutaneous insulin administration and carbohydrate consumption on blood glucose. J Diabetes 39(3):800–805
79. Percival M, Wang Y, Grosman B, Dassau E, Zisser H, Jovanovič L, Doyle F (2011) Development of a multi-parametric model predictive control algorithm for insulin delivery in type 1 diabetes mellitus using clinical parameters. J Proc Control 21(3):391–404. doi:10.1016/j.jprocont.2010.10.003
80. Pérez-Gandía C, Facchinetti A, Sparacino G, Cobelli C, Gómez EJ, Rigla M, Leiva AD, Hernando ME (2010) Artificial neural network algorithm for online glucose. Diabetes Technol Ther 12(1):81–88
81. Plougmann SR, Hejlesen O, Turner B, Kerr D, Cavan D (2003) The effect of alcohol on blood glucose in type 1 diabetes metabolic modelling and integration in a decision support system. Int J Med Inf 70(2–3):337–344. doi:10.1016/S1386-5056(03)00038-8
82. Poulsen J, Avogaro A, Chauchard F, Cobelli C, Johansson R, Nita L, Pogose M, del Re L, Renard E, Sampath S, Saudek F, Skillen M, Soendergaard J (2010) A diabetes management system empowering patients to reach optimised glucose control: from monitor to advisor. In: Proceedings of 2010 annual international conference of the IEEE engineering in medical and biology society (EMBC2010), pp 5270–5271. doi:10.1109/IEMBS.2010.5626313
83. Prigeon RL, Røder ME, Porte D, Kahn SE (1996) The effect of insulin dose on the measurement of insulin sensitivity by the minimal model technique. Evidence for saturable insulin transport in humans. J Clin Invest 97(2):501–507. doi:10.1172/JCI118441
84. Puckett WR (1992) Dynamic modeling of diabetes mellitus. PhD thesis. University ofWisconsin- Madison
85. Raftery AE, Gneiting T, Balabdaoui F, Pololakowski M (2005) Using Bayesian model averaging to calibrate forecast ensembles. Mon Weather Rev 133:1155–1174
86. Raftery AE, Kárný M, Ettler P (2010) Online prediction under model uncertainty via dynamic model averaging: application to a cold rolling mill. Technometrics 52(1):52–66
87. Rebrin K, Steil GM (2000) Can interstitial glucose assessment replace blood glucose measurements? Diabetes Technol Ther 2(3):461–472
88. Rizza R, Mandarino LJ, Gerich JE (1981) Dose-response characteristics for effects of insulin on production and utilization of glucose in man. Am J Phys Endocrinol Metab 240(6):E630–E639
89. Roy A, Parker RS (2006) Dynamic modeling of free fatty acid, glucose, and insulin: an extended minimal model. Diab Technol Ther 8(6):617–626

90. Roy A, Parker RS (2006) Mixed meal modeling and disturbance rejection in type I diabetes patients. In: Proceedings of 28th IEEE EMBS annual international conference, pp 323–326
91. Roy A, Parker RS (2007) Dynamic modeling of exercise effects on plasma glucose and insulin levels. J Diabetes Sci Technol 1(3):338–347
92. Salzsieder E, Albrecht G, Fischer U, Freyse EJ (1985) Kinetic modeling of the glucoregulatory system to improve insulin therapy. IEEE Trans Biomed Eng BME-32(10):846–855
93. Salzsieder E, Vogt L, Kohnert KD, Heinke P, Augstein P (2011) Model-based decision support in diabetes care. Comput Meth Prog Biomed 102(2):206–218. doi:10.1016/j.cmpb.2010.06.001
94. Schvarcz E, Palmer M, Aman J, Horowitz M, Stridsberg M, Berne C (1997) Physiological hyperglycemia slows gastric emptying in normal subjects and patients with insulin-dependent diabetes mellitus. Gastroenterology 113(1):60–66
95. Sorensen JT (1985) A physiologic model of glucose metabolism in man and its use to design and assess improved insulin therapies for diabetes. PhD thesis. Massachusetts Institute of Technology
96. Sparacino G, Zanderigo F, Corazza S, Maran A, Fachinetti A, Cobelli C (2007) Glucose concentration can be predicted ahead in time from continuous glucose monitoring sensor time-series. IEEE Trans Biomed Eng 54(5):931–937
97. Ståhl F (2003) Diabetes mellitus modelling based on blood glucose measurements. Master Thesis TFRT-5703, Department of Automatic Control, Lund University, Sweden
98. Ståhl F (2012) Diabetes mellitus glucose prediction by linear and Bayesian ensemble modeling. Licentiate Thesis TFRT–3255, Department of Automatic Control, Lund University, Sweden (2012)
99. Ståhl F, Johansson R (2009) Diabetes mellitus modeling and short-term prediction based on blood glucose measurements. Math Biosci 217:101–117
100. Takagi T, Sugeno M (1985) Fuzzy identification of system and its applications to modelling and control. IEEE Trans Syst Man Cybern SMC-15:116–132
101. Vaddiraju S, Burgess DJ, Tomazos I, Jain FC, Papadimitrakopoulos F (2010) Technologies for continuous glucose monitoring: current problems and future promises. J Diabetes Sci Technol 4(6):1540–1562
102. Van den Berghe G, Wouters P, Weekers F, Verwaest C, Bruyninckx F, Schetz M (2001) Vlas-selaers D, Ferdinande P, Lauwers P, Bouillon R (2001) Intensive insulin therapy in critically ill patients. N Engl J Med 345(19):1359–1367
103. Wilinska ME, Chassin LJ, Acerini CL, Allen JM, Dunger DB, Hovorka R (2010) Simulation environment to evaluate closed-loop insulin delivery systems in type 1 diabetes. J Diab Sci Technol 4(1):132–144
104. Wilinska ME, Chassin LJ, Schaller HC, Schaupp L, Pieber TR, Hovorka R (2005) Insulin kinetics in type-1 diabetes: continuous and bolus delivery of rapid acting insulin. IEEE Trans Biomed Eng 52(1):3–12
105. Worthington DRL (1997) Minimal model of food absorbtion in the gut. Med Inform 22(1):35–45
106. Zecchin C, Facchinetti A, Sparacino G, De Nicolao G, Cobelli C (2011) A new neural network approach for short-term glucose prediction using continuous glucose monitoring time-series and meal information. 2011 annual international conference of the IEEE engineering in medical and biology society (EMBC2011), pp 5653–5656. doi:10.1109/IEMBS.2011.6091368

Hypoglycemia Prevention Using Low Glucose Suspend Systems

B. Wayne Bequette

Abstract The fear of nocturnal hypoglycemia was one of the main motivations behind the development of continuous glucose monitors (CGM) and hypoglycemic alarms. Since many individuals do not awake to these alarms, the next step is to implement a low glucose suspend (LGS) algorithm that shuts off the insulin infusion pump to avoid hypoglycemia. A threshold-based LGS simply shuts off the pump when the CGM signal has violated a lower limit on glucose. A predictive LGS shuts off the pump when a glucose threshold is predicted to occur within a specified time limit, or horizon. A review of algorithms hypoglycemic predictors/alarms is followed by a presentation of threshold and predictive LGS algorithms. Finally, an overview of the implementation challenges that remain is provided.

1 Motivation and Background

One of the greatest concerns of the parent of a child with type 1 diabetes is overnight hypoglycemia. During the daytime it is reasonably easy to keep track of when a child's blood glucose may be going low, but using the current state of care there is no way to keep track of blood glucose overnight. Parents live in fear of the so-called "dead in bed" syndrome, where a person is in a severe state of hypoglycemia for an extended period of time, resulting in a diabetic coma and death; among patients with type 1 diabetes, there is a 6 % lifetime risk of "dead in bed" [41]. One of the main motivations behind continuous glucose monitoring (CGM) technology was to provide real-time alarms warning of overnight hypoglycemia. Buckingham et al. [10] present reports of four subjects, wearing CGM devices, that had low blood glucose (sensor values <60 mg/dL) for 2.25–4 h before a seizure

B. W. Bequette (✉)
Department of Chemical and Biological Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180-3590, USA
e-mail: bequette@rpi.edu

V. Marmarelis and G. Mitsis (eds.), *Data-driven Modeling for Diabetes*, Lecture Notes in Bioengineering, DOI: 10.1007/978-3-642-54464-4_3,

occurred; a CGM-based alarm could possibly wake the individual or their caregiver with adequate time to treat the hypoglycemia and avoid the seizure. Unfortunately, people do not adequately respond to alarms and early devices had a high false alarm rate [9].

An alternative to overnight hypoglycemic alarms is a low glucose suspend (LGS) or pump shut-off (PSO) algorithm that automatically shuts-off the pump to reduce the risk of extended periods of hypoglycemia. These devices can be either be *reactive* and shut-off the pump once a glycemic threshold is violated, or *predictive*, shutting off the pump when it is predicted that a hypoglycemic threshold will be violated within the near future (often 15–70 min). Note that either system is an example of a feedback or closed-loop system, since a sensor signal is being used to take a control action. In this example, the closed-loop algorithm is an on-off algorithm, with the pump simply in one of two-states: either on or off. In general, the application of feedback control to glucose regulation in individuals with type 1 diabetes is referred to as a closed-loop artificial pancreas. For an overview of recent developments towards a closed-loop artificial pancreas, see Bequette [6].

The Juvenile Diabetes Research Foundation (JDRF) has outlined a natural pathway towards an artificial pancreas, as shown in Fig. 1 [33]; each step requires a more advanced system with a higher level of knowledge than the previous step. A LGS system represents steps 1 or 2 (depending on whether it is reactive or predictive) on this six-step JDRF pathway.

In this chapter we first review algorithms used for hypoglycemia prediction and alarms, before covering low glucose suspend algorithms. In addition, we provide an overview of the challenges related to implementing closed-loop control in general, and LGS in specific. Finally, we provide an overview of current efforts to also reduce hyperglycemia. Our focus is on overnight blood glucose control, as a number of other chapters further develop fully closed-loop algorithms that can also be used during the daytime.

2 Hypoglycemia Prevention Alarms

Glucose values from initial commercial versions of some CGMs could only be analyzed retrospectively; the CGM would be downloaded by a clinician, who would then review the results with the patient. Current devices present glucose values in real-time, and can be set to alarm when the values are low (hypoglycemic) or high (hyperglycemic). An example of the use of threshold alarms set at 80 (low) and 200 mg/dL (high) are presented by Davey et al. [20]; hypoglycemia was defined as blood glucose values less than 65 mg/dL.

There has been a significant effort to develop predictive alarms, to enable individuals to take corrective action and avoid low (or high) glucose values. Bequette [5] provides a review of hypoglycemia prevention algorithms, which are concisely summarized in this section.

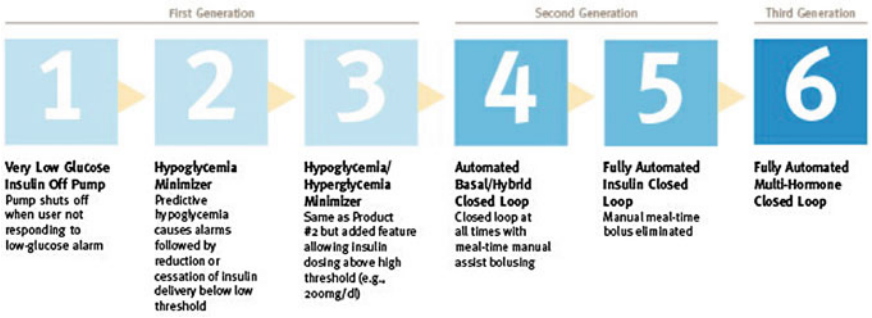


Fig. 1 Six steps on the pathway towards a closed-loop artificial pancreas [33]

A number of methods can be used to predict future glucose levels. It is undesirable to simply extrapolate linearly from the two most recent CGM measurements, e.g. because of sensor noise. The most common methods include using an autoregressive model and Kalman filtering using a state space model.

2.1 Auto-Regressive

An autoregressive (AR) model has the form

$$y_k = \sum_{i=1}^{n} a_i y_{k-i} + w_k \tag{1}$$

where y_k is the glucose value at time step k, w_k is a white noise sequence, and there are n coefficients, a_i (i.e. the model is n-th order). Bremer and Gough [7] apply autoregressive models to blood glucose data available at 10 min sample times, and compare predictions for 10, 20 and 30 min ahead. Reifman et al. [40] use tenth-order AR model and compare the performance of a single, cross-subject model with individual models, for 30-min prediction horizons. Gani et al. [26] use a smoothing procedure and regularization to minimize changes in the glucose first-derivatives. This approach, with a 30-th order model, is shown to reduce the prediction lag compared to Reifman et al. [40]. An adaptive first-order model, with the parameter updated at each time step, is used by Sparacino et al. [42]; results are presented for a 3-min sample time, with a 30-min (10 step) prediction horizon.

An auto-regressive, moving average (ARMA) model has a more general form than AR, with n a_i coefficients and m c_i coefficients

$$y_k = \sum_{i=1}^{n} a_i y_{k-i} + \sum_{i=0}^{m-1} c_i w_{k-i} \tag{2}$$

A tutorial overview of algorithms for CGM and glucose prediction for hypo- and hyperglycemica alarms is presented by Sparacino et al. [43]. They note that a limitation to the approach of Reifman et al. [40] is the substantial amount of training data required for estimation of a large number of parameters in a fixed-parameter formulation. Gani et al. [26] show, however, that a model based on data from one individual could be successfully used on other individuals.

Eren-Oruklu et al. [22, 23] include real-time adaptation of parameters, and find that the best model is second-order, with the form

$$y_k = a_1 y_{k-1} + a_2 y_{k-1} + w_k + c_1 w_{k-1} \tag{3}$$

where three parameters must be estimated. Sparacino et al. [42] use a first-order model

$$y_k = a_1 y_{k-1} + w_k \tag{4}$$

where only one parameter must be estimated. In these studies the model coefficients are estimated recursively, using weighted least squares, where data further in the past has less of an impact than more recent data.

2.2 Kalman Filtering

The Kalman filter is an optimal estimation technique that trades-off the probability that a change in a measured output is due to sensor error versus the probability that it is due to a real input change. For more background on Kalman filtering and optimal estimation, see Stengel [44].

The underlying discrete-time model used in a Kalman filter has the form

$$\begin{aligned} x_{k+1} &= \Phi x_k + \Gamma^w w_k \\ y_k &= C x_k + v_k \end{aligned} \tag{5}$$

where x is a vector of *states* and y is the measured *output*, w_k is the *process noise* (covariance Q), and v_k is the *measurement noise* (covariance R). If it is assumed that the process noise drives the first derivative of glucose with time, then the following relationships result

$$\begin{aligned} g_{k+1} &= g_k + d_k \\ d_{k+1} &= d_k + w_k \\ y_k &= g_k + v_k \end{aligned} \tag{6}$$

where g and d represent the glucose and the change in glucose from step-to-step, respectively. The state space model corresponding to Eq. (6) is

$$\begin{aligned}\underbrace{\begin{bmatrix} g \\ d \end{bmatrix}_{k+1}}_{x_{k+1}} &= \underbrace{\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}}_{\Phi} \underbrace{\begin{bmatrix} g \\ d \end{bmatrix}_{k}}_{x_k} + \underbrace{\begin{bmatrix} 0 \\ 1 \end{bmatrix}}_{\Gamma^w} w_k \\ y_k &= \underbrace{\begin{bmatrix} 1 & 0 \end{bmatrix}}_{C} \underbrace{\begin{bmatrix} g \\ d \end{bmatrix}_{k}}_{x_k} + v_k \end{aligned} \tag{7}$$

The process and measurement noise are considered stochastic processes, and the process noise covariance (Q) is only approximately known and often used as a tuning parameter. An implicit assumption out this 2-state model is that glucose tends to change at a constant rate for significant periods of time. This is equivalent to assuming that an automobile tends to maintain a constant velocity for periods of time, subject to minor perturbations in the velocity. Note that this assumption also holds for a case of glucose concentration being held relatively constant, which means that the rate-of-change is constant at zero.

The states are estimated using the following equations

Predictor (time update):

$$\hat{x}_{k|k-1} = \Phi \hat{x}_{k-1|k-1} \quad \text{predictor (time update)} \tag{8}$$

State estimate covariance:

$$P_k = \Phi P_{k-1} \Phi^T + \Gamma^w Q \Gamma^{wT} - \Phi P_{k-1} C^T \left(C P_{k-1} C^T + R \right)^{-1} C P_{k-1} \Phi^T \tag{9}$$

where the first two terms on the right-hand-side of the equals sign represent the propagation of the state covariance and the process noise, while the third term represents a correction due to the measurement update.

Kalman Gain

$$L_k = P_k C^T \left(C P_k C^T + R \right)^{-1} \tag{10}$$

Corrector (measurement update)

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + L_k \left(y_k - C \hat{x}_{k|k-1} \right) \tag{11}$$

where $\hat{x}$ represents an estimate of the states, and the subscript $k|k-1$ means the estimate at step k is based on measurements up (and including) step $k-1$. Note that a model is used to propagate the state estimate from the previous time step $(k-1)$ to the current time step (k). The measurement at the current time step is then used to update the state estimate, based on the Kalman Gain (L_k).

Future glucose predictions from the most recent measurement at time step k, to step k + j, are given by

$$\hat{x}_{k+j|k} = \Phi^j \hat{x}_{k|k} \tag{12}$$

For the two-state model in Eq. (7), this is equivalent to assuming that the rate-of-change of glucose is constant in the future. That is, for the jth future time step

$$\hat{g}_{k+j|k} = \hat{g}_{k|k} + j\hat{d}_{k|k} \tag{13}$$

where $\hat{d}_{k|k}$ is the estimated change in glucose from step $k-1$ to step k (the current measurement). The uncertainty in the future grows as [Eq. (9) without the measurement feedback term]

$$P_k = \Phi P_{k-1} \Phi^T + \Gamma^w Q \Gamma^{wT} \tag{14}$$

For this problem, the confidence interval grows with each step that is not followed by a measurement update, since there are two "integrators" in the glucose model; this behavior will be demonstrated in the simulations that follow.

Here we show how Kalman filtering is used to obtain glucose and rate-of-change of glucose, in real-time, despite substantial measurement noise. The state vector estimate is initialized with the measured glucose at t = −15 min, with an assumed rate-of-change of 0 mg/dL/min. Figure 2 displays the real-time estimates of glucose (top) and its rate-of-change (bottom), along with the uncertainty bounds of each. Although the actual rate-of-change was −2 mg/dL/min, the Kalman filter converges to this value within 15 min. Figure 3 shows the predicted glucose values based on the measurements available until t = 0 min. Because of the uncertainty about future rates-of-change, the confidence intervals of the glucose predictions increase each step into the future; simulation details are provided in Bequette [5].

Facchinetti et al. [25] have presented a two-state Kalman filter, which Bequette [5] has be shown to be identical to the 2-state model shown in Eq. (7).

The Kalman Filter estimates shown in Figs. 2 and 3 are based on a second-order (2-state) model. On the other hand, if it is assumed that the process noise drives the second derivative of glucose with time, the following third-order (3-state) model can be used:

$$\Phi = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \quad \Gamma^w = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad C = [1 \quad 0 \quad 0], \quad x = \begin{bmatrix} g \\ d \\ f \end{bmatrix}, \quad y = G \tag{15}$$

where g, d, and f, represent the glucose, rate-of-change of glucose, and the second-derivative of glucose with respect to time, respectively. In contrast to the two-state model in Eq. (7), an implicit assumption of this three-state model is that glucose rate-of-change tends to change at a constant rate for a period of time. This is equivalent to assuming that an automobile tends to maintain a constant acceleration for periods of time, subject to minor perturbations in the velocity. Naturally, this assumption will only hold for brief periods of time, such as when accelerating from, or braking before, a traffic light. The advantage of the 3-state model is that it captures dynamics near the maximum (peak) and minimum (valley) values of glucose. While the 3-state model yields better estimates for previous and current

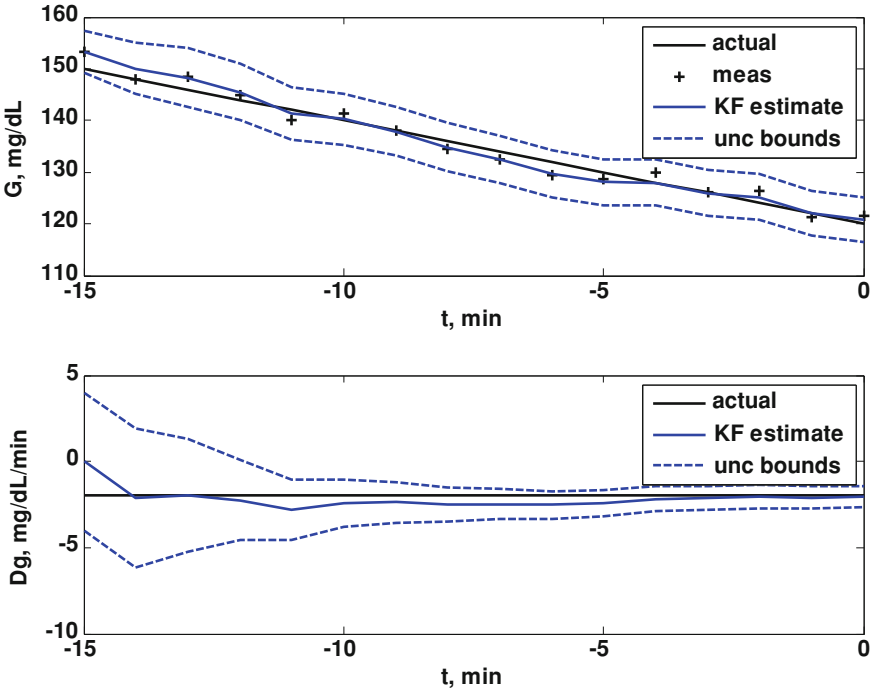


Fig. 2 Glucose (*top*) and rate-of-change of glucose (*bottom*). Actual (*black*), Kalman filter estimate (*blue*) and uncertainty bounds (–). The Kalman filter was initialized at t = −15 min, with uncertainty in the glucose and rate-of-change states. The estimated error variance improves with the measurement updates. Q = 0.01, R = 4. From Bequette [5]

glucose estimates, Palerm and Bequette [37] found that the assumption of a constant first-derivative (2-state) for the future predictions led to better performance for multistep-ahead predictions on clinical data, than assuming a constant second-derivative (3-state).

The Kalman filter equations can also be formulated in continuous-time, with discrete measurement updates [34]. A major advantage of this formulation is that non-constant sample times can be used, as well as nonlinear models.

2.3 Combinations and Other Methods for Glucose Prediction

Dassau et al. [19] incorporate (i) statistical prediction (ii) numerical logical (iii) hybrid infinite impulse response (HIIR) filter (iv) Kalman filter, and (v) linear prediction into a voting-based strategy to predict hypoglycemia. Retrospective testing on 18 CGM datasets (with a 1-min sensor sample time) was performed for different prediction horizons (35, 45 and 55 min) and alarm thresholds (70, 80 and

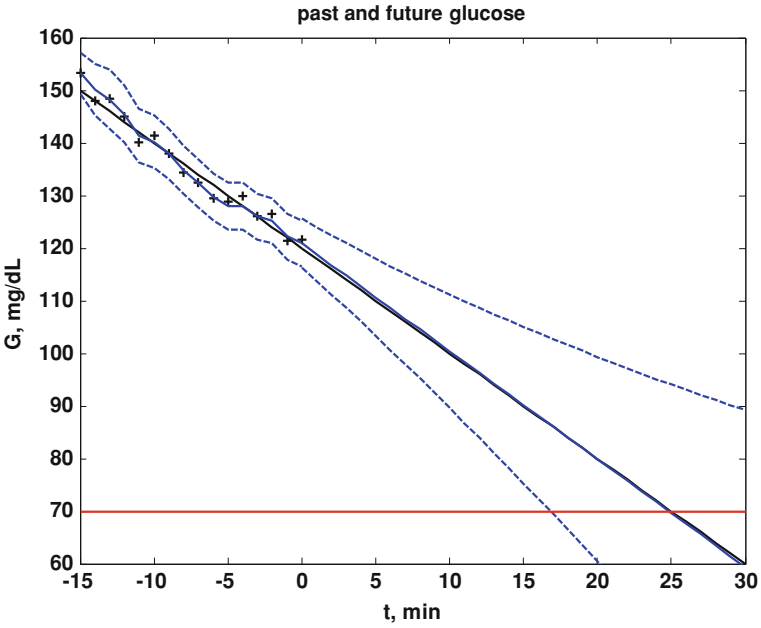


Fig. 3 Actual glucose (*black*), measured (+), estimated (*blue*) and uncertainty bounds (–). The Kalman filter was initialized at t = −15 min. After t = 0 the estimated error variance grows since there are no measurements to improve the estimates. From Bequette [5]

90 mg/dL). A 35 min horizon with an alarm threshold of 80 mg/dL, and a voting threshold of 3 out of 5 votes, resulted in a prediction of 91 % of the hypoglycemic events.

Harvey et al. [29] present a glucose prediction module as part of a health monitoring system (HMS). First, a spike filter is used to reduce the effect of noise spikes, then a low-pass, first-order filter is applied. The first-derivative of Lagrange interpolation polynomial is then used to predict the glucose concentration 15 min into the future. This method resulted in a 93.5 % detection rate with 2.9 false alarms/day, in retrospective analysis of 393 days of clinical data.

Bayrak et al. [2] develop an autoregressive partial least squares algorithm to model CGM data and predict future glucose concentrations. The models are updated recursively using a moving window at each time step, where the mean and variance are recalculated at each interval. The method is applied retrospectively to data from 17 subjects, as well as in simulation studies using the UVa-Padova simulator. A sensitivity of 86 %, with a false alarm rate of 0.42/day, was obtained based on a 30-min prediction horizon. The average detection time was 25.25 min before the hypoglycemic event.

The methods discussed thus far are based on CGM signals. Other information, such as insulin infusion rate (or boluses), meal glucose amount, or activity can also be used to improve glucose predictions. Zhao et al. [47] use a latent variable (LV)

based method to predict future CGM values from past glucose, and known meal glucose and insulin boluses, and apply the technique to simulator (10 subjects) and clinical data (seven subjects). They find that the LV-based method has a glucose prediction accuracy at least as good as methods based on AR and ARX models.

Additional measurements, such as those available from accelerometry, can be used to improve glucose predictions. Eren-Oruklu et al. [24] use a multi-sensor body monitor that provided seven signals related to activity and emotional conditions; they found that five of the seven signals provided a significant contribution to glucose changes. Improved predictive performance is obtained using an ARMAX model with the multiple inputs, compared to an ARMA model based only on glucose signals from the CGM.

3 Threshold-Based Low Glucose Suspend Algorithms

Hypoglycemic alarms, based on continuous glucose monitoring (CGM) technology have had limited success, because of a relatively high false alarm rate and the reality that many patients (and caregivers) sleep through the alarms; Buckingham et al. [9] found that 71 % of alarms were not responded to during sleep. In a clinical trial with 176 subjects for 36,467 nights using CGM, there was an 8.5 % incidence of nocturnal hypoglycemia, with a mean duration of 81 min [32]; hypoglycemia was defined as two consecutive CGM readings $\leq$60 mg/dL in 20 min. A pump shut-off or low glucose suspend (LGS) system, where an insulin pump is shut-off in response to hypoglycemia, is an intermediate step between a hypoglycemic alarm system and a fully closed-loop artificial pancreas. In this section we present simple threshold-based low glucose suspend systems, while the next section presents predictive low glucose suspend methods.

A threshold-based low glucose suspend system simply shuts off the insulin pump when the CGM goes below a desired threshold. Once the pump is turned off, different conditions could be used to turn the pump on. The method implemented in the Medtronic Veo system is to first alarm if a low threshold (set by the user) is violated; if there is no user response, then the pump is shut-off for 2 h (unless the user intervenes to resume basal insulin delivery). Choudhary et al. [17] present results for 31 adults tested for 3 week; the threshold to initiate LGS was individualized (ranging from 40 to 63, with a median of 43 mg/dL). Agrawal et al. [1] studied 7 months of data from 935 patients, with the LGS threshold commonly set between 50 and 60 mg/dL; most events were either <5 min or greater than 115 min. The shorter events tended to occur in the early afternoon, while the longer events were in the late-night and early-morning hours. Danne et al. [18] studied 21 patients, with LGS enabled for 6 weeks, and found that 43 and 57 % of LGS activations occurred during the night (10:00 p.m.–6:00 a.m.) and day (6:00 a.m.–10 p.m.), respectively. Garg et al. [27] studied 50 subjects, who fasted overnight and exercised until their blood glucose reached $\leq$85 mg/dL. The mean time for hypoglycemia (<70 mg/dL) with the LGS-on was reduced from 170.5 to 138.5 min

compared to LGS-off. Ly et al. [36] studied 24 patients and found that 25 % of participants had one full 2-h suspension overnight at least once per week. There were a total of 3128 LGS events during 2493 days, with a mean duration time of 11.2 min, and with 36 % of events occurring overnight.

It should be noted that algorithms other than on-off can be used to prevent/reduce hypoglycemia. For examples, Cengiz et al. [16] and Elleri et al. [21] show that PID and MPC algorithms, respectively, can result in pump suspension. A nice summary of early LGS results is presented by Pickup [39].

4 Predictive Low Glucose Suspend Algorithms

Because of subcutaneous insulin pharmacodynamics and a time-lag between blood glucose and the sensor signal, there is a limit to the performance of a threshold-based system, where glucose will continue to drop past the threshold. It is desirable, then, to use a prediction of blood glucose to turn-off the pump before a hypoglycemic threshold is violated, i.e. predictive low glucose suspend (PLGS). Zecchin et al. [46] have demonstrated the improved performance of predictive versus threshold low glucose suspend methods in simulation studies.

Buckingham et al. [11] report clinical pump shut-off results of 22 subjects using a linear prediction algorithm and a 90-min pump suspension; they found that a 45-min horizon and a 80 mg/dL threshold prevented hypoglycemia 80 % of the time. This approach was extended by Buckingham et al. [12], based on the hypoglycemia prediction strategy by Dassau et al. [19] discussed in Sect. 2.3, to incorporate (i) statistical prediction (ii) numerical logical (iii) hybrid infinite impulse response (HIIR) filter (iv) Kalman filter, and (v) linear prediction into a voting-based strategy that prevented 84 % of possible hypoglycemic events, in a clinical study involving 40 subjects. Initial studies required that three of the five algorithms agreed to suspend the pump, but better performance was achieved when the scheme was changed to only require two algorithms to suspend the pump.

The voting based strategy was based on a fixed sensor sample time of 1 min. To develop a more general framework, where the sensor sample time could be variable, the pump shut-off strategy developed by Cameron et al. [14] was based solely on predictions from a two-state Kalman filter.

In the study by Cameron et al. [14], the initial values used for states $\hat{x}_0$, and covariance, P_0, used in Eqs. (8–11) were:

$$\hat{x}_0 = \begin{bmatrix} 140 \\ 0 \end{bmatrix} \text{ or } \hat{x}_0 = \begin{bmatrix} CGM(0) \\ 0 \end{bmatrix}, \; P_0 = \begin{bmatrix} 36000 & 0 \\ 0 & 1000 \end{bmatrix} \tag{16}$$

where $CGM(0)$ refers to the first CGM reading, assuming it is available. The initial values of P_0 are largely chosen to encourage adherence to the first available glucose values.

Since q and R are scalars for this system model, q/R is a scalar tuning parameter that adjusts the response of the estimates to a new sensor value. In these studies we used a q/R ratio of 0.001 based on retrospective data and reflecting slowly changing nocturnal glucose levels. They additionally provide a steady-state glucose uncertainty of 10 mg/dL, ensuring somewhat reasonable uncertainty estimates. Future work may include adapting this parameter to match the noise characteristics of the sensor [25].

For brief periods where CGM measurements are not available, e.g. between regular measurements, the state and covariance estimates are propagated as in Eqs. (12) and (14). In the studies by Cameron et al. [14], estimates were propagated at 1-min intervals and sensor signals were available at 5-min intervals.

The pump shut-off algorithm consisted of a set of prioritized rules to assure patient safety. The pump is allowed to be off for a maximum of 120 min every 150 min. In addition, the pump can only be off for a maximum of 180 min each night. A 70-min prediction horizon and an 80 mg/dL threshold were used to shut-off the pump. The pump is turned back on when the 70-min prediction of the blood glucose rises above 100 mg/dL. The algorithm provides glucose (and rate-of-change) predictions at 1-min intervals, which are updated when a new sensor signal is available (typically at 5 min intervals). Indeed, the glucose predictions continue during sensor signal dropouts. After 20 min of signal dropout the predictions are gradually reset by assuming a 140 mg/dL sensor value in Eq. (11) and multiplying R by a large scalar (1000) in Eq. (9).

Initial testing of the algorithm was conducted by retrospectively using data from the studies by Buckingham et al. [12]; details are reported in the appendix of Cameron et al. [14]. Clinical studies were then conducted using 16 subjects. On the night of the CTRC admission visit, each participant ate dinner at 6:00 p.m. and checked into the clinic at 7:00 p.m.. The physician initiated a systematic increase in the basal insulin beginning at 9:00 p.m. to promote negative rate-of-change in the blood glucose; this same protocol was used in Buckingham et al. [12]. The study continued until 7:00 or 8:00 a.m., with the subjects in a recumbent position. Reference blood glucose samples were taken every 15–30 min, with blood ketone measurements at the study onset, after pump suspension and before breakfast. The glucose values of one subject remained elevated throughout the study, so data analysis was only performed on 15 subjects. Hypoglycemia was prevented in 11/15 subjects, that is, four subjects had reference glucose values below 60 mg/dL. It should be noted, however, that only one subject had CGM readings below 60 mg/dL, indicating that calibration and CGM accuracy is a problem. A successful example is shown in Fig. 4. Detailed metrics are reported by Cameron et al. [14].

Based on the clinical results, out-patient studies were initiated and results for 375 nights were presented by Buckingham et al. [13]. Roughly 1/3 (123) were "control" nights (PLGS algorithm not activated), and 2/3 (252) were "intervention" nights (PLGS algorithm activated). The prediction horizon 70 min for initial studies, revised to 50 min for the next set, and finally adjusted to 30 min for the final studies. The final tuning resulted in a reduction in nocturnal hypoglycemia of almost 50 %.

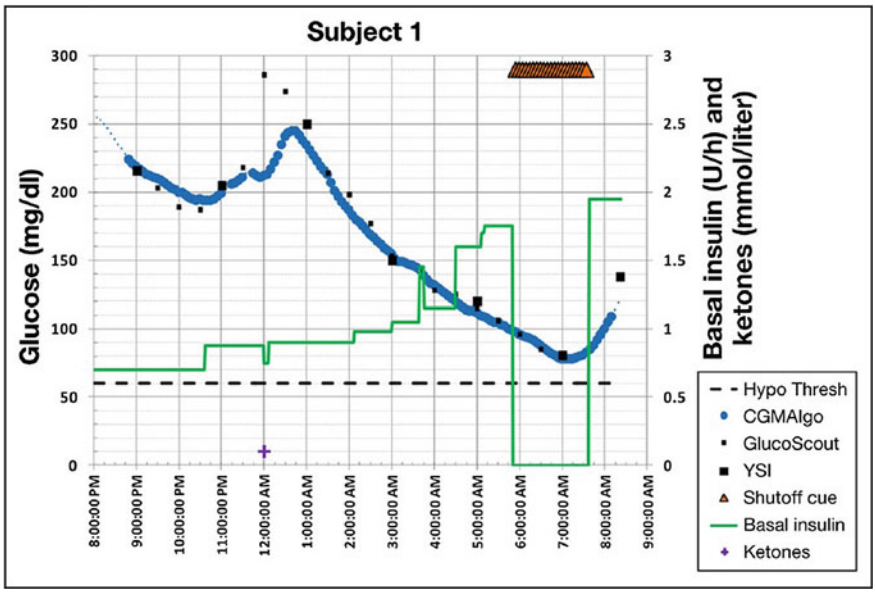


Fig. 4 Results from an in-patient study of a predictive low glucose suspend (PLGS) algorithm. Initially, the basal insulin rate is raised in steps, through 5:00 a.m., to force that subject's CGM to decrease. Then, the PLGS algorithm is switched on. The pump shut-off occurs in time to avoid hypoglycemia. From Cameron et al. [14]

The predictive low glucose suspend studies reviewed above make use of a minimum amount of information—simply real-time CGM signals. The use of additional information, such as insulin-on-board can improve glucose predictions and thus enhance low glucose suspend system performance. Hughes et al. [30] include insulin infusion information to improve hypoglycemia prediction, and incorporate an attenuation of insulin that does not necessarily fully shut-off the pump. Hughes-Karvetski et al. [31] extend this approach by using historical information to provide a time-dependent hypoglycemic risk, which enables a more aggressive insulin attenuation that is patient-specific.

5 Related Challenges

The current state of CGM technology creates a number of challenges when implementing hypo-hyperglycemic alarms and low glucose suspend systems. Sometimes the signals are not available, either due to a communication disruption or a error detected by the sensor that does not allow transmission of the signal. In these cases it is desirable to provide a prediction based on recently available data, using, for example, a Kalman filter without the measurement update.

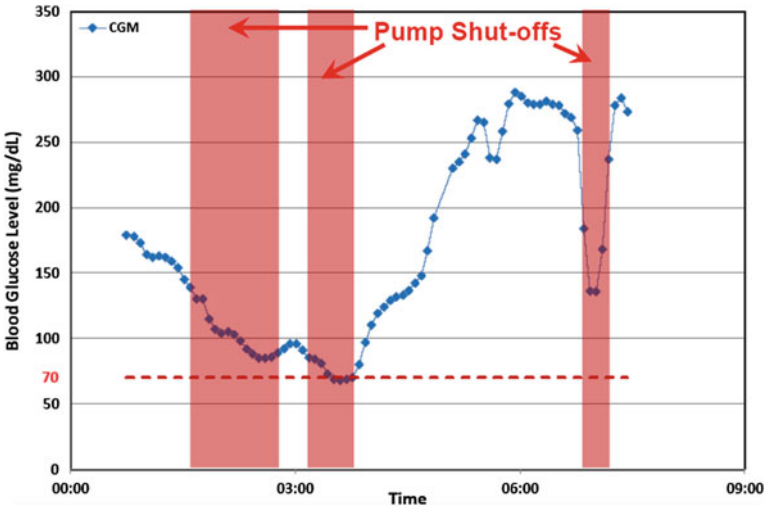


Fig. 5 Example of a PISA that caused an unnecessary pump shut-off at around 7:00 a.m. [3, 4]

A common artifact, particularly overnight, is when an individual rolls over on their sensor, causing the signal to attenuate. We refer to this phenomenon as a pressure-induced sensor attenuation (PISA). An example in the context of a predictive low glucose suspend system is shown in Fig. 5. While the first two pump shut-offs are clearly desired, the third is obviously due to a PISA and is an undesirable shut-off. In this case it is desirable to use a detection algorithm to declare the CGM signal to be invalid, say based on the non-physiologic rate-of-change, and to not allow the pump to be turned off; such a PISA detection algorithm has been presented by Baysal et al. [3, 4]. Characteristic PISA behavior includes a sudden non-physiologic rate-of-change, at least 15 min of signal degradation, followed by a CGM increase. The CGM signal can be considered valid again when the second derivative of the sensor signal begins to decrease.

Another anomaly, which can cause problems in open-loop, LGS, or fully closed-loop, is infusion set failure, which can occur due to blocked or dislodged sets, inflammation, or leakage back to the skin surface. Techniques for infusion set failure detection are presented by Baysal et al. [3, 4].

6 Hypo- and Hyperglycemia Mitigation

An obvious extension of a low glucose suspend system is to combine it with hyperglycemia mitigation, as in step three on the pathway shown in Fig. 1. If a hyperglycemic threshold is violated, or predicted to be violated, the insulin infusion rate can be increased from the current basal delivery rate. This approach is

more commonly called a "control-to-range" strategy. Kovatchev et al. [35] and Patek et al. [38] present the modular architecture for this approach, while Breton et al. [8] present clinical results for 38 subjects using two different incarnations of the strategy. A similar strategy using zone model predictive control, is presented by Grosman et al. [28] in simulation studies.

Cameron et al. [15] extend their low glucose suspend algorithm [14], to include a prediction of hyperglycemia. When a hyperglycemic threshold (180 mg/dL, for example) is predicted to be exceeded within a tunable prediction horizon (30 min, for example), the algorithm provides a correction dose of insulin based on the correction factor (which is based on the 1800 rule [45] if unknown) and knowledge of the current insulin-on-board.

7 Conclusions

A low glucose suspend system is a natural early step on the route to a fully closed-loop artificial pancreas. A threshold-based system suspends an insulin pump once the threshold is violated, while a predictive system suspends the pump in advance. Different conditions to turn the pump back on can be used, including simple timing (such as a 2-h period) or a positive rate of change of glucose (either above a pre-defined threshold or past a nadir).

Acknowledgments I have enjoyed many discussions with Bruce Buckingham, Darrell Wilson, Fraser Cameron and Nihat Baysal during the course of our on-going research effort towards the development of a closed-loop artificial pancreas. This work has been supported in part by the Juvenile Diabetes Research Foundation (JDRF), grants 22-2011-647, 22-2009-795 and 22-2007-1801, and NIH 5R01DK085591-03.

References

1. Agrawal P, Welsh JB, Kannard B, Askari S, Yang Q, Kaufman FR (2011) Usage and effectiveness of the low glucose suspend feature of the Medtronic Paradigm Veo insulin pump. J Diabetes Sci Technol 5(5):1137–1141
2. Bayrak ES, Turksoy K, Cinar A, Quinn L, Littlejohn E, Rollins D (2013) Hypoglycemia early alarm systems based on recursive autoregressive partial least squares. J Diabetes Sci Technol 7(1):206–214
3. Baysal N, Cameron F, Buckingham BA, Wilson DM, Chase HP, Bequette BW (2013a) Nocturnal continuous glucose monitor signal attenuation: an outpatient study. J Diabetes Sci Technol 7(1):A9 (Abstract from presentation at the 2012 diabetes technology meeting, Bethesda)
4. Baysal N, Cameron F, Buckingham BA, Wilson DM, Bequette BW (2013b) Detecting sensor and insulin infusion set anomalies in an artificial pancreas. In: Proceedings of the 2013 American control conference, Washington, DC, pp 2935–2939
5. Bequette BW (2010) Continuous glucose monitoring: real-time algorithms for calibration, filtering and alarms. J Diabetes Sci Technol 4(2):404–418

6. Bequette BW (2012) Challenges and progress in the development of a closed-loop artificial pancreas. Annu Rev Control 36:255–266
7. Bremer T, Gough DA (1999) Is blood glucose predictable from previous values? A solicitation for data. Diabetes 48(3):445–451
8. Breton M, Farret A, Bruttomesso D, Anderson S, Magni L, Patek S, Dalla Man C, Place J, DeMartini S, Del Favero S, Toffanin C, Hughes C, Dassau E, Zisser H, Doyle III FJ, de Nicolao G, Avogaro A, Cobelli C, Renard E, Kovatchev on behalf of the International Artificial Pancreas (IAP) Study Group (2012) Fully integrated artificial pancreas in type 1 diabetes: modular closed-loop glucose control maintains near normoglycemia. Diabetes 61(9):2230–2237
9. Buckingham B, Block J, Burdick J, Kalajian A, Kollman C, Choy M, Wilson DM, Chase HP (2005) Diabetes research in childrens network: response to nocturnal alarms using a real-time glucose sensor. Diabetes Technol Ther 7:440–447
10. Buckingham B, Wilson DM, Lecher T, Hanas R, Kaiserman K, Cameron F (2008) Duration of nocturnal hypoglycemia before seizures. Diabetes Care 31(11):2110–2112
11. Buckingham B, Cobry E, Clinton P, Gage V, Caswell K, Kunselman E, Cameron F, Chase HP (2009) Preventing hypoglycemia using predictive alarm algorithms and insulin pump suspension. Diabetes Tech Ther 11(2):93–97
12. Buckingham B, Chase HP, Dassau E, Cobry E, Clinton P, Gage V, Caswell K, Wilkinson J, Cameron F, Lee H, Bequette BW, Doyle FJ III (2010) Prevention of nocturnal hypoglycemia using predictive alarm algorithms and insulin pump suspension. Diabetes Care 33(5):1013–1018
13. Buckingham BA, Cameron F, Calhoun P, Maahs DM, Wilson DM, Chase HP, Bequette BW, Lum J, Sibayan J, Beck RW, Kollman C (2013) Outpatient safety assessment of an in-home predictive low-glucose suspend system with T1 D subjects at elevated risk of nocturnal hypoglycemia. Diabetes Technol Ther 15(8):622–627
14. Cameron F, Wilson DM, Buckingham BA, Arzumanyan H, Benzsi K, Chase HP, Lum J, Maahs DM, Calhoun PM, Bequette BW (2012) In-patient studies of a Kalman filter based predictive pump shut-off algorithm. J Diabetes Sci Technol 6(5):1142–1147
15. Cameron F, Buckingham BA, Wilson DM, Chase HP, Bequette BW (2013) Adding hyperglycemia mitigation to predictive low glucose suspension. In: Presented at the 73rd scientific sessions of the American Diabetes Association, Chicago
16. Cengiz E, Swan KL, Tamborlane WV, Steil GM, Steffen AT, Weinzimer SA (2009) Is an automatic pump suspension feature safe for children with type 1 diabetes An exploratory analysis with a closed-loop system. Diabetes Technol Ther 11(4):207–210
17. Choudhary P, Shin J, Wang Y, Evans ML, Hammond PJ, Kerr D, Shaw JAM, Pickup JC, Amiel SA (2011) Insulin pump therapy with automated insulin suspension in response to hypoglycemia. Reduction in nocturnal hypoglycemia in those at greatest risk. Diabetes Care 34:2023–2025
18. Danne T, Kordonouri O, Holder M, Haberland H, Golembowski S, Remus K, Blasig S, Wadien T, Zierow S, Hartmann R, Thomas A (2011) Prevention of hypoglycemia by using low glucose suspend function in sensor-augmented pump therapy. Diabetes Technol Ther 13(11):1129–1134
19. Dassau E, Cameron F, Lee H, Bequette BW, Zisser H, Jovanovic L, Chase HP, Wilson DM, Buckingham BA, Doyle FJ III (2010) Real-time hypoglycemia prediction suite using continuous glucose monitoring: a safety net for the artificial pancreas. Diabetes Care 33(6):1249–1254
20. Davey RJ, Jones TW, Fournier PA (2010) Effect of short-term use of a continuous glucose monitoring system with a real-time glucose display and a low glucose alarm on incidence and duration of hypoglycemia in a home setting in type 1 diabetes mellitus. J Diabetes Sci Technol 4(6):1457–1464
21. Elleri G, Allen JM, Nodale M, Wilinska ME, Acerini CL, Dunger DB, Hovorka R (2010) Suspended insulin infusion during overnight closed-loop glucose control in children and adolescents with Type 1 diabetes. Diabet Med 27:480–484

22. Eren-Oruklu M, Cinar A, Quinn L, Smith D (2009) Estimation of future glucose concentrations with subject-specific recursive linear models. Diabetes Technol Ther 11(4):243–253
23. Eren-Oruklu M, Cinar A, Quinn L (2010) Hypoglycemia prediction with subject-specific recursive time-series models. J Diabetes Sci Technol 4(1):25–33
24. Eren-Oruklu M, Cinar A, Rollins DK, Quinn L (2012) Adaptive system identification for estimating future glucose concentrations and hypoglycemia alarms. Automatica 48:1892–1897
25. Facchinetti A, Sparacino G, Cobelli C (2010) An on-line self-tuneable method to denoise CGM sensor data. IEEE Tran Biomed Eng 57(3):634–641
26. Gani A, Gribok AV, Rajaraman S, Ward WK, Reifman J (2009) Predicting subcutaneous glucose concentration in humans: data-driven glucose modeling. IEEE Trans Biomed Eng 56(2):246–254
27. Garg S, Brazg R, Bailey TS, Buckingham BA, Slover RH, Klonoff DC et al (2012) Reduction in duration of hypoglycemia by automatic suspension of insulin delivery: the in-clinic ASPIRE study. Diabetes Technol Ther 14(3):205–209
28. Grosman B, Dassau E, Zisser HC, Jovanovic L, Doyle FJ III (2010) Zone model predictive control: a strategy to minimize hyper- and hypoglycemic events. J Diabetes Sci Technol 4:961–975
29. Harvey RA, Dassau E, Zisser H, Seborg DE, Jovanovic L, Doyle FJ III (2012) Design of the health monitoring system for the artificial pancreas: low glucose prediction module. J Diabetes Sci Technol 6(6):1345–1354
30. Hughes CS, Patek SD, Breton MD, Kovatchev BP (2010) Hypoglycemia prevention via pump attenuation and red-yellow-green ''traffic'' lights using continuous glucose monitoring and infusion pump data. J Diabetes Sci Technol 4(5):1146–1155
31. Hughes-Karvetski C, Patek SD, Breton MD, Kovatchev BP (2013) Historical data enhances safety supervision system performance in T1 DM insulin therapy risk management. Comput Methods Programs Biomed 109(2):220–225
32. JDRF CGM Study Group (2010) Prolonged nocturnal hypoglycemia is common during 12 months of continuous glucose monitoring in children and adults with type 1 diabetes. Diabetes Care 33(5):1004–1008
33. JDRF (2013) Artificial pancreas project research. http://jdrf.org/research/treat/artificial-pancreas-project/
34. Knobbe EJ, Buckingham BA (2005) The extended Kalman filter for continuous glucose monitoring. Diabetes Technol Ther 7(1):15–27
35. Kovatchev B, Patek S, Dassau E, Doyle FJ III, Magni L, De Nicolao G, Cobelli C, The Juvenile Diabetes Research Foundation Artificial Pancreas Consortium (2009) Control to range for diabetes: functionality and modular architecture. J Diabetes Sci Technol 3(5):1058–1065
36. Ly TT, Nicholas JA, Retterath A, Davis EA, Jones TW (2012) Analysis of glucose responses to automated insulin suspension with sensor-augmented pump therapy. Diabetes Care 35:1462–1465
37. Palerm CC, Bequette BW (2007) Hypoglycemia detection and prediction using continuous glucose monitoring—a study on hypoglycemic clamp data. J Diabetes Sci Technol 1(5):624–629
38. Patek SD, Magni L, Dassau E, Karvetski CH, Toffanin C, de Nicolao G, del Favero S, Breton M, Dalla Man C, Renard E, Zisser H, Doyle III FJ, Cobelli C, Kovatchev BP (2012) Modular closed-loop control of diabetes. IEEE Trans Biomed Eng 59(11):2986–2999
39. Pickup JC (2011) Semi-closed-loop insulin delivery systems: early experience with low-glucose insulin suspend pumps. Diabetes Technol Ther 13(7):695–698
40. Reifman J, Rajaraman S, Gribok A, Ward WK (2007) Predictive monitoring for improved management of glucose levels. J Diabetes Sci Technol 1(4):478–486
41. Sovik O, Thordarson H (1999) Dead-in-bed syndrome in young diabetic patients. Diabetes Care 22(Suppl 2):B40–B42

42. Sparacino G, Zanderigo F, Corazza S, Maran A, Facchinetti A, Cobelli C (2007) Glucose concentration can be predicted ahead in time from continuous glucose monitoring sensor time series. IEEE Trans Biomed Eng 54(5):931–937
43. Sparacino G, Facchinetti A, Maran A, Cobelli C (2008) Continuous glucose monitoring time series and hypo/hyperglycemia prevention: requirements, methods, open problems. Curr Diab Rev 4(3):181–192
44. Stengel RF (1994) Optimal control and estimation. Dover, New York
45. Walsh J, Roberts R, Bailey T (2010) Guidelines for insulin dosing in continuous subcutaneous insulin infusion using new formulas from a retrospective study of individuals with optimal glucose levels. J Diabetes Sci Technol 4(5):1174–1181
46. Zecchin C, Facchinetti A, Sparacino G, Cobelli C (2013) Reduction of number and duration of hypoglycemic events by glucose prediction methods: a proof-of-concept in silico study. Diabetes Technol Ther 15(1):66–77
47. Zhao C, Dassau E, Jovanovic L, Zisser HC, Doyle JC III, Seborg DE (2012) Predicting subcutaneous glucose concentration using a latent-variable-based statistical method for type 1 diabetes mellitus. J Diabetes Sci Technol 6(3):617–633

Adaptive Algorithms for Personalized Diabetes Treatment

Elena Daskalaki, Peter Diem and Stavroula Mougiakakou

Abstract Dynamic systems, especially in real-life applications, are often determined by inter-/intra-variability, uncertainties and time-varying components. Physiological systems are probably the most representative example in which population variability, vital signal measurement noise and uncertain dynamics render their explicit representation and optimization a rather difficult task. Systems characterized by such challenges often require the use of adaptive algorithmic solutions able to perform an iterative structural and/or parametrical update process towards optimized behavior. Adaptive optimization presents the advantages of (i) individualization through learning of basic system characteristics, (ii) ability to follow time-varying dynamics and (iii) low computational cost. In this chapter, the use of online adaptive algorithms is investigated in two basic research areas related to diabetes management: (i) real-time glucose regulation and (ii) real-time prediction of hypo-/hyperglycemia. The applicability of these methods is illustrated through the design and development of an adaptive glucose control algorithm based on reinforcement learning and optimal control and an adaptive, personalized early-warning system for the recognition and alarm generation against hypo- and hyperglycemic events.

E. Daskalaki · S. Mougiakakou (✉)
Diabetes Technology Research Group, ARTORG Center for Biomedical Engineering Research, University of Bern, Bern, Switzerland
e-mail: stavroula.mougiakakou@artorg.unibe.ch

P. Diem
Department of Endocrinology, Diabetes and Clinical Nutrition, Bern University Hospital "Inselspital", Bern, Switzerland

V. Marmarelis and G. Mitsis (eds.), *Data-driven Modeling for Diabetes*, Lecture Notes in Bioengineering, DOI: 10.1007/978-3-642-54464-4_4,

1 Introduction

Individuals with type 1 diabetes (T1D) are dependent on daily insulin administration in order to maintain their glucose levels and reduce the risk for short-term and long-term complications. The development of an external artificial pancreas (AP) combining a continuous glucose monitor (CGM), a continuous insulin infusion pump and an automatic control system for optimal insulin delivery is to date in the forefront of research towards improved glycemic regulation. A control system to be used within an AP comprises a number of algorithmic components necessary for efficient, prompt and safe glucose regulation: (i) a control algorithm for optimized insulin infusion, (ii) a safety mechanism for prediction, warning and prevention against abnormal metabolic situations and (iii) various fault detection modules for supervision of the well-functioning of all involved devices and alarm generation in case of failure (e.g. wireless communication failures). All AP components base their functionality on continuous access to the individual's glucose concentration, measured by the CGM, and possibly other inputs or vital signals related to his/her daily activities such as meal intake and physical exercise. The design of the algorithmic components of an AP faces the challenges of inter- and intra-patient variability, glucose measurement and insulin action delays, system uncertainties and strict safety constraints. The glucose and insulin infusion profiles are tightly related to patient-specific characteristics (e.g. body mass index, insulin sensitivity), lifestyle (e.g. meals, physical activity) and environmental disturbances (e.g. stress, illness). In the presence of these challenges, adaptive and personalized algorithmic solutions able to learn the individual's dynamics and perform high-level and real-time optimization are necessary towards efficient diabetes management.

This chapter presents the basic characteristics of online adaptive algorithms and their application in the field of diabetes management. Emphasis is given in the research areas in which the Diabetes Technology Research Group (DTRG) of the University of Bern is currently active. These involve the development of adaptive control algorithms for insulin infusion and personalized prediction systems for early recognition and warning against upcoming hypo- and hyperglycemias.

2 Adaptive Algorithms and Diabetes Management

The basic characteristic of adaptive algorithms is their ability to change their structure and/or parameters over time according to the information received from their environment. The primary scope of this change is to gradually improve their performance, usually translated as the minimization of an appropriate cost equation. For this purpose an adaptive algorithm employs a learning process for the structural and/or parametrical update as new information from the environment becomes available.

Development of adaptive algorithms has been initiated by the need to address time-varying processes or systems characterized by uncertain dynamics. In recent years, the increasing computational power of modern computers has permitted research to focus on solutions of complex and computationally expensive optimization problems. This fact led to drastic advancements in the fields of artificial intelligence, machine learning and optimal control which have proved extremely successful in learning and performance optimization mainly based on data and a priori knowledge derived from the underlying processes. A step forward towards extension of these methods to address time-varying problems has led to their combination with online adaptive learning algorithms which are able to carry on the optimization process in real-time. In the biomedical field, where the challenge of complex and time-varying systems is crucial, adaptive algorithms have been used and proved successful in a wide range of applications. Representative examples can be found in:

- **Control**: Adaptive control algorithms have been extensively used in bioscience in order to handle the continuously changing dynamics of physiological processes such as internal organ motion or heart rate regulation [1, 2].
- **Modeling**: Adaptive models have been applied for the representation and/or prediction of non-stationary processes. Important applications can be found in the field of personalized risk assessment [3] and tumor tracking [4].
- **Segmentation**: Real-time segmentation is very crucial in medical image processing and image guided surgery. Adaptive algorithms have been used in order to segment moving elements such as vessels [5] and cells [6] or highly inhomogeneous magnetic resonance images [7].

Adaptive systems are now making their tentative appearance in diabetes management. The high variability of the diabetic population has raised the need for personalized diabetes treatment. Moreover, the intra-patient variability and the high degree of uncertainty that characterizes the glucoregulatory system require continuous learning and advanced optimization solutions. Within this framework, adaptive solutions have been proposed for closed-loop glucose control and glucose prediction in the near future.

Various control approaches have been developed for glucose regulation to be used within an AP system. A few recent studies have presented control algorithms with adaptive components. A Model Predictive Controller (MPC) using a nonlinear model for glucose prediction and adaptive techniques for individualized tuning of its parameters has been developed in [8]. Furthermore, a MPC enhanced with an individualized model and run-to-run tuning has been developed stressing the importance of personalization of the control algorithms in the face of inter- and intra- patient variability [9]. An algorithm for the automatic daily update of insulin basal rate and insulin:carbohydrates (IC) ratio has been also presented based on run-to-run techniques [10, 11]. Finally, a new algorithm based on MD-Logic (MDL) has been proposed enhanced with adaptive components for performance optimization [12].

In glucose prediction very limited studies have presented online adaptive modeling approaches. These include the development of online adaptive autoregressive (AR) models for prediction of glucose profile and specialization in hypoglycemia prediction [13, 14].

The DTRG has been extensively working on the design and development of online adaptive solutions for optimized and personalized diabetes management. The work of the DTRG involves the development of an adaptive control algorithm for glucose regulation based on reinforcement learning and optimal control and more specifically on the Actor-Critic (AC) learning approach [15]. In the field of glucose prediction, the current research efforts of the DTRG have resulted in development of online adaptive and personalized AR and artificial neural network (ANN) models for near future glucose prediction which have been further integrated into an early-warning system for recognition of upcoming hypo- and hyperglycemic events [16, 17].

3 Adaptive Glucose Control Based on Reinforcement Learning

3.1 Reinforcement Learning

RL belongs to the general class of machine learning and involves adaptive agents able to optimize their performance over time through interaction with the environment, which may include partially known or unknown dynamics [18]. AC is a RL-based algorithm which consists of two complementary adaptive agents: the Critic and the Actor, with the former being responsible for the control policy evaluation and the latter for the control policy optimization.

The system dynamics can be modeled as a Markov Decision Process (MDP) with finite state space X and action space U. The transition from the current state x to the next state y given an action u follows a transition probability matrix $p(y|x, u)$. The control policy is a deterministic or stochastic function $\mu(u|x, \theta)$ which maps an action u to a state x based on a policy parameter vector $\theta \in R^K$. A local cost $c(x, u)$ is associated with each state x and action u. The aim of the AC algorithm is to minimize the average expected cost per state defined as

$$\bar{a}(\theta) = \sum_{x \in X, u \in U} c(x,\ u)\eta_\theta(x,\ u) \tag{1}$$

where $\eta_\theta(x, u)$ is the stationary probability associated with the Markov chain $\{X_k, U_k\}$ dependent on θ. In order to proceed to an iterative estimation of the optimal control policy towards minimization of the above cost, a gradient method can be used requiring the estimation of $\nabla_\theta \bar{a}(\theta)$. Define the value-function $V_\theta(x)$ as the long-term expected cost when starting from state x and following control policy $\mu(u|x, \theta)$. Function $V_\theta(x)$ fulfills the following Poisson equation:

$$\bar{a}(\theta) + V_\theta(x) = \sum_u \mu(u|x,\ \theta)\left[c(x,\ u) + \sum_y p(y|x,\ u)V_\theta(y)\right] \tag{2}$$

Similarly, define the action-value function $Q_\theta(x,\ u)$ as the long-term expected cost starting at state x, taking control action u and following control policy $\mu(u|x,\ \theta)$:

$$Q_\theta(x,\ u) = c(x,\ u) - \bar{a}(\theta) + \sum_y p(y|x,\ u)V_\theta(y) \tag{3}$$

It has been proved in [19] that the gradient $\nabla_\theta \bar{a}(\theta)$ can be estimated by the following formula:

$$\nabla_\theta \bar{a}(\theta) = \sum_{x,y} \eta_\theta(x,\ u)Q_\theta(x,\ u)\psi_\theta(x,\ u) \tag{4}$$

where $\psi_\theta(x,\ u) = \nabla \ln[\mu(u|x,\ \theta)]$.

AC implementations may vary in the design of both the Actor and the Critic part. In most cases, the Critic estimates the Q-value function based on a function approximation method while the Actor follows a gradient descent approach for the iterative optimization of the control policy. A comprehensive review of RL and AC algorithms can be found in [20]. A scheme of a system controlled by an AC algorithm is shown in Fig. 1.

Critic: Policy evaluation. The critic evaluates the current control policy through the estimation of the Q-value function $Q_\theta(x,\ u)$. As (3) cannot be analytically solved in most cases, a function approximation method is usually employed in which $Q_\theta(x, u)$ is commonly approximated as a linear parameterized function of the form

$$\tilde{Q}^r_\theta(x, u) = \sum_{\iota=1}^{K} r_\iota \phi^\iota_\theta(x, u) = r^T \phi_\theta \tag{5}$$

where $r \in R^K$, ϕ_θ is a vector of basis functions dependent on the Actor's parameters θ and r^T denotes transpose. The method of Temporal Differences (TD) [21] is used for the update of the Critic's parameter vector r and has the following general form

$$r_{k+1} = r_k + \gamma^c_k d_k \sum_{i=0}^{k} \lambda^{k-i}\phi_\theta(x_i, u_i) \tag{6}$$

where the TD error d_k, defined as the error between consecutive predictions, is given by the following formula

$$d_k = c(x_k, u_k) - \bar{a}_k + r_k^T \phi_{\theta_k}(x_{k+1}, u_{k+1}) - r_k^T \phi_{\theta_k}(x_k, u_k) \tag{7}$$

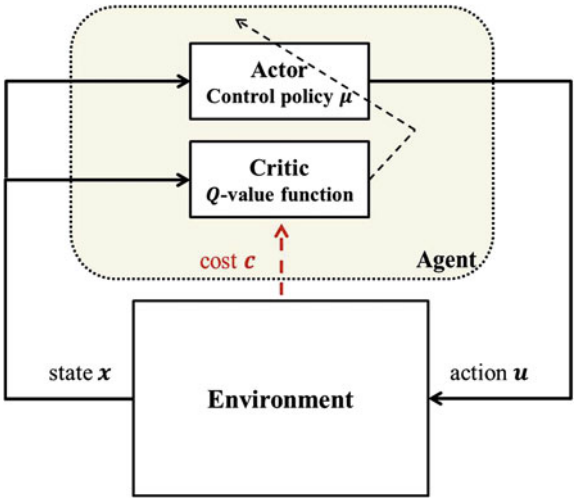


Fig. 1 System controlled by an AC algorithm

γ_k^c is a positive non-increasing function which defines the learning rate of the Critic, $0 < \lambda < 1$ is constant which stands as a forgetting factor. It is suggested in [21] that higher values of λ may result in improved approximations of the value function. A sequence of eligibility vectors z_k is defined as

$$z_k = \sum_{i=0}^{k} \lambda^{k-i} \phi_\theta(x_i, u_i) \tag{8}$$

in order to achieve the iterative form of the algorithm. The update rule for z_k is given by the following formula

$$z_{k+1} = \lambda z_k + \phi_{\theta_k}(x_{k+1}, u_{k+1}) \tag{9}$$

Thus, the final update procedure for Critic is formulated as

$$r_{k+1} = r_k + \gamma_k^c d_k z_k \tag{10}$$

Finally, the average cost is updated as

$$\bar{a}_{k+1} = \bar{a}_k + \gamma_k^c [c(x_{k+1}, u_{k+1}) - \bar{a}_k] \tag{11}$$

Actor: Policy improvement. The aim of the Actor is to minimize the average cost $\overline{a}$ through the iterative update of the control policy parameter vector θ based on the gradient $\nabla_\theta \bar{a}(\theta)$

$$\theta_{k+1} = \theta_k - \gamma_k^a \nabla_\theta \bar{a}(\theta) \tag{12}$$

where γ_k^a is a positive non-increasing function which defines the learning rate of the Actor. Based on (12), (4), and (5), the final update formula of the Actor policy parameter vector becomes

$$\theta_{k+1} = \theta_k - \gamma_k^a r_k^T \phi_{\theta_k}(x_{k+1},\ u_{k+1})\psi_{\theta_k}(x_{k+1},\ u_{k+1}) \tag{13}$$

3.2 AC-based Algorithm for Glucose Regulation

The AC algorithm implements a dual control policy for the optimization of the average daily BR and the IC ratio defined as:

$$IC = {}^{I_{bolus}}/_{CHO} \tag{14}$$

where I_{bolus} is the insulin bolus dose and CHO is the amount of carbohydrates contained in a meal. The Critic estimates the Q-value function based on the TD method as previously described while the Actor updates the two control policies on a daily basis. The dynamics of the glucoregulatory system are represented as a MDP where the state and control action reflect the system's status of one day. Define the glucose error EG as:

$$EG(t) = \begin{cases} G(t) & if\ G > G_h \\ G(t) & if\ G < G_l \\ 0 & else \end{cases} \tag{15}$$

where $G(t)$ is the glucose value at time t and $G_h = 180$ mg/dl, $G_l = 70$ mg/dl are the hyper- and hypoglycemia bounds respectively. The glycemic profile of day k is described by two features related to the hyperglycemic and hypoglycemic status of that day and more specifically to the average daily hypoglycemia and hyperglycemia error:

$$x_k^1 = \frac{1}{N_1} \sum_{t \in day\, k} \mathbf{H}(EG(t)) \tag{16}$$

$$x_k^2 = \frac{1}{N_2} \sum_{t \in day\, k} \mathbf{H}(-EG(t)) \tag{17}$$

where $\mathbf{H}(\cdot)$ is the Heaviside function and N_i is the number of time samples above the hyperglycemia (i = 1) or below the hypoglycemia (i = 2) threshold. First, the features are normalized between [0 1]. The normalized features formulate the state $x_k = [x_k^1\, x_k^2]^T$ of day k which is used by the algorithm for the estimation of the

long-term expected costs and the control policies parameters. The control policy for the average BR and the IC ratio is updated as:

$$\mu(u_k|x_k,\ \theta_k^S) = S_k = S_{k-1} + P_k^S S_{k-1} \tag{18}$$

where $S = \{BR, IC\}$, S_k is the control policy for day k and P_k^S is the rate of change of S_k from day $k - 1$ to day k estimated as a linear combination of the features x:

$$P_k^S = x_k \theta_k^S \tag{19}$$

with θ^S being the policy parameter vector of the respective control policy. The policy parameter vectors θ^S are updated based on (13). A detailed description of the design and implementation of the AC-based algorithm for glucose regulation can be found in [15].

A major challenge during the design of adaptive algorithms, especially for applications where safety matters, is to keep the learning period as short as possible. Furthermore, even during learning, the necessary safety constraints should be guaranteed. One way to achieve this goal in designing an AC-based algorithm is through the appropriate initialization of the policy parameter vectors θ^S, which regulate the optimization of the control policy over time. The parameters θ^S can be viewed intuitively as weights that define the percentage of change of the BR and IC ratio according to the daily hypoglycemia and hyperglycemia status. Setting these parameters away from their optimal values results in longer learning period, which can be crucial for the safety of the patient. One would expect that the percentage of change depends on the amount of information transfer (IT) from insulin to glucose in the sense that for high IT small adaptations of the insulin scheme may be sufficient whereas for low IT larger updates may be needed. Based on this reasoning, the IT from insulin to glucose has been estimated and used for the automatic, patient-specific initialization of θ^S.

3.2.1 Automatic Tuning of the AC-based Algorithm

Assessing causality and IT between signals has been extensively studied and various measures have been proposed. A comprehensive review can be found in [22]. Transfer entropy (TE) is a powerful measure of IT, mainly due to its nonlinear and directional structure [23, 24], and has found promising application in biomedical signal analysis [25]. TE measures the information flow from a signal Y (source) to a signal X (target) while it excludes redundant effects coming from other signals. Let $X = \{x_i, i = 1{:}n\}$, $Y = \{y_i, i = 1{:}n\}$ and $Z = \{z_i, i = 1{:}n\}$ be three observed random processes of length n. TE estimates the IT from process Y to X, which can be also translated as the amount of knowledge we gain about X when we already know Y, based on the following formula:

$$T_{Y\to X} = \sum_i p(x_i, y_i, z_i) \log \frac{p(x_i|y_i, z_i)}{p(x_i|z_i)} \tag{20}$$

where $p(\cdot)$ denotes probability density function (pdf) and *log* is the basis two logarithm. Division with the conditional probability of X to Z excludes the redundant information coming from both Y and Z without excluding, though, the possible synergistic contribution of the two signals on X [24]. Main challenge in computing (9) is the estimation of the involved pdfs. Several approaches have been proposed for this purpose [22]. One of the most commonly used methods is the fixed data partitioning in which the time-series are partitioned into equi-sized bins and the pdfs are approximated as histograms [26].

TE was estimated as the IT from a delayed active insulin value to the current glucose excluding the dependence on the previous glucose value. In this case (20) takes the following form:

$$TE_{IA\to G} = \sum_t p(G_t, G_{t-1}, IA_{t-d}) \log \frac{p(G_t|G_{t-1}, IA_{t-d})}{p(G_t|G_{t-1})} \tag{21}$$

where G_t is the current glucose value and IA is the insulin on board. The delay d in insulin was introduced in order to compensate on the physiological delay in insulin action due to absorption from the subcutaneous tissue to the blood and was chosen as $d = 20$ min. Expecting that high TE is related to smaller rates of change in the insulin scheme, the initial values of the policy parameter vectors θ^S are set to be inversely proportional to the estimated TE per patient as:

$$\theta_0^S(p) = \frac{W}{TE(p)} \tag{22}$$

where p denotes a specific patient and W is constant universally set as $W = +1$ for the elements of θ_0^S related to hyperglycemia and $W = -1$ for elements related to hypoglycemia.

3.3 Results and Discussion

3.3.1 Simulation Environment

The AC-based algorithm has been in silico evaluated on 28 virtual T1D patients (10 adults; 10 adolescents; 8 children) using the educational version of the FDA-accepted University of Virginia (UVa) T1D simulator [27]. Two children have been excluded due to excessive glucose responses. The simulator comprises a database of in silico patients based on the meal simulation model of the glucose-insulin subsystem [28], a simulated sensor that replicates the typical errors of continuous

glucose monitoring, and a simulated insulin pump. It allows for the creation of meal protocols and scenarios of varying insulin sensitivity (SI). Furthermore, the simulator provides optimized values for the BR and IC ratio which can be assumed as the standard SAP treatment of the patients as defined by his/her physician.

3.3.2 Meal Protocol

The meal protocol consisted of four kinds of meals: breakfast, lunch, dinner, and snack. In order to simulate the daily variation in terms of meal quantity and timing, the meal scenario per day was created in a randomized way as follows: each meal kind was assigned a range of possible quantities of grams of CHO and a range of timings as:

- Breakfast: [30–60] g of CHO at [07:00–11:00]
- Lunch: [70–100] g of CHO at [13:00–16:00]
- Dinner: [70–110] g of CHO at [20:00–22:00]
- Snack: [20–40] g of CHO at [23:00–00:00].

The accurate amount of grams of CHO and the timing for each meal and each trial day were chosen randomly from the respective ranges. As a second step, several dinners and snacks were omitted in order to represent the common case of missing these meals. The missed meals were randomly distributed within the trial days. The average duration of each meal was considered 15 min and the meals were announced to the controller 30 min prior to intake.

In order to simulate the errors when real patients estimate the CHO content of their meal, a random meal uncertainty uniformly distributed between -50 and $+50$ % has been introduced. The total trial duration was 10 days. The initial values of BR and IC ratio have been set equal to their optimized values as provided by the UVa simulator. For adults and adolescents, these values are close to the optimal ones, whereas in the case of children they are too high and, when applied in an open-loop scenario, they lead to excessive insulin infusion and frequent hypoglycemic events [15]. Consequently, the AC algorithm must perform significant updates of the BR and IC ratio in order to optimize glucose regulation, a fact that renders the duration of the learning period in children very challenging.

Three different initialization scenarios of the policy parameter vectors θ^S have been investigated:

- **S1**. The policy parameter vectors θ^S are initialized to zero values.
- **S2**. The policy parameter vectors θ^S are initialized to random values with magnitude ranging in (0, 1).
- **S3**. The policy parameter vectors θ^S are initialized based on the estimated insulin-to-glucose transfer entropy per patient as in (22).

The Control Variability Grid Analysis (CVGA) [29] has been used for the evaluation of the AC algorithm, while the risk of hypoglycemia has been estimated based on the Low Blood Glucose Index (LBGI) [30].

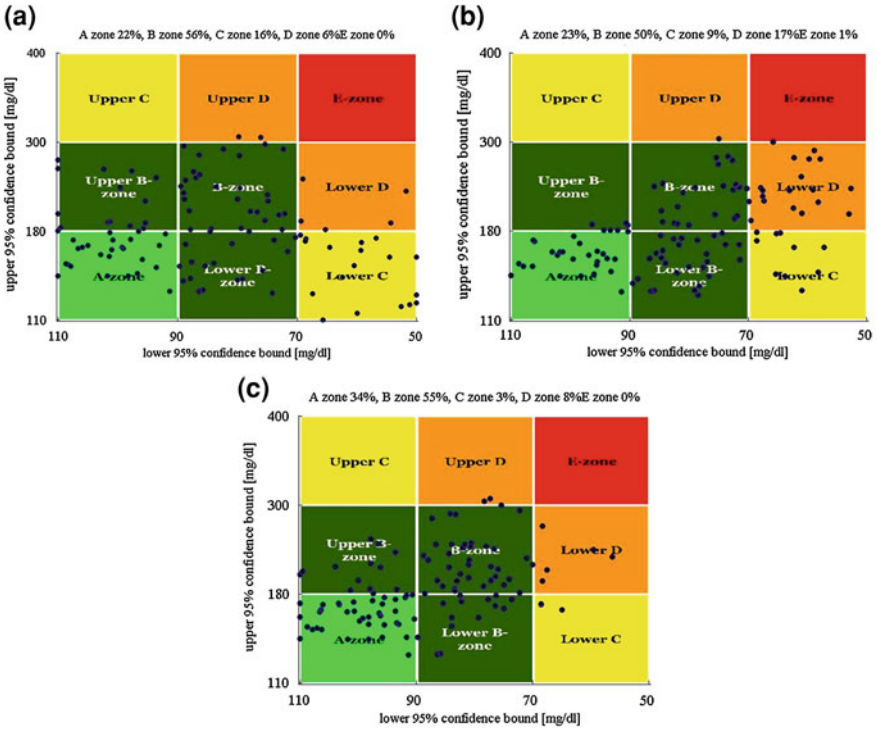


Fig. 2 CVGA plots for all patients and the last five trial days when AC initialization is based on scenario **(a)** S1, **(b)** S2 and **(c)** S3

Table 1 Percentages (%) in the A + B zones of the CVGA for the three age groups and scenarios S1, S2, S3

Patient age group	S1	S2	S3
Adults	93.00	95.00	98.00
Adolescents	88.00	78.00	90.00
Children	41.00	41.00	73.00

Figure 2 presents the CVGA plots for all patients and scenarios S1–S3. Table 1 presents the percentage of values within the A + B zones of the CVGA separately for each age group and the three scenarios. Setting the maximum acceptable duration of the learning period to five days, these results refer to the last five days of the trial. Finally, the daily evolution of the LBGI for the three groups when following scenarios S1–S3 is presented in Fig. 3. This result refers to the total trial duration.

Figure 2 shows that, when the AC initialization is based on the patient-specific TE (S3), the general performance of the algorithm after five days of learning is increased with 89 % of the values in the A + B zones of the CVGA, compared to 78 % for scenario S1 and 73 % for S2. The same result can be observed from Table 1, where the values are separately presented for each patient group. Furthermore, from Table 1 it is clear that most of the hypoglycemic events present in

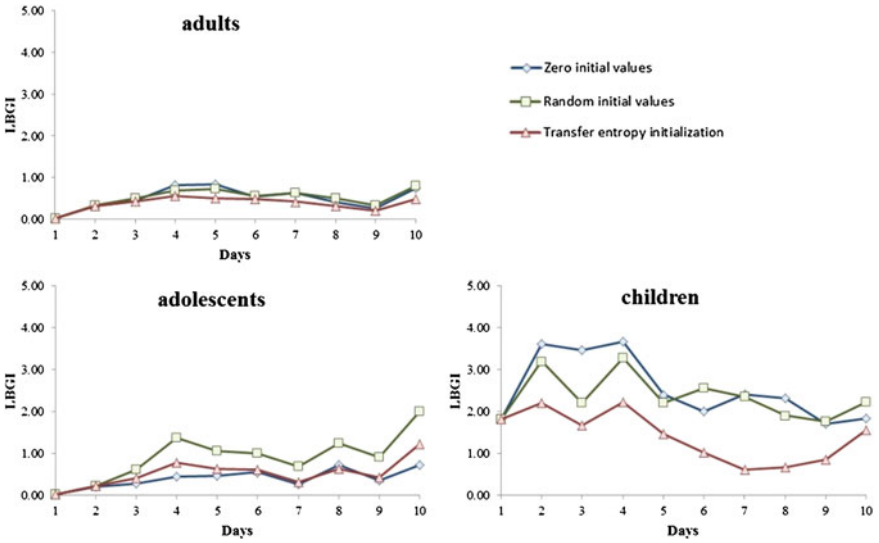


Fig. 3 Evolution of LBGI during the 10 days of the in silico trial for adults, adolescents and children and scenarios S1 (*blue*), S2 (*green*) and S3 (*red*)

Fig. 2 belong to the age group of children. This is expected since, as mentioned earlier, the initial simulator-suggested values of BR and IC ratio are far from being optimal. The contribution of the TE-based tuning especially in the case of children is thus critical as it significantly reduces the duration of the learning period and achieves increased overall performance. Figure 3 supports this above remark presenting the evolution of the daily LBGI over the total trial duration. As expected, children start from much higher LBGI values compared to adults and adolescents. It can be further seen that the daily LBGI is kept to low and comparable values among the three scenarios for adults and adolescents, however, in the case of children, LBGI reduces much faster when the AC algorithm is initialized based on the TE compared to the zero or random initialization.

3.3.3 Computational Cost

Due to the iterative process followed for the approximation of the optimal control policy, the AC algorithm presented very low computational cost. Each update of the BR and IC ratio needed less than one second. To this end, the AC algorithm can be used for real time insulin infusion.

3.4 Conclusions

An adaptive control algorithm for glucose regulation in T1D has been designed and developed based on RL and more specifically on the AC learning approach. The primary aim of the algorithm is the optimization of the daily insulin infusion policy through iterative learning of important patient-specific characteristics captured by his/her daily glucose profile. A novel method for the automatic and personalized tuning of the algorithm was proposed based on the estimation of TE from insulin to glucose time-series. The AC algorithm was evaluated in silico using an FDA accepted T1D simulator under a complex meal protocol and meal uncertainty. The results showed that AC is able to perform efficient and real-time optimization of insulin infusion and achieve high percentages in the clinically accepted range of the CVGA for all patient age groups with very low hypoglycemias. The performance of the AC algorithm was significantly improved when the TE-based initialization was used. This study illustrates that RL methods are able to provide online adaptive and optimized glucose regulation and compensate on the inter-/intra-variability and the system's uncertainties. Finally, the results indicate that RL should be employed and extensively investigated in diabetes control towards personalized insulin infusion treatment.

4 Adaptive Prediction of Hypo-/Hyperglycemic Events

4.1 System Modeling

System modeling refers to the design of a mathematical representation which captures the system's properties and input–output responses. In recent years, data-driven models have gained increased acceptance due to their simpler structure, ability to process significant amount of data in real-time and ease of personalization.

Model development is a three-step process and includes the (i) data collection, (ii) identification and (iii) evaluation phases. Model identification refers to the estimation of the modeling mathematical functions and parameters for the optimal representation of the system based on minimization of a pre-defined cost function relative to real and model predicted data. For the identification phase, a part of the collected dataset is used referred as the training dataset. Evaluation of the resulting model is conducted on a different part of the dataset denoted as the testing or evaluation dataset. It is important that both datasets are equivalent in the sense that they express the same properties of the system.

During evaluation, the models may remain static using the optimal representation and parameter values identified during training, or continue to optimize their performance based on the new coming information. When dynamic, time-varying systems are considered, static models may not be adequate for the continuous

representation of their changing dynamics. In these cases, online adaptive data-driven models can be employed.

In order to illustrate the advantages of adaptive data-driven models for glucose prediction, two types of modeling strategies able to perform real time glucose predictions in prediction horizons (PH) up to 45 min are discussed; (i) linear autoregressive (AR) models and (ii) nonlinear recurrent artificial neural networks (RNN). Special emphasis is given to their personalized identification and online adaptive ability. The developed models are subsequently integrated into an early-warning system (EWS) for recognition and alarm generation against upcoming hypo- and hyperglycemic events. The performance of the EWS is evaluated using a dataset of individuals with T1D.

4.2 Autoregressive Models

AR models are statistical models for the representation of a random process as a linear combination of its past values and possibly one or more external inputs (ARX models). AR models assume that the value of a process y at time t can be described by the following formula:

$$y(t) = \sum_{i=1}^{k} a_i y(t-i) + \varepsilon(t) \tag{23}$$

where k is the AR order signifying the length of the past period used for the estimation of $y(t)$, a_i are the model's parameters and $\varepsilon(t)$ is a stationary purely random process that represents the one-step prediction error. In the case that external inputs are also considered, the general representation of the ARX model is as follows:

$$y(t) = \sum_{i=1}^{k} a_i y(t-i) + \sum_{j=s}^{f} b_j u(t-j) + \varepsilon(t) \tag{24}$$

where k, s and f are the ARX orders describing the range of past signal and external input values, a_i and b_j are the model's parameters and $\varepsilon(t)$ is the one-step prediction error. As shown in (24), the external input may cover a different range of past input values, from s to f, not necessarily starting from time $t-1$. This is significant in time-delay systems where the effect of an input may become active after a certain time period.

Identification of AR and ARX models involves the estimation of the optimal orders and parameters given a set of input-output training data. One of the most commonly used methods for the selection of the optimal AR/ARX order is the Minimum Description Length (MDL) criterion which combines the number of model parameters (i.e. the model order) and the mean square prediction error as shown below:

$$MDL(q) = \left(1 + q\frac{\log n}{n}\right)\sigma_\varepsilon^2 \tag{25}$$

where q is the number of free parameters of the model to be identified ($q = k$ for AR and $q = k + f - s$ for ARX), n is the number of samples used for the identification and σ_ε^2 is the mean sum of squares function of the one-step prediction errors:

$$\sigma_\varepsilon^2 = \frac{1}{n}\sum_t (y(t) - \phi(t)\theta)^2 = \frac{1}{n}\sum_t \varepsilon(t)^2 \tag{26}$$

4.2.1 Online adaptive AR/ARX models

Let the training dataset be of size N. Rewriting (22) in a vector format, an AR model can be represented by the following formula:

$$Y = \Phi\theta \tag{27}$$

where $Y = [y(1)\, y(2) \ldots y(N)]$ is a vector containing all the time samples of the process $y(t)$, $\Phi \in \Re^{N\times k}$ is an array containing the past values of y so that each row $\phi(t) = [y(t-1) \ldots y(t-k)]$ and $\theta \in \Re^k$ is a vector of the parameters a_i. Since negative time values appear in this formulation, the first k time-steps for Y and Φ should be omitted. Assuming that the AR order k is known, the parameter vector θ can be estimated using least-squares (LS) as:

$$\theta^* = \arg\min_\theta \|Y - \Phi\theta\|^2 \tag{28}$$

where $\|\cdot\|^2$ is the Euclidean norm and θ^* is the parameter vector which minimizes the right-hand side of (27). A similar approach can be followed for the LS parameter identification of the ARX model with $\Phi(t) = [y(t-1) \ldots y(t-k)\, u(t-s) \ldots u(t-f)]$ and $\theta = [a_1 \ldots a_k\, b_s \ldots b_f]^T$. In this case the first $\max(k, f)$ time-steps should be omitted. Despite the high popularity of LS in optimization problems, they are not necessarily the most efficient way to go. LS may be inefficient for time-varying problems and in these cases online adaptive approaches should be explored. For this purpose, the recursive LS (RECLS) have been proposed as an extension to the regular LS. In contrast to LS which perform a batch optimization using the whole provided dataset and need to solve a matrix inversion problem, RECLS optimize the target parameters θ iteratively as new samples become available. The advantage of RECLS is that they can be used in cases of time-varying systems in order to provide online adaptive identification. The update process is described in the following equations:

$$\varepsilon(t+1) = y(t+1) - \phi(t+1)\theta(t) \tag{29a}$$

$$P(t+1) = P(t)\left[I_m - \frac{\phi^T(t+1)\phi(t+1)P(t)}{1+\phi(t+1)P(t)\phi^T(t+1)}\right] \tag{29b}$$

$$\theta(t+1) = \theta(t) + P(t+1)\phi^T(t+1)\varepsilon(t+1) \tag{29c}$$

where $\theta(t) = [a_1(t) \ldots a_k(t)\, b_s(t) \ldots b_f(t)]^T$ is the vector of the model's parameters, now time-varying, $\varepsilon(t)$ is the one-step prediction error and P(t) is the estimation of the error covariance.

4.2.2 Output Correction Module

A novel module has been developed in order to provide correction of the AR/ARX model's output based on the estimated prediction error. The principle behind model output correction lies on the development of a sub-model for the association of the prediction error of a specified *PH* to current glucose features and the use of this sub-model during evaluation to modify the respective model's output [31]. After identification, the model was applied to predict glucose in the three *PH* on the training dataset. Let $\hat{G}(t+PH|t^p)$ be the prediction of glucose at time $t + PH$ by the model given the glucose and insulin information of the period $t^p = \{t, \ldots, t-p\}$ where $t - p$ is the oldest time-sample used and t is the time when the prediction is performed. For each PH the prediction error E_{PH} is computed as:

$$E_{PH}(t+PH) = G(t+PH) - \hat{G}(t+PH|t^p) \tag{30}$$

where $G(t+PH)$ is the real glucose value at time $t + PH$. The prediction error was modeled as a linear combination of three glucose features: current glucose $G(t)$ value, first and second order glucose derivatives, $DG(t)$ and $D^2G(t)$ respectively as shown in (30). A different model was used for each PH.

$$\hat{E}(t+PH|t) = a_{PH}G(t) + b_{PH}DG(t) + c_{PH}D^2G(t) \tag{31}$$

where $\hat{E}(t+PH|t)$ is the estimation of the prediction error at time $t + PH$ based on the glucose features at time t when the prediction is performed. The parameters a_{PH}, b_{PH}, c_{PH} were identified in least-squares sense for each patient and each *PH* based on the paired training glucose and prediction error values. The first order glucose derivative was estimated as:

$$DG(t) = \frac{G(t) - G(t-s)}{s} \tag{32}$$

where s is the sampling time of glucose time-series. The second order derivative was estimated similarly. During evaluation, the ARX model provides predictions of glucose for the three PHs. Each prediction is subsequently corrected based on the estimated expected prediction error $\hat{E}_{PH}$ as:

$$\hat{G}_c(t+PH|t^p) = \hat{G}(t+PH|t^p) + \hat{E}_{PH}(t+PH|t) \tag{33}$$

and $\hat{G}_c$ is the corrected model output.

4.3 Artificial Neural Networks

ANNs are mathematical models used for the mapping of input to output data. ANNs are inspired by the principles of biological neural networks and aim to simulate the human brain function. One of their main advantages is that they can be highly nonlinear and incorporate dynamic behaviour and thus they can capture the structure and dynamics of intrinsically nonlinear processes.

4.3.1 Architecture

The architecture of an ANN model comprises a number of neurons, organized in layers, and synaptic weights for the interconnection of the neurons. Usually, the model consists of an input layer, one or more hidden layers and an output layer each involving a different number of neurons. Each neuron is characterized by an activation function which can be linear or nonlinear. The latter type of activation functions is the source of nonlinearity of the ANN model. Common nonlinear activation functions are the hyperbolic sigmoid and sigmoid tangent. ANN models are mainly classified as feedforward (FFANN) and feedback or recurrent (RNN). Other types of ANN include the radial basis function (RBF) network and Kohonen self-organizing network. In FFANNs, the neurons of one layer are connected through synaptic weights only with the neurons of the next layer, thus the signal is travelling in one direction from the input to the output. In contrast, RNNs can involve feedback loops through synaptic weights connecting neurons to previous layers or themselves. For the choice of the optimal ANN architecture the following parameters should be investigated:

- Type of ANN (FFANN, RNN, RBF etc.)
- Number of input neurons
- Number of hidden layers and neurons per layer
- Types of activation functions.

The ANN type is usually chosen empirically, depending on the problem and the purpose of the application. The same holds for the choice of the activation functions. Regarding the structural parameters (number of inputs, hidden layers

and neurons per layer), several paths exist including rule of thumb methods, exhaustive search and dynamical approaches such as the prune method. In recent years, adaptive architecture schemes have also been proposed based on evolutionary algorithms [32].

4.3.2 Dynamic Artificial Neural Networks

An important aspect in time-series analysis in general and in ANN design in particular is the incorporation of time, in order to represent the dynamic behaviour of the underlying process. This can be achieved by adding some kind of memory which is the primary ingredient that transforms a static network into a dynamic one [33]. Memory can be added by introducing time delays that is past values of the input signal to the ANN. Yet, another way to introduce memory to the network is by adding feedback. The first approach is used in order to incorporate dynamic behaviour in a simple FFANN resulting in the time-lagged FFANN. The use of feedback introduces the new category of RNN. In this study, a RNN based model has been used for the prediction of glucose in the near future. To this end, the analysis that follows is focused on the main principles and training algorithms of RNN models.

A dynamic system with external input can be mathematically described by its state-space representation as:

$$\frac{dy(t+1)}{dt} = F(y(t), u(t)) \tag{34}$$

where $y(t)$ is the current measured system output, $\frac{dy(t+1)}{dt}$ is the future signal change, $u(t)$ is the current value of the external input and F is a nonlinear continuous function. A RNN able to perform highly nonlinear dynamic mapping can be applied to model the above dynamic system by approximating the function F. The one-step-ahead signal value is given by:

$$y(t+1) = y(t) + \hat{F}(y(t), u(t)) \tag{35}$$

where $y(t+1)$ and $y(t)$ are the next and current signal values predicted by the model, and $\hat{F}$ is the approximation of F implemented by the RNN.

As in all types of ANN, identification of an RNN model involves the estimation of the optimal architecture (number of hidden layers, number of neurons per layer, activation functions) and synaptic weights. One of the most commonly used approaches for the estimation of the weights is back-propagation through time (BBTT) which is an extension of the common BP algorithm for RNN [34]. The common BP algorithm cannot be directly used in RNN as it presupposes the absence of cycles within the units of the network. For this reason, the BBTT algorithm performs unfolding of the RNN by replacing it with identical copies and redirecting the feedback loops in subsequent copies. Yet another successful

approach is the real-time recurrent learning (RTRL) which is suitable for online learning based on the continuous adaptation of the weights as new samples are fed to the network. A detailed description of the teacher-forced real-time recurrent learning algorithm can be found in [35, 36].

4.4 A Personalized System for the Early Recognition of Hypo-/Hyperglycemic Events based on Adaptive Glucose Prediction Models

An EWS for the prediction of hypo- and hyperglycemic events and the triggering of respective warnings has been developed and evaluated using sensor glucose and insulin pump data from a dataset of 23 individuals with T1D. The ultimate aim of the EWS is to be integrated to the control algorithm in order to be used as a safety mechanism. The EWS incorporates a glucose prediction model and a warning algorithm for processing the model's output and issuing an alert if an upcoming hypo- or hyperglycemic event is expected to start in the near future. The EWS was evaluated for its ability to predict upcoming hypoglycemic and hyperglycemic events using the following criteria:

1. **Percentage of correct warnings**: A warning is defined as correct if it was issued at most 45 min before the start of an event.
2. **Event detection time**: Refers only to correct warnings and denotes the time between the issued warning and the start of the event.
3. **Daily false alarms**: A false alarm is a warning that could not be matched to a real event within 45 min from its triggering time.

Four candidate models have been investigated within the EWS all able to predict glucose in $PH = 15$, 30 and 45 min. Common base of all models is their personalized architecture and online adaptive parameters.

i. *ARX*

An online adaptive ARX model was developed for each candidate T1D individual using his/her glucose and insulin data. During training, the model was identified by selection of the orders and the parameter vectors that best fit the actual to the one-step-ahead predicted time series samples. First, the model order was chosen based on the minimization of the MDL criterion defined in (25). The parameter s was chosen to be equal to 45 (maximum PH) to enable only insulin information already provided to be used. This did not influence the model's accuracy as the PHs investigated were short, and the maximum action of the insulin analogs takes place between 30 and 120 min after infusion. For the parameter estimation, the Tikhonov regularization approach [37] was used. During evaluation, the models proceeded to continuous adaptation of their parameters based on the RECLS algorithm (4.29a–c). The initial RECLS parameter vector was set to the optimal parameter values estimated during training.

ii. *cARX*
The ARX models where enhanced with the previously described output correction module for each *PH*. The resulting ARX model with corrected output is referred hereafter as cARX.

iii. *RNN*
A fully connected, multilayered RNN with two feedback loops was developed for each T1D individual [38]. It consisted of one input layer with one external input corresponding to the last insulin infusion value and one state variable. The RNN has an output layer with one neuron and a different number of hidden layers and neurons per layer for each patient. The feedback loops are formed by the one-to-one connection from the network output to the state variable and the feedback of the state variable into itself. The hyperbolic sigmoid and linear functions are used as activation functions in the hidden and output layer, respectively. Before being fed into the RNN, the input values were normalized in the range [−1.0, 1.0]. The state variable was initialized as the first glucose value available, and the RNN's weight matrices were initially assigned to random values in the range [−0.5, +0.5]. During training, the synaptic weights and biases were online updated based on the teacher-forced RTRL algorithm for every new input of the RNN as described above. The weights and biases update continued in the same way during evaluation as in the case of the ARX models in order to provide individualized and online adaptive predictions.

iv. *Hybrid cARN*
A hybrid model based on fusion of cARX and RNN models was investigated with the aim to explore the improvement of the aforementioned evaluation criteria when two different modeling strategies are combined. The fusion was implemented as a linear combination of the two models' outputs. The final output of the fused model is formed as shown in (35).

$$\hat{G}_{cARN}=a\hat{G}_{cARX}+(1-a)\hat{G}_{RNN} \tag{36}$$

where $0 \leq a \leq 1$ is a balancing factor between the two outputs. Increasing the value of α, moves the fused model closer to cARX and away from RNN. The balancing factor a was chosen individually for each candidate patient based on the maximization of the following cost function:

$$J=\frac{(Percentage\ of\ Correct\ Warnings)^2+(Event\ Detection\ Time)^2}{1+(Daily\ False\ Alarms)^2} \tag{37}$$

All three values, before inserted in the cost function are normalized in [0 1].

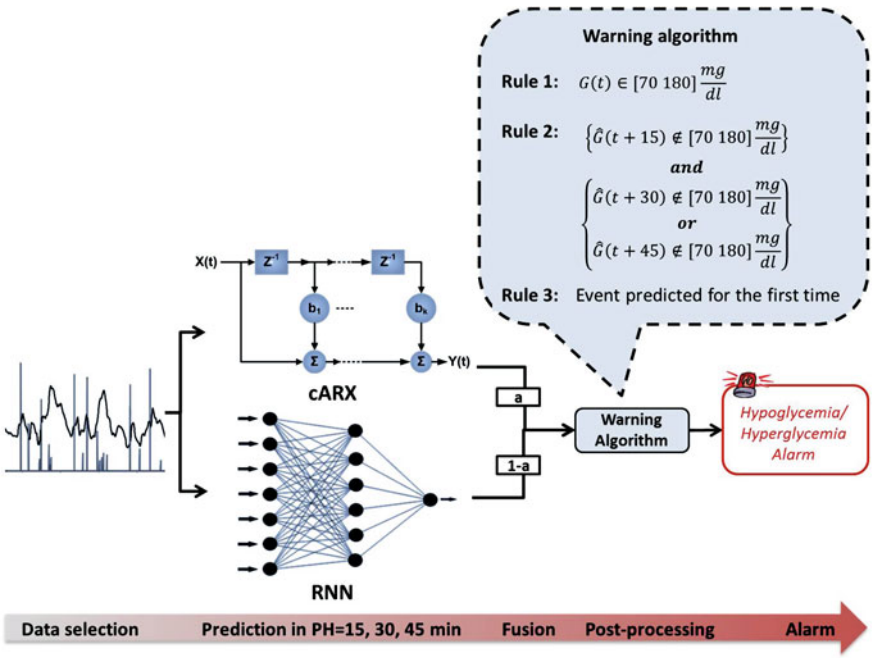


Fig. 4 Flowchart of the hypo-/hyperglycemia prediction process with combination of glucose prediction models and a warning algorithm

4.4.1 Warning Algorithm

The prediction models naturally present errors and time-lags (TL) which cannot be avoided neither neglected. The presence of noise in the sensor glucose measurements is an important deteriorating factor of the models' performance. It has been further observed that increase in the *PH* results in higher prediction errors and TLs. On the other hand, the higher accuracy of short *PH*s is counteracted by the inadequately short-term notice for the patient, which may not leave enough time for the needed actions (insulin injection or carbohydrate intake). To this end, in order to enhance the reliability of the predictions and the prompt event detection, the a rule-based warning algorithm is designed to issue warnings combining the predictions of all three *PH*. Figure 4 presents the warning algorithm and the flowchart of the hypoglycemia and hyperglycemia prediction process and warning generation. More details regarding the warning algorithm can be found in [17].

Table 2 Statistical parameters of the patient dataset in mean ± standard deviation values

Data statistics per patient	Training set	Evaluation set
Data collection time (days)	5.30 ± 1.40	4.83 ± 1.80
Number of hypoglycemic events[a]	7.43 ± 6.64	6.57 ± 5.66
Number of hyperglycemic events[a]	14.00 ± 9.22	13.22 ± 7.17
Hypoglycemic event duration (min)	51.55 ± 26.31	51.82 ± 36.16
Hyperglycemic event duration (min)	122.60 ± 55.86	137.73 ± 69.98
Data statistics total dataset		
Data collection time (days)	122	111
# of hypoglycemic events in dataset	171	151
# of hyperglycemic events in dataset	322	304

[a] A hypoglycemic (hyperglycemic) event was defined as sensor glucose values below 70 mg/dl (over 180 mg/dl) for at least 10 min (2 consecutive sensor glucose measurements)

Table 3 Performance of the ARX, cARX, RNN and fused cARN-based systems in median (5–95th percentiles) for **(A)** hypoglycemia and **(B)** hyperglycemia event prediction

Evaluation criteria	ARX	cARX	RNN	cARN
(A) Hypoglycemia				
Correct alarms (%)	100 (94.0–100)	100.00 (100–100)	100.00 (58.1–100)	100 (100–100)
Detection time (min)	10.0 (5.0–24.5)	17.5 (11.8–31.0)	8.4 (0.3–14.9)	16.7 (10.0–25.0)
Daily false alarms	0.7 (0.0–1.94)	1.5 (0.5–4.4)	0.1 (0.0–0.6)	0.8 (0.0–1.2)
(B) Hyperglycemia				
Correct Alarms (%)	100 (90.1–100)	100 (93.8–100)	92.0 (70.5–100)	100 (95.3–100)
Detection time (min)	8.0 (1.7–13.9)	14.8 (8.8–20.6)	7.0 (4.7–15.2)	14.7 (5.1–19.25)
Daily false alarms	0.5 (0.0–1.2)	1.3 (0.4–3.2)	0.2 (0.0–0.7)	0.8 (0.0–1.4)

4.5 Results and Discussion

Subject Data: Sensor glucose and insulin pump data from 23 T1D individuals [age 17–70 years, HbA1c (%) 7.3 ± 0.7, BMI (kg/m^2) 24.2 ± 3.0] under SAP therapy were utilized in the study. All patients used Medtronic insulin pumps (Medtronic MiniMed Inc., Northridge, CA, USA) combined with a real-time, CGM system during everyday living conditions. The sensor glucose values were equally sampled every 5 min. For each patient, half of the dataset was used as training set for the identification of the models' architecture and parameters. The other half was used for the evaluation of the models. Table 2 presents the statistical parameters of the dataset.

The EWS based on the each of the four modeling approaches (ARX, cARX, RNN, cARN) has been evaluated using the aforementioned patient dataset. The results are presented in Table 3 in median (5–95th percentiles) values.

From Table 3 it can be seen that the EWS based on all models presents high accuracy and detection times and limited daily false alarms. The cARX model presents lower TL, increased RMSE and comparable CC to the ARX model. The correction module increased the responsiveness of the model especially to fast

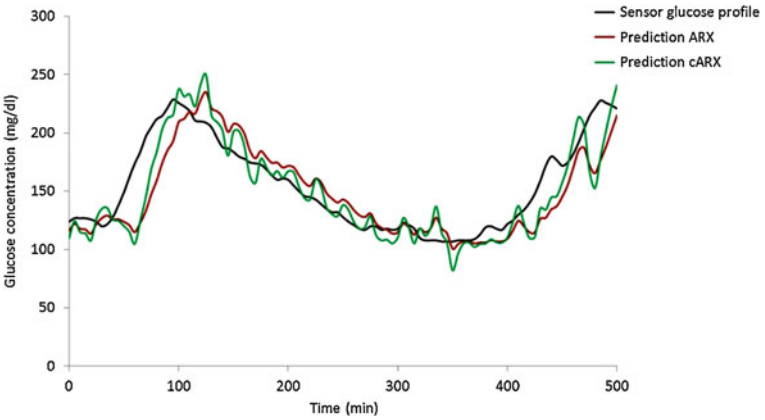


Fig. 5 Example of prediction profile from the ARX (*red*) and cARX (*green*) model in PH = 30 min

glucose changes which resulted in faster reactions but at the same time larger fluctuations. This observation can be more easily illustrated in Fig. 5 where an example of glucose prediction profile by the two models is plotted. The ARX- and cARX-based EWS are more accurate than RNN-based as the later presents high variability among the patients and cases where many hypo- and hyperglycemic events were missed. In terms of detection time, the cARX-based EWS presents the best performance while the RNN-based EWS outperforms the others with respect to daily false alarms. These results show that the cARX- and RNN-based EWS present complementary performances. The aim of the hybrid cARN-based EWS is to combine the advantages of both approaches and achieve competent performance in all evaluation criteria. Figure 5 presents an example of prediction profile for the ARX and cARX models in PH = 30 min.

The performance of the cARN-based EWS is presented in the fourth column of Table 3 for the optimal choice of the balancing factor a per patient. The cARN-based EWS presents the highest accuracy of correct alarms, very good detection times, approaching the performance of the cARX-based EWS, and very low daily false alarms. Considering the pharmacokinetics of insulin and glucose, the detection times achieved by the cARN-based EWS can be sufficient for the avoidance of the upcoming hypo- or hyperglycemic event if the patient takes the needed actions for each occasion. The above results clearly indicate that combination of complementary modeling strategies preserve the qualities of each approach and achieve accurate and prompt prediction performance.

4.6 Conclusions

Combination of online adaptive ARX, cARX and RNN models with a warning algorithm for the generation of hypo- and hyperglycemic alarms resulted in the development of a personalized EWS. In all cases the EWS presented efficient and prompt recognition of upcoming hypo- and hyperglycemic events for all patients. The cARX-based and RNN-based EWS presented complementary qualities leading to the development of the hybrid cARN-based EWS. The hybrid system managed to preserve the high accuracy and detection times of the cARX-based and the low false alarms of the RNN-based EWS combining both safety and comfort. The results indicate that online adaptive data-driven models can effectively represent the time-varying glucose process and detect future metabolic disturbances. Furthermore, different modeling techniques should be extensively investigated and comparatively assessed for the determination of complementary properties. Successful combination of complementary approaches can lead to more accurate representation of the glucose time-series and significant improvement of the prediction performance.

5 Final Remarks and Future Trends

The applicability of adaptive algorithmic solutions in the study of complex systems characterized by variability, uncertainties and time-varying dynamics has been investigated in the two principle fields of diabetes management, namely in glucose regulation and glucose prediction. To this aim, an online adaptive control algorithm for optimized insulin infusion was designed and developed based on the principles of reinforcement learning and more specifically of the AC learning approach. Furthermore, two online adaptive glucose prediction models have been developed and subsequently integrated into an EWS for the recognition of future hypo- and hyperglycemic events. The results of the AC algorithm showed that personalized insulin treatment through iterative learning of patient-specific characteristics can lead to real-time improvement of glycemic regulation. The performance of the EWS enhances this result showing that personalized and adaptive prediction models can render the early recognition of abnormal metabolic situations accurate and safe.

This work offers significant room for improvement and future research development. The direct next step is the integration of the EWS into the AC control algorithm in order to function as a safety-supervision mechanism towards the development of an adaptive control system for personalized and safe insulin treatment. The control system will be clinically evaluated in order to reinforce the significance of the in silico results and provide important feedback for future improvement. On the algorithmic side, additional RL-based control approaches and alternative adaptive prediction models will be investigated along with more advanced fusion techniques for the optimal representation of glucose dynamics.

Overall, the results of this chapter indicate the applicability and efficiency of online adaptive algorithms towards optimization of diabetes management.

References

1. Lof J et al (1998) An adaptive control algorithm for optimization of intensity modulated radiotherapy considering uncertainties in beam profiles, patient set-up and internal organ motion. Phys Med Biol 43:1605–1628
2. Yu C et al (2011) A model-free adaptive control to a blood pump based on heart rate. ASAIO J 57(4):262–267
3. Jiang X et al (2012) A patient-driven adaptive prediction technique to improve personalized risk estimation for clinical decision support. J Am Med Inform Assoc 19:e137–e144
4. Torshabi AE (2013) An adaptive fuzzy prediction model for real time tumor tracking in radiotherapy via external surrogates. J Appl Clin Med Phys 14(1):4008
5. Wu X et al (2011) Segmentation and reconstruction of vascular structures for 3D real-time simulation. Med Image Anal 15(1):22–34
6. Möller B et al (2011) Adaptive segmentation of particles and cells for fluorescent microscope imaging. Computer vision, imaging and computer graphics. Theory and applications. Springer, Heidelberg, pp 154–167
7. Liu L et al (2013) Adaptive segmentation of magnetic resonance images with intensity inhomogeneity using level set method. Magn Reson Imaging 31(4):567–574
8. Hovorka R et al (2004) Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes. Physiol Meas 25:905–920
9. Magni L et al (2009) Run-to-run tuning of model predictive control for type 1 diabetes subjects. Silico Trial J Diabetes Sci Technol 3(5):1091–1098
10. Zisser H et al (2005) Run-to-run control of meal-related insulin dosing. Diab Technol Ther 7(1):48–57
11. Wand Y et al (2010) Automatic bolus and adaptive basal algorithm for the artificial pancreas b-cell. Diab Technol Ther 12:879–887
12. Miller S et al (2011) Automatic learning algorithm for the MD-Logic artificial pancreas system. Diab Technol Ther 13:983–990
13. Eren-Oruklu M et al (2010) Hypoglycemia prediction with subject-specific recursive time-series models. J Diab Sci Technol 4(1):25–33
14. Meriyan EO et al (2012) Adaptive system identification for estimating future glucose concentrations and hypoglycemia alarms. Automatica 48:1892–1897
15. Daskalaki E, Diem P, Mougiakakou S (2013) An Actor-Critic based controller for glucose regulation in type 1 diabetes. Comput Methods Programs Biomed 109(2):116–125
16. Daskalaki E, Prountzou A, Diem P, Mougiakakou S (2012) Real-time adaptive models for the personalized prediction of glycemic profile in type 1 diabetes patients. Diab Technol Ther 14(2):168–174
17. Daskalaki E, Nørgaard K, Prountzou A, Züger T, Diem P, Mougiakakou S (2013) An early-warning system for hypoglycemic/hyperglycemic events based on fusion of adaptive prediction models. J Diab Sci Technol 7(3):689–698
18. Sutton RS, Barto AG (1998) Reinforcement learning. MIT Press, Cambridge
19. Marbach P, Tsitsiklis JN (2001) Simulation-based optimization of Markov reward processes. IEEE Trans Autom Control 46:191–209
20. Szepesvari C (2010) Algorithms for reinforcement learning - Synthesis lectures on artificial intelligence and machine learning. Morgan & Claypool Publishers
21. Tsitsiklis JN, Van Roy B (1997) An analysis of temporal-difference learning with function approximation. IEEE Trans Autom Control 42(5):674–690

22. Hlaváčková-Schindler K, Paluš M, Vejmelka M, Bhattacharya J (2007) Causality detection based on information-theoretic approaches in time series analysis. Phys Rep 441(1):1–46
23. Schreiber T (2000) Measuring information transfer. Phys Rev Lett 85(2):461–464
24. Williams PL, Beer RD (2011) Generalized measures of information transfer, preprint arXiv. 1102.1507
25. Lee J, Nemati S, Silva I, Edwards BA, Butler JP, Malhotra A (2012) Transfer entropy estimation and directional coupling change detection in biomedical time series. Biomed Eng Online 11:19
26. Butte AJ, Kohane IS (2000) Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. Pac Symp Biocomput 5:418–429
27. Patek SD, Bequette BW, Breton M, Buckingham BA, Dassau E, Doyle FJ III, Lum J, Magni L, Zisser H (2009) In silico preclinical trials: methodology and engineering guide to closed-loop control in type 1 diabetes mellitus. J Diab Sci Technol 3:269–282
28. Dalla Man C et al (2007) Meal simulation model of the glucose-insulin system. IEEE Trans Biomed Eng 54(10):1740–1749
29. Magni L, Raimondo DM, Dalla Man C, Breton M, Patek S, De Nicolao G, Cobelli C, Kovatchev BP (2008) Evaluating the efficacy of closed-loop glucose regulation via control-variability grid analysis. J Diab Sci Technol (Online) 2:630–635
30. Kovatchev BP, Cox DJ, Gonder-Frederick LA, Young-Hyman D, Schlundt D, Clarke W (1998) Assessment of risk for severe hypoglycemia among adults with IDDM: validation of the low blood glucose index. Diabetes Care 21(11):1870–1875
31. Bosnic Z, Kononenko I (2010) Correction of regression predictions using the secondary learner on the sensitivity analysis outputs. Comput Inform 29:929–946
32. Zhang C et al (2000) Particle swarm optimization for evolving artificial neural network. IEEE Int Conf Syst Man Cybern 4:2487–2490
33. Haykin S (1999) Neural Networks: a comprehensive foundation, 2nd edn. Prentice-Hall Inc., New Jersey, p 07458
34. Werbos PJ (1990) Back propagation through time, what it does and how to do it. Proc IEEE 78:1550–1560
35. Williams R, Zipser D (1995) Gradient based algorithms for recurrent NN and their computational complexity. In: Chauvin Y, Rumelhart DE (eds) Back-propagation: theory, architecture, and applications. Lawrence Erlbaum, Hillsdale
36. Williams R, Zipser D (1989) A learning algorithm for continually running fully recurrent NN. Neural Comput 1:270–280
37. Tikhonov AN, Arsenin VY (1977) Solutions of Ill-posed problems. Winston, Washington
38. Mougiakakou S, Prountzou A, Iliopoulou D, Nikita K, Vazeou A, Bartsokas C (2006) Neural network based glucose–insulin metabolism models for children with type 1 diabetes. Conf Proc IEEE Eng Med Biol Soc 1:3545–3548

Pitfalls in Model Identification: Examples from Glucose-Insulin Modelling

Simona Panunzi and Andrea DeGaetano

Abstract Two important statistical parameter estimation pitfalls, examples of which can be found in the literature, are here reviewed and discussed. The first concerns the lack of model qualitative behaviour analysis before proceeding to the actual parameter estimation phase: this may give rise in the worst cases to aberrant model behaviour and to meaningless parameter estimates. The second concerns the use of interpolated noisy observations taken to represent the real input or driving variable into a model: this gives rise to the artifactual reproduction of meaningful features of the output variables, based on data errors and hence inherently non-reproducible. This is particularly dangerous when using noisy observations instead of model predictions in coupled systems. Examples of these pitfalls are drawn from existing glucose-insulin modelling literature and recommendations are made.

1 Introduction

Insulin Resistance (IR), the impaired metabolic response to circulating insulin resulting in a decreased ability of the body to respond to the hormone by suppressing Hepatic Glucose Output and enhancing tissue glucose uptake, plays a central role in the development of Type 2 Diabetes Mellitus. It is therefore of considerable interest to have an accurate measurement of the degree of IR by tests that are objective and easy to perform.

While the Euglycemic Hyperinsulinemic Clamp (EHC) [1] is often considered the "*gold standard*" for the determination of insulin resistance, the Intra-Venous

S. Panunzi (✉) · A. DeGaetano
CNR-IASI BioMatLab, Largo A. Gemelli 8, 00168 Rome, Italy
e-mail: simona.panunzi@biomatematica.it

A. DeGaetano
e-mail: andrea.degaetano@biomatematica.it

V. Marmarelis and G. Mitsis (eds.), *Data-driven Modeling for Diabetes*,
Lecture Notes in Bioengineering, DOI: 10.1007/978-3-642-54464-4_5,

Glucose Tolerance Test (IVGTT) is simpler to perform, carries no significant associated risk and delivers potentially richer information content. What prevents many experimental diabetologists from using the IVGTT is its interpretation, for which it is necessary to apply a mathematical model of the status of the negative feedback regulation of glucose and insulin on each other in the studied experimental subject. A reliable mathematical model for the representation of the glucose/insulin control system during the IVGTT is therefore necessary. Such a model, like each model used to explain a physiological phenomenon, should be simple enough to allow precise estimation of its structural parameters (in the specific case it should allow precise estimation of insulin sensitivity on a single patient), yet exhibit stable dynamics and reproduce accepted physiological behavior. Further, the model formulation, while applicable to the standard IVGTT, should logically and easily extend to model other often envisaged experimental procedures, like repeated glucose boli, or infusions.

Due to its relatively simple structure and to its great clinical importance, the glucose/insulin system has been the object of repeated mathematical modeling attempts [2–20]. The mere observation that several models have been proposed points to the fact that it is not trivial to satisfactorily integrate mathematical, statistical and physiological aspects in order to realistically and reliably represent the glucose/insulin system.

Moving from the need of modeling this concrete and relevant physiological problem, the present chapter has the goal of discussing two important model parameter estimation pitfalls, to which modellers should pay attention when involved in the mathematical representation of a physiological system.

The first pitfall is to neglect to study the mathematical formulation of the model, even before collecting or analyzing any data, in order to establish that the physiological assumptions underlying the model formulation determine a qualitative behavior of its solutions which is coherent with known physiology and with the expected behavior of the phenomenon under study. The consequence of this pitfall is that estimated model parameter values refer in fact to a model structure inconsistent with known physiology, and are therefore of limited usefulness for quantifying and better understanding this physiology.

The second pitfall consists of using interpolated noisy observations of some ‘driving’ variable to represent the real input into a modeled system. The parameters obtained in this way, estimated by fitting model output to observed system output, are apparently effective in capturing meaningful features of the output variables, but do so by exploiting random errors in the input data: this strategy gives rise therefore to a merely apparent, artifactual exhibition of some desired model’s behavior, based on data errors and hence inherently non-reproducible. A clear example of this pitfall can be seen in the artificial ‘decomposition’ of an integrated feedback loop, where the parameters, separately estimated, determine an integrated behavior that widely diverges from what is expected and from recorded data.

While the points discussed are general and apply to general modeling practice, both for what regards the necessary preliminary qualitative study of the model

solutions and for what regards the incorrect use of interpolated noisy observations as input function, the examples brought forward in the present chapter concern the modeling of the Intra-Venous Glucose Tolerance Test (IVGTT). In particular the Minimal Model [21], Bolie's model [2] and the recently proposed Single Delay Model [22] of the glucose-insulin system have been used as examples in this discussion.

The chapter is structured as follows: the experimental procedure and the physiological principles on which the IVGTT is based are briefly presented, then each of the two statistical pitfalls is discussed, using the above published modeling attempts of the IVGTT as example.

2 The IVGTT Experiment

The standard IVGTT (without either Tolbutamide or insulin injections) [21] consists in a rapid injection (less than 3 min) at time 0 ($0'$) of a 33 % glucose solution (0.33 g Glucose /kg Body Weight) through one arm line. Blood samples (3 ml each, in lithium heparin) are then obtained at frequent time intervals (typically at $-30'$, $-15'$, $0'$, $2'$, $4'$, $6'$, $8'$, $10'$, $12'$, $15'$, $20'$, $25'$, $30'$, $35'$, $40'$, $50'$, $60'$, $80'$, $100'$, $120'$, $140'$, $160'$ and $180'$) through a contralateral arm vein. Each sample is immediately centrifuged and plasma is separated. The plasma levels of glucose and insulin obtained at $-30'$, $-15'$ and $0'$ are averaged to yield the baseline values referred to $0'$.

In extreme synthesis, the sudden appearance of a large amount of glucose in peripheral blood stimulates the pancreas to secrete insulin, which is in fact released (in the normal subject) in two phases, an immediate first phase and a delayed second phase, probably deriving, respectively, from insulin granules immediately available for release at the β-cell membrane and to insulin that has to be transferred to the membrane and 'docked' before secretion. Peripheral tissues respond to the insulin concentration increase by accelerating glucose extraction from plasma, and glucose concentrations are brought back to normal. As the hyperglycaemic stimulus decays, so does insulin secretion, which also goes back to normal. Pre-infusion steady state is then reached after approximately two to three hours in the normal subject.

3 First Pitfall: Failing to Study the Qualitative Behaviour of the Model's Solutions

One of the very first attempts to understand and express mathematically the determinants of glucose concentrations in blood dates back to 1961 with Bolie's model, which can be written as follows:

$$\frac{dG(t)}{dt} = -p_1\, G - p_2\, I(t) + p_3, \quad G(0) = G_0 s$$

where p_1, p_2 and p_3 are three positive parameters. The model is very simple and each term on the right-hand side would appear to make sense: variations of glycemia are due, first, to the level of glycemia itself (the higher the glycemia, the faster its decrease); second, to insulinemia (the higher the insulinemia, the faster the decrease in glycemia); and, third, some necessary glucose production (say, by the liver), which we may assume constant throughout a short experiment.

The qualitative analysis of the model reveals however that negative solutions are permissible: at zero glycemia, its derivative may well be strictly negative, depending on the level of insulinemia.

The fact that solutions of the model, i.e. predicted glycemia, can be negative is evidently contrary to common sense. There are then at least two possible philosophical stances, which the modeller can take in the face of such paradoxical result. The first is to assert that the model has some limited validity in some circumscribed region of its state space: in the present case, that the model is still good, if insulin is not too high and glycemia is not too low. Oftentimes, modellers subscribing to this philosophical position defend it by quoting G.E.P. Cox: "Essentially, all models are wrong, but some are useful" [23]. We believe this position to be somewhat simplistic, and potentially misleading: it aggregates all models, irrespective of their merits, into the class of "wrong" models, and elects a model as 'appropriate' depending merely on contingent utility. We may offer a two-tiered philosophical approach, by defining a model to be "inconsistent" when it does not agree with irrefutable available data, and further by defining a model as "obsolete" when it is inconsistent and another model exists, which does explain the irrefutable available data. If the qualitative behaviour of a model's solutions contradicts common knowledge, then the model's construction must imply some fundamental mistake: for example, from the very moment that Bolie's model was published, it was clear that it could predict glucose concentrations to become negative, hence it was inconsistent from the very start. However, at the time it was not obsolete, because no other model existing at the time (there were none) was consistent: indeed, Bolie's model spurred mathematical investigation of the Glucose-Insulin system, and deserves our affection and respect for having been a pioneer. The fundamental mistake of Bolie's model concerns the form of the second term, the decrement in glycemia depending on insulinemia. Figure 1 summarizes the problem: the model asserts that, no matter how little glucose there is in blood (even zero), we would be able to make the tissues extract as much glucose as we want, just by increasing insulin levels. Since we cannot take out of plasma the glucose that is not there, the second term in Bolie's model is wrong and determines in fact the possible negativity of predicted glycemia.

A consequence of flawed model construction, as detectable by theoretical behaviour inconsistent with basic observations, is that the parameters of the model may well be devoid of meaning, in so far as they refer to phenomena which either

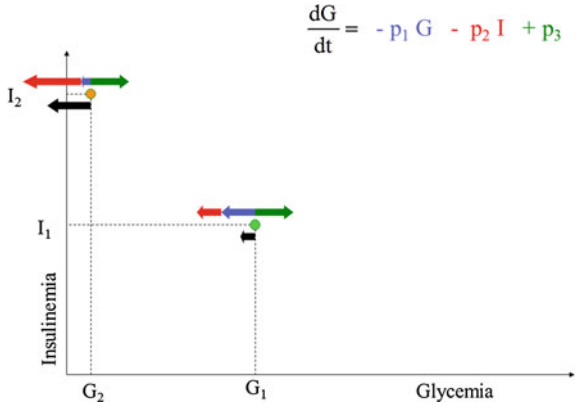


Fig. 1 Schematic representation of Bolie's model behaviour. *Black arrows* represent the variation in blood glucose concentration as a result of glucose tissue uptake mediated by insulin (*red arrows*), glucose produced by liver (*green arrows*) and spontaneous glucose elimination. While the model works correctly around the point (G_1, I_1), when plasma glucose is low and serum insulin is high (G_2, I_2) the model would predict negative glycemias, by still (incorrectly) allowing arbitrarily high glucose extraction

do not exist or are incorrectly represented in the model. For example, the insulin-dependent glucose elimination rate p_2 in Bolie's model, which would quantify the number of mM/min of glycemia decrement per pM of insulin concentration, has no meaning: insulin-stimulated glucose uptake has some value at high glycemia and is in fact zero at zero glycemia, so that giving it a single value over the entire allowable range of glycemias is, at a minimum, misleading.

It is clear that models capture only certain features of the phenomenon under investigation, and that ultimately all models will fail to agree with data from sufficiently advanced and detailed experiments. All models are simplifications of reality, will eventually be proven inconsistent by further experimentation, and will eventually be made obsolete by better models.

When designing a new model to interpret a physiological phenomenon, studying the qualitative properties of the solutions ensures that the model is not already inconsistent at the moment it is proposed. As another example from the literature on the IVGTT we may mention the Minimal Model [21], which is still the most widely employed model to assess insulin resistance, even in recent research applications [24–31]. This model describes the time-course of glucose plasma concentrations, depending upon serum insulin concentrations and on a variable X, representing the 'Insulin activity in a remote compartment' [32]. While in later years different versions of the Minimal Model appeared [33, 34], the following discussion refers to the original formulation [21, 35] as presented in [36] with detailed instructions on how to identify the model from data.

A formal, qualitative study of the solution of the model appeared in the year 2000 [37]. The model is non-autonomous, and it can be proven that, mainly depending on the non-autonomous time term, when glucose concentrations are

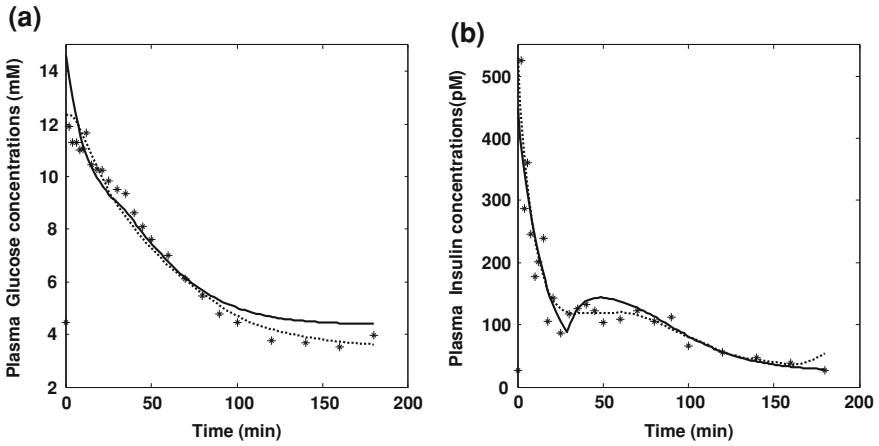


Fig. 2 Glucose and Insulin predictions with the SDM and MM models. *Panels* **a** and **b** respectively show observed plasma glucose concentrations and observed plasma insulin concentrations (*asterisks*) following an IVGTT experiment, along with predictions by the Single Delay Model (*continuous lines*) and by the Minimal Model (*dotted lines*)

higher than a threshold value (lower than observed basal glycemia) the variable X goes to infinity and the system does not admit an equilibrium. Proponents of the model specifically state that the model is 'valid' only for the 3 h or so of duration of a typical IVGTT: on the other hand the Insulin Sensitivity index, as derived from the model, is defined for time that goes to infinity. Figure 2, panel b (dotted lines) shows the anomalous behaviour in the insulinemia predictions which are visibly increasing at the end of the observation period and which would be predicted to increase to extremely high levels within a few hours instead of tending to the equilibrium value corresponding to basal insulin concentration IB_b.

We find ourselves in essentially the same situation as with Bolie's model: in that case, solutions could become negative, in this case solutions can explode to infinity. Limiting the operation domain of Bolie's model to not-too-high insulinemias, or limiting the operation domain of the Minimal Model to 'about three hours' does not address the issue, which is that if an abnormal behaviour is produced by the model this means that the model's construction incorporates flawed assumptions. Both models therefore could have been diagnosed as inconsistent from the start. It remains to be seen if they are now obsolete.

Recent work [22] presented in fact an alternative model for the interpretation of glucose and insulin concentrations observed during an IVGTT: the so-called discrete Single Delay Model (SDM). In the context of the present discussion, the relevant thing to notice is that this model has been proven to have solutions mathematically consistent with expected behavior [38], admitting the fasting state as its single equilibrium point and converging back to it from the perturbed state. Moreover, the model aimed at being consistent with known physiology by incorporating a limited pancreatic insulin secretion in response to increasing

glucose concentrations: Fig. 2 (continuous line) reports the fitting of the SDM over an IVGTT experiment. While no claim is made as to the optimality of the SDM in any sense (indeed, better models are expected to be formulated in the future), its existence makes the other two models obsolete, in the sense defined above.

Model builders should therefore study the qualitative behaviour of the solutions of their models in order to assure that these are capable of describing consistently the relevant physiology, as a pre-condition to obtaining reliable quantitative estimates of the parameters of interest.

4 Second Pitfall: Using Interpolated Noisy Observations as Input Functions, Particularly When Fitting Coupled Systems

The idea that interpolated observed data, used in place of theoretically reconstructed curves, are a reliable approximation of the true signal for the purpose of parameter estimation is rather widespread in the domain of insulin glucose model (*ex multis* [36, 39]).

What is generally not considered is that, during an estimation process for models depending on input or forcing functions, the optimizer algorithm attempts to obtain the best possible reconstruction of the output signal changing the parameter values *conditional on the input signal*. This determines, in particular, the possible leverage of accidental data oscillations in the input signal, through *ad-hoc* parameter values, in order to build an output signal that matches output observations. The result is, indeed, an apparently good match to the observed output, and what is sometimes not appreciated is that this apparently good fit is contingent on finding parameter values exploiting accidental variations in input: the result is artifactual because it does not build on the expected behavior of the model but on stochastic realizations of the errors; it is inherently non-reproducible, because the parameters found depend directly on such random variations in input; and the corresponding parameters have a high variability, because different random configurations of the input will produce in general different parameter values attempting to reconcile the variable input with the output observations.

As a clear example of the misleading results that can be obtained when incurring in the pitfall of using interpolated observations as input into the model, consider once again the Minimal Model of glucose-insulin dynamics, in the form and with the estimation procedure proposed in 1986 [36]. In this case, two sub-models are used, one for glucose kinetics depending on insulinemia as driving function, and one for insulin kinetics depending on glycemia as driving function: the proposed estimation procedure consists of using observed insulinemias to represent the true input for the purpose of estimating glucose kinetics and then using observed glycemias for the estimation of insulin kinetics (instead of performing a single optimization on both feedback control arms of the integrated glucose/insulin system).

Following this approach, the estimated parameters are optimal in predicting observed glucose assuming the erratic observed insulin as the true value of the insulin concentration (and *viceversa*), but would not be optimal in describing the modeled effects of expected glucose on expected insulin and *viceversa*. This fitting strategy produces sets of estimated parameters such that the expected time course of glucose using the expected time course of insulin as input may differ markedly from both the actual glucose observations and from the expected glucose obtained using the noisy insulin observations as input. In other words, this strategy produces parameter values which do not make model predictions of glucose and insulin consistent with each other. Furthermore, the exploitation of random variations in input determines the very ability of the algorithm to reproduce characteristic features of the output signal, which the model may have difficulty reproducing when moving from the theoretical or expected input signal.

In order to concretely exemplify the general remarks made so far, Fig. 3 shows four sets of model-reconstructed curves associated with the same data set (in all four panels squares indicate observed insulin concentrations, while circles indicate observed glucose concentrations; solid lines indicate model-predicted insulinemias, dotted lines indicate model-predicted glycemias).

In Fig. 3a the observed points are fitted with the SDM model mentioned above (one-pass fitting, minimizing errors in glycemias and insulinemias simultaneously) and the resulting SDM-predicted time courses are superimposed to the data. The fitting, with six parameters, is reasonably good and the classical second-phase insulin secretion "hump", characteristics of insulin secretion following an IVGTT, is clearly visible.

In Fig. 3b the observed points are fitted with Minimal Model using the 'decoupling approach' [36] (two-pass fitting, using interpolated insulinemia as input for the fit of glycemias, then interpolated glycemia as input for the fit of insulinemia). The predicted curves lie close to the observations (in this set-up eight parameters are free) and second-phase insulin secretion is readily apparent. So far, it would seem that the Minimal Model and the SDM are approximately equivalent in their ability to explain this dataset.

In Fig. 3c the observations are fitted with exactly the same Minimal Model formulation, this time however using a simultaneous, one-pass procedure: glycemias and insulinemias are simultaneously predicted from the model and parameters are adjusted to provide the best overall weighted fit. While the predicted curves still pass through the observed points, no second-phase insulin secretion hump is visible. In fact, this exercise was repeated in 72 subjects [40] and in none of them second-phase insulin secretion was reproduced by the model.

Finally, Fig. 3d shows the original observations and the curves obtained when using simultaneously the classical MM parameter estimates (those obtained by means of the decoupled procedure estimation). In other words, in Fig. 3d the same parameter values obtained in the classical 'decoupling' Minimal Model fit of Fig. 3b are employed; this time, however, instead of using the recorded noisy observations to provide feedback regulation, the actual predictions of the model are used, so that predicted glycemia influences the prediction of insulinemia and

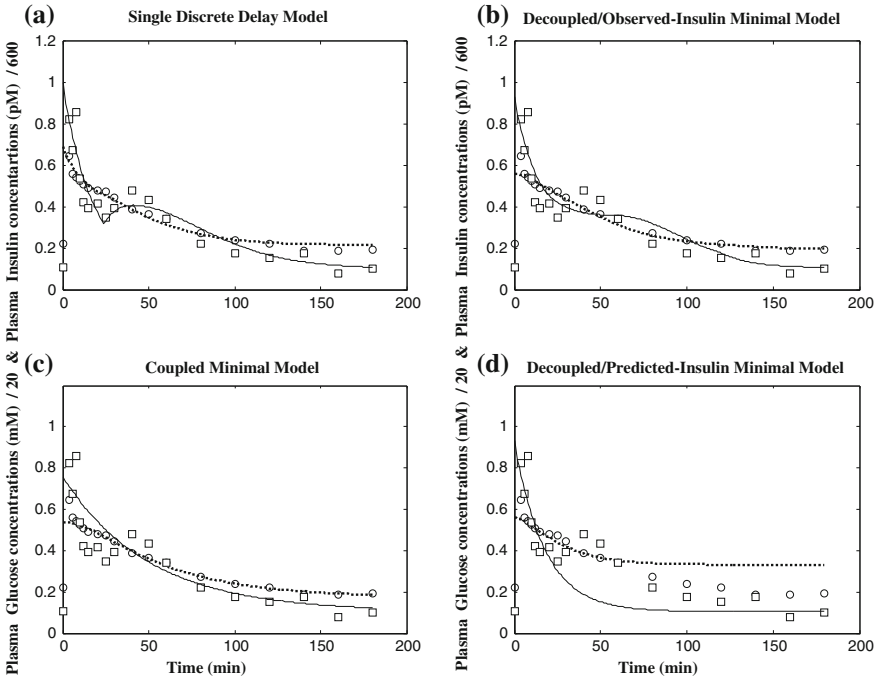


Fig. 3 Comparison among SDM fitting, 'decoupling approach' and one-pass fit for MM. **a** This part figure reports the observed points (*circles* for glycemia and *squares* for insulinemia) of a single experimental subject, fitted with the Single Delay Model (*dotted line* for glucose predictions and *continuous line* for insulin predictions) with one-pass fitting (minimizing errors in glycemias and insulinemias simultaneously). **b** This part figure reports the same observed points as in *panel* **a**, fitted with the Minimal Model using the 'decoupling approach' as suggested by the original Authors (two-pass fitting, using interpolated insulinemia as input for the fit of glycemias, and interpolated glycemia as input for the fit of insulinemia). **c** This part figure reports again the observations fitted with the Minimal Model, this time using a simultaneous, one-pass procedure (minimizing errors in glycemias and insulinemias simultaneously, similarly to what was done in *panel* **a** for the Single Delay Model). **d** This part figure shows, along with the observations, the curves obtained when using the Minimal Model with the parameter values obtained in the 'decoupling approach' of (**b**), using however, instead of the recorded noisy observations to provide feedback regulation, the actual predictions of the Minimal Model itself (predicted glycemia influences the prediction of insulinemia and vice versa), thereby showing the actual dynamics expressed by the Minimal Model given the parameters of (**b**)

vice versa. It can be appreciated how, in this case, predicted curves not only still do not exhibit secondary insulin secretion, but actually fail to approximate the observations.

When fitting a model to observations it is required that the identified model be consistent, i.e. that the functional form of the model, together with the estimated parameter values, reproduce the dynamics actually observed. In this example it can be appreciated that decoupling the feedback and estimating separately its two arms from interpolated observations of the driving variable provides misleading results:

while it would seem that the fit is good (like in Fig. 3b), such good fit actually relies on the specific realization of chance occurrence of errors in the observations. In this way parameters are obtained which can apparently reproduce features (like in this case the second-phase insulin secretion hump), but can do so only by exploiting that experiment's specific observation errors. When these same parameters are used to model the interaction of predicted glycemias and insulinemias on each other (as in Fig. 3d), no such features appear and indeed, actual data fit is very poor, which indicates that we have not been able to identify the dynamics underlying the observations.

In fact, two sources of estimate variability are simultaneously present in this situation: the usual variability due to error in the observations of the modeled (dependent) variable and an additional variability due to random variations of the observations of the observations of the input or forcing function. Moreover, features may appear from model predictions (e.g. in the present case, a secondary insulin secretion hump), which would not appear if the input function were smooth (as is the case in coupled estimation), and which may depend on chance peaks in the input being translated (through unreliable parameter values) into the derived features in output.

5 Conclusions

When attempting to mathematically explain a physiological phenomenon, the numerical parameter estimation phase (easy to conduct by means of standard software) should be preceded by a careful consideration of the qualitative behavior of the model itself, from a purely theoretical, data-independent standpoint. It should be required that the model exhibits behavior consistent with what is already known of the phenomenon under study. For example, it should be checked whether equilibrium points exist and what they are, that the solutions are limited and positive and that they converge back to the equilibrium points from the perturbed state, that oscillations are or are not present, that continuous dependence on parameter values and on initial conditions exists, that the system exhibits the degree of stability that is expected. Good data fitting, even of a plausible-looking model (such as Bolie's model above), is no guarantee that the estimated parameters have the desired meaning relative to the physiological phenomenon under study, unless the model structure is shown to be consistent with that physiology.

The use of interpolated observation data as a substitute for the theoretical driving function of a given model is a risky procedure, which may well mislead the investigator into believing that the model equations themselves, and not accidental error realizations, are able to reproduce characteristic features of the physiology under study. Using interpolated observations in place of actually available model predictions in a coupled system is particularly hazardous, because the decoupling thus introduced will achieve misleadingly close empirical data fitting at the expense of inconsistency of behavior of the (un)coupled variables relative to one

another: one may indeed incorrectly believe to have identified the dynamics of interest, while in fact the estimated parameter values determine another, different system behavior.

A further consequence of using interpolated observations as input or forcing function is the high variability of the parameter estimates obtained, since parameters are chosen, with which the optimizer algorithm will seek to reproduce expected output conditionally on random input.

References

1. Defronzo RA, Tobin JD, Andres R (1979) Glucose clamp technique: a method for quantifying insulin secretion and resistance. Am J Physiol 237:E214–E223
2. Bolie VW (1961) Coefficients of normal blood glucose regulation. J Appl Physiol 16:783–788
3. Ackerman E, Gatewood LC, Rosevear JW, Molnar GD (1965) Model studies of blood-glucose regulation. Bull Math Biophys 27(Suppl:21–37)
4. Subba Rao G, Bajaj JS, Subba Rao J (1990) A mathematical model for insulin kinetics II. Extension of the model to include response to oral glucose administration and application to insulin-dependent diabetes mellitus (IDDM). J Theor Biol 142:473–483
5. Turner RC, Rudenski AS, Matthews DR, Levy JC, O'Rahilly SP, Hosker JP (1990) Application of structural model of glucose-insulin relations to assess beta-cell function and insulin sensitivity. Horm Metab Res Suppl 24:66–71
6. Proietto J (1990) Estimation of glucose kinetics following an oral glucose load. methods and applications. Horm Metab Res Suppl 24:25–30
7. Nomura M, Shichiri M, Kawamori R, Yamasaki Y, Iwama N, Abe H (1984) A mathematical insulin-secretion model and its validation in isolated rat pancreatic islets perifusion. Comput Biomed Res 17:570–579
8. Sluiter WJ, Erkelens DW, Terpstra P, Reitsma WD, Doorenbos H (1976) Glucose tolerance and insulin release, a mathematical approach II. Approximation of the peripheral insulin resistance after oral glucose loading. Diabetes 25:245–249
9. Gatewood LC, Ackerman E, Rosevear JW, Molnar GD, Burns TW (1968) Tests of a mathematical model of the blood-glucose regulatory system. Comput Biomed Res 2:1–14
10. Ceresa F, Ghemi F, Martini PF, Martino P, Segre G, Vitelli A (1968) Control of blood glucose in normal and in diabetic subjects. Studies by compartmental analysis and digital computer techniques. Diabetes 17:570–578
11. Jansson L, Lindskog L, Norden NE (1980) Diagnostic value of the oral glucose tolerance test evaluated with a mathematical model. Comput Biomed Res 13:512–521
12. O'Connor MD, Landahl H, Grodsky GM (1980) Comparison of storage- and signal-limited models of pancreatic insulin secretion. Am J Physiol 238:R378–R389
13. Sturis J, Polonsky KS, Mosekilde E, Van Cauter E (1991) Computer model for mechanisms underlying ultradian oscillations of insulin and glucose. Am J Physiol 260:E801–E809
14. Topp BG, Promislow K, deVries G, Miura RM, Finegood DTA (2000) Model of beta-cell mass, insulin, and glucose kinetics: pathways to diabetes. J Theor Biol 206:605–619
15. Tolic IM, Mosekilde E, Sturis J (2000) Modeling the insulin-glucose feedback system: the significance of pulsatile insulin secretion. J Theor Biol 207:361–375
16. Lenbury Y, Ruktamatakul S, Amornsamarnkul S (2001) Modeling insulin kinetics: responses to a single oral glucose administration or ambulatory-fed conditions. Biosystems 59:15–25
17. Makroglou A, Li J, Kuang Y (2006) Mathematical models and software tools for the glucose-insulin regulatory system and diabetes: an overview. Appl Numer Math 56(3):559–573

18. Bennett L, Gourley S (2004) Asymptotic properties of a delay differential equation model for the interaction of glucose with plasma and interstitial insulin. Appl Math Comput 151:189–207
19. Prager R, Wallace P, Olefsky JM (1986) In vivo kinetics of insulin action on peripheral glucose disposal and hepatic glucose output in normal and obese subjects. J Clin Invest 78:472–481
20. Shapiro ET, Tillil H, Polonsky KS, Fang VS, Rubenstein AH, Van Cauter E (1988) Oscillations in insulin secretion during constant glucose infusion in normal man: relationship to changes in plasma glucose. J Clin Endocrinol Metab 67:307–314
21. Bergman RN, Ider YZ, Bowden CR, Cobelli C (1979) Quantitative estimation of insulin sensitivity. Am J Physiol 236:E667–E677
22. Panunzi S, Palumbo P, De Gaetano A (2007) A discrete single delay model for the intravenous glucose tolerance test. Theor Biol Med Model 4:35
23. Box GEP, Draper NR (1987) Empirical model building and response surfaces. Wiley, New York, p 424
24. Pedrini MT, Niederwanger A, Kranebitter M, Tautermann C, Ciardi C, Tatarczyk T, Patsch JR (2006) Postprandial lipaemia induces an acute decrease of insulin sensitivity in healthy men independently of plasma NEFA levels. Diabetologia 49:1612–1618
25. Penesova A, Rovensky J, Zlnay M, Dedik L, Radikova Z, Koska J, Vigas M, Imrich R (2005) Attenuated insulin response and normal insulin sensitivity in lean patients with ankylosing spondylitis. Int J Clin Pharmacol Res 25:107–114
26. Hsieh CH, Hung J, He CT, Lee CH, Hung SC, Wang YK, Kuo SW, Pei D (2005) The capability of glucose toxicity on severe type 2 diabetes. Endocr Res 31:149–158
27. Ranganathan G, Unal R, Pokrovskaya I, Yao-Borengasser A, Phanavanh B, Lecka-Czernik B, Rasouli N, Kern PA (2006) The lipogenic enzymes DGAT1, FAS, and LPL in adipose tissue: effects of obesity, insulin resistance, and TZD treatment. J Lipid Res 47:2444–2450
28. Fernandez-Real JM, Lopez-Bermejo A, Castro A, Casamitjana R, Ricart W (2006) Thyroid function is intrinsically linked to insulin sensitivity and endothelium-dependent vasodilation in healthy euthyroid subjects. J Clin Endocrinol Metab 91:3337–3343
29. Henderson DC, Copeland PM, Borba CP, Daley TB, Nguyen DD, Cagliero E, Evins AE, Zhang H, Hayden DL, Freudenreich O, Cather C, Schoenfeld DA, Goff DC (2006) Glucose metabolism in patients with schizophrenia treated with olanzapine or quetiapine: a frequently sampled intravenous glucose tolerance test and minimal model analysis. J Clin Psychiatry 67:789–797
30. Cagnacci A, Tirelli A, Renzi A, Paoletti AM, Volpe A (2006) Effects of two different oral contraceptives on homocysteine metabolism in women with polycystic ovary syndrome. Contraception 73:348–351
31. Treiber KH, Boston RC, Kronfeld DS, Staniar WB, Harris PA (2005) Insulin resistance and compensation in thoroughbred weanlings adapted to high-glycemic meals. J Anim Sci 83:2357–2364
32. Bergman RN, Steil GM, Bradley DC, Watanabe RM (1992) Modeling of insulin action in vivo. Annu Rev Physiol 54:861–883
33. Caumo A, Giacca A, Morgese M, Pozza G, Micossi P, Cobelli C (1991) Minimal models of glucose disappearance: lessons from the labelled IVGTT. Diabet Med 8:822–832
34. Callegari T, Caumo A, Cobelli C (2003) Bayesian two-compartment and classic single-compartment minimal models: comparison on insulin modified IVGTT and effect of experiment reduction. IEEE Trans Biomed Eng 50:1301–1309
35. Toffolo G, Bergman RN, Finegood DT, Bowden CR, Cobelli C (1980) Quantitative estimation of beta cell sensitivity to glucose in the intact organism: a minimal model of insulin kinetics in the dog. Diabetes 29:979–990
36. Pacini G, Bergman RN (1986) MINMOD: a computer program to calculate insulin sensitivity and pancreatic responsivity from the frequently sampled intravenous glucose tolerance test 172. Comput Methods Programs Biomed 23:113–122

37. De Gaetano A, Arino O (2000) Mathematical modelling of the intravenous glucose tolerance test. J Math Biol 40:136–168
38. Palumbo P, Panunzi S, De Gaetano A (2007) Qualitative behavior of a family of delay-differential models of the glucose-insulin system. Discrete Continuous Dyn Syst Ser B 7:399–424
39. Dalla Man C, Caumo A, Cobelli C (2002) The oral glucose minimal model: estimation of insulin sensitivity from a meal test 38. IEEE Trans Biomed Eng 49:419–429
40. Panunzi S, De Gaetano A, Mingrone G (2010) Advantages of the single delay model for the assessment of insulin sensitivity from the intravenous glucose tolerance test. Theor Biol Med Model 7:9

Simulation Models for In-Silico Evaluation of Closed-Loop Insulin Delivery Systems in Type 1 Diabetes

Malgorzata E. Wilinska and Roman Hovorka

Abstract This chapter presents simulation models created to support the development of closed-loop insulin delivery systems in type 1 diabetes. Such models, also known as 'virtual patient' models, represent an input-output relationship between insulin delivery and other external inputs such as meals or exercise, and the resulting glucose response. It is argued that these simulation models are an essential prerequisite for an accelerated development of the artificial pancreas systems in various populations of type 1 diabetes ranging from children to adults and pregnancies. The present review provides a general introduction to the models of glucose regulation in type 1 diabetes and then proceeds to discussing the individual submodels of glucose kinetics and insulin action, subcutaneous insulin kinetics, subcutaneous glucose kinetics, glucose absorption from the gut, and the exercise effect on the glucose kinetics. Finally, several important virtual-patient models used for in silico testing of glucose controllers are reviewed.

1 Introduction

Mathematical models of glucose regulation have been studied since 1960s [1] utilising early experimental data to measure glucose production and glucose disposal [2]. The complex dynamic behaviour of the pancreatic beta cell and the

M. E. Wilinska · R. Hovorka (✉)
Department of Paediatrics, University of Cambridge, P.O. Box 289, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK
e-mail: rh347@cam.ac.uk

M. E. Wilinska
e-mail: mew37@cam.ac.uk

M. E. Wilinska · R. Hovorka
Wellcome Trust-MRC Institute of Metabolic Science, University of Cambridge Metabolic Research Laboratories, P.O. Box 289, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK

V. Marmarelis and G. Mitsis (eds.), *Data-driven Modeling for Diabetes*, Lecture Notes in Bioengineering, DOI: 10.1007/978-3-642-54464-4_6,

intricacy and efficiency of the glucose-insulin feedback loop inspired mathematicians to describe, in a mathematical formalism, pathophysiological pathways and their interactions, and to simulate the responses of the glucoregulatory system to various stimuli. By 1980s, a wide range of models had been developed improving our understanding of glucose homeostasis and supporting the diagnosis and management of diabetes [3, 4].

Mathematical models in diabetes are found in a variety of applications. Many are used for educational purposes. Interacting with the models, healthcare professionals, physiologists but also subjects with diabetes, gain quantitative understanding of glucose regulation. The AIDA educational package [5] or the KADIS system [6] are typical examples of such models. The Diabetes Advisory System (DIAS) [7], or long-term treatment/health policy outcomes such as the Archimedes model [8], have been developed to support treatment decisions in diabetes, insulin-dosing decisions in particular. Models have been used to estimate insulin sensitivity and beta cell function. The widely used Bergman's minimal model [9] is the most popular method to estimate insulin sensitivity during the intravenous glucose tolerance test (IVGTT) complementing the more experimentally demanding glucose clamp technique [10]. Models can also be used to support the development of new treatment targets and new drugs for the management of diabetes [11].

Fuelled by the availability of continuous glucose monitors [12], a focused research is underway to develop closed-loop insulin delivery systems also known as the artificial pancreas [13, 14]. The artificial pancreas consists of three components, a continuous glucose monitor (CGM), a control algorithm, and an insulin pump. While the insulin pump technology is well advanced, and widely adopted, the accuracy, reliability, and limited experience with currently available continuous glucose monitors is still a potential roadblock for the commercialisation of closed-loop systems [15]. Existing control algorithms also need to be refined to allow reliable adaptation to an individual subject and to assure a safe operation with a minimal risk of low and high glucose levels. The artificial pancreas has the potential to revolutionise the treatment of type 1 diabetes within a decade [15–17]. However, substantial clinical testing has to be done before commercial closed-loop systems are made available and become widespread.

The present review provides a general introduction to the glucoregulatory models. This introduction is followed by a description of submodels of the glucose kinetics and insulin action, subcutaneous insulin kinetics, subcutaneous glucose kinetics, glucose absorption from the gut, and the exercise effect on the glucose kinetics. Finally, several important simulation models used for in silico testing of glucose controllers in type 1 diabetes are described. It is argued that these simulation models are an essential prerequisite for an accelerated development of the artificial pancreas in various populations of type 1 diabetes ranging from children to adults and pregnancies.

For interested readers, models of glucose regulation have been reviewed previously by Mari [18] and Parker and Doyle [19], the latter focusing on control-relevant models. Makroglou et al. [20] provided an interesting overview of mathematical models of the glucose-insulin regulatory system from the point of

view of their mathematical formulations and parameter estimation techniques. The review also includes a survey of available software packages for the numerical analysis and simulation. Readers interested in closed-loop control algorithms are referred to reviews by Parker et al. [21], Bellazzi et al. [22], Steil et al. [23] and Hovorka [13, 14].

2 Models and Virtual Patients

Mathematical models of glucose regulation can be broadly classified into two categories, data-driven models and conceptual models, although the distinction is not always clear cut. Generally, the data-driven models are simple. The number of model equations is deduced and parameters are estimated on individual basis from experimental data using the principle of parsimony. The conceptual models utilise holistic knowledge of the physiology and metabolic processes. Each organ/tissues can be treated as a separate compartment with its own mass-balance equations and a detailed description of the metabolic and kinetic processes. These models are more complex, numerically demanding, and the estimation of individual parameters from the experimental data is hindered by identifiability issues [3, 4].

A simulation model structure is usually motivated and constrained by physiological considerations and a set of model parameters. The model parameters can be determined in different ways influencing the capabilities of the simulation model. In some cases, average population values have been derived from the literature as in models by Bolie [1], Tiran et al. [24] or Sorensen [25] allowing population average but not individual simulations to be carried out. In other cases such as the minimal model [9] or the model by Hovorka et al. [26], the majority of individual parameters are identified by fitting subject's experimental data and selecting physiologically feasible parameters that result in the best fit to the data. Such models allow individual predictions to be made and population variability to be inferred.

In certain situations, such as when testing treatment concepts or representing physiological knowledge, a prediction of an average subject with diabetes is sufficient. Then, a single parameter set representing the average subject is required as is often the case with conceptual models of glucose-insulin dynamics, see Tiran and colleagues [24, 27], Guyton et al. [28], and Sorensen [25].

In order to estimate model parameters from individual data sets, models with a lower complexity are required. Cramp and Carson [29] and then Cobelli et al. [30] reduced the structural complexity by proposing lumped compartment models utilising the principle of parsimony for model selection. Other models in the spectrum represent a trade-off between the detailed physiological insights of conceptual models and the relative ease of parameter estimation associated with lower-order models.

More recently developed simulation models are built from smaller, already existing submodels that represent subsystems of glucose regulation. At the core, such simulation models often include a submodel of the glucose kinetics and

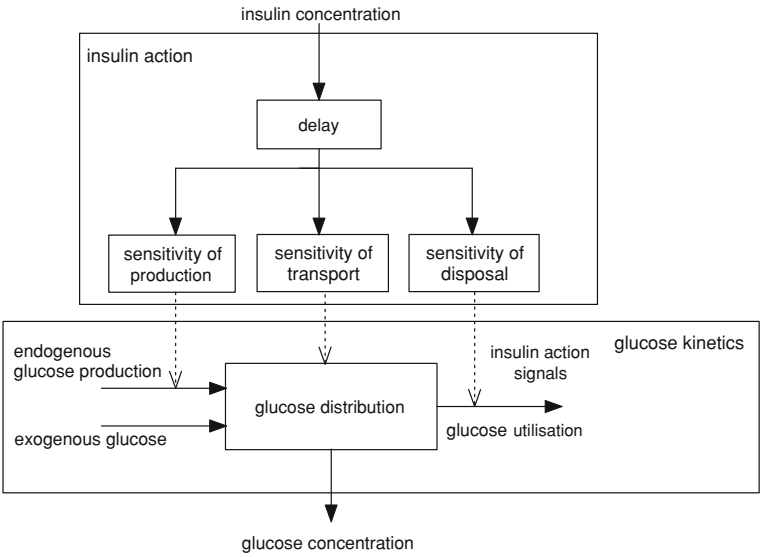


Fig. 1 Representation of the glucose-insulin system; *dotted line arrows* indicate insulin control; adapted from [17] and [25]

insulin action, see Fig. 1. The number of other interacting submodels depends on the input/output requirements. If the simulation model is to be used for evaluating a closed-loop controller that drives a subcutaneous insulin infusion pump and is informed by subcutaneous glucose measurements, the necessary submodels are a model of the subcutaneous insulin kinetics and a model of the interstitial glucose kinetics. Other submodels may include a meal model to represent glucose absorption from the gastrointestinal tract and an exercise model to represent the effect of physical activity on the glucose concentration.

A simulation model that represents an input-output relationship between insulin delivery and other external inputs such as meals or exercise, and the resulting glucose response is often referred to as a 'virtual patient'. This type of a model constitutes an essential part of a computer simulation environment referred to as the simulator. A simulation model can utilise a single parameter set representing an 'average patient' or a multiple parameter set representing a population of 'virtual subjects'.

The concept of the 'virtual population' is illustrated in Fig. 2. The design is driven by the simulator's intended purpose, i.e. to predict an outcome of a clinical trial conducted in a number of subjects with diabetes producing population-relevant results.

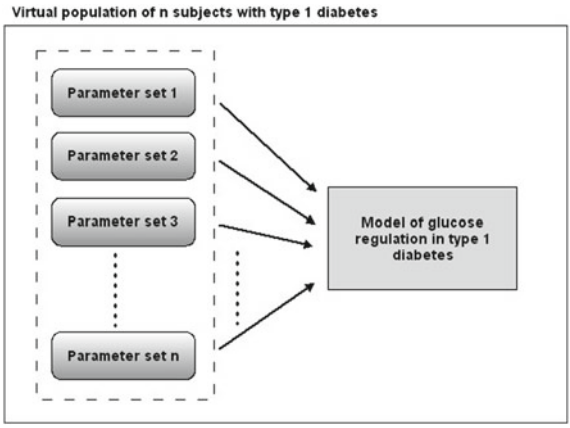


Fig. 2 A population of 'virtual subjects' with type 1 diabetes comprises a simulation model of the glucose regulation accompanied by n parameter sets representing n 'virtual subjects'; virtual populations can be further subdivided into subgroups of adults, children, pregnant women etc. with type 1 diabetes

3 Models of Physiological Subsystems of Glucose Regulation

3.1 Models of Glucose Kinetics and Insulin Action

The conceptual paradigm of models representing the glucose kinetics and insulin action is illustrated in Fig. 1. The model describes the distribution, production and utilisation of glucose, and the control by insulin. The model of insulin action typically includes a remote compartment to represent the delay in insulin action.

One of the first models of the glucose-insulin system to be used in simulations was a linear model developed by Ackerman et al. [31]. Although this four-parameter, two ordinary differential equations model was a clear oversimplification of the whole body glucose regulatory system, the strength of the model was that its parameters could be identified from clinical data [32].

An early example of a low-order model is the minimal model by Bergman et al. [9]. It was developed to estimate insulin sensitivity and glucose effectiveness but the model has also found numerous applications in the domain of glucose control. The structural simplicity and the ability to estimate significant physiological parameters led to its popularity among the clinicians. The model is widely used to estimate insulin sensitivity in clinical and epidemiological studies [33–36] and is referred to in over 1,000 publications [37]. Its usefulness and popularity sparked great interest in mathematical modelling of glucose regulation. This model with two state variables describes plasma glucose dynamics based on plasma insulin feeding into a remote insulin compartment. The original model was developed using data obtained from the intravenous glucose tolerance test (IVGTT). The model has two compartments representing plasma glucose and remote insulin. The rate change of glucose in this model is the difference between the net hepatic glucose balance and the disappearance of glucose into peripheral tissues. A nonlinear regression analysis is applied to estimate four model parameters with

insulin sensitivity (S_I) and glucose effectiveness (S_G) being the main indices of interest.

Several authors have documented the shortcomings of the minimal model [38–41] such as its inability to separate glucose production from glucose disposal. Only the net hepatic glucose balance can be described. As a consequence, the metabolic indices S_I and S_G measure not only the effect of insulin and glucose on glucose disposal but also their inhibitory effect on hepatic glucose production. Other concerns regarding the minimal model include poor precision of parameter estimates and unsatisfactory reproducibility of S_I [41]. In order to overcome these limitations, the labelled IVGTT has been devised and tracer minimal models developed [42, 43]. The tracer models allowed the estimation of tissue sensitivity to insulin and glucose effectiveness without distortion of glucose and insulin effects on hepatic glucose production. Despite these improvements, Caumo and Cobelli [44] observed that the minimal model gives a physiologically implausible prediction of the hepatic glucose production. The source of the error is likely due to under-modelling of the glucose kinetics by a single compartment [38].

From a simulation point of view, the glucose effectiveness of approximately 0.02 per min, as estimated in healthy subjects, results in an overestimated self-regulatory capability of the glucoregulatory system especially when used in the original formulation with basal glucose set at normoglycemia. Figure 3 illustrates the exaggerated glucose self-regulation at this nominal value of glucose effectiveness. When used to evaluate glucose controllers, the overestimated self-regulatory capability provides an overoptimistic assessment of the controller as most of the control is exerted through glucose effectiveness and not through the controller directed insulin delivery. Glucose effectiveness at or below 0.01 per min is needed to represent appropriately the self-regulatory capability of the glucoregulatory system.

Glucose kinetic models with two glucose compartments have been proposed in the last decade [44–46] to improve the minimal model. The total number of parameters increased to six and the parameters could no longer be uniquely estimated from the data. The problem can be solved by using tracer glucose bolus and physiological constraints [44, 45] or by applying the Bayesian estimation [46]. The estimation of the three insulin sensitivities of glucose uptake in the peripheral tissues, the sensitivity of hepatic glucose production and the sensitivity of glucose transport/distribution was first attempted by Ferranini et al. [47]. The partitioning of the three sensitivities was fully accomplished by Hovorka et al. [26] using a two-compartment glucose model for the dual-tracer IVGTT. The model described the steady state as well as dynamic conditions and was validated in healthy subjects.

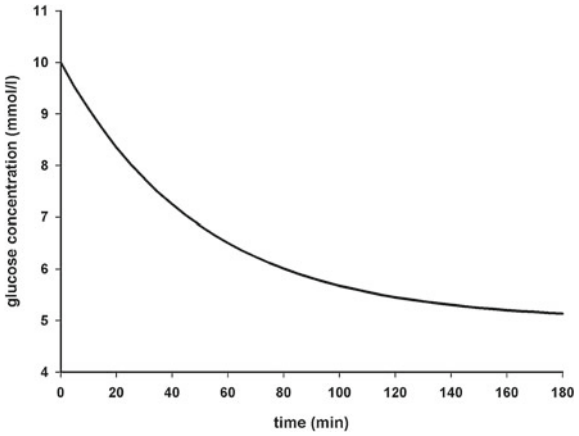


Fig. 3 Glucose concentration curve obtained using the minimal model with glucose effectiveness of 0.02 min^{-1}, starting glucose value of 10 mmol/l, basal glucose of 5 mmol/l, and a constant insulin concentration at the basal insulin level. At this nominal value of glucose effectiveness the minimal model predicts that elevated glucose rapidly returns, with a half time of 35 min, to the 'basal glucose level' without the need to increase the insulin concentration. This ill fits with glucose concentration curves recorded in subjects with type 1 diabetes

3.2 Models of Subcutaneous Insulin Kinetics

Subcutaneous insulin absorption is a complex process influenced by factors including the associated state of insulin, its concentration, injected volume, injection site and depth, and tissue blood flow [48–51]. Several mostly compartmental models have been proposed for various insulin preparations. A comprehensive review of these models by Nucci and Cobelli [52] compares six models of the subcutaneous insulin kinetics ranging from simple compartmental models [53–56] through empirically derived [57] to complex conceptual models [58, 59]. In all models except that by Hovorka et al. [60], plasma insulin is assumed to be a single compartment. The compartmental models differ mainly in the way they consider the absorption of subcutaneous insulin.

Kobayashi et al. [53] proposed a two-compartment plasma and subcutaneous tissue model with a delay to represent insulin absorption. Kraegen et al. [54] proposed a model with two subcutaneous compartments and five parameters including the insulin degradation rate in the subcutaneous tissue. A simple three-compartment model by Puckett et al. [55] is widely used. The model's three parameters were estimated from clinical data collected in subjects with diabetes. All three models were identified for short acting soluble insulin. Rapid acting insulin analogues were not available at the time of these publications. Few years later, Shimoda et al. [56] proposed a three-compartment model of a soluble insulin and monomeric insulin analogue to be used in the artificial pancreas. The model was fitted using a 3-h block of data obtained from ten subjects with type 1 diabetes.

Contradictory to Kraegen et al. [54], the insulin degradation in the subcutaneous tissue estimated by this model was not negligible. The effect of insulin degradation was corroborated by Wilinska et al. [61] who evaluated ten alternative models of the subcutaneous insulin kinetics of varying complexity. Model selection was based on the principle of parsimony and physiological plausibility using data collected in subjects with type 1 diabetes and rapid acting insulin analogue. The selected model indicated the existence of a fast and slow absorption channel and the presence of the local insulin degradation in the subcutaneous tissue.

3.3 Models of Subcutaneous Glucose Kinetics

Understanding subcutaneous glucose dynamics is important as CGM devices employed in the closed-loop systems measure glucose concentration in the subcutaneous tissue. The ability to describe the dynamics of the relationship between the subcutaneous and plasma glucose depends on the understanding and quantification of the physiological processes in the interstitial fluid (ISF) surrounding the adipose tissue. Plasma glucose is separated from the interstitial glucose by a capillary wall, and hence changes in ISF glucose are related to changes in plasma glucose by the rate of diffusion across the capillary wall and by the rate of glucose removal from the ISF representing the glucose uptake by adipose tissue. Studies with the microdialysis [62] and open-flow microperfusion [63] have demonstrated the presence of a gradient between plasma and interstitial glucose of a varying magnitude [64]. An equilibration time delay between plasma and ISF glucose has been reported by a number of studies [65, 66]. Models of subcutaneous glucose kinetics have been postulated by Bonnecaze and colleagues [67, 68]. Wilinska et al. [69] postulated nine alternative models to account for the temporal variations of the plasma to ISF glucose gradient. All models were fitted to experimental data collected in subjects with type 1 diabetes. Model which best represented the clinical data included zero-order glucose disposal from the ISF and an insulin effect on the glucose transfer from the plasma to the interstitial fluid.

3.4 Models of Glucose Absorption from the Gut

Following a meal, glucose is absorbed in the upper gastrointestinal tract, transported to the splanchnic bed, mostly the liver, and then to the peripheral circulation. Worthington [70] postulated a simple two-compartment model of glucose absorption from the gut with identical transfer rate coefficients between the compartments. The model was shown to be adequate in representing the glucose rate of appearance from the gut. Arleth et al. [71] split the gut absorption into three terms, each one corresponding to a class of carbohydrates with different absorption rates: sugars, a fast absorbing starch, and a slow absorbing starch. More recently,

Dalla Man et al. [72] proposed a three compartment non-linear model with two compartments representing the stomach (solid and liquid phase) and the third compartment representing the intestine. The model assumes a constant rate of the intestinal absorption but describes gastric emptying rate to be dependent on the total amount of nutrient in the stomach.

3.5 Models of Exercise Effect on Glucose Kinetics

Exercise has been associated with a transient increase in insulin sensitivity and insulin-dependent glucose uptake in the muscle through augmented availability of glucose transporters [73–75]. In type 1 diabetes, a higher glucose variability associated with exercise can be difficult to manage leading to an increased risk of hypoglycaemia [76]. This difficulty poses a major concern in the field of closed-loop glucose control where mathematical models able to predict dynamic changes of glucose during exercise could prove very useful.

To date very few system-level models have been proposed [77–81]. The two models with identifiable parameters and a potential to be used in closed-loop glucose control are those by Derouich et al. [78] and Breton [81]. Both models are extensions of the minimal model of glucose kinetics [82]. Breton's novel approach was to detect and quantify exercise through changes in the heart rate. Hernandez–Ordonez [79] extended the ability of a well-known Sorensen model [25] to reproduce variations in the glucose concentration induced by exercise in type 1 diabetes. The extended model was verified using experimental data collected during a light and moderate intensity exercise. The authors found a good agreement with experimental data from the literature. The effect of high intensity exercise was simulated by extrapolation. However, the physiological differences between moderate versus intense exercise are still poorly understood [83]. Chassin et al. [84] advocate a wider use of real-time continuous glucose monitoring devices to improve current modelling efforts.

4 Simulation Models in Type 1 Diabetes

Recent technological advances in real-time continuous glucose monitoring fuel the development of a wearable artificial pancreas [23, 85–87], which consists of a CGM device to measure glucose concentration, a titrating control algorithm to compute the amount of insulin to be delivered, and an insulin pump delivering the advised insulin doses.

Algorithms for closed loop insulin delivery in subjects with type 1 diabetes can be designed and tuned empirically, and evaluated during clinical testing. However, a validated simulation model of glucose regulation in type 1 diabetes accelerates the design and the evaluation process. Recent efforts focus around the development of a

simulation environment based on a validated model of type 1 diabetes and designed specifically for testing and evaluating closed-loop glucose controllers. Examples of such simulation environments and their related models are presented below.

4.1 Hovorka Simulation Model

The European Commission funded Advanced Insulin Infusion using a Control Loop (Adicol) Project which run from 2000 for 3 years and adopted an adaptive nonlinear model predictive controller (MPC) to develop a minimally invasive closed-loop system with meal announcement [88]. The development of the MPC-based glucose controller was facilitated and greatly accelerated by a metabolic simulator [89].

The simulation model described by Hovorka and colleagues [90] is based on a compartment model of glucose kinetics and insulin action published by the same author [26]. Other subsystems include two compartment models of the subcutaneous insulin and subcutaneous glucose kinetics as well as a two compartment model of the glucose absorption from the gastro-intestinal tract. An important property of this simulation environment is its ability to represent between and within subject variability. The between subject variability is represented by a population of 18 virtual subjects with type 1 diabetes. The model parameters were obtained either from clinical studies in subjects with type 1 diabetes or from informed probability distributions. The within individual variability of the glucoregulatory system was implemented by superimposing sinusoidal oscillations on a subset of model parameters. A possible weakness of Hovorka virtual patient model is its simple representation of glucose absorption from the gut which may need to be refined. In addition, the within subject variability may also require further refinement. A clear strength of this virtual population of 18 subjects with type 1 diabetes is that it was validated by comparing the simulation study predictions against a clinical study evaluating overnight closed-loop insulin delivery in young people with T1D using a model predictive controller [91]. The simulation model is being used by Hovorka's group in the development of a prototype closed-loop insulin delivery system for children and adolescents [92, 93], adults [94] and pregnant women [95] with type 1 diabetes.

4.2 Medtronic Virtual Patient Model

Medtronic's closed loop project adopts the proportional integral derivative (PID) approach to develop an external physiological insulin delivery (ePID) system based on the Guardian Real-Time Medtronic subcutaneous glucose monitor (Medtronic Minimed Northridge CA, USA) [96] and subcutaneous delivery of a rapid acting insulin analogue. Using data from previously completed closed-loop

studies in subjects with type 1 diabetes, the Medtronic team derived a simulation model such that the original study could be recreated using the computer simulation alone [97]. The simulation model, referred to as the Identifiable Virtual Patient (IVP) model, used Bergman's minimal model at its core interacting with two compartment models of the subcutaneous insulin kinetics and the meal absorption. Diurnal variations of minimal model parameters such as S_I (insulin sensitivity), S_G (glucose effectiveness at zero insulin concentration) and the endogenous glucose production were introduced whenever the root mean square error between the model-predicted and the measured glucose exceeded 22^2 mg/dl. Ten glucose profiles fitted in this study were identified as the 'virtual population'. Validation of the IVP model was carried out by simulating previously conducted clinical trials in adults and children with type 1 diabetes (independent subjects) and reproducing the outcomes of these studies [98]. Medtronic plans to use this relatively small in size virtual population to evaluate and optimise new closed loop insulin delivery algorithms. Although the time-variant parameters can be viewed as the strength of this simulation model, it is limited by the rather simplistic representation of the glucose kinetics by Bergman's minimal model including a short duration of insulin action and overestimation of glucose effectiveness.

4.3 Dalla Man Simulation Model

Cobelli's group in Padova, Italy, developed a simulation model of glucose-insulin interactions utilising data collected in 204 normal subjects who underwent a triple tracer meal protocol [99]. The application of glucose tracers allowed glucose and insulin fluxes during a meal to be calculated [100]. The simulation model is made up of a number of parsimonious submodels describing the various unit processes that have been identified using a forcing function strategy. There are two main subsystems in the model. The glucose subsystem is described by a two compartment model [101] as is the insulin subsystem [102]. The unit process models were identified from average data with a forcing function strategy; 35 parameters of the normal subject were estimated.

The simulation model has been employed to simulate a typical day of a normal subject with three meals. To account for diurnal variations in insulin sensitivity and beta cell responsiveness, it was assumed that insulin sensitivity is 25 % lower during the evening meal compared to the breakfast and the lunch and beta-cell responsiveness is 25 % lower during the lunch and the evening meal compared to the breakfast. The main novelty of this simulation model is a more detailed representation of glucose transit through the gastro-intestinal tract [72]. The main weakness however is that the diurnal variations of certain model parameters have not been modelled.

Although the original model was identified using data collected in normal subjects, it is also being used to simulate subjects with type 1 diabetes [37]. In this model, described by 26 parameters, the authors substituted the insulin secretion

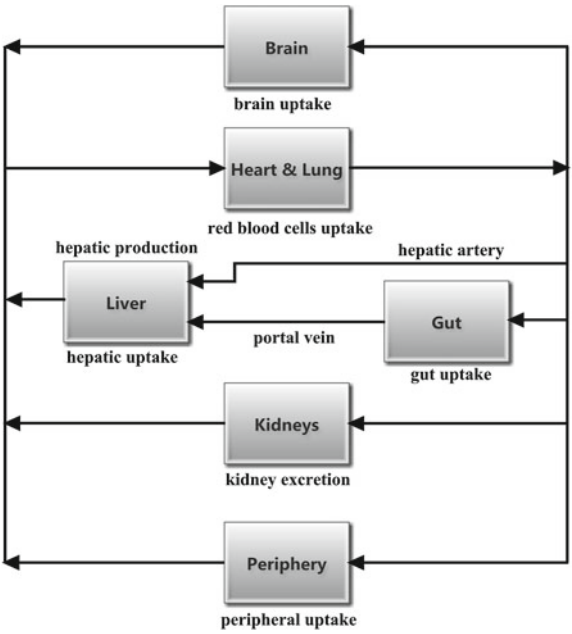


Fig. 4 A flow diagram of the Sorensen model with its six physiological compartments

model with the model of subcutaneous insulin kinetics. In order to account for the higher basal glucose in type 1 diabetes, the endogenous glucose production was increased. Other parameters were kept identical. The virtual patient model, also known as the UVA (University of Virginia) simulator, has been accepted by the Federal Drug Administration to replace animal testing of glucose controllers.

4.4 Sorensen Model

Sorensen model [25] belongs to the class of complex physiologically-based compartment models. The model, based on earlier work by Guyton et al. [28], divides the body into six physiologic compartments: (1) the brain representing the central nervous system (2) the heart and the lung, which represent the rapidly mixing volumes of the heart, the lung and the arteries (3) the periphery, which includes the skeletal muscle and adipose tissue (4) the gut (5) the liver and (6) the kidneys, see Fig. 4. Glucose and insulin subsystems are considered separately, with coupling through metabolic effects.

The model was originally developed to represent the healthy subject utilising 22 nonlinear differential equations including three equations to describe the endogenous insulin secretion. In order to simulate a subject with type 1 diabetes, the insulin secretion term is omitted resulting in a model comprising 19 differential equations and 44 parameters. The parameter values were derived from the

literature and hence could only represent a nominal "average" virtual subject with type 1 diabetes. As all the parameters of this model are time-invariant the model fails to represent the within subject variability. Sorensen model was further developed by Parker et al. [103, 104] to test glucose controllers.

4.5 Fabietti Model

Another group from Italy based at the University of Perugia developed a model of the insulin and glucose dynamics in type 1 diabetes to facilitate the design and evaluation of control algorithms for external artificial pancreas using the subcutaneous route [105]. The model is based on a modified Bergman's minimal model. The endogenous insulin secretion is substituted by subcutaneously delivered exogenous insulin and the glucose kinetics is represented by two instead of a single compartment. External inputs of the model such as meals and intravenous glucose boluses have been added together with the submodel of the glucose absorption from the gastro-intestinal tract.

An interesting feature of the model is the sinusoidal representation of the circadian variability of insulin sensitivity. The amplitude and the phase of the circadian rhythm are estimated 'off-line' to characterise an individual subject. The circadian variation is not considered if the model is to be used in a self-tuning controller. Another 4 out of 14 model parameters are estimated from clinical data. These include insulin sensitivity, a constant related to the plasma insulin distribution volume, and a time constants of the insulin diffusion in the plasma and the remote insulin compartments. Most of the remaining parameters are obtained from the literature or by fitting published data. Although circadian variability of insulin sensitivity can be viewed as the strength of this simulation model, the model is yet another example of a nominal average virtual patient model.

4.6 Comparison of Virtual Patient Models

Five well known simulation models have been presented. The models vary in their complexity and the parameter estimation methods. Two of the models represent the 'average virtual patient' with type 1 diabetes while the remaining simulation models comprise a population of 'virtual patients' with the ability to represent between subject variability. Table 1 summarises parameter derivation methods and identifies the time variant parameters of the five 'virtual patient' simulation models.

Table 1 Virtual patient models and their parameters

Model	Parameters	Parameters			Reference
		From clinical data	From literature	Time-variant	
Hovorka	21	8	13	11	[90]
Medtronic	9	9	–	3	[97]
Dalla Man	30	26	4	–	[37, 99]
Sorensen	44	–	44	–	[25]
Fabietti	14	6	8	1	[105]

5 Summary

A variety of models have been proposed to describe the glucose and insulin kinetics, and the glucose regulation in subjects with type 1 diabetes. Building on this experience, research groups are developing 'virtual patient' models used to support and accelerate the development of an artificial pancreas. Several models exist but most are still being improved and refined. Further research is needed to increase physiological understanding of and to describe more accurately in mathematical terms within subject variability over a short (hours to a day) to longer (days to weeks) time periods. Specific challenges include improving the model of glucose appearance from the gut following ingestion of complex, slowly and rapidly absorbed meals including day to day variability of gut absorption. Validated virtual parameter sets, which characterise different subpopulations of type 1 diabetes, such as children, adolescents, and diabetes pregnancies are also needed. Finally, validation methods need to be clarified and adopted by all research groups developing virtual patient models for the simulators to gain recognition by the clinical community.

References

1. Bolie VW (1960) Coefficients of normal blood glucose regulation. J Clin Invest 39:783–788
2. Steele R (1959) Influences of glucose loading and injected insulin on hepatic glucose output. Ann NY Acad Sci 82:420–430
3. Cobelli C, Bergman RN (1981) Carbohydrate Metabolism: quantitative physiology and mathematical modelling. Wiley, New York
4. Carson ER, Cobelli C, Finkelstein L (1983) The mathematical modeling of metabolic and endocrine systems, 1st edn. Wiley, New York
5. Lehmann ED, Deutsch T, Roudsari AV, Carson ER, Sonksen PH (1993) Validation of a metabolic prototype to assist in the treatment of insulin-dependent diabetes mellitus. Med Inform 18:83–101
6. Rutscher A, Salzsieder E, Fischer U (1994) KADIS—model-aided education in type-i diabetes. Comput Methods Programs Biomed 41:205–215

7. Andreassen S, Benn JJ, Hovorka R, Olesen KG, Carson ER (1994) A probabilistic approach to glucose prediction and insulin dose adjustment: description of metabolic model and pilot evaluation study. Comput Methods Programs Biomed 41:153–165
8. Eddy DM, Schlessinger L (2003) Archimedes—a trial-validated model of diabetes. Diab Care 26:3093–3101
9. Bergman RN, Phillips LS, Cobelli C (1981) Physiologic evaluation of factors controlling glucose tolerance in man: measurement of insulin sensitivity and β-cell glucose sensitivity from the response to intravenous glucose. J Clin Invest 68:1456–1467
10. DeFronzo RA, Tobin JD, Andres R (1979) Glucose clamp technique: a method for quantifying insulin secretion and resistance. Am J Physiol 237:E214–E223
11. Zheng Y, Kreuwel HTC, Young DL, Shoda LKM, Ramanujan S, Gadkar KG et al (2007) The virtual NOD mouse—applying predictive biosimulation to research in type 1 diabetes. How do we best employ animal models for type 1 diabetes and multiple sclerosis? 1103: 45–62
12. Klonoff DC (2005) Continuous glucose monitoring—roadmap for 21st century diabetes therapy. Diab Care 28:1231–1239
13. Hovorka R (2006) Continuous glucose monitoring and closed-loop systems. Diabet Med 23:1–12
14. Hovorka R (2008) The future of continuous glucose monitoring: closed loop. Curr Diab Rev 4:269–279
15. Hovorka R, Wilinska ME, Chassin LJ, Dunger DB (2006) Roadmap to the artificial pancreas. Diab Res Clin Pract 74:S178–S182
16. Cobelli C, Renard E, Kovatchev B (2011) Artificial pancreas: past, present, future. Diabetes 60:2672–2682
17. Kumareswaran K, Evans ML, Hovorka R (2012) Closed-loop insulin delivery: towards improved diabetes care. Discov Med 13:159–170
18. Mari A (2002) Mathematical modeling in glucose metabolism and insulin secretion. Curr Opin Clin Nutr Metab Care 5:495–501
19. Parker RS, Doyle FJ (2001) Control-relevant modeling in drug delivery. Adv Drug Deliv Rev 48:211–228
20. Makroglou A, Li J, Kuang Y (2006) Mathematical models and software tools for the glucose-insulin regulatory system and diabetes: an overview. Appl Numer Math 56:559–573
21. Parker RS, Doyle FJ III, Peppas NA (2001) The intravenous route to blood glucose control. IEEE Eng Med Biol Mag 20:65–73
22. Bellazzi R, Nucci G, Cobelli C (2001) The subcutaneous route to insulin-dependent diabetes therapy. IEEE Eng Med Biol Mag 20:54–64
23. Steil GM, Panteleon AE, Rebrin K (2004) Closed-loop insulin delivery-the path to physiological glucose control. Adv Drug Deliv Rev 56:125–144
24. Tiran J, Galle KR, Porte D (1975) Simulation-model of extracellular glucose distribution in human body. Ann Biomed Eng 3:34–46
25. Sorensen JT (1985) A physiologic model of glucose metabolism in man and its use to design and assess improved insulin therapies for diabetes. Massachusetts Institute of Technology Ph.D, Cambridge
26. Hovorka R, Shojaee-Moradie F, Carroll PV, Chassin LJ, Gowrie IJ, Jackson NC et al (2002) Partitioning glucose distribution/transport, disposal, and endogenous production during IVGTT. Am J Physiol 282:E992–E1007
27. Tiran J, Avruch LI, Albisser AM (1979) A circulation and organs model for insulin dynamics. Am J Physiol 237:E331–E339
28. Guyton JR, Foster RO, Soeldner JS, Tan MH, Kahn CB, Koncz L et al (1978) A model of glucose-insulin homeostasis in man that incorporates the heterogeneous fast pool theory of pancreatic insulin release. Diabetes 27:1027–1042
29. Cramp DG, Carson ER (1981) The dynamics of short-term blood glucose regulation. In: Cobelli C, Bergman RN (eds) Carbohydrate metabolism: quantitative physiology and mathematical modelling. Wiley, Chichester, pp 349–367

30. Cobelli C, Federspil G, Pacini G, Salvan A, Scandellari C (1982) An integrated mathematical model of the dynamics of blood glucose and its hormonal control. Math Biosc 58:27–60
31. Ackerman E, Gatewood LC, Rosevear JW, Molnar GD (1965) Model studies of blood-glucose regulation. Bull Math Biophys 27:21–37
32. Yipintsoi T, Gatewood LC, Ckerman E, Spivak PL, Molnar GD, Rosevear JW et al (1973) Mathematical analysis of blood glucose and plasma insulin responses to insulin infusion in healthy and diabetic subjects. Comput Biol Med 3:71–78
33. Rewers M, Zaccaro D, D'Agostino R, Haffner S, Saad MF, Selby JV et al (2004) Insulin sensitivity, insulinemia, and coronary artery disease: the insulin resistance atherosclerosis study. Diab Care 27:781–787
34. Hong J, Gui MH, Gu WQ, Zhang YF, Xu M, Chi ZN et al (2008) Differences in insulin resistance and pancreatic B-cell function in obese subjects with isolated impaired glucose tolerance and isolated impaired fasting glucose. Diabetic Med 25:73–79
35. Bloem CJ, Chang AM (2008) Short-term exercise improves beta-cell function and insulin resistance in older people with impaired glucose tolerance. J Clin Endocrinol Metab 93:387–392
36. Salinari S, Bertuzzi A, Asnaghi S, Guidone C, Manco M, Mingrone G (2009) First-phase insulin secretion restoration and differential response to glucose load depending on the route of administration in type 2 diabetic subjects after bariatric surgery. Diab Care 32:375–380
37. Kovatchev BP, Breton M, Man CD, Cobelli C (2009) In Silico preclinical trials: a proof of concept in closed-loop control of type 1 diabetes. J Diab Sci Technol 3:44–55
38. Regittnig W, Trajanoski Z, Leis HJ, Ellmerer M, Wutte A, Sendlhofer G et al (1999) Plasma and interstitial glucose dynamics after intravenous glucose injection—evaluation of the single-compartment glucose distribution assumption in the minimal models. Diabetes 48:1070–1081
39. Finegood DT, Tzur D (1996) Reduced glucose effectiveness associated with reduced insulin release—an artifact of the minimal-model method. Am J Physiol 34:E485–E495
40. Quon MJ, Cochran C, Taylor SI, Eastman RC (1994) Non-insulin-mediated glucose disappearance in subjects with IDDM—discordance between experimental results and minimal model analysis. Diabetes 43:890–896
41. Weber KM, Martin IK, Best JD, Alford FP, Boston RC (1989) Alternative method for minimal model analysis of intravenous glucose tolerance data. Am J Physiol 256:E524–E535
42. Cobelli C, Pacini G, Toffolo G, Sacca L (1986) Estimation of insulin sensitivity and glucose clearance from minimal model: new insights from labeled IVGTT. Am J Physiol 250:E591–E598
43. Avogaro A, Bristow JD, Bier DM, Cobelli C, Toffolo G (1989) Stable-label intravenous glucose tolerance test minimal model. Diabetes 38:1048–1055
44. Caumo A, Cobelli C (1993) Hepatic glucose production during the labeled IVGTT: estimation by deconvolution with a new minimal model. Am J Physiol 264:E829–E841
45. Vicini P, Caumo A, Cobelli C (1997) The hot IVGTT two-compartment minimal model: indexes of glucose effectiveness and insulin sensitivity. Am J Physiol 273:E1024–E1032
46. Cobelli C, Caumo A, Omenetto M (1999) Minimal model S-G overestimation and S-I underestimation: improved accuracy by a Bayesian two-compartment model. Am J Physiol 277:E481–E488
47. Ferrannini E, Smith JD, Cobelli C, Toffolo G, Pilo A, DeFronzo RA (1985) Effect of insulin on the distribution and disposition of glucose in man. J Clin Invest 76:357–364
48. Berger M, Cuppers HJ, Hegner H, Jorgens V, Berchtold P (1982) Absorption kinetics and biologic effects of subcutaneously injected insulin preparations. Diab Care 5:77–91
49. Binder C, Lauritzen T, Faber O, Pramming S (1984) Insulin pharmacokinetics. Diabetes 7:188–199

50. Owens DR, Jones MK, Birtwell AJ, Burge CT, Jones IR, Heyburn PJ et al (1984) Pharmacokinetics of subcutaneously administered human, porcine and bovine neutral soluble insulin to normal man. Horm Metab Res 16(Suppl 1):195–199
51. Kang S, Brange J, Burch A, Volund A, Owens DR (1991) Absorption kinetics and action profiles of subcutaneously administered insulin analogues (AspB9GluB27, AspB10, AspB28) in healthy subjects. Diab Care 14:1057–1065
52. Nucci G, Cobelli C (2000) Models of subcutaneous insulin kinetics. A critical review. Comput Methods Programs Biomed 62:249–257
53. Kobayashi T, Sawano S, Itoh T, Kosaka K, Hirayama H, Kasuya Y (1983) The pharmacokinetics of insulin after continuous subcutaneous infusion or bolus subcutaneous injection in diabetic patients. Diabetes 32:331–336
54. Kraegen EW, Chisholm DJ (1984) Insulin responses to varying profiles of subcutaneous insulin infusion: kinetic modelling studies. Diabetologia 26:208–213
55. Puckett WR (1995) LIGHTFOOT M: a model for multiple subcutaneous insulin injections developed from individual diabetic patient data. Am J Physiol 269:E1115–E1124
56. Shimoda S, Nishida K, Sakakida M, Konno Y, Ichinose K, Uehara M et al (1997) Closed-loop subcutaneous insulin infusion algorithm with a short-acting insulin analog for long-term clinical application of a wearable artificial endocrine pancreas. Front Med Biol Eng 8:197–211
57. Berger M, Rodbard D (1989) Computer-simulation of plasma-insulin and glucose dynamics after subcutaneous insulin injection. Diab Care 12:725–736
58. Mosekilde E, Jensen KS, Binder C, Pramming S, Thorsteinsson B (1989) Modeling absorption kinetics of subcutaneous injected soluble insulin. J Pharmacokinet Biopharm 17:67–87
59. Trajanoski Z, Wach P, Kotanko P, Ott A, Skraba F (1993) Pharmacokinetic model for the absorption of subcutaneously injected soluble insulin and monomeric insulin analogs. Biomed Tech 38:224–231
60. Hovorka R, Powrie JK, Smith GD, Sonksen PH, Carson ER, Jones RH (1993) Five-compartment model of insulin kinetics and its use to investigate action of chloroquine in NIDDM. Am J Physiol 265:E162–E175
61. Wilinska ME, Chassin LJ, Schaller HC, Schaupp L, Pieber TR, Hovorka R (2005) Insulin kinetics in type-1 diabetes: continuous and bolus delivery of rapid acting insulin. IEEE Trans Biomed Eng 52:3–12
62. Bolinder J, Ungerstedt U, Arner P (1993) Long-term continuous glucose monitoring with microdialysis in ambulatory insulin-dependent diabetic-patients. Lancet 342:1080–1085
63. Trajanoski Z, Brunner GA, Schaupp L, Ellmerer M, Wach P, Pieber TR et al (1997) Open-flow microperfusion of subcutaneous adipose tissue for on- line continuous ex vivo measurement of glucose concentration. Diab Care 20:1114–1121
64. Schaupp L, Brunner GA, Schaller H, Bodelenz M, Wutte A, Wach P et al (2001) Glucose monitoring in the adipose tissue of type 1 diabetic patients using open-flow microperfusion and microdialysis. Diabetologia 44:A46
65. Pfeiffer EF, Meyerhoff C, Bischof F, Keck FS, Kerner W (1993) On line continuous monitoring of subcutaneous tissue glucose is feasible by combining portable glucosensor with microdialysis. Horm Metab Res 25:121–124
66. Sternberg F, Meyerhoff C, Mennel FJ, Mayer H, Bischof F, Pfeiffer EF (1996) Does fall in tissue glucose precede fall in blood glucose? Diabetologia 39:609–612
67. Freeland AC, Bonnecaze RT (1999) Inference of blood glucose concentrations from subcutaneous glucose concentrations: applications to glucose biosensors. Ann Biomed Eng 27:525–537
68. Schmidtke DW, Freeland AC, Heller A, Bonnecaze RT (1998) Measurement and modeling of the transient difference between blood and subcutaneous glucose concentrations in the rat after injection of insulin. Proc Natl Acad Sci USA 95:294–299

69. Wilinska ME, Bodenlenz M, Chassin LJ, Schaller HC, Schaupp LA, Pieber TR et al (2004) Interstitial glucose kinetics in subjects with type 1 diabetes under physiologic conditions. Metab Clin Exp 53:1484–1491
70. Worthington DR (1997) Minimal model of food absorption in the gut. Med Inform (Lond) 22:35–45
71. Arleth T, Andreassen S, Orsini-Federici M, Timi A, Massi-Benedetti M (2000) A model of glucose absorption from mixed meals. Model Control Biomed Syst 2000:307–312
72. Dalla Man C, Rizza RA, Cobelli C (2006) Mixed meal simulation model of glucose-insulin system. In: 2006 28th annual international conference of the IEEE engineering in medicine and biology society, Vols 1–15. 3769–3772
73. Goodyear LJ, Hirshman MF, Horton ES (1991) Exercise-induced translocation of skeletal-muscle glucose transporters. Am J Physiol 261:E795–E799
74. Thorell A, Hirshman MF, Nygren J, Jorfeldt L, Wojtaszewski JFP, Dufresne SD et al (1999) Exercise and insulin cause GLUT-4 translocation in human skeletal muscle. Am J Physiol Endocrinol Metab 277:E733–E741
75. Ren JM, Semenkovich CF, Gulve EA, Gao JP, Holloszy JO (1994) Exercise induces rapid increases in Glut4 expression, glucose-transport capacity, and insulin-stimulated glycogen-storage in muscle. J Biol Chem 269:14396–14401
76. DCCT Research Group (1991) Epidemiology of severe hypoglycemia in the diabetes control and complications trial. Am J Med 90:450–459
77. Fischer U, Salzsieder E, Stowhas H, Rutscher A (1994) Model-based prevention in iddm of exercise-induced hypoglycemia. Diabetes 43:A218
78. Derouich M, Boutayeb A (2002) The effect of physical exercise on the dynamics of glucose and insulin. J Biomech 35:911–917
79. Hernandez-Ordonez M, Campos-Delgado DU (2008) An extension to the compartmental model of type 1 diabetic patients to reproduce exercise periods with glycogen depletion and replenishment. J Biomech 41:744–752
80. Kim J, Saidel GM, Cabrera ME (2007) Multi-scale computational model of fuel homeostasis during exercise: effect of hormonal control. Ann Biomed Eng 35:69–90
81. Breton MD (2008) Physical activity-the major unaccounted impediment to closed loop control. J Diab Sci Technol 2:169–174
82. Bergman RN, Ider YZ, Bowden CR, Cobelli C (1979) Quantitative estimation of insulin sensitivity. Am J Physiol 236:E667–E677
83. Guelfi KJ, Jones TW, Fournier PA (2005) The decline in blood glucose levels is less with intermittent high-intensity compared with moderate exercise in individuals with type 1 diabetes. Diab Care 28:1289–1294
84. Chassin LJ, Wilinska ME, Hovorka R (2007) Intense exercise in type 1 diabetes: exploring the role of continuous glucose monitoring. J Diab Sci Technol 1:164–167
85. Shichiri M, Kawamori R, Yamasaki Y, Hakui N, Abe H (1982) Wearable artificial endocrine pancreas with needle-type glucose sensor. Lancet 2:1129–1131
86. Vering T, Beyer U, Both M, Heiniger H, Hutzli I, Kalt L et al (2003) Minimally invasive control loop system for SC-SC control on patients with type 1 diabetes. In: Third annual diabetes technology meeting. San Francisco, p A21
87. Galley P, Wagner R, Buck H, Weinert S, Bousamra S, Long J (2004) Use of subcutaneous glucose measurements to drive real-time algorithm-directed insulin infusion recommendations. Diab Technol Ther 6:245–246
88. Hovorka R, Chassin LJ, Wilinska ME, Canonico V, Akwi JA, Federici MO et al (2004) Closing the loop: the adicol experience. Diab Technol Ther 6:307–318
89. Chassin LJ, Wilinska ME, Hovorka R (2004) Evaluation of glucose controllers in virtual environment: methodology and sample application. Artif Intell Med 32:171–181
90. Hovorka R, Canonico V, Chassin LJ, Haueter U, Massi-Benedetti M, Orsini FM et al (2004) Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes. Physiol Meas 25:905–920

91. Wilinska ME, Chassin LJ, Acerini CL, Allen JM, Dunger DB, Hovorka R (2010) Simulation environment to evaluate closed-loop insulin delivery systems in type 1 diabetes. J Diab Sci Technol 4:132–144
92. Hovorka R, Allen JM, Elleri D, Chassin LJ, Harris J, Xing D et al (2010) Manual closed-loop insulin delivery in children and adolescents with type 1 diabetes: a phase 2 randomised crossover trial. Lancet 375:743–751
93. Elleri D, Allen JM, Kumareswaran K, Leelarathna L, Nodale M, Caldwell K et al (2013) Closed-loop basal insulin delivery over 36 h in adolescents with type 1 diabetes: randomized clinical trial. Diab Care 36:838–844
94. Hovorka R, Kumareswaran K, Harris J, Allen JM, Elleri D, Xing D et al (2011) Overnight closed loop insulin delivery (artificial pancreas) in adults with type 1 diabetes: crossover randomised controlled studies. Br Med J 342:d1855
95. Murphy HR, Elleri D, Allen JM, Harris J, Simmons D, Rayman G et al (2011) Closed-loop insulin delivery during pregnancy complicated by type 1 diabetes. Diab Care 34:406–411
96. Bode B, Gross K, Rikalo N, Schwartz S, Wahl T, Page C et al (2004) Alarms based on real-time sensor glucose values alert patients to hypo-and hyperglycemia: the guardian continuous monitoring system. Diab Technol Ther 6:105–113
97. Kanderian S, Saad MF, Rebrin K, Steil GM (2006) Modeling glucose profiles obtained using closed loop insulin delivery—implications for controller cptimization. Diabetes 55[Suppl 1]:A98
98. Kanderian SS, Weinzimer SA, Steil GM (2012) The identifiable virtual patient model: comparison of simulation and clinical closed-loop study results. J Diab Sci Technol 6:371–379
99. Dalla Man C, Rizza RA, Cobelli C (2007) Meal simulation model of the glucose-insulin system. IEEE Trans Biomed Eng 54:1740–1749
100. Basu R, Di Camillo B, Toffolo G, Basu A, Shah P, Vella A et al (2003) Use of a novel triple-tracer approach to assess postprandial glucose metabolism. Am J Physiol Endocrinol Metab 284:E55–E69
101. Vicini P, Sparacino G, Caumo A, Cobelli C (1997) Estimation of endogenous glucose production after a glucose perturbation by nonparametric stochastic deconvolution. Comput Methods Programs Biomed 52:147–156
102. Ferrannini E, Cobelli C (1987) The kinetics of insulin in man. I. General aspects. Diab Metab Rev 3:335–363
103. Parker RS, Doyle FJ, III, Peppas NA (1999) A model-based algorithm for blood glucose control in type I diabetic patients. IEEE Trans Biomed Eng 46:148–157
104. Parker RS, Doyle FJ, Ward JH, Peppas NA (2000) Robust H-infinity glucose control in diabetes using a physiological model. AIChE J 46:2537–2549
105. Fabietti PG, Canonico V, Federici MO, Benedetti MM, Sarti E (2006) Control oriented model of insulin and glucose dynamics in type 1 diabetics. Med Biol Eng Comput 44:69–78

Simple Parameters Describing Gut Absorption and Lipid Dynamics in Relation to Glucose Metabolism During a Routine Oral Glucose Test

Andrea Tura and Giovanni Pacini

Abstract The oral glucose tolerance test (OGTT) is a simple and physiological test used for the diagnosis of diabetes and, more generally, for the assessment of the metabolic condition of an individual. For a deep analysis of the OGTT data, the exploitation of mathematical models for the interpretation of the test results is necessary. The focus of this chapter is on some recent models based on OGTT data, describing the glucose absorption from the gut, and the effects of insulin on the dynamics of non-esterified free fatty acids (NEFA). As regards the glucose absorption model, it requires measure of plasma glucose and insulin during the OGTT. The model allows the estimation of absorption rates, the total glucose absorbed, and half-life in the gastrointestinal tract. In fact, the increase of post-prandial circulating plasma glucose over time is the result of gain from gut glucose absorption and liver endogenous glucose production, and loss because of glucose uptake, predominantly by skeletal muscle. Endogenous glucose production can be assessed by some mathematical expressions depending on plasma insulin values, whereas the glucose loss is calculated as a function of the endogenous glucose production and of the insulin sensitivity in each individual. Once these variables have been determined, glucose absorption can be determined by solving a nonlinear least squares problem fitting the plasma glucose values during the OGTT. The model was used to analyze sex related differences in OGTT glucose metabolism, including gut absorption, in healthy humans. We found that, in the early phase of the OGTT, males had markedly increased glucose absorption rates by approximately 200 mg/min from the gastrointestinal tract, whereas in the final phase of the OGTT, females absorbed approximately 60 mg/min more glucose. In another study, the model was used to assess glucose absorption in 15 pregnant women with gestational diabetes. As regards the model of NEFA dynamics, it was postulated

A. Tura · G. Pacini (✉)
Metabolic Unit, Institute of Biomedical Engineering, Italian National Research Council, Padova, Italy
e-mail: Giovanni.pacini@isib.cnr.it

V. Marmarelis and G. Mitsis (eds.), *Data-driven Modeling for Diabetes*, Lecture Notes in Bioengineering, DOI: 10.1007/978-3-642-54464-4_7,

that NEFA kinetics could be described by first order (single compartment) kinetics, with NEFA production controlled by insulin in remote compartment. It is assumed that plasma insulin enters a remote compartment, before having an inhibitory effect on NEFA production. Starting from basal production at basal insulin, NEFA production decreases, for suprabasal increase in remote insulin up to a prescribed value. The identification of the unknown parameters of the model was performed by applying Genetic Algorithms. The model was used to study NEFA kinetics in a group of women with a history of gestational diabetes (fGDM). The fGDM women were divided into normal glucose tolerance group (NGT) and a impaired glucose metabolism (IGM). We also studied 15 control (CNT) women. We found that, while fasting NEFA were not different between groups, IGM exhibited slower decline in plasma NEFA during the OGTT. We conclude that appropriate mathematical modeling allows sophisticated analyses of the OGTT data, thus possibly providing a comprehensive picture of the metabolic condition.

1 Introduction

The oral glucose tolerance test (OGTT) is the most commonly used dynamic test in studies that deal with the metabolic assessment of an individual [1]. It is a *dynamic* test, since the fasting steady state is perturbed with a glucose load and the dynamics of the glucose concentration in the peripheral circulation provides information on the ability of the subjects to dispose of the exogenously administered glucose. It is a *simple* test because it only requires a two-hour time after the ingestion of the water solution containing usually 75 g of glucose, during which a few blood samples are collected from an easily accessible brachial vein. No particular setting is required and it could be even performed in the general practitioner's office. It is a *physiological* test, since it mimics how carbohydrates contained within the food are normally ingested. It is a test accepted by the diabetologists' community as a *diagnostic* test, being used to assess the glucose tolerance of a subject and to establish the presence of diabetes [2].

To achieve this goal, only glucose at fasting and after 2 h is measured. However, this provides a limited amount of information; therefore, in order to know more on the status of the subjects additional measurements are necessary. A typical protocol is to collect a blood sample after an overnight (12 h) fast, give the drink with the glucose solution and gather blood samples 30, 60, 90 and 120 min afterwards. A minimum requirement is then to assess plasma concentration of glucose and insulin at those specific time points. In such a way, a great deal of indications can be achieved on the general metabolic picture of the individual. However, a quite sophisticated analysis of the data is necessary to extract evidences on those processes responsible for the maintenance or modification of the clinical condition. To this aim the exploitation of mathematical models for the

interpretation of the test results revealed necessary, if not mandatory, to extract some of the facts hidden within the data-set.

This chapter illustrates mathematical models which describe the evolution of the application of modeling to OGTT for the assessment of other physiological processes strictly involved in the characterization of the glucose dynamics in response to an oral glucose load.

2 Physiological Background and Existing Methods

Once glucose is given orally, it goes through the stomach and starts its intestinal journey. Passaging through the early part of the gut (duodenum, jejunum, ileum), it triggers the release of incretin hormones [3], which augment the sensitivity of pancreatic beta cell to the glucose stimulus, leading to the production and release of insulin. From the entire gut, in fact, the glucose is absorbed into the peripheral blood circulation, where its concentration markedly increases; glucose spreads through all peripheral tissues and organs, including the pancreatic beta cells from which insulin, equimolarly with C-peptide, is secreted. *Beta cell function* defines and quantifies the ability of the beta cells to release insulin in response to changes in glucose concentration. After passing through the liver, where it is in part degraded, insulin reaches muscle and adipose tissues; here it allows glucose incorporation into these tissues, lowering circulating glucose concentration until it reaches again the pre-load levels. *Insulin sensitivity* defines and quantifies the ability of insulin to promote glucose disposal. In addition, other insulin effects are those of inhibiting hepatic glucose production and lipolysis: i.e., the production of non-esterified free fatty acids (NEFA) from precursors. A well-tuned performance of these processes and a good balance among the levels of the various substances keep the subject metabolically healthy.

Some of the above processes have already been mathematically described during OGTT. In particular, the action of insulin of inhibiting liver glucose production and promoting glucose uptake by peripheral tissues has been the main component of a glucose kinetics model yielding a parameter (OGIS, oral glucose insulin sensitivity index) of glucose clearance, strictly related to the degree of insulin resistance [4]. A few composite models have been introduced [5, 6], but as they require difficult implementation by skilled operators, complex calculus and exhibit a limited validation, they have not been generally used. Simple empirical formulas, basically usable by anyone with a spread sheet have been very successful [e.g., 7, 8, 9], including the same OGIS [4]. These methods have been tested and validated against the glucose clamp, which is the gold standard for insulin sensitivity assessment. Other model-based assessments from OGTT of insulin secretion and beta cell function, exploit again complex [6, 10, 11] or simpler [12, 13] mathematical systems. For instance, insulin secretion has been represented with a model that describes the three main components involved in the insulin release process (first phase, dose response to glucose, potentiation) [11].

This model yields parameters on beta cell function that are fundamental to the understanding of possible impairments of the pancreatic release of insulin [14]. The most widely used technique for assessing beta cell function, however, relies on a simple formula (insulinogenic index [15]). All the above models analyze the relationships between the relative levels of insulin (and/or C-peptide) and glucose. Another method includes in the analysis also the relative shapes of the concentration time courses as a discriminant for impaired beta cell function [16], with special reference to the C-peptide curve. In fact, C-peptide is more informative than insulin for the assessment of real, pancreatic insulin secretion and beta-cell function, since C-peptide does not undergo partial degradation by the liver (as insulin does). Nonetheless, recently, methods have been proposed for the assessment of insulin secretion directly from insulin [17]. For detailed descriptions on methods and modeling for insulin sensitivity and insulin secretion, refer to specific reviews [e.g., 18].

3 New Models for the Analysis of Other OGTT Aspects

As seen above, the OGTT comprises different processes which interact with each other at different, but all important, levels. The most detailed possible analysis of them will be very useful to characterize in full the metabolic status of a single individual. To this aim, recently other models have been presented, and two more pieces of this multifaceted puzzle are reported in the following, which focus on the glucose absorption from the gut and on the effects of insulin on the dynamics of NEFA.

3.1 Glucose Absorption

3.1.1 The Model

To assess glucose absorption, which in humans is not directly measurable, we developed a novel approach, based on mathematical modeling to mimic absorptive conditions at the intestine site. Our method allows the estimation of absorption rates, the total glucose absorbed, and half-life in the gastrointestinal tract. The validity of our novel model was checked by using manifold-labeled glucose tracers, the gold standard of glucose absorption measurement [19]. The model requires only the measurement of both glucose and insulin concentration during the OGTT.

The increase of postprandial circulating plasma glucose (dG_{circ}) over time (dt) is the result of gain from gut glucose absorption (ABS) and liver endogenous glucose production (EGP), and loss because of glucose uptake (Rd), predominantly

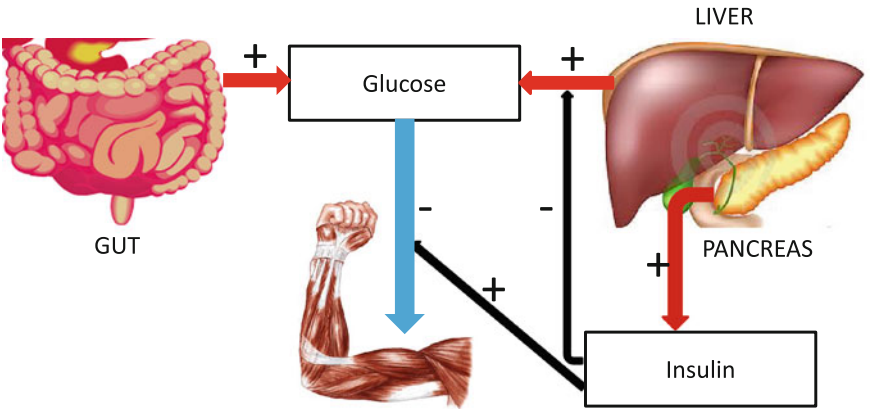


Fig. 1 Schematics of the main processes of glucose homeostasis. Glucose is produced by the liver (endogenous glucose production) and absorbed by the gut (glucose absorption); insulin is produced by pancreatic beta-cells (insulin secretion); then, insulin inhibits endogenous glucose production, and promotes glucose uptake by the muscles

by skeletal muscle (Fig. 1). Thus, changes in glucose concentration during the OGTT can be expressed as:

$$dG_{circ}/dt = 1/VG \times [BW \times (EGP(t) - Rd(t)) + ABS(t)] \quad (1)$$

with initial condition: G(0) = fasting glucose concentration; ABS(0) = 0 and EGP(0) = Rd(0). BW is the body weight and VG is the oral glucose distribution volume. VG is in deciliters, BW in kilograms; G_{circ} is in mg/dl; EGP and Rd are in mg min^{-1} kg^{-1}, and ABS in mg min^{-1}. VG is assumed as 9–15 % of BW, according to previous investigations [20].

It was demonstrated that EGP and Rd are predominantly regulated by circulating insulin concentrations [19]. During the OGTT, EGP in women can be calculated using the following relationship:

$$EGP(t) = 1.889 - 0.342 \cdot \log(\text{insulin}(t)) \quad (2)$$

while in men, the relationship is equal to:

$$EGP(t) = 2.542 - 0.496 \cdot \log(\text{insulin}(t)) \quad (3)$$

The constants in Eqs. (2) and (3) have been fixed for insulin in μU/ml.

Rd is assessed in the single individual. Specifically, a logarithmical relationship is computed between two points in the (Rd, insulin) space. The first point is (Rd(0), insulin(0)), with Rd(0) calculated by the Eq. (2) (it should be reminded that Rd(0) = EGP(0)), and insulin(0) equal to basal, fasting insulin. The second point is (Rd(maximum), insulin(maximum)). Thus, it is assumed that Rd is maximum when plasma insulin is maximum. Of course, the maximum value of insulin during the

OGTT is easily identified. As regards Rd(maximum), it is assumed that it is equal to the glucose disposal represented by the parameter M of an euglycemic-hyperinsulinemic clamp (that is, the glucose infusion rate during the last part of the clamp [21]). Alternatively, when the clamp is not performed, the clamp-like index [9] is used, or another OGTT insulin sensitivity index which can be considered a surrogate of M. Figure 2 shows the described approach for calculating Rd: the logarithmic function identified in each subject is used to assess Rd in each time sample of the OGTT. Thus, since dG_{circ}/dt is calculated from the plasma glucose concentration during the OGTT, in Eq. (1) the only unknown is ABS. Once the model (1) is integrated in time and ABS estimated, for each subject total glucose absorbed is calculated by integrating glucose absorption rates over the OGTT duration. Moreover, Glucose half-life (t½) in the gastrointestinal tract is individually determined by linear curve interpolation of relative glucose retention during the OGTT, by using the closest time points to cross the 50 % threshold [19].

In the original formulation [19], Eq. (1) was implemented using Simulink for Matlab (MathWorks Inc., Boston, MA), and ABS was estimated by solving the corresponding nonlinear least squares problem fitting the plasma glucose values during the OGTT (MatLab function used LSQNONLIN with 10,000 iterations at maximum). Every time sample of ABS was treated as an independent parameter to be estimated. Common criteria of model performance (best fit, residuals, variance-covariance Fisher's matrix) were evaluated for accepting the final model's prediction.

3.1.2 Applications and Outcomes

In a first study [19], the model (1) was used to analyze sex related differences in OGTT glucose metabolism, including gut absorption, in healthy humans. We analyzed 48 females and 26 males, with comparable age and body mass index but different height (males being taller). We found that, in the early phase of the OGTT, males had markedly increased glucose absorption rates by approximately 200 mg/min from the gastrointestinal tract, whereas in the final phase of the OGTT, females absorbed approximately 60 mg/min more glucose. These differences in glucose absorption were accompanied by an approximately 14 min longer gut glucose half-life in women, which could explain the higher glucose levels at the end of the OGTT in females. However, we also found that females have lower peripheral insulin release during the first 60 min of the OGTT, and this may result in a lower glucose disappearance rate after a time interval, given the known delay of insulin action on peripheral tissues. In fact, the period of lower glucose disappearance rate included the 2 h data sample, characterized by higher glucose level in females. Therefore, impaired glucose tolerance (IGT) higher prevalence in women may depend on a concomitant effect of insulin release and glucose absorption.

In another study [22], the model (1) was used to assess glucose absorption in 15 pregnant women with gestational diabetes mellitus (GDM), which were compared

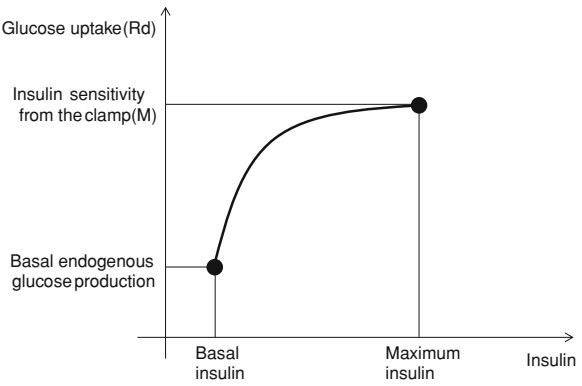


Fig. 2 Schematics of the method for the calculation of rate of glucose disappearance (glucose uptake, Rd)

to 7 normal glucose tolerance women (NGT), matched for major anthropometric characteristics. After delivery (6–7 months later), both groups were studied the same way. It was found that in GDM gut glucose absorption rates were lower than in NGT in the 30–120 min interval of the OGTT. In addition, in GDM glucose absorption rates also were lower during pregnancy than after delivery (where all the GDM women returned to non-diabetic glycemic levels). In contrast, glucose absorption rates in NGT were comparable during and after pregnancy. Based on these findings, it could be concluded that the hyperglycemia of the GDM women cannot be ascribed to excessively rapid or increased glucose absorption.

3.2 NEFA Kinetic

3.2.1 The Model

We have developed a model of and non-esterified fatty acids (NEFA) kinetics. It was postulated that NEFA kinetics could be described by first order (single compartment) kinetics, with NEFA production controlled by insulin in remote compartment [23]. In detail, the compartmental representation of NEFA kinetic model is shown in Fig. 3. It is assumed that plasma insulin enters a remote compartment, before having an inhibitory effect on NEFA production. The compartmental state variable Y(t) is a delayed profile of I(t), characterized by the fractional clearance P_2 in min^{-1}. NEFA production was considered to be: (i) linearly dependent on deviations from the basal value of remote insulin; (ii) with saturable inhibition. Thus, starting from basal production at basal insulin, NEFA production decreases with slope S_{NEFA}, in ml/mU, for suprabasal increase in remote insulin up to a value beyond which NEFA production becomes constant. The maximum inhibitory capacity of insulin on NEFA production is expressed as a

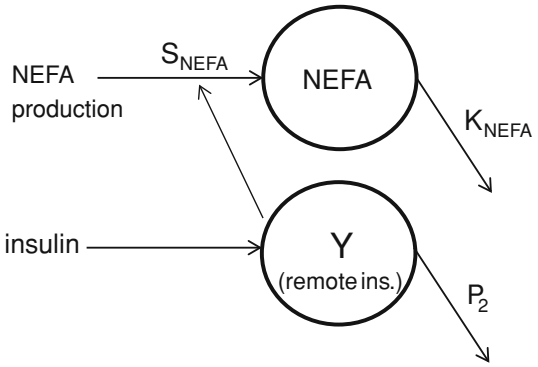


Fig. 3 Schematics of the NEFA kinetics model and of the insulin action on NEFA

non-dimensional fraction (ρ_{NEFA}) of the basal production. The differential equations modelling the NEFA kinetics can be written as

$$\begin{cases} \dfrac{dNEFA(t)}{dt} = -K_{NEFA} \cdot (NEFA(t) - NEFA_b \cdot \max\left\{1 - S_{NEFA} \cdot Y'(t), \rho_{NEFA}\right\} \\ NEFA(0) = NEFA_b \end{cases} \tag{4}$$

$$\begin{cases} \dfrac{dY'(t)}{dt} = -P_2 \cdot [Y'(t) - (I(t) - I_b)] \\ Y'(t) = Y(t) - I_b \end{cases} \tag{5}$$

The subscript b identifies the basal values of plasma NEFA and insulin, I, In Eq. (5),the quantity Y' is the remote insulin level above basal insulin $Y'(t) = Y(t) - I_b$. Therefore, the value assumed by the quantity S_{NEFA} in Eq. (1) provides measure of the lipolysis inhibition due to suprabasal insulin variations, whereas the quantity ρ_{NEFA} accounts for circulating NEFA at complete HSL inhibition, and it is related to the fraction of NEFA that, produced within the intravascular system, escapes cellular uptake. The quantity K_{NEFA}, in min^{-1}, represents the fractional turnover rate for NEFA concentration. For given values of K_{NEFA} and basal concentration $NEFA_b$, the basal production rate of NEFA is calculated as $NEFA_b \times K_{NEFA}$ in mmol/l/min.

The model identification process allows the determination of an unknown parameter vector $\mathbf{P} = [NEFA_b, K_{NEFA}, S_{NEFA}, \rho_{NEFA}, P_2]^T$, the identification of which usually resorts to traditional gradient-based methods, in particular the weighted nonlinear least squares approach, using the Levenberg-Marquardt algorithm for the minimization procedure. However, the identification of the parameters **P** of the NEFA model of Eqs. (4) and (5) cannot be carried out by applying gradient-based methods straightforward. In fact, in gradient-based deterministic optimization methods, in every iteration step the calculation of the gradient of the least-square functional, with respect to the parameters **P**, must be calculated. Thus,

the differentiability of the least-square functional with respect to the parameters **P** must be guaranteed. However, in this, case, this criterium is not satisfied by Eq. (1), because it includes a maximum function, whose differentiability is not guaranteed everywhere. Thus, the identification problem could be ill-posed.

For the reasons mentioned above, the identification of the unknown parameters of the model was performed by applying Genetic Algorithms (GAs), which are theoretically and empirically proven to provide a robust search in a complex space [24, 25]. GAs are in fact derivative free random search algorithms for nonlinear problems based on the rules of natural selection. Their implementation requires the five components of vector **P** to be encoded through a proper representation method. In the study, floating-point representation was chosen. Each individual was composed by 5 genes, i.e., the 5 components of vector **P**. The limits of the parameters to be estimated, defining the borders for the field of existence of the solution, were set as:

$$NEFA_b = 200 - 1200; K_{NEFA} = 0.0001 - 1; S_{NEFA} = 0.0001 - 0.1; \rho_{NEFA} = 0 - 1; p_2 = 0 - 1$$

To overcome the problem of the possible dependence of the identification procedure on the initial values of the parameters, the genetic algorithm was initiated with 150 randomly generated individuals. An appropriate cost function, the so-called "fitness function", was built to control the GA search progress in an acceptable direction, and a possible solution $\hat{\mathrm{P}}$ is then reached. Afterwards, a new generation of individuals (i.e., possible solutions) is constructed from the actual population by applying three genetic operators: reproduction, mutation and crossover), and the search for the best solution is repeated. The number of iterations required for convergence to the desired solution was set at 2,000 after a sensitivity study. Furthermore, to avoid converging to a local minimum, the cycle was repeated 200 times, and the best solution was then finally chosen. Like in previous investigations [25], the number of parameters to be estimated, together with the specific genetic operators applied, made the choice adopted adequate for the population size. The numerical implementation of the methodology described above was performed using an in-house developed code written with MATLAB programming language (The MathWorks, Natick, MA).

3.2.2 Applications and Outcomes

The model presented here was used to study NEFA kinetics in a group of women with a history of gestational diabetes (fGDM) that underwent an oral glucose tolerance test (OGTT) [26]. In fact, we studied a large group of fGDM immediately after partum, and a subgroup also one to two years afterwards, to test the hypothesis that women with former GDM are characterized by impaired NEFA dynamics. The fGDM women were divided into normal glucose tolerance group

(NGT) and a group with impaired glucose metabolism (IGM), including subjects with impaired glucose tolerance and/or impaired fasting glucose. We also studied 15 control (CNT) women without known risks for diabetes, and with normal glucose tolerance during pregnancy and at the time of study. We found that, while fasting NEFA were not different between groups, IGM exhibited slower decline in plasma NEFA during the OGTT, with a corresponding area under the curve of plasma NEFA, AUC_{NEFA}, greater than in NGT and CNT during the OGTT. Parameters of NEFA kinetics, i.e., inhibition of lipolysis (S_{NEFA}) and the fractional turnover rate (K_{NEFA}), were typically lower in fGDM (P range 0.018−0.041); parameters $p2_{NEFA}$ and ρ_{NEFA} were not different among the groups. We also analyzed the possible effect of obesity (BMI $\geq$ 30 kg/m^2) on NEFA variables. Thus, we divided the fGDM in obese ($n = 15$) and non-obese ($n = 57$) subjects. AUC_{NEFA} was greater in obese than non-obese subjects ($P = 0.006$), but fasting NEFA was not different. S_{NEFA} and K_{NEFA} tended to be lower in the obese group but the difference did not reached statistical significance. As regards the subgroup of 28 women that underwent a follow-up examination, they showed no differences in the majority of the measurements and parameters analyzed in the study [26]. In fact, glucose concentration values did not change from basal to follow-up condition, neither in terms of mean glucose ($P = 0.75$), nor in terms of the single OGTT glucose samples (P range 0.23–0.90); similarly, NEFA concentration values were very similar ($P > 0.3$).

4 Conclusions

The OGTT is the test of choice for the determination of the metabolic state of an individual due to the several advantages of its performance. Given its wide use, it has also been the target for the development of mathematical models describing the OGTT outcomes in different conditions and for describing the interrelationships among different compounds which could be measured after an oral glucose load. We and others have produced simple models that integrated in the main glucose/insulin control the effect of other substances such as C-peptide [11, 12, 27], proinsulin [28], glucagon [29], amylin [30], incretins [31]. In this chapter, we have presented two further recent developments of this broad description of the OGTT, focusing on glucose appearance into the systemic circulation after the load and on the macro processes regulating NEFA kinetics following changes in the concentration of glucose and insulin. These two new models revealed very useful to add further insights on the metabolic conditions of different subjects in particular pathophysiological situations.

OGTT is considered a test that mimics the normal way of assuming carbohydrates, and in addition, it is used for the diagnosis of type 2 diabetes [2]: these aspects make OGTT extremely relevant. However, currently, also the meal test (MTT) has become increasingly popular, since it is assumed to be the most physiologically realistic test available. In fact, in addition to glucose also other

nutrients (proteins, fats) are administered, so that it resembles a real meal: i.e., the way humans naturally assume energy. Many of the existing models and empirical formulas which were developed for the OGTT can be safely applied to MTT [32–34], others have been developed on purpose [4, 10]. In relation to the two models presented in this chapter, it is worth saying that it would be necessary testing whether the integration of the role played by other nutrients such as proteins and fats may indeed increase the information content of the models' outcome. This is particularly worthwhile for the NEFA model, since it is clear that exogenously given fats are definitely interfering with those produced endogenously, in terms of appearance and inhibition.

In conclusion, given all the advantages of performing an OGTT in a single subject, this test must be progressively exploited to gather an increased amount of information. The model presented here are two more steps ahead toward a widest possible description of the OGTT results.

References

1. American Diabetes Association (2013) Standards of medical care in diabetes—2013. Diabetes Care 36(Suppl. 1):S11–S66
2. American Diabetes Association (2013) Diagnosis and classification of diabetes mellitus. Diabetes Care 36(Suppl. 1):S67–S74
3. Drucker D, Nauck M (2006) The incretin system: glucagon like peptide-1 receptor agonists and dipeptidyl peptidase-4 inhibitors in type 2 diabetes. Lancet 368:1696–1705
4. Mari A, Pacini G, Murphy E, Ludvik B, Nolan JJ (2001) A model-based method for assessing insulin sensitivity from the oral glucose tolerance test. Diabetes Care 24(3):539–548
5. Caumo A, Bergman RN, Cobelli C (2000) Insulin sensitivity from meal tolerance tests in normal subjects: a minimal model index. J Clin Endocrinol Metab 85:4396–4402
6. Breda E, Cavaghan MK, Toffolo G, Polonsky KS, Cobelli C (2001) Oral glucose tolerance test minimal model indexes of beta-cell function and insulin sensitivity. Diabetes 50:150–158
7. Matsuda M, DeFronzo RA (1999) Insulin sensitivity indices obtained from oral glucose tolerance testing: comparison with the euglycemic insulin clamp. Diabetes Care 22(9):1462–1470
8. Stumvoll M, Mitrakou A, Pimenta W, Jenssen T, Yki-Järvinen H, Van Haeften T, Häring H, Fritsche A, Gerich J (2000) Assessment of insulin secretion from the oral glucose tolerance test in white patients with type 2 diabetes. Diabetes Care 23(9):1440–1441
9. Anderwald C, Anderwald-Stadler M, Promintzer M, Prager G, Mandl M, Nowotny P, Bischof MG, Wolzt M, Ludvik B, Kästenbauer T, Pacini G, Luger A, Krebs M (2007) The Clamp-like index: a novel and highly sensitive insulin sensitivity index to calculate hyperinsulinemic clamp glucose infusion rates from oral glucose tolerance tests in nondiabetic subjects. Diabetes Care 30(9):2374–2380
10. Hovorka R, Chassin L, Luzio SD, Playle R, Owens DR (1998) Pancreatic beta-cell responsiveness during meal tolerance test: model assessment in normal subjects and subjects with newly diagnosed noninsulin-dependent diabetes mellitus. J Clin Endocrinol Metab 83:744–750
11. Mari A, Tura A, Gastaldelli A, Ferrannini E (2002) Assessing insulin secretion by modelling in multiple-meal tests: role of potentiation. Diabetes 51(Suppl. 1):S221–S226

12. Thomaseth K, Kautzky-Willer A, Ludvik B, Prager R, Pacini G (1996) Integrated mathematical model to assess beta-cell activity during the oral glucose test. Am J Physiol 270(3 Pt 1):E522–E531
13. Cretti A, Lehtovirta M, Bonora E, Brunato B, Zenti MG, Tosi F, Caputo M, Caruso B, Groop LC, Muggeo M, Bonadonna RC (2001) Assessment of beta-cell function during the oral glucose tolerance test by a minimal model of insulin secretion. Eur J Clin Invest 31:405–416
14. Mari A, Ferrannini E (2008) Beta-cell function assessment from modelling of oral tests: an effective approach. Diab Obes Metab 10(Suppl 4):77–87
15. Tura A, Kautzky-Willer A, Pacini G (2006) Insulinogenic indices from insulin and C-peptide: comparison of beta-cell function from OGTT and IVGTT. Diab Res Clin Pract 72:298–301
16. Tura A, Morbiducci U, Sbrignadello S, Winhofer Y, Pacini G, Kautzky-Willer A (2011) Shape of glucose, insulin, C-peptide curves during a 3 h oral glucose tolerance test: any relationship with the degree of glucose tolerance? Am J Physiol Regul Integr Comp Physiol 300:R941–R948
17. Tura A, Pacini G, Kautzky-Willer A, Gastaldelli A, DeFronzo RA, Ferrannini E, Mari A (2012) Estimation of prehepatic insulin secretion: comparison between standardized C-peptide and insulin kinetic models. Metabolism 61(3):434–443
18. Pacini G, Mari A (2003) Methods for clinical assessment of insulin sensitivity and beta-cell function. Best Pract Res Clin Endocrinol Metab 17(3):305–322
19. Anderwald C, Gastaldelli A, Tura A, Krebs M, Promintzer-Schifferl M, Kautzky-Willer A, Stadler M, DeFronzo RA, Pacini G, Bischof MG (2011) Mechanism and effects of glucose absorption during an oral glucose tolerance test among females and males. J Clin Endocrinol Metab 96(2):515–524
20. van Tulder L, Michaeli B, Chiole′ro R, Berger MM, Revelly JP (2005) An evaluation of the initial distribution volume of glucose to assess plasma volume during a fluid challenge. Anesth Analg 101:1089–1093
21. Tura A, Sbrignadello S, Succurro E, Groop L, Sesti G, Pacini G (2010) An empirical index of insulin sensitivity from short IVGTT: validation against the minimal model and glucose clamp indices in patients with different clinical characteristics. Diabetologia 53(1):144–152
22. Anderwald C, Tura A, Winhofer Y, Krebs M, Winzer C, Bischof MG, Luger A, Pacini G, Kautzky-Willer A (2011) Glucose absorption in gestational diabetes mellitus during an oral glucose tolerance test. Diabetes Care 34(7):1475–1480
23. Morbiducci U, Di Benedetto G, Kautzky-Willer A, Deriu MA, Pacini G, Tura A (2011) Identification of a model of non-esterified fatty acids dynamics through genetic algorithms: the case of women with a history of gestational diabetes. Comput Biol Med 41(3):146–153
24. Goldberg DE (1989) Genetic algorithms in search, optimization and machine learning. Addison Wesley, New York
25. Morbiducci U, Tura A, Grigioni M (2005) Genetic algorithms for parameter estimation in mathematical modeling of glucose metabolism. Comput Biol Med 35(10):862–874
26. Tura A, Pacini G, Winhofer Y, Bozkurt L, Di Benedetto G, Morbiducci U, Roden M, Kautzky-Willer A (2012) Non-esterified fatty acid dynamics during oral glucose tolerance test in women with former gestational diabetes. Diabet Med 29(3):351–358
27. Tura A, Ludvik B, Nolan JJ, Pacini G, Thomaseth K (2001) Insulin and C-peptide secretion and kinetics in humans: direct and model-based measurements during OGTT. Am J Physiol Endocrinol Metab 281(5):E966–E974
28. Tura A, Pacini G, Kautzky-Willer A, Ludvik B, Prager R, Thomaseth K (2003) Basal and dynamic proinsulin-insulin relationship to assess beta-cell function during OGTT in metabolic disorders. Am J Physiol Endocrinol Metab 285(1):E155–E162
29. Ludvik B, Thomaseth K, Nolan JJ, Clodi M, Prager R, Pacini G (2003) Inverse relation between amylin and glucagon secretion in healthy and diabetic human subjects. Eur J Clin Invest 33(4):316–322
30. Thomaseth K, Pacini G, Clodi M, Kautzky-Willer A, Nolan JJ, Prager R, Olefsky JM, Ludvik B (1997) Amylin release during oral glucose tolerance test. Diabet Med 14(Suppl 2):S29–S34

31. Salinari S, Bertuzzi A, Mingrone G (2011) Intestinal transit of a glucose bolus and incretin kinetics: a mathematical model with application to the oral glucose tolerance test. Am J Physiol Endocrinol Metab 300(6):E955–E965
32. Mari A, Gastaldelli A, Foley JE, Pratley RE, Ferrannini E (2005) Beta-cell function in mild type 2 diabetic patients. Effects of 6-month glucose lowering with nateglinide. Diabetes Care 28:1132–1138
33. Ahrén B, Pacini G, Foley JE, Schweizer A (2005) Improved meal-related beta-cell function and insulin sensitivity by the dipeptidyl peptidase-IV inhibitor vildagliptin in metformin-treated patients with type 2 diabetes over 1 year. Diabetes Care 28:1936–1940
34. Ohlsson L, Alsalim W, Carr RD, Tura A, Pacini G, Mari A, Ahrén B (2013) Glucose-lowering effect of the DPP-4 inhibitor sitagliptin after glucose and non-glucose macronutrient ingestion in non-diabetic subjects. Diabetes Obes Metab 15(6):531–537

Data-Driven Modeling of Diabetes Progression

Andrea DeGaetano, Simona Panunzi, Pasquale Palumbo, Claudio Gaz and Thomas Hardy

Abstract A realistic representation of the long-term physiologic adaptation to developing insulin resistance would facilitate the effective design of clinical trials evaluating diabetes prevention or disease modification therapies. In the present work, a realistic, robust description of the evolution of the compensation of the glucose-insulin system in healthy and diabetic individuals, with particular attention to the physiological compensation to worsening insulin resistance is formulated, its physiological assumptions are presented, and its performance over the span of a lifetime is simulated. Model-based simulations of the long-term evolution of the disease and of its response to therapeutic interventions are consistent with the transient benefits observed with conventional therapies, and with promising effects of radical improvement of insulin sensitivity (as by metabolic surgery) or of β-cell protection. The mechanistic Diabetes Progression Model provides a credible tool by which long-term implications of anti-diabetic interventions can be evaluated.

Keywords Glucose · Insulin resistance · Beta cell mass · Mathematical models · Type 2 diabetes mellitus

A. DeGaetano · S. Panunzi (✉) · C. Gaz
Consiglio Nazionale delle Ricerche, Istituto di Analisi dei Sistemi ed Informatica "A. Ruberti" - Laboratorio di Biomatematica,
UCSC-Largo A. Gemelli, 8, 00168 Roma, Italy
e-mail: simona.panunzi@biomatematica.it

P. Palumbo
Consiglio Nazionale delle Ricerche, Istituto di Analisi dei Sistemi ed Informatica "A. Ruberti", Viale Manzoni, 30, Roma 00185, Italy

T. Hardy
Lilly Research Laboratories, Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN, USA

V. Marmarelis and G. Mitsis (eds.), *Data-driven Modeling for Diabetes*, Lecture Notes in Bioengineering, DOI: 10.1007/978-3-642-54464-4_8,

1 Introduction

The glucose-insulin system is a time-honored topic for the development of mathematical models, starting from the early days of Bolie [1] and Ackerman et al. [2]. Part of the attention devoted to this specific physiological system over the years is due to its embodying, in its most skeletal interpretation, a relatively simple, clear-cut feedback mechanism: when glycemia increases, the pancreas secretes insulin which brings glycemia down. If the system were this simple, however, mathematical physiologists would probably not have found material for continuing scientific investigation for more than 50 years to date. In fact, the control is complicated by internal nonlinearities and delays, by external influences (notably through the link with energy consumption and lipid metabolism) and by the superposition of different levels of neural and hormonal regulation (e.g. through the sympathetic system and through the incretins mechanism).

Modeling the regulation of glycemia through its main controller, the hormone insulin, may satisfy different purposes. On one hand, understanding short-term (minutes to hours) regulation is relevant to the conduction of perturbation experiments (such as the Intra-Venous Glucose Tolerance Test, IVGTT, or the Oral Glucose Tolerance Test, OGTT), whose goal is to identify characteristics of the experimental subject (e.g. insulin sensitivity), of direct interest to the clinician or to the investigator; other Authors deal with modeling of short-term glucose-insulin regulation within the context of the present volume. On the other hand, understanding the long-term regulation of the system (months to years or decades) should help devising therapeutic schemes, or assessing the likely impact and ultimate benefit of drugs, for which only the short term effects are known. As will be discussed at length in the present chapter, depending on the assumptions that the physiologist is willing to accept, modeling may clarify quantitatively the ultimate results, which can be expected if the assumptions hold true.

The present chapter is structured in the following way: in the continuation of this Introduction, a brief survey of the physiology of the glucose-insulin system is sketched, as is relevant to long-term diabetes progression. A model for long-term glucose-insulin-betacell population dynamics, previously introduced and validated by the present authors, is then briefly described. Relevant model features are then discussed, with emphasis on the applicability of this kind of modeling to clinical trial design and simulation.

1.1 Physiological Remarks

The short-term regulation of glycemia involves a hormonal response based on glucagon and adrenalin (responding to hypoglycemia) as well as on insulin (responding to hyperglycemia). In the context of a discussion on diabetes, the main metabolic defect consists either in a failure of pancreatic β-cells to secrete

sufficient insulin in response to hyperglycemia (Type 1 Diabetes Mellitus, T1DM) or in a failure of the secreted insulin to determine an increase in tissue glucose uptake (peripheral insulin resistance) and/or decreased liver gluconeogenesis and glycogenolysis (central insulin resistance) [3].

There is evidence from rodents that another negative feedback loop acts over a longer time-span: sustained or sub-acute hyperglycemia may increase the net rate of β-cell replication (β-cell replication or differentiation from progenitor ductal cells, minus β-cell apoptosis) [4–8], so that the active β-cell population increases, the capacity for insulin secretion increases, and insulin resistance is somewhat overcome.

However, when insulin resistance continues for a long time, it is common to observe clinically the progressive 'exhaustion' of insulin secreting ability, the establishment of a relative insufficiency of elevated circulating insulin levels to compensate for insulin resistance, and the emergence of frank Type 2 Diabetes Mellitus (T2DM). It has in fact been observed that after prolonged severe hyperglycemia β-cell replication rates may be depressed, possibly due to glucose toxicity [9]. It is therefore the interplay of variations in insulin sensitivity and the resulting changes in pancreatic β-cell responsiveness and β-cell population size that determines the development of T2DM. In most individuals, the development of insulin resistance and the loss of β-cell function develops over many years and reflects genetic and environmental influences that differ widely among affected individuals.

Whatever the genetic or environmental determinants, a progressively worsening insulin resistance translates at first in a slowly increasing glycemia, due to the early compensation by increased circulating insulin levels. This compensatory process is known to involve an expansion of β-cell mass in rodents [4, 10], and similar conclusions can be drawn from studies of human autopsy specimens [11]. If this mechanism were able to continue indefinitely, higher and higher insulin concentrations would indefinitely compensate for lower and lower insulin sensitivity, but what is observed clinically is that compensation is not complete, and indeed in the later stages of the disease a deficit of β-cells develops: glucose toxicity on β-cell replication/apoptosis rates has been strongly advocated as a possible cause of this eventual failure [12]. As soon as the hyperinsulinemic compensation mechanisms fails, glycemia increases rapidly and T2DM develops [13].

It is important to realize that the sequence of events described above appears to underlie the progression from the normal state, to the pre-diabetic states of Impaired Glucose Tolerance (IGT) or Impaired Fasting Glucose (IFG), to frank T2DM. In fact, isolated insulin resistance, as can be seen in severely obese individuals and in pregnant women, or simply reduced β-cell mass, occurring for instance in pre-clinical T1DM [14] and in pancreatectomized animals after islet transplantation [15], do not lead to appreciable hyperglycemia by themselves.

Prompted by the clinical and experimental need to devise procedures for the assessment of the degree of insulin sensitivity in a given subject, perturbation protocols like the Intra-Venous Glucose Tolerance Test (IVGTT), the Oral Glucose Tolerance Test (OGTT) or the Euglycemic Hyperinsulinemic Clamp (EHC) have been devised. The glycemia and insulinemia observations during a perturbation

study reflect of course the short-term control mechanisms, and mathematical models have been proposed aiming at the quantitative understanding of such mechanisms [1, 16–22]. If the short-term model is correct and the dataset is sufficiently rich, statistical estimation of the model's parameters allows the snapshot determination of the degree of insulin resistance or beta cell function at that particular time in the life of the subject.

It is on the other hand very difficult to assess quantitatively the natural history of the development of diabetes, since its progression is slow (over decades), it is difficult to measure relevant variables (like β-cell mass) and variability (also due to diverse lifestyle and therapeutic interventions) is high. In fact, we do not have long-term studies of cohorts of individual subjects (beyond a few years). In this context, a mathematical model of diabetes progression of the disease offers the opportunity to simulate the evolution of the disease, depending on the assumptions made, and to explore in silico different scenarios, with different primary defects (in insulin sensitivity or secretion) associated with different therapeutic schemes. In this as in other areas of physiology, the utility of a model is to force the investigator to unambiguously specify the assumptions made, delivering quantitative, verifiable predictions corresponding to those assumptions.

Different approaches to the long-tem forecasting of the evolution of Diabetes have been pursued: Mechanistic models for the progression of diabetes [23, 24] attempt to coherently integrate commonly accepted physiological mechanisms in order to predict the long-range behavior of disease processes. Other models extend mass-balance principles to long-term situations [25]. Discrete-state dynamical systems with covariate-dependent transition probabilities among states [26–28] depict the statistically observed evolutions over populations without depending on the identification of the underlying physiology. The quest for a good long-term model of diabetes progression has practical implications: a valid disease and therapeutic model could be used to inform clinical trial design, influencing decisions about study population, sample size, and treatment duration through model simulations.

In the following we will describe a recently introduced [24] and validated [29] model for the long-term progression of diabetes mellitus, based upon the interplay of glycemia, insulinemia and β-cell mass. This model explicitly represents such phenomena involved in glucose homeostasis as glucose-dependent insulin secretion, insulin mediated glucose disposal, short-term stimulation and long-term depression of insulin production by hyperglycemia (glucose toxicity), allowing the simulation of progressive worsening of insulin resistance and of the resulting changes in β-cell mass.

Based upon this model, sample life trajectories of subjects with different combinations of primary defects will be explored: disease heterogeneity in T2DM is manifest in multiple ways, including age of onset, relative contribution of insulin resistance and β-cell dysfunction, rate of disease progression, risk of complications, and response to available therapies. Tailored therapeutics (TTx) involves matching treatment options to individual patient characteristics with the goal of maximizing the benefit/risk ratio. In this context, the DPM mechanistic mathematical model of T2DM long-term evolution is used to explore TTx hypotheses.

In the real world patients show variegated responses to different therapies due to many interacting factors. The model allows us to define primary 'pure' defects and well-defined combinations thereof: the main goal in the examples will be to use model predictions to investigate the interaction between specific pathophysiologies and specific therapies.

2 The Diabetes Progression Model

The main model equations are reported below. For a thorough description of the structure of the model and a detailed report of the assumptions made in order to assess model parameter values see De Gaetano et al. [24] and Hardy et al. [29]:

$$\frac{dB(t)}{dt} = \lambda(G)B\left(1 - \frac{B}{B_{max}}\right), \quad B(t_0) = B_0 \tag{1}$$

$$\frac{d\eta}{dt} = -K_{\eta G}\, G\eta + T_\eta, \quad \eta(t_0) = \eta_0 \tag{2}$$

$$G = \frac{\gamma}{\rho + I} \tag{3}$$

$$I = h(G)\ \ I_{max\,B} B \tag{4}$$

$$h(G) = \frac{(G/G_h)^{\nu_h}}{1 + (G/G_h)^{\nu_h}} = \frac{G^{\nu_h}}{\alpha_h + G^{\nu_h}}, \text{ letting } \alpha_h = (G_h)^{\nu_h} \tag{5}$$

$$\lambda(G) = \lambda_{min} + \eta\frac{x^3}{1 + x^3},\ x(G) = x_0\frac{G}{G_\lambda} \tag{6}$$

$$K_{xi}(t) = K_{xiStart} + \frac{t - t_0}{t_{max} - t_0}(K_{xiEnd} - K_{xiStart}) \tag{7}$$

$$\frac{dA}{dt} = -K_{xa}A + K_{ag}G\frac{(100 - A)}{100}, A(t_0) = A_0 \tag{8}$$

with $\gamma = \frac{T_{gI}}{K_{xgI}}$, $\rho = \frac{K_{xg}}{K_{xgI}}$, $I_{max\,B} = \frac{T_{igB}}{K_{xi}}$, $T_{igB} = \frac{K_{xiStart}\, I_0}{h(G_0)B_0}$, $K_{ag} = \frac{K_{xa}\, A_0}{G_0\frac{(100-A_0)}{100}}$, and $\lambda_{max} = \lambda_{min} + \eta$.

The definitions and units of all state variables and parameters are reported respectively in Tables 1 and 2.

Equations 3 and 4 simply state that prevailing glycemias at any historical moment are inversely related to the prevailing insulinemias, while prevailing insulinemias are directly related to (an increasing function of) prevailing glycemias as well as being directly related to available β-cell mass and secretory capacity. The "prevailing" glycemia and insulinemia levels described in the model do not strictly refer to either fasting or post-prandial values, but are in fact representative

Table 1 Definition of state variables

State variables: symbol and meaning	Units
t, time	[mo]
B, β-cell mass as Millions of active β-cells	[Mc]
I, fasting serum insulin concentration (monthly average)	[pM]
G, fasting plasma glucose concentration (monthly average)	[mM]
h, glucose effect on pancreas	[#]
x, replaces Glucose variations with variations of a scale-free pure number	[#]
λ, net rate constant for β-cell growth (or decay), resulting from the difference between production (replication) rate and mortality (apoptosis) rate	[mo^{-1}]
η, possible lambda excursion, represents pancreatic replication 'reserve'	[mo^{-1}]
K_{xgI}, second-order insulin-dependent glucose tissue uptake rate per pM insulin, represents insulin sensitivity	[min^{-1}/pM]
γ, resistance to insulin as the number of mM of glucose at which glycemia stabilizes for a single pM of insulinemia	[mM * pM]
λ_{max}, maximum (positive) value of λ, i.e. the maximum possible difference between replication rate and apoptosis rate	[mo^{-1}]
I_{maxB}, insulin attainable levels, expressed as the maximal contribution of a million β-cells to fasting insulin plasma concentration	[pM/Mc]
A, glycosylated Haemoglobin, HbA1c (monthly average)	[%]

of daily steady-state blood concentrations, hourly oscillations being averaged out when considering the evolution of the system over a time span of months or years. Prevailing glycemia and insulinemia are derived as equilibrium values, given the existing level of insulin sensitivity and the current functional β-cell mass available for insulin secretion: they substantially reflect real-life post-absorptive values.

The model hypothesizes, in a straightforward population dynamics fashion, that the variation in β-cell mass B (Eq. 1) is proportional to current mass in logistic growth fashion, with carrying capacity B_{max}, multiplied by a net-growth rate λ.

In turn, λ depends (Eq. 6) on prevailing glucose concentrations, in the sense that it varies from a minimum negative value (λ_{min}) to a maximum value that is dependent both on the prevailing pancreatic reserve described below (i.e. $\lambda_{max} = \lambda_{min} + \eta$) and on glucose level according to a sigmoidal 3rd-degree Hill function, with $\lambda = \lambda_{min}$ when G = 0, and λ tending to λ_{max} as G tends to infinity.

The Hill function is normalized (it is a function of x rather than G) so as to have maximum slope at the glycemia value G_λ. If G_λ is approximately equal to normal fasting glycemia (e.g. around 5 mM), this is the same as assuming that the regulation mechanisms of the pancreas work optimally in a neighborhood of the target fasting glycemia.

The term η represents the 'pancreatic reserve', i.e. the current ability of the pancreas to increase its β-cell proliferation rate if sufficiently stimulated by the ambient glucose concentration. Pancreatic reserve, however, is not a fixed quantity: it varies with time (Eq. 2), in such a way that it always tends to an equilibrium level. This equilibrium is the result of two competing forces: glucose-toxicity-induced shrinkage of pancreatic reserve and spontaneous recovery of the pancreas. Pancreatic reserve (at equilibrium) is thus inversely proportional to prevailing

Table 2 Definition of model parameters

Model parameters: symbol and meaning	Units
$B_0 = B(t_0)$: initial condition on β-cell mass	[Mc]
B_{max}: maximal size of β-cell mass	[Mc]
I_0: insulinema at age t_0	[pM]
G_0: glycemia at age t_0	[mM]
G_λ: the glycemia of maximal sensitivity of the regulation of β-cell population dynamics	[mM]
T_{gl}: liver glucose output. This parameter can be considered as a constant basal, insulin-independent glucose output, while insulin-dependent glucose output (decreasing with increasing levels of insulin) is taken into account through the action of KxgI	[mM/min]
K_{xg} (FAST): first-order insulin-independent glucose tissue uptake rate	$[min^{-1}]$
T_{igB} (FAST): maximal pancreatic insulin secretion per million β-cells	[pM/min/Mc]
K_{xi} (FAST): apparent first-order elimination rate constant for insulin	$[min^{-1}]$
$K_{xiStart}$: apparent first-order elimination rate constant for insulin at baseline (e.g. at age 18 years)	$[min^{-1}]$
K_{xiEnd}: Apparent first-order elimination rate constant for insulin at the end of a normal life (e.g. at age 90 years)	$[min^{-1}]$
K_{xgI0}: value of K_{xgI} at t_0	$[min^{-1}/pM]$
t_I: time of midpoint K_{xgI} decrease (time at 0.5 * K_{xgI0})	[mo]
ν_I : exponent for the decrease in K_{xgI}	[#]
ρ: ratio of 1st order to 2nd order (Insulin Dependent) rate constants for tissue glucose uptake from plasma	$[min^{-1}/(min^{-1}/pM)] = pM$
I_{maxB0}: value of I_{maxB} at t_0	[pM/Mc]
λ_{min}: the minimum (negative) value of λ, i.e. the maximum net apoptosis rate	$[mo^{-1}]$
η_0: value of η at t_0 (determined)	$[mo^{-1}]$
T_η: the spontaneous recovery rate of the pancreas	$[mo^{-2}]$
$K_{\eta G}$: Rate constant expressing pancreatic glucose toxicity	$[mo^{-1}mM^{-1}]$
G_h: centering glucose concentration for Hill-shaped glycemia effect on pancreatic insulin release, set to 9 mM	[mM]
ν_h: power coefficient for Hill-shaped glycemia effect on pancreatic insulin release, set to 4 in order to achieve linearity between 5 and 10 mM, essentially zero before 3.5 mM, essentially saturated above 25 mM	[#]
h_0: The value of the h(G) function at t_0 (determined)	[#]
ξ: the scaling factor for the X support variable, computed so as to center the X sigmoid curve on the value $\lambda = 0$	[#]
$A_0 = A(t_0)$: initial condition on glycosylated Haemoglobin (HbA1c).	[%]
K_{ag} : the rate constant of production of glycosylated haemoglobin from circulating glucose	[%/mo/mM]
K_{xa}: the spontaneous elimination rate constant of (glycosylated) Haemoglobin	$[mo^{-1}]$
t_0: starting age, with system at equilibrium, in months	[mo]
$t_{\eta 10}$, Half-life of Pancreatic reserve at 10 mM Glucose	[mo]

glycemias. In other words, Eq. 6 prescribes that hyperglycemia will always acutely stimulate pancreatic β-cell replication (to the extent that this is allowed by the current pancreatic reserve level), while Eq. 2 prescribes that, chronically, sustained

hyperglycemia will make pancreatic reserve decrease. Notice that, if pancreatic reserve is sufficiently damaged, i.e. if $\eta < -\lambda_{min}$, then $\lambda_{min} + \eta < 0$ and β-cell population would decline, in the absence of therapy, no matter what the prevailing glucose concentration. It should be kept in mind that this effect could well stem both from an anti-proliferative and from a pro-apoptotic original lesion.

Equations 3 and 4 derive, as hinted to above, from the equilibrium condition for glycemia and insulinemia respectively, given that the model focuses on changes over months or years. At each point in slow time, therefore, the fast dynamics of glucose and insulin are assumed to be at equilibrium. This equilibrium may be conceived as being the average glycemia produced by the average insulinemias and by the current levels of insulin sensitivity. Conversely, average insulinemia is determined by the prevailing glycemias, by the β-cell mass available at the time, and by a coefficient of insulin production at maximal stimulation per β-cell mass unit. It should be pointed out that using fasting or average glucose and insulin concentrations amounts conceptually to the same thing, as we only need indicator quantities for the prevailing state of the system over a period of a day or a week.

The function h (Eq. 5) captures the hypothesis that glycemia stimulates insulin production by the pancreas in a sigmoidal fashion, starting at zero, sloping up approximately linearly between 5 and 10 mM glycemia, and approaching asymptotically a maximum as glycemia increases.

We found it necessary to stipulate that, as observed in the literature [30], the apparent first order rate constant of elimination of insulin from plasma decreases with age (Eq. 7): this allows the model to replicate the observed decline of insulin secretory function, associated with preserved or even increased levels of serum insulin, with advancing age [31].

The variable γ (inversely related to the index of insulin sensitivity K_{xgI}) expresses resistance to insulin, as the concentration at which glucose stabilizes for each pM of insulin concentration.

For the purposes of the present simulations, it has been assumed that the insulin-independent glucose tissue uptake rate constant K_{xg} is small (within the range of considered glycemias) compared to insulin-dependent glucose uptake [19], and that consequently the ratio ρ is also approximately zero.

The variable I_{maxB} is the maximal contribution of a million β-cells to fasting plasma insulin concentration. Its value is determined by the maximal insulin secretion rate per million β-cells T_{igB} (determined by equilibrium conditions at t_0), and by the actual first-order apparent rate of elimination of insulin from plasma, K_{xi}, which is considered here to decrease with age (linearly, in first approximation), as mentioned above. It should be kept in mind that T_{igB} in our simulations is made to decrease when supposing primary defects of insulin secretory function with advancing age, and conversely it is made to increase in response to secretagogues.

In order to link observable glycated hemoglobin with the glucose dynamics, a simple linear model of the kinetics of HbA1c (indicated as A in Eq. 8) is part of DPM, with increase determined by prevailing glycemias and by the concentration of native HbA0, and decrease linearly determined by the continuous destruction of red blood cells.

2.1 Parameter Assessment

Parameters for the model have been carefully assessed from literature data and from the results of directed experiments, with the goal of maintaining consistency among the several sources. Only human data have been used. Although limited information was available for some parameters, most were based on empirical observations derived using well-established methodologies. The majority of parameters were obtained from in vivo studies but in a few cases only in vitro or post-mortem data were available. Preference was generally given to studies with larger numbers of subjects, but study quality and design were also considered (e.g. age range of study population, level of validation of key measurements). For each parameter, we sought to identify a typical value, reflecting a non-diseased population, keeping in mind that the range of reasonable values for each parameter would include those seen in diabetic individuals. Typical values and ranges at time $= t_0$ were generally taken from data obtained in young, healthy adults (age 18–30).

Parameter values used in this simulation were as reported in [29], with the following specifications. Moderate secretion and sensitivity defects were assumed to have a minimum reachable value of the corresponding coefficients at 15 % normal, whereas severe defects were supposed to cause a decrease of the coefficient down to 2.5 % normal. In both cases the age at 50 % defect was assumed to be 25 years. The threshold of glycemia for the diagnosis of diabetes was 7 mM. Proportional improvement by classical sensitizers or secretagogues was assumed to be 50 % with respect to the current "natural" value as determined by the progression of disease. Definitive improvement of insulin sensitivity produced by bariatric surgery was assumed to 7.5 % of normal (i.e. a small fraction of supposedly perfect insulin sensitivity as seen in young age) above the current "natural" value due to age and disease. Definitive decrease of glucose toxicity produced by β-cell protectors was assumed to be 20 % of normal.

The model's numerical integration, starting with given parameter values, has been implemented in a mixed Matlab (© 1994–2007 The MathWorks, Inc.) and C/C++ (GCC, © Free Software Foundation, Inc.) environment, and is freely available as a service to academic users through the CNR IASI BioMatLab website [32].

3 Results

3.1 Qualitative Analysis of the Behavior of Solutions

The original published model [24] was proven to admit three equilibrium points: a physiological steady-state, which is locally asymptotically stable, consisting of the (normal) basal (i.e. 'early-life') values of glycemia and insulinemia G_b and I_b; a pathological 'severe diabetes' steady-state, which is also locally asymptotically stable, consisting of zero levels of β-cell mass and insulinemia, to which there

corresponds a hyperglycemic state ($G_h > G_b$); and an unstable saddle point, with intermediate glycemia ($G_b < G_s < G_h$).

In hypoglycemic ($G(t) < G_b$) or moderate hyperglycemic conditions ($G_b < G(t) < G_s$), plasma glycemia tends to return to the physiological equilibrium point, $G(t) \rightarrow G_b$. Imposing a moderate external increase in glycemia (as can happen by prolonged inappropriate dietary behavior or by independently worsening insulin sensitivity) pushes towards an increase in β-cell mass and insulinemia; when the saddle point threshold is exceeded ($G(t) > G_s$) glucose toxicity prevails and β-cell mass eventually disappears, leading to the pathological steady-state with zero insulin production.

Concerning the effects of parameter changes on model behavior, decreasing insulin sensitivity does not change the existence (and stability properties) of the three equilibrium points, nor does it change the relative equilibrium glycemias. Conversely, a decrease in insulin sensitivity produces an increase of both equilibrium β-cell mass and equilibrium insulinemia (needed to maintain the same equilibrium glycemia).

Increasing glucose production does not change the existence (and stability properties) of the three equilibria. While an increase in glucose production does not increase the glycemia of the physiological steady-state (which is stabilized by larger β-cell mass and corresponding insulinemia), it does increase the pathological (diabetic) glycemia (corresponding to the absence of β-cells and zero insulinemia).

Further detailed qualitative analysis results are proven in the original paper [24].

3.2 Numerical Simulations

The evolution of fasting glycemia, fasting insulinemia and beta-cell mass were simulated for virtual patients with a range of defects in insulin secretion, insulin sensitivity, or both. For each hypothesized time course of the primary defects or their combination, the natural history of the disease as well as the response to several interventions or combinations of interventions were simulated. Among the therapeutic approaches tested are pharmacologic interventions directed at improving insulin sensitivity and insulin secretion proportionally (50 %) to underlying metabolic status and secretory ability, as well as radical maneuvers hypothetically restoring a fraction of normal sensitivity or secretion, such as metabolic/bariatric surgery (sensitivity to at least 10 % normal) or partial protection from inhibition of β-cell replication (reduction of glucose toxicity by 20 %). Results from simulation are reported in Figs. 1, 2 and 3. Figure 1 depict the evolution of the insulin sensitivity (Fig. 1a), beta-cell insulin secretory function (Fig. 1b), as well as the corresponding model-derived time-courses of beta-cell mass (Fig. 1c), insulinemia (Fig. 1d) and glycemia (Fig. 1e) for a patient with a combined, moderate defect in insulin sensitivity and insulin secretion, undergoing the four different type of therapeutic approaches.

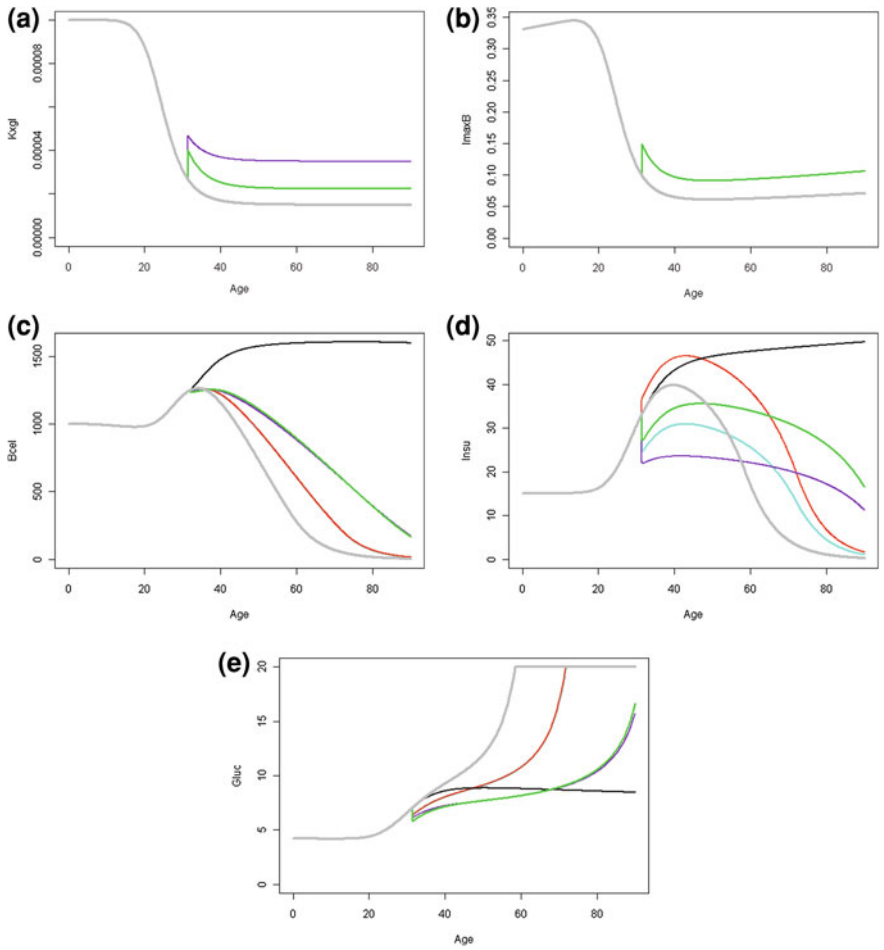


Fig. 1 Combined moderate sensitivity and moderate secretion defects. Patient simulated metabolic variable trends under "Moderate" defects in both insulin sensitivity and insulin secretion, that is an eventual development of an 85 % defect (i.e. sensitivity or secretion down to 15 % of normal). Panel **a** shows the assumed time courses of insulin sensitivity (KxgI); panel **b** shows the beta-cell insulin secretory function (ImaxB,); Panel **c** shows the corresponding model-derived time-courses of beta-cell mass (Bcel); Panel **d** shows the insulinemia time course (Insu) and Panel **e** shows glycemia (Gluc). Each panel reports the time-course of the corresponding variable under no intervention, that is the natural course of the disease (*grey line*), under treatment with an insulin sensitizer (*cyan line*), under treatment with an insulin secretagogue (*red line*), under treatment with insulin sensitizer plus secretagogue (*green line*), after bariatric surgery intervention (*purple line*) and after a hypothetical radical treatment offering glucose toxicity protection (*black line*)

A "Moderate" defection, in the present simulation, means the eventual development of an 85 % defect (i.e. sensitivity or secretion down to 15 % of normal) which leads to the development of Diabetes Mellitus (DM, Glycemia > 7 mM) if no intervention is made (see Fig. 1e). It is to be noted that the effects of

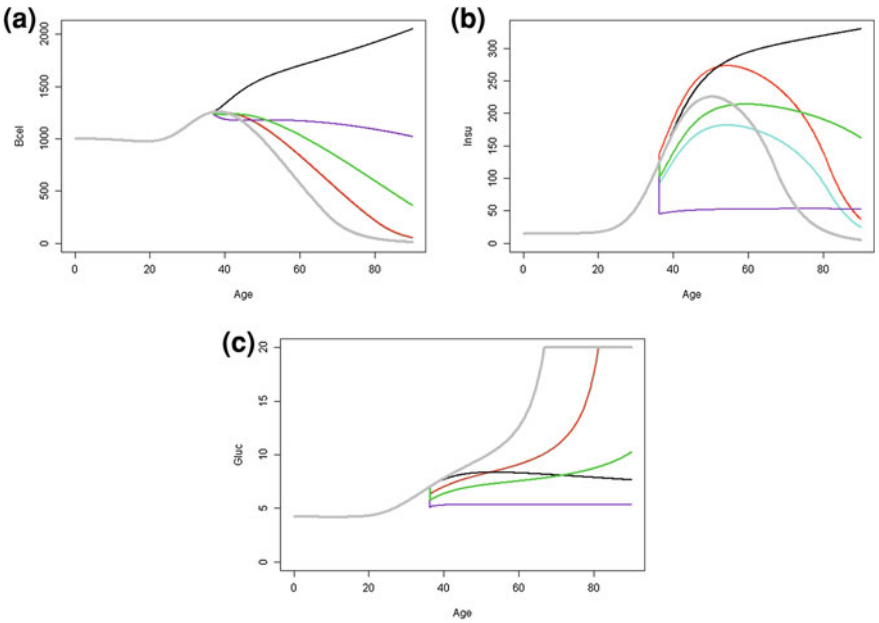


Fig. 2 Severe isolated insulin sensitivity defect. Patient simulated metabolic variable trends under an isolated "Severe" defect of insulin sensitivity, which means an eventual development of a 97.5 % defect (sensitivity down to 2.5 % of normal). Panel **a** shows the model-derived time-courses of beta-cell mass (Bcel); Panel **b** shows the insulinemia time course (Insu) and Panel **c** shows glycemia (Gluc)

conventional insulin sensitizer and secretagogue on glycemia, hence on β-cell mass are identical: the cyan and red curves are superimposed and only the red one is visible (see Fig. 1e and c). Figures 2a–c and 3a–c, report results related to a severe isolated insulin sensitivity defect and to a severe isolated insulin secretion defect respectively. The term "Severe", in the present simulation, means the eventual development of a 97.5 % defect (i.e. sensitivity or secretion down to 2.5 % of normal). If no intervention has taken, either a severe isolated insulin sensitivity defect (Fig. 2) or a severe isolated insulin secretion defect (Fig. 3) can lead to the development of Diabetes Mellitus (DM, Glycemia >7 mM) by age 40 (Figs. 2c and 3c). The clear difference between the isolated sensitivity defect in Fig. 2 and the isolated secretion defect in Fig. 3 consists in the respective "b" panels, insulin secretion being supra-normal in the first case and sub-normal in the second. The time courses of glycemia and, correspondingly, of beta-cell mass do not differ much in the two situations. If only isolated moderate defects in either sensitivity or secretion are considered an increase of glycemia (which however stabilizes at moderately hyperglycemic levels at steady state) (results are not shown) is observed but it does not lead to the development of DM.

As can be seen from Figs. 1, 2 and 3, response to individual therapies is influenced not only by the degree of insulin resistance or insulin secretion defects,

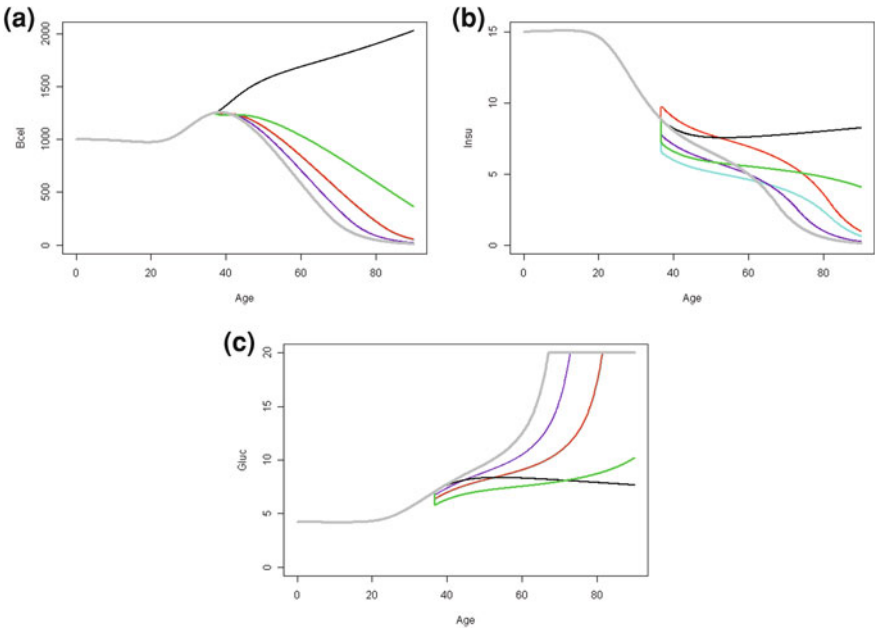


Fig. 3 Severe isolated insulin secretion defect. Patient simulated metabolic variable trends under an isolated "Severe" defect of insulin secretion, which means an eventual development of a 97.5 % defect (secretion down to 2.5 % of normal). Panel **a** shows the model-derived time-courses of beta-cell mass (Bcel); Panel **b** shows the insulinemia time course (Insu) and Panel **c** shows glycemia (Gluc)

which are function of the type and stage of progression of the original abnormality, but also by the accumulated injury to beta-cell replicating ability, which depends on the duration and severity of the primary defect.

Therapy directed at proportional improvement of the current insulin sensitivity or secretion levels has largely equivalent effects, independent of the nature of the original defect. This effect lacks durability if started after development of hyperglycemia. On the other hand, combinations of these moderately effective interventions, or single interventions with bariatric surgery or protection of beta cells from glucose toxicity, are predicted to have much more pronounced and long-lived effects.

3.3 Model Use Scenarios

3.3.1 Scenario 1

As mentioned above, the importance of the modelling approach resides in the ability of depict possible disease evolutions for a specific patient, under no treatment or under one or a combination of treatments. In general, disease

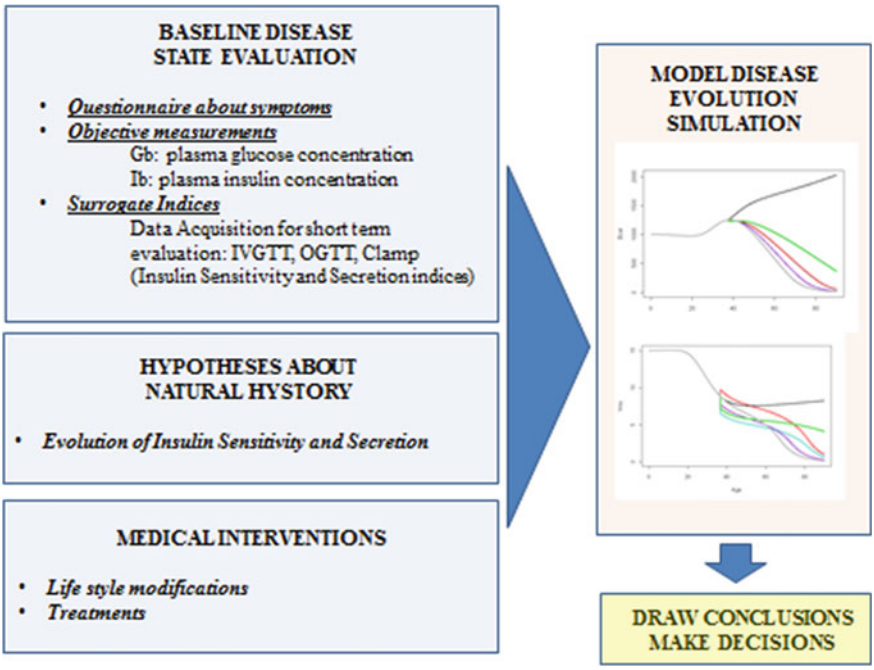


Fig. 4 DPM for tailored patient therapy. Schematic representation of the connection of information items for the exploitation of the Diabetes Progression Model (DPM) towards individually tailored diabetes therapy

progression models require a measure of the "disease status" at the evaluation time t, described by present and recent symptoms, by measurements of physiological and biological conditions (in the specific case basal glucose and insulin plasma concentrations) as well as by measures of surrogate endpoints associated with or predictive of the disease state (in this case, for example, short-term model-derived indices of insulin sensitivity or insulin secretion). After the Baseline Disease State description, the investigators have to make hypotheses about the evolution of the disease, according to its natural history, for example hypotheses about the likely time-course of insulin sensitivity or secretion under no treatment. The model can then be used to simulate the effect of different treatments and the corresponding responses under different scenarios (life style interventions, different disease evolutions) and to compare the effectiveness of the different manoeuvres in order to draw conclusions. Figure 4 schematically summarizes this scenario.

3.3.2 Scenario 2

The model presented here is a comprehensive model of the global functioning of the insulin/glucose system. It includes in fact equations describing the evolution of insulin secretion, β-cell mass dynamics and insulin sensitivity evolution, feedback

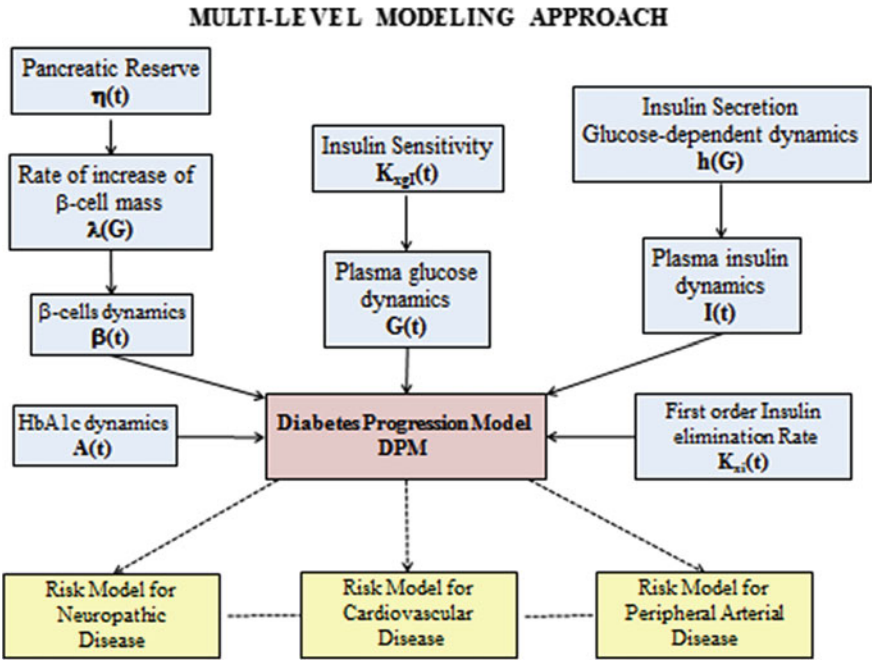


Fig. 5 Mult-level modeling approach. Schematic representation of the Diabetes Progression Model (DPM) in the framework of a multi-level approach

mechanisms describing the interaction between insulin and glucose plasma concentrations, as well as an equation describing HbA1c dynamics. Some of the model equations and the evolution over time of a series of parameters (such as the rate of increase of the β-cell mass or the first-order insulin elimination rate) enter the model as forcing or input functions: their form is free and is the result of a number of hypotheses, reflects literature information and current physiological knowledge. The presence of freely modifiable input functions also allows the modeler to represent different pathophysiological situations and different interventions and therapeutic maneuvers. This flexibility opens the possibility of improvement of each single piece of the model, as a consequence of future specific clinical trials, of clinical or in vitro experiments, of the use of advanced technical devices for monitoring (calories intake, physical activity, glycemic control). In this view the model acts as an "Output" into which an array of diverse "Inputs" can be incorporated. On the other hand the model can be seen as representing one possible "Input" in the framework of a wider consideration of the constellation of clinical issues affecting the presentation and outcome of diabetes mellitus. For example, the Diabetes Progression Model can represent an input into further modeling of the risks and complications of diabetes mellitus (neuropathy, stroke, peripheral arterial disease, hypertension, in the sense that models for the risk of developing associated complications could taking in input the simulated evolution of glycemias as

predicted by the DPM. This approach is consistent with current frameworks for multi-level modeling, such as those centering around the virtual physiological patient (see for example http://www.mosaicproject.eu or http://www.vph-noe.eu), or it could represent one important element of a future comprehensive diabetes modelling project. Figure 5 summaries the concept.

4 Discussion

The need of a quantitative description of the likely long-term evolution of the compensation state of the glucose-insulin system is a worthwhile goal for diabetological research. Such a description would be very valuable, for instance, in the design of trials testing the efficacy of drugs delaying or preventing the occurrence of T2DM.

Obtaining accurate estimates of the parameters defining the long-term evolution of this system is very difficult, due to the long time-scale of the phenomenon, which makes comprehensive longitudinal studies impractical, to the ethical issues connected with observing subjects with increasing severity of the disease without intervening, and to the difficulty of obtaining serial measures of β-cell mass, which is a key determinant of the compensation (even though imaging and functional tests may deliver surrogate measures [33]).

Much of the information which we can use to estimate the β-cell population response to chronic hyperglycemia stems from chronic glucose infusion studies [34–36]. We have discussed elsewhere [24] at length the assessment of the relative implications of β-cell apoptosis and replication rate changes in the overall adaptation to changes in (average) glycemia, but it seems clear that while short-term hyperglycemia induces an increased net replication rate of β-cells (at the very least in rodents [4]), long-term hyperglycemia impairs the same replication rate [9]. While the concept of 'glucose toxicity' may still have to be clarified, in its exact mode of operation at the cellular and subcellular level, there do not seem to be any substantial doubts that long-term hyperglycemia is associated with decreased β-cell mass in humans. DPM successfully predicts this decline in beta cell mass by introducing the term "pancreatic reserve" (η), which is assumed to decline with chronic hyperglycemia.

It can be noticed that one physiologic implication of the present definition of pancreatic reserve η is that persisting long-term hyperglycemia could in fact lead to such a decrease in η that the net replication rate would be negative, and consequently that pancreatic β-cell mass would decrease at a given moment (say, when the subject comes to medical attention), no matter what normal glycemias could be restored in the short term by aggressive therapy. This situation corresponds to a terminal state of pancreatic insufficiency, and would naturally lead to a clinical picture of frank diabetes, given an adequate time frame. However, in this as well as in less dramatic pre-diabetic conditions, early administration of glycemia normalizing therapy (whatever the approach, whether by insulin sensitizers,

or secretagogues, or by important and persistent lifestyle changes) would be conducive to the restoration of pancreatic reserve and to the delay or avoidance of the final self-sustaining, positive-feedback deterioration of diabetes. Whether this mathematical interpretation of events actually well represents physiological and pharmacological changes is open to experimental verification, but as a first-line approximation this definition of pancreatic reserve seems to conform well to commonly accepted physiological notions.

Another point worth of notice is that the DPM predicts that, in the natural course of the progression of the decompensation due to increasingly severe insulin resistance (such as, for instance, that due to dietary intemperance and massive increase in BMI), glycemias are maintained near normal levels for a long time, before compensation fails and a rapidly increasing glycemia determines the clinical picture of T2DM. It is such a common clinical experience to observe patients, asymptomatic and well compensated for decades (possibly with modest increases in glycemia), suddenly turning diabetic, that a concomitant pathologic cause is often hypothesized, such as an infection, surgery, or some severe stress episode. According to the DPM, no such concomitant cause is actually needed: it is the natural behavior of the described compensation system to be able to maintain normal or acceptable glycemia up to some late failure time, when the disease becomes clinically manifested.

In the introduction of the present model, a very substantial effort has been made to assess, if not to statistically estimate, typical parameter values which are coherent with accepted physiology. To this end, the literature has been extensively consulted and different sources of direct or indirect information on the same model parameter have been critically reconciled. While, clearly, different opinions are perfectly acceptable (indeed ameliorations are to be expected), it is felt that the values obtained here represent, as far as possible, prudent and informed choices. In fact, the end result of such choices is that the model now reproduces naturally such time courses of glycemia and insulinemia over many years as are felt to be compatible with the observed clinical history in representative cohorts of subjects [29].

While we feel that this work represents an improvement in the direction of Topp's approach [23], we believe that there are substantial differences with respect to another model recently proposed by de Winter et al. [25, 37].

The stated modeling goal of both the present approach and de Winter's is to represent the long-term variations in the ability of the individual to adapt to glucose loads maintaining acceptable glycemias. However, de Winter's approach extends previous mass balance representations over a more extended period of time. It is however questionable that mass balance concepts apply in the present situation, where basal insulinemia and basal glycemia are sampled infrequently (say, once a day), over a period of many years. The automatic transposition of the formalism of mass balance, which is justified in the short term (e.g. 3 h), when transiently elevated glucose concentrations may themselves be responsible for their progressive decline due to insulin-dependent or insulin-independent mass transfer out of the central compartment, is in fact unwarranted when modeling the

long run of basal glycemia. An in-depth discussion of this issue is presented elsewhere [24]. It is of interest in this context that the Diabetes Progression Model can be shown to be the slow equilibrium model for the fast models of a wide class, including, by way of example, both de Winter's mass balance formulation (considered in fast time) and Bergman's Minimal Model [16].

That the model, as described, has a substantial degree of physiological plausibility can be argued from the fact that DPM has been demonstrated [29] to produce predictions in very good agreement with both the Diabetes Prevention Program [38, 39] and the CANOE [40] trial results. One of the interesting uses to which a robust model of diabetes progression may be put is the design of clinical trials and the evaluation of therapeutic regimens. One fundamental question regarding diabetes management and, possibly, prevention is whether there exists a possibility to intervene early enough in the history of the disease, so that the eventual crisis and the development of the frank clinical picture of diabetes mellitus are averted. Judging from the Diabetes Progression Model, the answer may be affirmative: if the interaction of glycemia, glucose toxicity and β-cell replication is qualitatively similar to what has been expressed in the proposed equations, a sufficiently early and effective intervention would spare pancreatic reserve and actually prevent the positive-feedback chain of events leading to the manifestation of the disease. The earlier the intervention, the less drastic it may need to be: reasonable dietary habits and lifestyle since the early teens may be as effective as drastic bariatric surgery later on.

In general, the advantage of using the model, in these circumstances, is that once a quantitative estimate of the effect of a putative treatment is obtained from short-term observations, its long-term effects may be forecast. These need not be, and in general are not, naively proportional to the short-term gain, given the highly nonlinear interplay of the several factors concurring in determining glucose homeostasis. This can be well appreciated from Fig. 1 (Panels d and e), where the model predicts that long-term glycemia and insulinemia eventually follow the same time course (albeit somewhat delayed) in the presence or absence of what seems to be, in the short term, a highly effective regimen (such as conventional sensitizers or secretagogues). Over time, the natural progression of the disease effectively overcomes short-term advantages and cancels the specific advantages of one regimen over the others. This is consistent with clinical practice, where it is common, in order to control glycemia, to administer progressively increasing dosages of eventually multiple agent combinations. A major qualitative difference is only obtained when the fundamental mechanisms of the progression of disease are altered, for instance by guaranteeing maintenance of insulin sensitivity to a sizeable fraction of normal, as could be the case of drastic fat tissue loss after major metabolic surgery procedures, or by restoring β-cell replicating ability by suppressing inflammation.

Supposing the model to be reliable, within the evident limits of such a schematic representation of reality, it is of interest to use it in order to answer some basic questions about disease evolution and effectiveness of therapy. Since this is a theoretical model of disease evolution, which does not take into account a host of

concurrent life events, in particular therapy, when the model is left to run freely for many years in the life history of a virtual patient, state variables may assume values that in real life would be prevented by medical intervention (or by death). This applies particularly to glycemia, which the model may predict to increase to more than 20 mM, while in reality the subject would come to medical attention well before reaching these levels. In turn, such persistent high glycemias negatively affect β-cell population (given the glucose toxicity mechanism incorporated in the model) until potential disappearance of insulin secretion. Keeping in mind therefore that the model depicts theoretical, unconstrained evolutions, it can be seen from the simulations in Figs. 1, 2 and 3 how limited-impact pharmacological or lifestyle interventions (directed to improving either insulin sensitivity or insulin secretion), occurring at the moment of clinical evidence of the consequences of long years of progressive homeostatic derangement, may offer only modest and temporary slowing of the progression of the disease, while conversely the long-term maintenance of even a modest fraction of the normal levels of insulin sensitivity, or the suppression of a fraction of inflammation-mediated glucose toxicity, could substantially alter the evolution of the clinical picture. In these simulations, β-cell mass is predicted to eventually disappear if glycemia is left to soar uncontrolled. It is conversely maintained if a new equilibrium at modest, but non-vanishing, insulin sensitivity is reached, with compensating hyperinsulinemia and modest, non-diabetic increase in glycemia. In this sense, the model supports the use of early, preventive measures directed at limiting the occurrence of the dangerous positive-feedback loop of high glycemia–glucose toxicity–diminished pancreatic reserve–and still higher glycemia. The model seems also to indicate that the existing compensation mechanisms are indeed rather effective in preventing substantial hyperglycemia, and that either very severe insulin resistance, or a combination of rather intense insulin resistance and insulin secretory defects must be present in order for the subject to develop diabetes.

Tailored therapeutic approaches can be explored prospectively, for a given patient, by the use of diabetes progression model simulations, detailing response to interventions based on both the underlying pathology and the postulated therapeutic mechanism. For patients with advanced degrees of insulin sensitivity or secretion defects, simulations are consistent with clinical observation in predicting short-lived efficacy of conventional pharmacologic treatment. More articulate tailored therapeutic schemes can however be designed and their likely efficacy evaluated in silico over a realistically heterogeneous population of virtual subjects. Promising approaches, with improved benefit/risk ratios, can thus be preliminarily assessed and optimized employing an accurate mathematical model, in order to then test them in real-world clinical trials. According to simulations performed with the Diabetes Progression Model, durable improvements of the clinical picture can only be obtained with drastic and permanent lifestyle changes or highly effective therapies targeted at the underlying pathophysiology. The model suggests that it is unlikely that permanent results can be obtained by administering conventional therapy as soon as reasonable suspicion of deteriorating insulin sensitivity or secretion is entertained.

The model as formulated has of course some important limitations, which will have to be addressed in future research on Diabetes Progression Modeling.

For example, in its description of long-term prevailing glycemia and insulinemia adjustments it does not consider acute incidents such as the fast development of (relative) insulin secretion impairment, which is seen during acute hyperglycemia, in conditions like e.g. diabetic ketoacidosis. The excessive hyperglycemia, which is present during these episodes, may block insulin secretion from β-cells and determine a fast positive feedback loop resulting in the acute clinical crisis of ketoacidosis. Insulin administration and rehydration are however effective, in this case, in restoring euglycemia and normal β-cell function. These events are however so fast, in the time-frame of the progressive establishment of prediabetes and diabetes over decades, that from the point of view of the model they may be considered instantaneous and therefore neglected.

A structural problem with the current version of the model is the lack of separate mechanisms representing central and peripheral insulin sensitivity. Associated with this, the current model does not track independently fasting and post-prandial glycemias, FPG being more closely dependent on central and PPG more dependent on peripheral insulin sensitivity. As a consequence, the model is not able to discriminate between Impaired Fasting Glycemia (IFG) and Impaired Glucose Tolerance (IGT) in the progression towards T2DM.

Yet, another limitation of the current model is the lack of an explicit representation of renal glucose elimination. Indeed, the fact that simulated glycemias may increase over the years to artificially high levels is due, in addition to the absence of medical intervention, also to the absence of the linear glucose elimination with above-threshold plasma glycemia, which determines, together with polyuria and polydipsia, the sweet taste of urine to which the disease owes its name. On the other hand, the renal threshold for glucose (typically around 10 mM) is substantially higher than the 7 mM threshold for the diagnosis of diabetes considered here, so that significant departures of the model from correct physiology should be expected to occur only when glycemias increase much above diabetic levels.

References

1. Bolie VW (1961) Coefficients of normal blood glucose regulation. J Appl Physiol 16:783–788
2. Ackerman E, Gatewood LC, Rosevear JW, Molnar GD (1965) Model studies of blood-glucose regulation. Bull Math Biophys 27:Suppl:21–Suppl:37
3. Defronzo RA (1988) The triumvirate: beta cell, muscle, liver. a collusion responsible for NIDDM. Diabetes 37:667–687
4. Bonner-Weir S, Deery D, Leahy JL, Weir GC (1989) Compensatory growth of pancreatic beta-cells in adult rats after short-term glucose infusion. Diabetes 38:49–53
5. Swenne I (1982) The role of glucose in the in vitro regulation of cell cycle kinetics and proliferation of fetal pancreatic B-cells. Diabetes 31:754–760

6. Hugl SR, White MF, Rhodes CJ (1998) Insulin-Like Growth Factor I (IGF-I)-stimulated pancreatic beta-cell growth is glucose-dependent. synergistic activation of insulin receptor substrate-mediated signal transduction pathways by glucose and IGF-I in INS-1 cells. J Biol Chem 273:17771–17779
7. Efanova IB, Zaitsev SV, Zhivotovsky B, Kohler M, Efendic S, Orrenius S, Berggren PO (1998) Glucose and Tolbutamide induce apoptosis in pancreatic beta-cells. a process dependent on intracellular Ca2+ concentration. J Biol Chem 273:33501–33507
8. Hoorens A, Van de Casteele M, Kloppel G, Pipeleers D (1996) Glucose promotes survival of rat pancreatic beta cells by activating synthesis of proteins which suppress a constitutive apoptotic program. J Clin Invest 98:1568–1574
9. Yki-Jarvinen H (1992) Glucose toxicity. Endo Rev 13:415–431
10. Bonner-Weir S (2000) Islet growth and development in the adult. J Mol Endocrinol 24:297–302
11. Butler AE, Janson J, Bonner-Weir S, Ritzel R, Rizza RA, Butler PC (2003) Beta-cell deficit and increased beta-cell apoptosis in humans with type 2 diabetes. Diabetes 52:102–110
12. Maedler K, Spinas GA, Lehmann R, Sergeev P, Weber M, Fontana A, Kaiser N, Donath MY (2001) Glucose induces beta-cell apoptosis via upregulation of the Fas receptor in human islets. Diabetes 50:1683–1690
13. Mason CC, Hanson RL, Knowler WC (2007) Progression to type 2 diabetes characterized by moderate then rapid glucose increases. Diabetes 56:2054–2061
14. Eisenbarth GS (1986) Type I diabetes mellitus. A chronic autoimmune disease. N Engl J Med 314:1360–1368
15. Tobin BW, Lewis JT, Tobin BL, Rajotte RV, Finegood DT (1992) Markedly reduced beta-cell function does not result in insulin resistance in islet autografted dogs. Diabetes 41:1172–1181
16. Bergman RN, Ider YZ, Bowden CR, Cobelli C (1979) Quantitative estimation of insulin sensitivity. Am J Physiol 236:E667–E677
17. Makroglou A, Li J, Kuaang Y (2006) Mathematical models and software tools for the glucose-insulin regulatory system and diabetes: an overview. Appl Numer Math 56(3):559–573
18. Boutayeb A, Chetouani A (2006) A critical review of mathematical models and data used in diabetology. Biomed Eng Online 5:43
19. Panunzi S, Palumbo P, De Gaetano A (2007) A discrete single delay model for the intravenous glucose tolerance test. Theor Biol Med Model 4:35
20. Palumbo P, Panunzi S, De Gaetano A (2007) Qualitative behavior of a family of delay-differential models of the glucose-insulin system. Discret Continuous Dyn Syst Series B 7:399–424
21. Li J, Kuang Y, Mason CC (2006) Modeling the glucose-insulin regulatory system and ultradian insulin secretory oscillations with two explicit time delays. J Theor Biol 242:722–735
22. Walton C, Godsland IF, Proudler AJ, Felton C, Wynn V (1992) Evaluation of four mathematical models of glucose and insulin dynamics with analysis of effects of age and obesity. Am J Physiol Endocrinol Metab 262:E755–E762
23. Topp BG, Promislow K, de Vries G, Miura RM, Finegood DT (2000) A model of beta-cell mass, insulin, and glucose kinetics: pathways to diabetes. J Theor Biol 206:605–619
24. De Gaetano A, Hardy T, Beck B, Abu-Raddad E, Palumbo P, Bue-Valleskey J, Porksen N (2008) Mathematical models of diabetes progression. Am J Physiol Endocrinol Metab 295:E1462–E1479
25. de Winter W, DeJongh J, Post T, Ploeger B, Urquhart R, Moules I, Eckland D, Danhof M (2006) A mechanism-based disease progression model for comparison of long-term effects of pioglitazone, metformin and gliclazide on disease processes underlying type 2 diabetes mellitus. J Pharmacokinet Pharmacodyn 33:313–343
26. Eddy DM, Schlessinger L (2003) Validation of the Archimedes diabetes model. Diab Care 26:3102–3110

27. Schlessinger L, Eddy DM (2002) Archimedes: a new model for simulating health care systems—the mathematical formulation. J Biomed Inform 35:37–50
28. Zhou H, Isaman DJ, Messinger S, Brown MB, Klein R, Brandle M, Herman WH (2005) A computer simulation model of diabetes progression, quality of life, and cost. Diab Care 28:2856–2863
29. Hardy T, Abu-Raddad E, Porksen N, De Gaetano A (2012) Evaluation of a mathematical model of diabetes progression against observations in the diabetes prevention program. Am J Physiol Endocrinol Metab 303:E200–E212
30. Iozzo P, Beck-Nielsen H, Laakso M, Smith U, Yki-Jarvinen H, Ferrannini E (1999) Independent influence of age on basal insulin secretion in nondiabetic humans. European group for the study of insulin resistance. J Clin Endocrinol Metab 84:863–868
31. Haffner SM, D'Agostino R Jr, Festa A, Bergman RN, Mykkanen L, Karter A, Saad MF, Wagenknecht LE (2003) Low insulin sensitivity (S(i) = 0) in diabetic and nondiabetic subjects in the insulin resistance atherosclerosis study: is it associated with components of the metabolic syndrome and nontraditional risk factors? Diab Care 26:2796–2803
32. CNR IASI BioMatLab (2007) BIBIF Model Fitting Webservice. http://www.biomatematica.it/bibif/bibif.html. Ref type: internet communication
33. Robertson RP (2007) Estimation of beta-cell mass by metabolic tests: necessary, but how sufficient? Diabetes 56:2420–2424
34. Leahy JL, Cooper HE, Deal DA, Weir GC (1986) Chronic hyperglycemia is associated with impaired glucose influence on insulin secretion. a study in normal rats using chronic in vivo glucose infusions. J Clin Invest 77:908–915
35. Hager SR, Jochen AL, Kalkhoff RK (1991) Insulin resistance in normal rats infused with glucose for 72 H. Am J Physiol Endocrinol Metab 260:E353–E362
36. Laybutt DR, Cordery DV, Kraegen E (1997) Specific adaptations in muscle and adipose tissue in response to chronic systemic glucose oversupply in rats. Am J Physiol Endocrinol Metab 273:E1–E9
37. de Winter W, Post T, DeJongh J, Moules I, Urquhart J, Eckland D, Danhof M (2004) A mechanistic disease progression model for type 2 diabetes mellitus and pioglitazone treatment effects. PAGE 2004 meeting, Poster, 17
38. Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, Nathan DM (2002) Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. N Engl J Med 346:393–403
39. Knowler WC, Hamman RF, Edelstein SL, Barrett-Connor E, Ehrmann DA, Walker EA, Fowler SE, Nathan DM, Kahn SE (2005) Prevention of type 2 diabetes with troglitazone in the diabetes prevention program. Diabetes 54:1150–1156
40. Retnakaran R, Qi Y, Harris SB, Hanley AJ, Zinman B (2011) Changes over time in glycemic control, insulin sensitivity, and beta-cell function in response to low-dose metformin and thiazolidinedione combination therapy in patients with impaired glucose tolerance. Diab Care 34:1601–1604

Linear Modeling and Prediction in Diabetes Physiology

Marzia Cescon and Rolf Johansson

Abstract Diabetes Mellitus is a chronic disease characterized by the inability of the organism to autonomously regulate the blood glucose level due to insulin deficiency or resistance, leading to serious health damages. The therapy is essentially based on insulin injections and depends strongly on patient daily decisions, being mainly based upon empirical experience and rules of thumb. This chapter presents work on data-driven glucose metabolism modeling and short-term, that is, up to 120 min, blood-glucose prediction in Type 1 Diabetes Mellitus (T1DM) subjects. Low-order, individualized, data-driven, stable, physiologically relevant models were identified from a population of 9 datasets from T1DM patients. Model structures include: autoregressive moving average with exogenous inputs (ARMAX) models and state-space models. The performances of the model-based predictors were compared to those achieved by the zero-order hold (ZOH).

1 Introduction

1.1 Context and Motivation

Both from quality of life and economic perspectives it is critical for diabetes patients to regulate their blood glucose tightly, keeping its level within the target range, i.e., 70–140 [mg/dL] [64], through intensive insulin therapy. The strategy comprises test of blood glucose levels at least four times a day, taking insulin with

M. Cescon (✉) · R. Johansson
Department of Automatic Control, Lund University, 22100 Lund, Sweden
e-mail: marzia.cescon@control.lth.se

R. Johansson
e-mail: rolf.johansson@control.lth.se

V. Marmarelis and G. Mitsis (eds.), *Data-driven Modeling for Diabetes*,
Lecture Notes in Bioengineering, DOI: 10.1007/978-3-642-54464-4_9,

every meal by injections or subcutaneous infusions by a pump and patient assistance by health care team.

Although the standard tools in diabetes care improved significantly during the last decades, including the availability of self-monitoring of blood glucose (SMBG), insulin pen injectors, pumps and insulin analogues, insulin therapy remains one of the most difficult to manage. As a matter of fact, the treatments still strongly depend on the patient's daily decisions about insulin delivery adaptations. Many factors have to be considered in this decision process: health status, current blood glucose level, blood glucose target, insulin sensitivity, diet, meal composition, foreseen activities and individual experience of insulin effects on blood glucose level. Meanwhile, failure in management of insulin therapy has significant impact on short-, medium- and long-term prospects.

The availability of a blood glucose predictor that would inform the patient on the near future blood glucose and offer advices on how to modulate insulin therapy in relation to food intake and out-of-target glucose deviations would therefore be highly valuable.

1.2 The DIAdvisor™ Project

Against this background, the European FP7-IST research project DIAdvisor™ [22] aimed at developing a personalized blood glucose predicting system and an advisory control system, i.e., the DIAdvisor™ tool, to be used on the spot to assist the users in different daily situations, predicting hyperglycemic deviations following meals and stressful events, and giving them advices about how to adjust their treatments. The DIAdvisor™ tool would constitute a mobile short-term blood glucose predictor and treatment advisor. Figure 1 clarifies the concept. The predictor system would need user input concerning patient characteristics (e.g., insulin sensitivity or resistance), patient condition (e.g., fasting, meal time, rest or physical activity, illness, stress), therapeutic mode (type of insulin delivery route, type of insulin preparations), time and size of meals, inputs from non-invasive glucose sensors, wearable vital signs sensors and blood glucose meter measurements, and would be expected to produce short-term blood glucose predictions to be graphically shown to the patients and suggestions to the user from a decision support module. The users would have the opportunity to accept or reject the advice, thereby assuring safety.

1.3 Statement of the Problem

The development of a safe predictive and advisory system, such as the DIAdvisor™ tool, would require patient-specific dynamical models of the glucose metabolism able to describe the blood glucose evolution based on the most significant inputs, namely, meal carbohydrates, exogenous injected insulin and

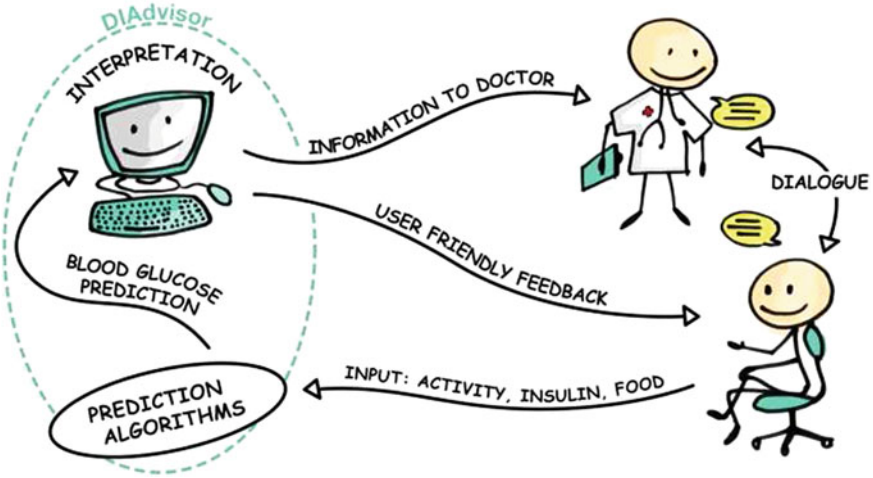


Fig. 1 DIAdvisor™ modeling and prediction [22]

possibly energy expenditure due to physical activity [60] to be used in a model predictive control setup [44]. Characteristics required in such models include a good trade-off between simplicity and accuracy, stability, qualitative correct responses to inputs and predictive capabilities with particular emphasis on hypo- and hyperglycemia detection. In addition, predictors targeting blood glucose forecasting need to be developed in parallel, as well.

Motivated by the above, this contribution addresses the questions of how to identify personalized models from individual patient data and how to provide individualized short-term blood glucose predictors.

1.4 Inherent Challenges in T1DM Modeling

Despite significant efforts devoted to the problem of blood glucose regulation in type 1 diabetic patients over the last decades (see e.g. [15] for a comprehensive review), many inherent challenges that must be overcome still remain. At the most basic level, the disease can be viewed as a process having one output, namely, glucose concentration in plasma, and two inputs, namely, meal carbohydrates and administered insulin. The first and perhaps most crucial challenge to overcome in modeling is that of poor data excitation: often the inputs are simultaneous and in the same ratio, the so-called insulin-to-carbohydrate ratio, precluding the possibility of distinguishing their relative effects. It is still an open issue how to strongly excite the system in order to obtain data meaningful for identification purposes preserving at the same time the patients from the risk of serious clinical events. Second, the most widespread way of treating diabetes comprises a series of impulse-like control actions, i.e., insulin injections and food intakes, applied

several times during the day at irregular sampling instants, typically at wake up, meal and bed times, the decisions being based on scarce assays of the controlled variable, i.e., blood glucose. This rises the problem of non-uniformly sampled and infrequent data and, since the signals in play interact in the bloodstream, introduces assumptions on the subcutaneous-to-intravenous insulin absorption and gastro-intestinal carbohydrates absorption dynamics.

In the last ten years, advances in sensor technology saw the advent of continuous glucose monitors (CGM), systems capable of measuring glucose concentration frequently (e.g., every 5 min) for several days, providing the patient with well-sampled data in real time. However, it is important to stress that together with the benefits, they introduce yet another limitation. Indeed, those devices measure glucose concentration in the interstitium and not in plasma. Interstitial glucose (IG) fluctuations are related to BG presumably via diffusion process [37, 63]. This leads to a number of issues, including distortion (which incorporates a time lag) and calibration errors, and necessitates the development of methods for their mitigation. In particular, it is necessary to consider that, since the BG-to-IG kinetics act as a low-pass filter, the frequency content of interstitial glucose is different from that of blood glucose [9, 47].

As for the inputs, when taking into account the appearance of insulin in the bloodstream from subcutaneous delivery and that of glucose in plasma after a meal, new time-lags and dynamics are introduced; further, subcutaneous insulin infusion involves degradation at the site of delivery. In addition, meals must be recorded by the patients, and the actual amount of carbohydrates must be estimated, a process that is prone to errors. Also, in practice, the combination of simple and complex carbohydrates, fats and proteins can affect the glucose absorption in the digestive system. Unrepresented inputs, such as stress and illness constitute another challenge to diabetes modeling. Furthermore, it is a well known fact that physical activity, apart from having a glycemia-lowering effect due to utilization of glucose by the muscle cells, enhances insulin sensitivity, playing a substantial role in the picture, but the magnitude and duration of such effects are hard to consider. Another important aspect is the degree of variability of the overall system dynamics over the day (the so-called “dawn phenomenon”, for instance, is characterized by increased insulin resistance during the morning hours [70]).

A priori knowledge of the diabetes process indicates two fundamental properties that should be satisfied by any sound and valid model:

- the gain associated with the insulin input should be negative (i.e., an increase in insulin results in a decrease in glucose concentration)
- the gain associated with the meal input should be positive (i.e., a meal results in an increase in glucose concentration).

However, the values of the above mentioned gains are related to age of the subject, disease duration, body mass index (BMI), insulin sensitivity, β-cells responsiveness and probably many more unknown factors so that it is not clear how to take them into account in the modeling process.

1.5 Models in Diabetes

In diabetes research and therapy, modeling of the glucose-insulin control system has received significant attention for more than 40 years [15]. Several types of models serving different purposes were proposed, most of these efforts being first-principles based descriptions of the physiological relationships associated with T1DM, and only to a lesser extent mathematical modeling by means of system identification.

1.6 Physiological Models

The first pioneering work describing the relationships between insulin and glucose utilization was that of Bolie [7], later modified by Ackerman and McGucking [2] in order to provide a model of the glucose metabolism during an oral glucose tolerance test (OGTT). Greater attention was received by the so-called minimal model [5, 6], developed for the specific purpose of quantifying pancreatic responsiveness and insulin sensitivity during an intravenous glucose-tolerance test (IVGTT) in non-diabetic individuals. The model consists of three differential equations describing plasma glucose and plasma insulin in a remote compartment, accounting for neither the dynamics of subcutaneous insulin infusion nor the dynamics of gut glucose absorption from a carbohydrates meal. In [40], a simulation model was presented, and later validated on a set of 24 subjects [41]. Glucose-insulin pharmacokinetics/pharmacodynamics in non-diabetic subjects was described by a 19-state model developed by Sorensen [57], the major shortcoming of this model being the failure in capturing the hyperglycemic events characteristic of type 1 diabetes [43]. The nonlinear model proposed by Hovorka et al. [35] describes glucose-insulin dynamics with several differential equations in three subsystems: a glucose subsystem, an insulin subsystem and an insulin action subsystem. Model inputs are the rate of subcutaneously infused fast acting insulin, meal carbohydrates amount and time of ingestion and its outputs are plasma glucose and insulin concentrations. The absorption kinetics associated with subcutaneous insulin delivery of the above mentioned model was later modified [69], the model replacing the original subcutaneous insulin subsystem consisting in two parallel fast and slow channels for insulin absorption as well as a degree of insulin degradation at the injection site, and integrated in a simulation environment dedicated to closed-loop evaluation of insulin delivery systems [68]. The efforts by Dalla Man and co-workers presented in [16–18, 20] lead to the meal simulation model of the glucose-insulin system [19] which has been accepted by the Food and Drug Administration (FDA) to be used as a substitute for animal trials in pre-clinical tests of closed-loop development [38, 39]. Dalla Man and co-workers [19] used a sophisticated triple tracer method to estimate important meal-related quantities such as the rates of appearance of glucose in the blood from the meal,

endogenous glucose production, utilization of glucose, and insulin secretion. Reviews of physiological diabetes models include that of Nucci and Cobelli [49], who specifically examined several models of subcutaneous-to-intravenous insulin kinetics, Makroglou and co-workers [45] presenting an overview of existing software packages specific to diabetes modeling and finally Cobelli and co-workers [15], discussing the main contribution to both modeling and control in diabetes from the early 1960s.

1.7 Predictive Models

Although seemingly simple in concept, the problem of glucose prediction in an active individual has to date proved intractable. Currently, continuous glucose monitoring (CGM) devices are the available technology able to provide high/low glucose alarms when certain user specified preset threshold levels have been crossed and to deliver warnings of events that are likely to occur if the current trend continues. However, patients will benefit more from an early alarm that predicts the hypoglycemic episode before it occurs. A review of hypoglycemia prevention algorithms is reported in [4]. To date many studies have investigated the possibility of predicting blood glucose concentration for the purpose of regulating glucose intervention, in order to enable individuals to take corrective actions and avoid low or high glucose values. Bremer and Gough [8] originally developed the idea of T1DM CGM time-series analysis using 10 min sampled data from ambulatory T1DM patients to identify autoregressive (AR) models. They explored 10, 20 and 30 min prediction horizons, and report that the 10 min predictions are accurate and the 20 or 30 min predictions may also be acceptable for a limited set of data only. However, no quantification of the accuracy of the model predictions was provided. In [55] tenth-order autoregressive models were identified from 5-days long glucose time-series belonging to a population of five in-hospital subjects. Performances of a single, cross-subject model compared with individual models were evaluated on 30 and 60 min prediction horizons. Gani et al. [30] used a 30th-order AR model with data smoothing and a regularization procedure to minimize the changes in the glucose first-derivative, reducing the prediction lag for 60 and 90 min-ahead prediction compared to [55]. Sparacino et al. [59] collected 48 h of continuous (3 min) glucose data from 28 T1DM subjects in ambulatory conditions. In their retrospective analysis, they recursively identified simple polynomial and AR models from the CGM time-series data, prefiltered to removed noise spikes. They investigated prediction horizons of 10 and 15 steps (i.e., 30 and 45 min) and concluded that hypoglycemia can be detected 25 min before the hypoglycemic threshold is passed. In [51] a Kalman filter approach was proposed, which only used information on past CGM readings by assuming a double integrated random walk as prior for glucose dynamics and estimating the states corresponding to the interstitial glucose level, and the first and second derivative thereof, i.e., rate of glucose change and acceleration.

The method was evaluated for 13 hypoglycemic clamp data sets in [50]. Using a hypoglycemic threshold of 70 [mg/dL], the sensitivity and specificity were 90 and 79 %, respectively, with unknown alarm time. A tutorial overview of algorithms for CGM time series analysis to the purpose of alarm generation is provided in [58]. Eren-Oruklu et al. [23] proposed a recursive second-order AR and ARMA model identification strategy with an adjustable forgetting factor for healthy and type II diabetics. Their models utilized only recent glucose history from a CGM device, achieving 3–5 % error for 30 min ahead prediction. In [48] a kernel-based regularization learning algorithm, in which the kernel and the regularization parameter are adaptively chosen on the basis of previous similar learning tasks, using past glucose information, was presented. The past few years witnessed the investigation of neural networks (NN) models for short-term glucose prediction proving it to be a competitive approach. Pappada et al. [54] created NN models with variable predictive windows in the range 50–180 min, trained using 18 patients CGM datasets and evaluated on patient data not included in the neural network formulation. They concluded that their models performed adequately in the normo- and hyperglycemic ranges, whereas hypoglycemic events were overestimated, a potential reason for that being due to the minimal occurrences of hypoglycemia in the training data. In [25] a feed-forward NN whose inputs were CGM samples in the previous 20 min and the current time instant, and whose output was the glucose concentration 15, 30 and 45 min ahead was tested on 15 actual data sets. The root-mean-squared error (RMSE) was 10, 18 and 2 [mg/dL] for the 15, 30 and 45 min prediction, with a delay of around 4, 9, and 14 min for upward trends, and 5, 15, and 26 min for downward trends. Dassau and co-workers [17] developed a real-time hypoglycemia prediction suite combining five individual algorithms, namely, linear prediction, Kalman filtering, hybrid impulse response (HIIR) filtering, statistical prediction and numerical logical into a voting-based system to predict hypoglycemia from 1 min sampled CGM data. A 35 min prediction horizon with an alarm threshold of 80 [mg/dL] and a voting threshold of three to five algorithms to predict hypoglycemia resulted in a 91 % correct predictions. A short-coming of the methods listed above is the lack of exploitation of the dynamic interplay between previously injected insulin, meal intake and eventually exercise to the purpose of improving glucose prediction. Hovorka and his group [35] performed experiments with ten T1DM patients under clinical conditions, using their physiological model to make predictions of glucose data up to 60 min into the future. The glucose was measured intravenously, but delayed by 30 min to mimic subcutaneous measurement. The model parameters were recursively estimated using a Bayesian method. The predictions of the resulting models had RMSE values of 8.6, 13.0, and 17.3 [mg/dL] for 2, 3, and 4-step (i.e., 30, 45, and 60 min) predictions, respectively. Finan et al. [27] analyzed both batch-wise and recursively identified patient-specific ARX models for 9 patients with a mean 30 min prediction error RMSE of 26 [mg/dL]. An ARX model with a nonlinear forgetting factor scaled according to the glucose range was considered in [11, 12], and a 45 min prediction horizon showed good results. Finally, in [31] it was asserted that a universal data-driven model identified from

CGMS time-series from a patient applying the algorithm previously published by the same authors in 2009 [30] could be used to make near-future glucose concentration predictions for other patients without any model customization procedure. They used regularization techniques to filter data from 34 subjects, then, using the filtered data, they develop auto-regressive models of order 30 to the purpose of making short-term, 30-min-ahead predictions. A feed forward NN was also exploited in [54] and tested on 10 real datasets, incorporating, in addition to CGM data, other inputs such as SMBG readings, information on insulin, meal, hypo- and hyperglycemia symptoms, lifestyle, activity and emotions and predict glucose values up to 75 min. In [71–73], 30-min-ahead prediction was performed with a feed-forward NN in cascade with the first-order polynomial model in [59]. The inputs to the linear predictor were the past CGM values weighted using a forgetting factor, while the inputs to the NN were current CGM and its trend, information on the past error committed by the polynomial model and information on meal, supplied as plasma glucose rate of appearance obtained from the physiological model of [19]. Zhao et al. [74] used a latent variable-based approach to predict future CGM values from past CGM and known carbohydrate and insulin boluses, transformed into time-smoothed inputs using second-order transfer functions. The method was applied to collected clinical data and simulated data generated by the model described in [38, 39]. They concluded that their LV-based method resulted in models whose prediction accuracy was as least as good as the accuracies of standard AR/ARX models. In [24] a multi-sensor body monitor providing seven signals related to activity and emotional conditions was used in addition to a CGM monitor to improve glucose prediction. A multivariate ARMAX model with weighted recursive least-squares estimation of the unknown parameters using a variable forgetting factor was proposed. Results showed that the prediction error is significantly reduced with the addition of the vital signs measurements, as compared to an ARMA model based only on CGM signals.

2 Experimental Conditions and Clinical Data Acquisition

In the framework of DIAdvisor^{TM} [22], acquisition of bioclinical data linked or potentially involved in blood glucose control from insulin-treated diabetic subjects was accomplished in a series of experimental sessions. The investigations focused on a population of basal-bolus regimen treated subjects, either as combination of multiple daily insulin injections (MDI) including long-acting and fast-acting analogues or as continuous subcutaneous insulin infusion (CSII) of fast-acting analogues from a pump. The clinical study was performed at the Clinical Investigation Center CIC-CHU in Montpellier, France. Collected data include: specific patient parameters (e.g., gender = male, age = 43 years old, BMI = 23.7, weight = 67 kg), characteristics related to diabetes (e.g., disease duration = 10 years, insulin delivery = external pump), associated health conditions and therapies, food intakes and administered insulin doses registered

in a logbook, capillary glucose strips, interstitial glucose levels, plasma glucose and plasma insulin concentrations from drawn blood samples as well as vital signs.

2.1 Equipment

For the whole duration of the study, the subjects were equipped with state-of-the-art devices provided by the DIAdvisor™ Consortium complying with the study protocol as explained in the following subsections.

2.1.1 HemoCue™ Glucose Analyzer

The HemoCue™ Glucose Analyzer [34] is a blood glucose meter based on a glucose dehydrogenase method and consisting of a pocket size handheld analyzer and a unique disposable microcuvette. The analyzer was factory calibrated and no calibration was needed between cuvette batches. The device was used by each patient as a reference glucose meter, assessing plasma glucose levels from finger-stick samples.

2.1.2 Abbott Freestyle Navigator™

The Abbott Freestyle Navigator™ [1] is a Continuous Glucose Monitoring System (CGMS) consisting of an amperometric electrochemical sensor placed under the skin, a wireless transmitter connected to the sensor and a wireless receiver collecting the sensor signals. The subcutaneous sensor was inserted about 5 mm into the subcutaneous tissues and could be worn for up to five days before replacement. Calibration relative to capillary glucose was required at specific times, namely, 10, 12, 24 and 72 h after insertion. Using the WIRED ENZYME™ technology, the sensor converted glucose concentration to electrical current. Once every minute the transmitter sent the estimate of the interstitial glucose concentration to the receiver, which displayed the final values once every 10 min.

2.2 Study Protocol and Experiments

The clinical study consisted of three visits: *Visit 0* for patients screening, *Visit 1* for sensors initialization and *Visit 2* for 75 h in-hospital tests. Figure 2 gives the flow chart of the trial.

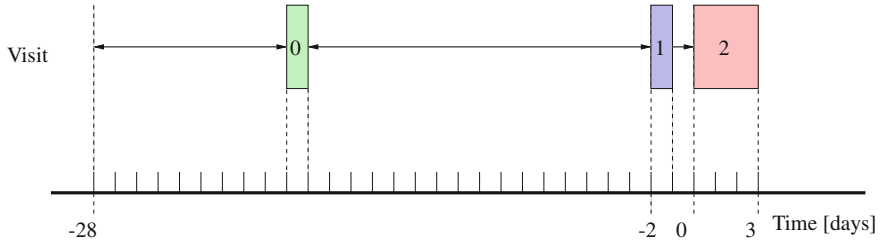


Fig. 2 DIAdvisor™ Data Acquisition Trial. *Visit 0* Screening visit (*green box*), *Visit 1* sensor initialization visit (*blue box*), *Visit 2* in-hospital tests (*red box*)

2.2.1 Visit 0

Prior to any study-related procedure, the purpose of this visit was to perform a screening examination of the patient, the outcome of it being recorded in the clinician's sheet.

2.2.2 Visit 1

Within 4 weeks of *visit 0* the patient was admitted at the clinic to initialize the Abbott Freestyle Navigator™ device. The sensor was inserted subcutaneously into the patient's skin and calibrated against capillary glucose by a nurse. The subject was hospitalized at best 48 h after sensor insertion, in order to begin the tests with a well-calibrated device.

2.2.3 Visit 2

During the whole 3-days-long visit, the patient was permanently equipped with the Abbott Freestyle Navigator™ [1] device. Standard meals for breakfast (8:00 a.m.), lunch (1:00 p.m.) and dinner (7:00 p.m.) were served, the amount of administered carbohydrates being 42, 70 and 70 g, respectively. Blood samples were collected by nurses to measure plasma glucose and plasma insulin concentrations: every hour during day, every 2 h during night, every 15 min after meals for 2 h. A specific sampling scheduled was adopted after breakfasts: 30 min before, mealtime, 10, 20, 30, 60, 90, 120, 150, 180, 240 and 300 min after, for a total of 37 blood samples per day. No specific intervention on usual diabetes treatment was scheduled during the period. The patients decided their insulin needs according to the HemoCue™ Glucose Analyzer [34] measurements as usually in activities of daily life. Throughout the chapter, modeling results obtained with *visit 2* data will be presented.

Table 1 Patient characteristics

Patient ID	Insulin therapy	HBA1c [%]	BMI
102	MDI	6.5	26.5
103	CSII	9.1	23.7
104	MDI	7.6	20
105	CSII	7.8	24.1
106	CSII	7.8	21.2
107	CSII	8.9	25.3
115	MDI	8.5	19.7
120	MDI	9	22.4
130	MDI	8.8	29.4

2.3 Patients Selection Criteria

A total of 30 diabetic subjects, male and female adults, were included in the study to allow availability of data from a wide spectrum of patients under basal-bolus regimen. Among these, a population of nine patients was chosen, the selection criteria being the quantity of data collected during *visit 2* (>80 % of the expected), no sensor failures and logbook correctly filled in. Exclusion criteria were data partially collected, laboratory results not available and/or patients not observant in annotating insulin/meal intakes. Table 1 reports the characteristics of the selected patients. Data from a representative patient are shown in Fig. 3.

3 Modeling of the Gluco-Regulatory System

The physiology of glucose metabolism in diabetes can be thought of as having one output, i.e., glucose level in the bloodstream y_{BG}, and two main inputs, i.e., carbohydrate intake u_{carb} and administered insulin I_{ir}. Further, given that physical activity has been proven to decrease plasma glucose levels due to increased glucose uptake by the exercising muscles [70], the effect of exercise, i.e., increased heart rate, respiration rate and body movements, is therefore to be regarded as an additional input or load disturbance. Hence, for modeling purposes, based on current knowledge of the overall physiological model, four subsystems have to be considered (Fig. 4):

- the glucose subsystem (GS), describing glucose intestinal absorption following a food intake;
- the insulin subsystem (IS), accounting for the pharmacokinetics of the exogenously administered insulin;
- the physical activity and energy expenditure subsystem (EES), measuring the rate of physical activity intensity;
- the glucose-insulin interaction subsystem (GIIS).

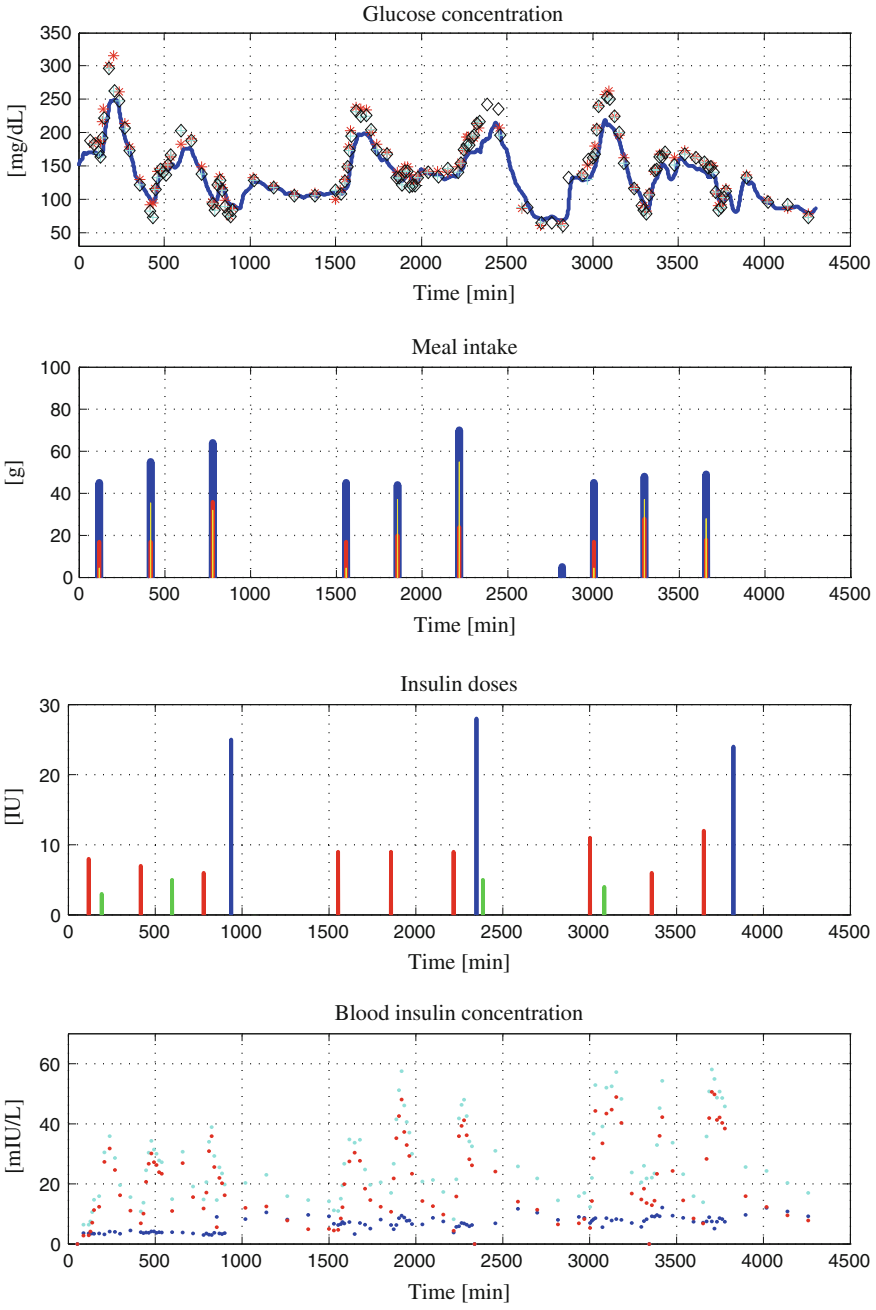


Fig. 3 Patient CHU0102 data versus Time [min]. *Top* Glucose concentration [mg/dL]: interstitial (*blue*), plasma (*red*), finger stick (*cyan* and *black*); *Upper Center* Meal intake [g]: carbohydrates (*blue*), lipids (*red*), proteins (*yellow*); *Lower Center* Insulin doses [IU]: basal (*blue*), bolus (*red*), correction (*green*); *Bottom* Blood insulin concentration [mIU/L]: basal (*blue*), bolus (*red*), total (*cyan*)

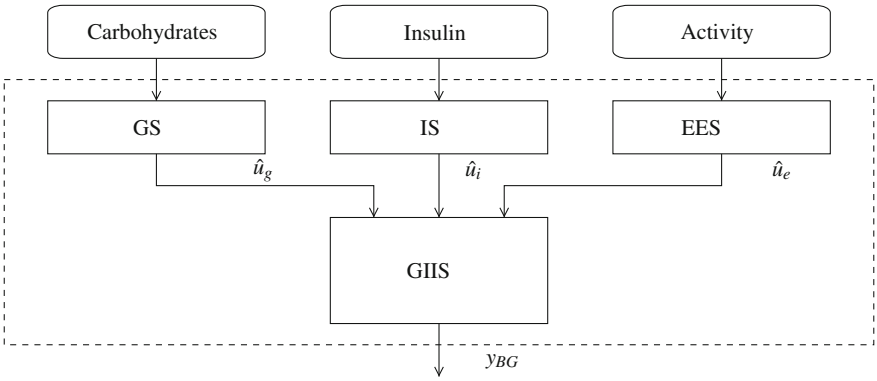


Fig. 4 Physiological model describing diabetic blood glucose dynamics. Notation: GS glucose subsystem, IS insulin subsystem, EES energy expenditure subsystem, GIIS glucose insulin interaction subsystem; $\hat{u}_g$ glucose rate of appearance following a meal, $\hat{u}_i$ insulin in plasma after subcutaneous injection, $\hat{u}_e$ energy expenditure, y_{BG} blood glucose

In this chapter, compartment models from the literature are exploited to describe the GS and IS. Data-based system identification is used, instead, to model the GIIS.

3.1 Glucose Subsystem

Glucose transit through the stomach and upper small intestine was described by a nonlinear chain of three compartments, where the first two compartments represent the stomach (solid and liquid phases) and the third one depicts the intestine. The left plot in Fig. 5 illustrates the model. Equations are [16]:

$$\begin{aligned} q_{sto}(t) &= q_{sto1}(t) + q_{sto2}(t) \\ \dot{q}_{sto1}(t) &= -k_{gri} \cdot q_{sto1}(t) + u_{carb} \cdot \delta(t), \qquad q_{sto1}(0) = 0 \\ \dot{q}_{sto2}(t) &= -k_{empt} \cdot q_{sto2}(t) + k_{gri} \cdot q_{sto1}(t), \quad q_{sto2}(0) = 0 \\ \dot{q}_{gut}(t) &= -k_{abs} \cdot q_{gut}(t) + k_{empt} \cdot q_{sto2}(t), \quad q_{gut}(0) = 0 \\ \hat{u}_g(t) &= \frac{f \cdot k_{abs} \cdot q_{gut}(t)}{m_b}, \qquad\qquad u_g(0) = 0 \end{aligned} \tag{1}$$

where $\hat{u}_g$ [mg/kg/min] denotes the rate of appearance of glucose in plasma, q_{sto1} and q_{sto2} [mg] are the amounts of carbohydrates in the stomach (solid and liquid phase, respectively), u_{carb} [mg] is the amount of ingested carbohydrates, q_{gut} [mg] is the carbohydrate mass in the intestine, k_{gri} is the rate of grinding, k_{empt} the rate of gastric emptying, m_b [kg] the subject's body weight, k_{abs} the rate of absorption and f the fraction of intestinal absorption that actually appears in plasma. The rate of

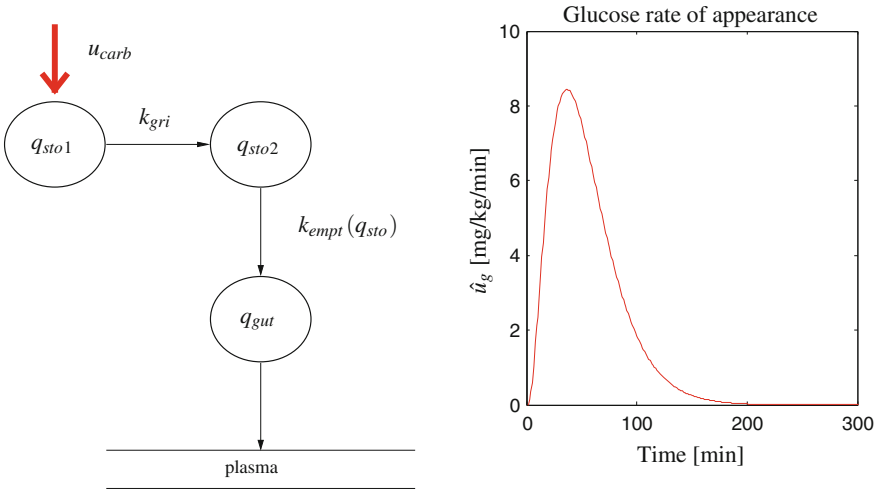


Fig. 5 *Left* Glucose intestinal absorption model [16]. *Right* Glucose rate of appearance $\hat{u}_g$ after ingestion of 40 [g] carbohydrate by a patient with $m_b = 65$ [kg] at time t = 0, simulated with the model in Eq. (1) and parameters in Table 2

Table 2 Parameter values in the glucose intestinal absorption model

Parameter	Value	Measurement unit
k_{gri}	0.055	[min^{-1}]
k_{max}	0.055	[min^{-1}]
k_{min}	0.008	[min^{-1}]
k_{abs}	0.056	[min^{-1}]
b	0.82	dimensionless
c	0.01	dimensionless
f	0.9	dimensionless

gastric emptying was a non-linear function of the amount of carbohydrates in the stomach q_{sto} according to the following relationship:

$$k_{empt}(q_{sto}) = k_{min} + \frac{k_{max} - k_{min}}{2} \cdot \{\tanh[\alpha(q_{sto} - b \cdot D)] - \tanh[\beta(q_{sto} - c \cdot D)] + 2\} \tag{2}$$

$$\alpha = \frac{5}{2 \cdot D \cdot (1 - b)}, \quad \beta = \frac{5}{2 \cdot D \cdot c} \tag{3}$$

Dalla Man and co-workers provided us with the mean population values for the parameters appearing in Eqs. 1, 2, 3 used throughout the thesis. Table 2 reports such values.

3.2 Insulin Subsystem

The insulin flow $s(t)$ entering the bloodstream from the subcutaneous depots in the diabetic subject is described by a subcutaneous insulin infusion model (Fig. 6), whose model equations are:

$$\begin{aligned} \dot{I}_{sc1}(t) &= -(k_d + k_{a_1})I_{sc1}(t) + I_{ir}(t), \quad I_{sc1}(0) = I_{sc1b} \\ \dot{I}_{sc2}(t) &= k_d I_{sc1}(t) - k_{a2} I_{sc2}(t), \qquad I_{sc2}(0) = I_{sc2b} \\ s(t) &= k_{a1} I_{sc1}(t) + k_{a2} I_{sc2}(t) \end{aligned} \tag{4}$$

with I_{sc1}, I_{sc2} pmol/kg the amount of nonmonomeric and monomeric insulin in the subcutaneous space, respectively, k_d [min^{-1}] the rate constant of insulin dissociation, k_{a1} [min^{-1}] and k_{a2} [min^{-1}] the rate constants of nonmonomeric and monomeric insulin absorption, respectively, and I_{ir} [pmol/kg/min] the exogenous insulin infusion rate. The insulin flow $s(t)$ which entered the bloodstream is degraded in the liver and the periphery according to the model equations [26]:

$$\begin{aligned} \dot{I}_p(t) &= -(m_2 + m_4)I_p(t) + m_1 I_1(t) + s(t), \quad I_p(0) = I_{ph} \\ \dot{I}_l(t) &= -(m_1 + m_3)I_l(t) + m_2 I_p(t), \qquad I_l(0) = I_{lb} \\ \hat{u}_i(t) &= \frac{I_p(t)}{V_i} \end{aligned} \tag{5}$$

where I_p [pmol/kg] and I_l [pmol/kg] are insulin masses in plasma and liver, respectively, V_i [L/kg] is the distribution volume of insulin, while $\hat{u}_i$ [pmol/L] accounts for the total plasma insulin concentration. m_1 is the rate of hepatic clearance, m_2, m_3, m_4 [min^{-1}], instead, are rate parameters as follows:

$$\begin{aligned} m_2 &= \frac{3}{5}\frac{l_{CL}}{HE_b(V_i m_b)} \\ m_3 &= m_1 \frac{H_{Eb}}{1 - HE_b} \\ m_4 &= \frac{2}{5}\frac{I_{CL}}{V_i m_b} \end{aligned} \tag{6}$$

where HE_b [dimensionless] is the basal hepatic insulin extraction, while I_{CL} [L/min] is the insulin clearance. At steady-state

$$\begin{aligned} 0 &= -(k_d + k_{a1})I_{sc1b} + I_{irb} \\ 0 &= k_d I_{sc1b} - k_{a_2} I_{sc2b} \\ s_b &= k_{a1} I_{sc1b} + k_{a_2} I_{sc2b} \end{aligned} \tag{7}$$

so that the basal value of insulin in the subcutaneous compartments, i.e., I_{sc1b} and I_{sc2b}

$$\begin{aligned} I_{sc1b} &= \frac{I_{irb}}{k_d + k_{a1}} \\ I_{sc2b} &= \frac{k_d}{k_{a2}} \cdot I_{sc1b} \end{aligned} \tag{8}$$

and $s_b = I_{irb}$. Further,

$$\begin{aligned} 0 &= -(m_2 + m_4)I_{pb} + m_1 I_{lb} + s_b \\ 0 &= -(m_1 + m_3)I_{lb} + m_2 l_{pb} \end{aligned} \tag{9}$$

leading to the expressions for the amount of insulin in the liver compartment at basal state:

$$I_{lb} = I_{pb} \frac{m_2}{m_1 + m_3} \tag{10}$$

and the amount of insulin in plasma at basal steady state:

$$I_{pb} = \frac{I_{irb}}{m_2 + m_4 - \frac{m_1 m_2}{m_1 + m_3}} \tag{11}$$

Model parameters used in the thesis were provided by Dalla Man and co-workers and are given in Table 3.

3.3 Data Analysis

Data analysis was performed in the following order [36]:

- autospectrum of inputs

$$S_{uu}(i\omega) = \mathscr{F}\left\{ \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} u(t)u^*(t-\tau)dt \right\}$$

- cross spectrum between inputs and output

$$S_{uy}(i\omega) = \mathscr{F}\left\{ \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} u(t)y^*(t-\tau)dt \right\}$$

Table 3 Parameter values for the subcutaneous insulin infusion model

Parameter	Fast insulin	Slow insulin	Unit
k_{a1}	0.004	0.0002	[min^{-1}]
k_{a2}	0.0182	0.00091	[min^{-1}]
k_d	0.0164	0.00164	[min^{-1}]
m_1	0.1766	0.1766	[min^{-1}]
V_i	0.05	0.05	[L/kg]
I_{CL}	1.1069	1.1069	[L/min]
HE_b	0.6	0.6	dimensionless

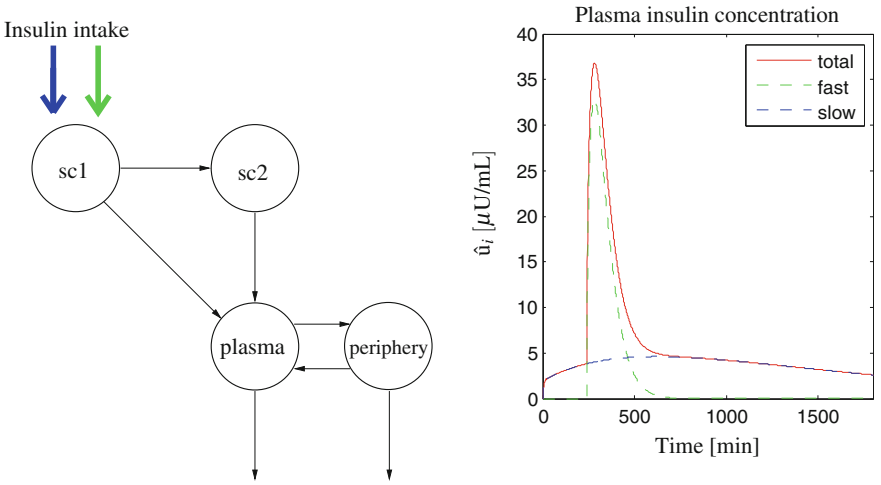


Fig. 6 *Left* Insulin pharmacokinetics model [38]. It accounts for both slow- and fast-acting insulin. Compartments scl and sc2 represents the subcutaneous insulin infusion module. Plasma insulin concentration: total $\hat{u}_i$ [μU/mL] (*red*), slow-acting (*blue*) and fast-acting (*green*) resulting from a basal dose of 20 [IU] at t = 0 and a bolus of 5 [IU] at t = 240 min, taken by a patient with $mb = 65$ kg, simulated with the model in Eqs. (4, 5) and parameters in Table 3

- quadratic coherence spectrum between inputs and output

$$\gamma^2_{uy}(\omega) = \frac{\left|S_{uy}(i\omega)\right|^2}{S_{uu}(i\omega)S_{yy}(i\omega)}$$

3.3.1 Data Pre-processing

For purposes of model identification, removal of the mean value of the data series was done as part of standard data pre-processing [42]. In addition, originally non-uniformly sampled, plasma glucose concentration and plasma total insulin concentration from laboratory results were linearly interpolated and uniformly resampled, the resampling period being 1 min.

3.4 Problem Formulation

Given the inputs:

- interpolated total plasma insulin concentration u_i [mlU/L] from drawn blood samples;
- plasma glucose rate of appearance u_g [mg/kg/min] after carbohydrate intestinal absorption;
  and the output:
- interpolated blood glucose y_{BG} [mg/dL] from drawn blood samples

the objective was to find an individual-specific and physiological relevant model of the glucose-insulin interaction for each of the subjects in the selected population. Minimum requirements on the model were:

- stability;
- white residuals;
- qualitative correct blood glucose responses to
  - 1 [IU] fast-acting insulin;
  - 10 [g] carbohydrates;

Performances were evaluated according to:

- Percentage FIT

$$\text{FIT} = \left(1 - \frac{|y_k - \hat{y}_k|}{|y_k - \bar{y}_k|}\right) \times 100\,\%$$

- Percentage Variance Accounted For (VAF)

$$\text{VAF} = 1 - \frac{E\left[(y_k - \hat{y}_k)(y_k - \hat{y}_k)^{\mathrm{T}}\right]}{E\left[y_k y_k^{\mathrm{T}}\right]} \times 100\,\%$$

so that additional requirement on the model were:

- FIT $\geq$ 50 % on 60 min-ahead model-based prediction on validation data;
- VAF $\geq$ 50 % on 60 min-ahead model-based prediction on validation data.

3.5 Model Estimation

The approach considered for modeling was system identification of discrete-time, time invariant linear models [42]. The data belonging to each of the selected patients records was equally divided into two parts: the first one for the calibration procedure of obtaining the optimal model structure and model parameters, and the second one for validation of the chosen configuration.

Assuming that input-output data $\{u_k, y_k\}$, $k = 1,\ldots, N$ were available, model structures describing the *GII* dynamical system were:

- autoregressive moving average with exogenous inputs (ARMAX) model

$$A\left(z^{-1}\right)y_k = z^{-k_1}B_1\left(z^{-1}\right)u_k^1 + \cdots + z^{-k_m}B_m\left(z^{-1}\right)u_k^m + C\left(z^{-1}\right)w_k \tag{12}$$

where

$$A\left(z^{-1}\right) = 1 + a_1 z^{-1} + \cdots + a_{n_a} z^{-n_a} \tag{13}$$

$$B_i\left(z^{-1}\right) = b_{0,I} + b_{1,i} z^{-1} + \ldots + b_{n_{bm,i}} z^{-n_{bm}} \tag{14}$$

$$C\left(z^{-1}\right) = 1 + c_1 z^{-1} + \ldots + c_{n_c} z^{-n_c} \tag{15}$$

n_a, n_{bi}, n_c integers representing the orders of the polynomials, $u_k^1, \ldots, u_k^m$, are the inputs, y_k is the output, w_k denotes the coloured noise and z^{-1} is the backward shift operator;

- state-space model in innovation form

$$\begin{cases} x_{k+1} = \mathscr{A}x_k + \mathscr{B}u_k + \mathscr{K}e_k \\ y_k = \mathscr{C}x_k + \mathscr{D}u_k + e_k \end{cases} \tag{16}$$

denoting with n the dimension of the state-space, m the number of inputs, $\mathscr{A} \in \mathbb{R}^{n\times n}$, $\mathscr{B} \in \mathbb{R}^{n\times n}$, $\mathscr{C} \in \mathbb{R}^{1\times n}$, $\mathscr{D} \in \mathbb{R}^{1\times m}$, $\mathscr{K} \in \mathbb{R}^{n\times 1}$ and $\{e_k\}$ the noise process, i.e., the one-step-ahead prediction error, which is a zero-mean white noise.

For each of the calibration dataset different methods were used for the estimation of the model parameters:

- prediction-error identification methods (PEM) [42] for identification of the ARMAX structure
- subspace-based methods, namely:
  - N4SID [65];
  - PO-MOESP [67];
  - PBSID [13, 14];

for the identification of the state-space model.

The identification procedure is outlined in Algorithm 1. Throughout the work Matlab® System Identification Toolbox [46] and the SMI Toolbox [33] were used.

3.6 Model Evaluation and Selection Criteria

The system identification procedure provided a plethora of models for each of the subjects in the population. However, to the purpose of model-based controller

design, it suffices to select one model per patient only. Necessary requirements on a model suitable for inclusion and exploitation in the DIAdvisor^{TM} tool were:

- stability;
- uncorrelated (white) residuals;
- physiologically sensible responses to insulin and food intake, i.e., blood glucose concentration should decrease in response to insulin and increase in response to food intake.

Algorithm 1 *Procedure for system identification*

ARMAX:

- for $1 \leq n_a \leq 10$, $1 \leq n_{bm,i} \leq 3$, $1 \leq n_c \leq 10$, $1 \leq n_{ki} \leq 3$
  - estimate model
  - compute Akaike Final Prediction Error
- rank the models according to their FPE in increasing order [61]

N4SID:

- choose the method for the estimation of the state-space
  - CVA [57]
  - MOESP [94]
- set the past horizon $p = 120$ and the future horizon $f = 60$
- set the model order $1 \leq n \leq 10$
- select the model order $\bar{n}$
- estimate models with order $\bar{n} - 1$, $\bar{n}$, $\bar{n} + 1$

MOESP:

- set the past and future horizons $s = 30$
- select the model order $\bar{n}$
- estimate models with order $\bar{n} - 1$, $\bar{n}$, $\bar{n} + 1$

PBSID:

- set the past horizon $p = 60$, and the future horizon $f = 20$
- select the model order $\bar{n}$
- estimate models with order $\bar{n} - 1$, $\bar{n}$, $\bar{n} + 1$

In particular, in order to assess whether the model showed correct responses to inputs, the simulated blood glucose reactions to a

- 1 [IU] fast insulin injection
- 10 [g] carbohydrates intake

were compared, the appearance of insulin in blood after subcutaneous injection being obtained with Eqs. (4, 5, 8 and the parameters listed in Table 3, and similarly, the glucose rate of appearance in plasma after an oral glucose ingestion being calculated using Eq. (1) and parameters in Table 2.

Each of the estimated models, i.e. the ARMAX models ranked in ascending FPE, the N4SID, MOESP and PBSID models ranked in ascending model order, were evaluated according to the diagram in Fig. 8. When a requirement was not fulfilled, a model with higher FPE in the case of ARMAX models or higher order in the case of the state-space models, was taken for evaluation.

Those models passing the tests depicted in Fig. 8 were compared on the basis of their prediction performances on 30, 60, 90 and 120-min-ahead prediction. In particular, the performances of model-based predictors obtained with the Matlab® System Identification Toolbox [46] command *predict.m* were evaluated according to:

- FIT [%]
- Prediction error variance

$$E\left\{(y_k - \hat{y}_k)(y_k - \hat{y}_k)^{\mathrm{T}}\right\}$$

- VAF [%]

and compared to those achieved with the zero-order hold (ZOH) $\hat{y}_{k+\tau|k} = y_k$, with $\tau = 30, 60, 90, 120$.

3.7 Results

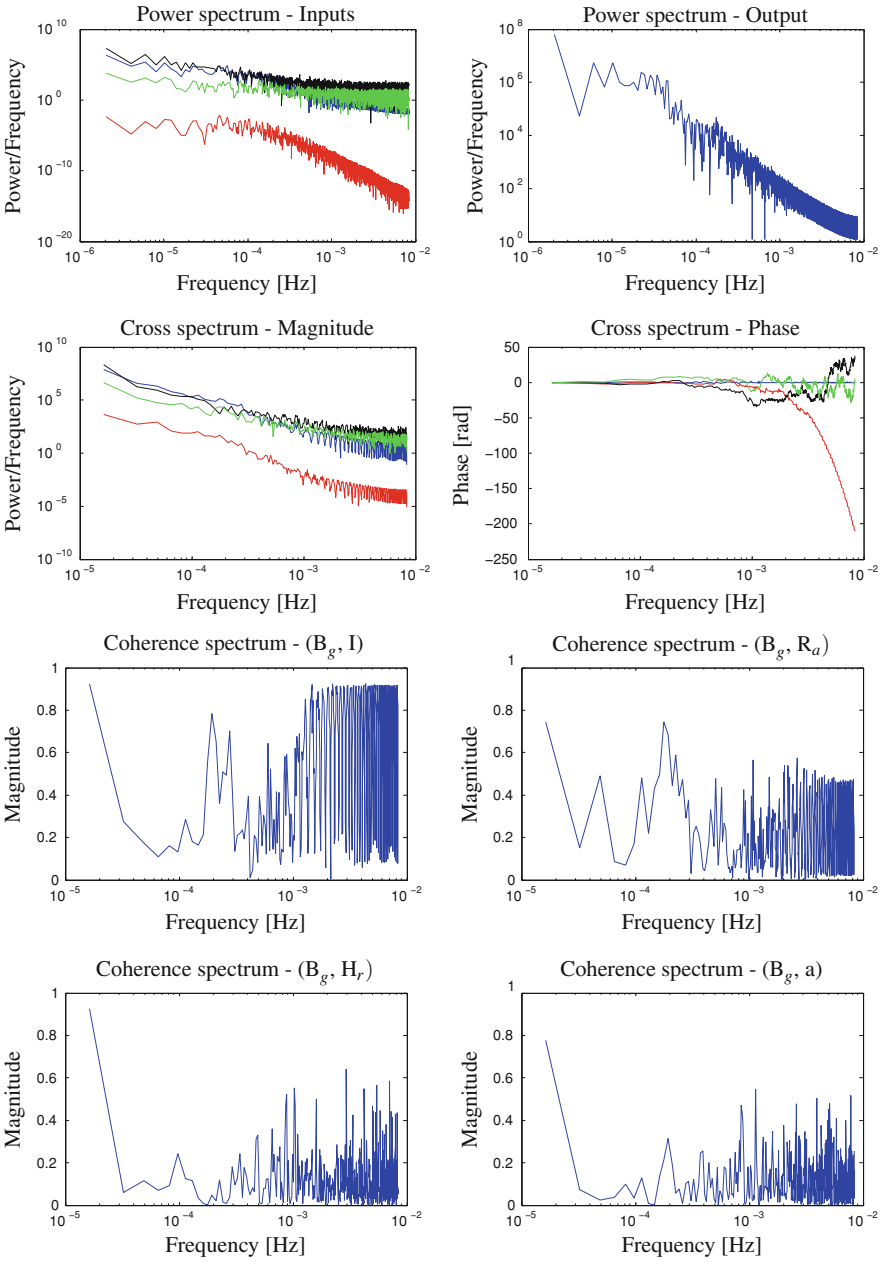
3.7.1 Data Analysis

The autospectra (power spectra) of inputs and output showing the frequency contents of the signals investigated as well as the coherence spectrum between the inputs and the controlled variable are shown in Fig. 7 for the representative patients 102 (Fig. 8).

3.7.2 Models

The methods failing to provide a model complying with the criteria in Sect. 3.6 were the subspace-based identified models. Figure 9 reports step and impulse responses as well as the model output to 1 [IU] of insulin and 10 [g] carbohydrates.

Performances on short-term predictions, i.e., up to 120 min are displayed in Fig. 10, while comparisons of the model-based predictors with the projection of the current glucose value in the future, i.e., the ZOH, are reported in Fig. 12 and quantitatively in Tables 4, 5, 6 (notice that the table were edited from Matlab output, the number of decimal places not representing numerical accuracy or significance) while the box plots in Fig. 11 show mean population performances.


Power spectrum - Inputs
Power/Frequency
Frequency [Hz]
Power spectrum - Output
Power/Frequency
Frequency [Hz]
Cross spectrum - Magnitude
Power/Frequency
Frequency [Hz]
Cross spectrum - Phase
Phase [rad]
Frequency [Hz]
Coherence spectrum - (B$_g$, I)
Magnitude
Frequency [Hz]
Coherence spectrum - (B$_g$, R$_a$)
Magnitude
Frequency [Hz]
Coherence spectrum - (B$_g$, H$_r$)
Magnitude
Frequency [Hz]
Coherence spectrum - (B$_g$, a)
Magnitude
Frequency [Hz]

◀ **Fig. 7** Patient CHU0102. *Top Panels Top Left* Magnitude of Power spectrum of inputs: total plasma insulin [$(mIU/L)^2/(Hz)$] (*blue*), plasma glucose rate of appearance [$(mg/kg/min)^2/(Hz)$] (*red*), heart rate [$(beats/min)^2/(Hz)$] (*black*), activity (a.u./(Hz)) (*green*); *Top Right* Magnitude of Power spectrum of output: blood glucose [$(mg/dL)^2/(Hz)$]; *Bottom Left* Magnitude of cross spectrum: total plasma insulin, blood glucose [$(mIU/L)^2(mg/dL)^2/(Hz)$] (*blue*), plasma glucose rate of appearance, blood glucose [$(mg/kg/min)^2(mg/dL)^2/(Hz)$] (*red*), heart rate, blood glucose [$(beats/min)^2(mg/dL)^2/(Hz)$] (*black*), activity, blood glucose [$(a.u.)(mg/dL)^2$ (Hz)] (*green*); *Bottom Right* Phase of cross spectrum rad: total plasma insulin (*blue*), plasma glucose rate of appearance (*red*), heart rate (*black*), activity (*green*). *Bottom Panels* Coherence spectra between blood glucose and *Top Left* total plasma insulin; *Top Right* plasma glucose rate of appearance; *Bottom Left* Heart Rate; *Bottom Right* Activity. All the spectra versus Frequency [Hz]

4 Discussion

4.1 Result of Experiments

This work exploited a unique dataset which has being collected within a major European project [22]. Albeit being in a controlled environment, the subjects participating in the study experienced both hypo- and hyperglycemic events, proving glucose regulation hard to accomplish. The glucose curves showed diurnal variations and wide excursions over the 75 h. When hypoglycemia occurred, this situation being common to all the patients, extra carbohydrates apart from the standard meals were given. Similarly, hyperglycemia was treated with extra insulin intakes. As it is common practice, the majority of the subjects bloused just before being served the meal. Since meal intake and the insulin injections have opposite effects on the blood glucose level, the estimation of the contribution from each of these inputs is difficult when both are active at the same time.

As for the sensor signals, the FreeStyle Navigator^{TM} continuous glucose monitor traces resulted often in poor agreement with the blood glucose reference obtained from the laboratory analysis. In particular, important offsets were present. The system measures glucose in the interstitial fluid (ISF), i.e., in between the body cells. Movements of nutrients, oxygen and glucose from the blood into the cells happen across the ISF; therefore, during times of rapid change in blood glucose, e.g. after eating, dosing insulin, or exercising, differences in glucose measurement between interstitial fluid and finger-stick measurements are expected to be observed. In the examined trial, however, it was noted that in 40–50 % of the traces the sensor was accurate at low glucose levels but inaccurate at high levels, a fact that cannot be explained by the plasma-to-interstitium dynamics only and may be explained by poor/difficult device calibration. The manufacturer also made available the intermediate signal calculated by the system at each minute. Whether or not this may overcome the limitations introduced by the physiological interstitium-to-plasma dynamics it is still unclear, since a model of how the signal relates to plasma glucose is missing.

Finally, the rich collection of blood samples provided reference plasma glucose values and plasma insulin values not available elsewhere.

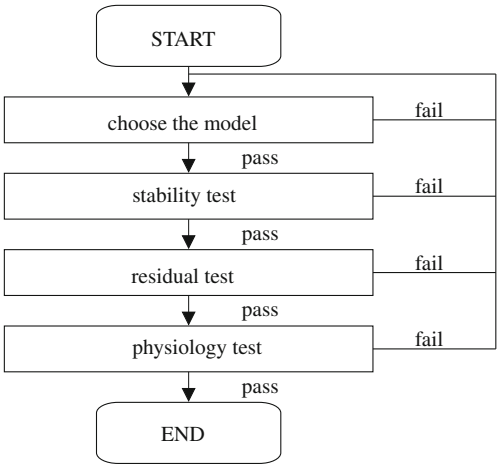


Fig. 8 Diagram for model evaluation

4.2 Modeling

Over the last decades, models describing the insulin-to-glucose system dynamics were developed for the purpose of simulation and glycemic control, the approach being physiology based [19, 45, 56]. Recently the problem of identifying such a model has been tackled from a data-driven perspective mainly using simulated data from models in the literature [28, 52]. Indeed, fitting actual T1DM subject data to the models has been treated to a much less extent (e.g., [29, 61, 62]) given the difficulties in gathering appropriate patient records.

The autospectra (power spectra) showing the frequency contents of the signals investigated and the coherence spectra between the inputs and the controlled variable were calculated [36], Recall that a coherence spectrum can be interpreted as frequency-resolved correlation analysis (or signal-to-noise analysis). A large absolute value close to 1 indicates that the input and output are correlated. A coherence value of 0.5 denotes that half of the output variation may be explained by variations in the stimulus input. As predictable, the data collected offered poor model input excitation despite the careful selection of the subjects, because of the correlation between food intake and consequent insulin injection.

Throughout the work, the glucose flux in the bloodstream after intestinal absorption and the total, i.e., fast-acting and slow-acting, insulin flux in the bloodstream were considered as input variables. As far as the glucose absorption modeling is concerned, it is a well known fact that not only the size of the meal but also the composition of the meal affects the digestion dynamics (see e.g. [10]). Unfortunately, detailed quantitative information on meal composition was not available in the data set considered for this chapter. In absence of such information, all sources of carbohydrates were assumed to be equal. Population mean values were used for the parameters appearing in the meal model and in the insulin kinetics model thus disregarding the inter-personal variability. Given the frequently drawn blood samples, it was decided to use the actual (interpolated and

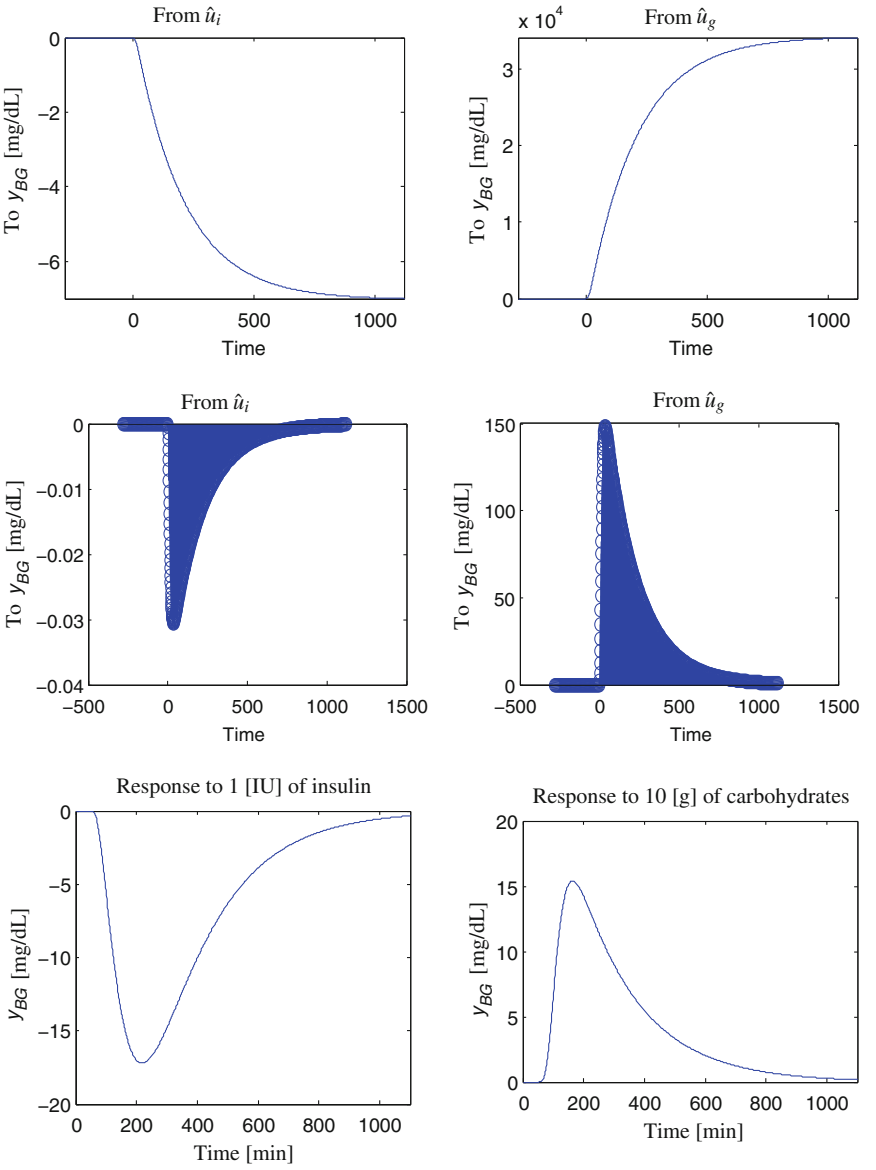


Fig. 9 Patient CHU0102. ARMAX 3rd-order model. *Top* Step Responses; *Center* Impulse responses; *Bottom Left* Blood glucose response to 1 [IU] of fast-acting insulin; *Bottom Right* Blood glucose response to 10 [g] of carbohydrates

uniformly resampled) insulin assays for the identification of both models and predictors, and to use the physiological insulin kinetics model at a later stage, to test the blood glucose response to 1 [IU] of fast acting insulin.

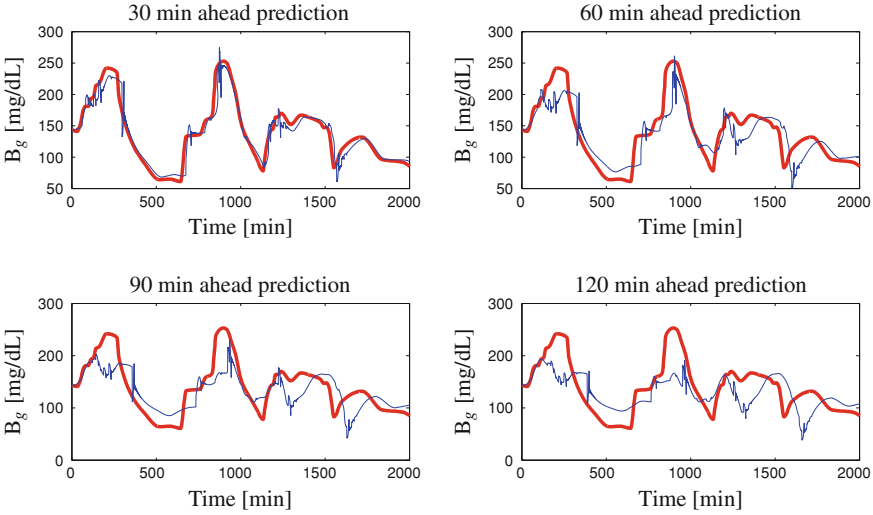


Fig. 10 Patient CHU0102. Evaluation on validation data: 3rd-order ARMAX-based predictor (*thin*) and measured plasma glucose (*thick*) [mg/dL] versus time [min]. *Top left* 30-min-ahead; *Top right* 60-min-ahead; *Bottom left* 90-min-ahead; *Bottom right* 120-min-ahead prediction

Individual-specific models of low-complexity were identified from the collected data. Estimated model structures included autoregressive with exogenous inputs (ARMAX) models and state-space models. As far as the ARMAX structure is concerned, identification of the model parameters was accomplished by minimization of a quadratic prediction error criterion using the Matlab® System Identification Toolbox routine *armax.m.* The range of the orders n_a, n_b, $\mathrm{n_c}$ was empirically set to the interval [1:10], while that of the inputs-output delays was [1:3], Regarding the subspace identification techniques, two parameters having substantial influence on the quality of the resulting model needed to be chosen, namely the lengths of past and future horizons, representing the dimension of certain Hankel matrices constructed with the data. There are no simple rules for choosing them, however, the knowledge of the application, that is blood glucose prediction up to 120 min ahead, provided an initial guess. The parameters were then tuned empirically: $p = f = 120$ in the N4SID algorithm [65], $s = 30$ in the PO-MOESP algorithm [67] and $p = 60, f = 20$ in PBSID [14],

The estimated models were tested according to various validation criteria. Stability was the first requirement that a model needed to fulfill. Residual tests with the purpose of finding remaining correlations which indicate whether the model order is adequate were carried out. With adequate model order, the residual process is white only and of sufficiently small magnitude. The residual autocorrelation and cross correlation between the prediction errors and the input tests needed to give significant (99 % confidence) validation with respect to changes of sign, independence of residuals, normality, and independence between residuals and input in order for the test to be passed. Finally, qualitatively correct responses

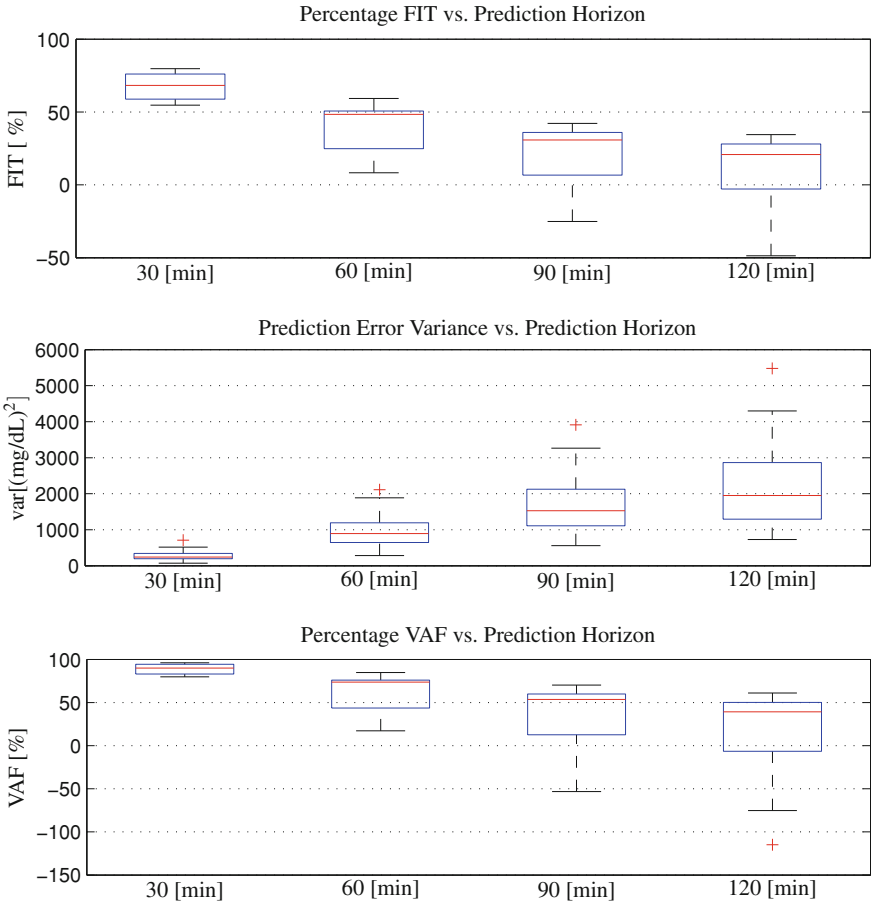


Fig. 11 Model-based predictor performance evaluation. Population results on validation data. *Top* Percentage FIT versus Prediction Horizon [min]; *Center* Prediction Error Variance [(mg/dL)^2] versus Prediction Horizon [min]; *Bottom* Percentage VAF versus Prediction Horizon [min]. Each box presents results over the population considered. The central mark is the median, the edges of the box are the 25th and the 75th percentiles

to inputs were guaranteed. From a quantitative point of view, according to clinicians and their experience gained from clinical trials, the average lowering effect of 1 [IU] of fast insulin falls within 25–60 [mg/dL], with peak time 60–240 [min], depending upon the subject's resistance or sensitivity to insulin, whereas an ingestion of 10 [g] pure dextrose makes the blood glucose rising 15 [mg/dL], in 20 [min] at best. However, these requirements seemed hard to achieve and were not fulfilled by all models.

Overall, the main difficulties encountered while carrying out the modeling task were assuring white residuals and estimating physiologically correct inputs to output transfer functions.

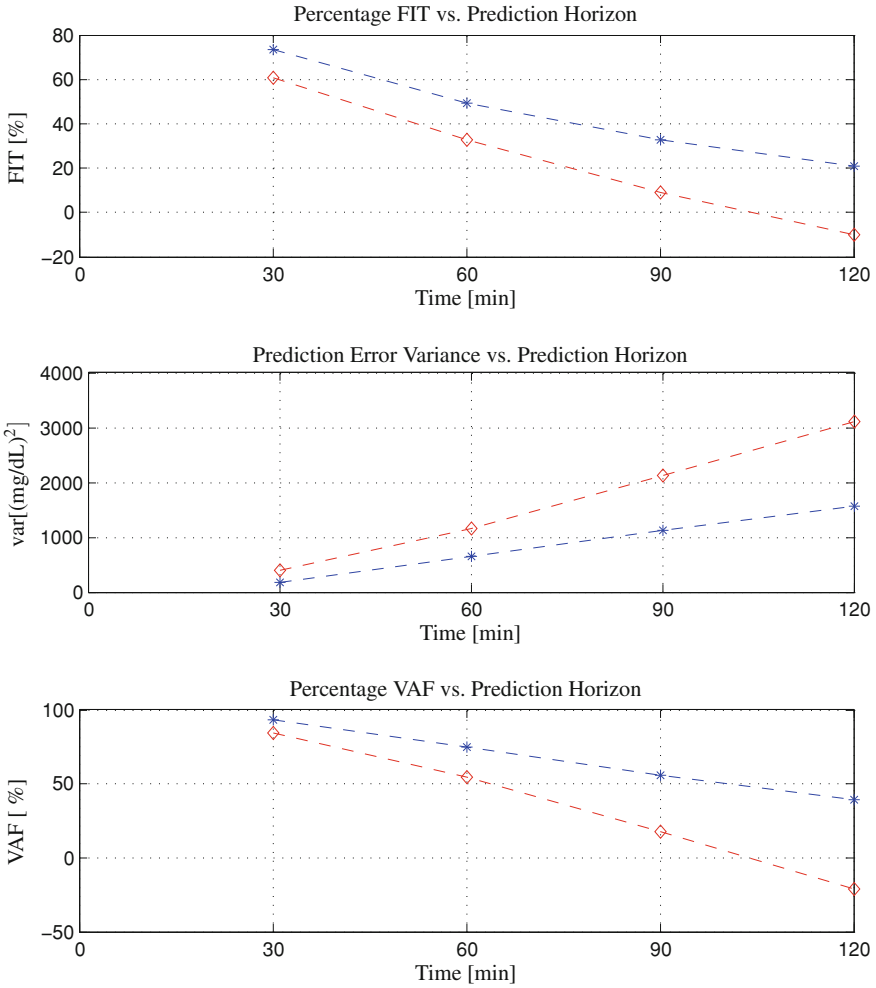


Fig. 12 Patient CHU0102. 3rd-order ARMAX-based predictor (*star*), ZOH (*diamond*). *Top* Percentage FIT *Center* Prediction Error Variance [$(mg/dL)^2$]; *Bottom* Percentage VAF. All the metrics on validation data versus Prediction Horizon [min]

4.3 Prediction

The quality of the predictors developed was assessed by mathematical metrics in order to quantify the error between the predicted blood glucose profile versus the actual ones. Specifically, predictions were evaluated with respect to

- FIT [%]
- prediction error variance [mg/dL]
- VAF [%]

Table 4 Model-based predictor performance evaluation

Patient ID	predictor	30 [min]	60 [min]	90 [min]	120 [min]
102	ARMAX	73.5358	49.4739	32.8066	20.7979
	ZOH	60.6483	32.5733	9.0154	−10.1487
103	ARMAX	65.0763	33.7277	19.71	10.3026
	ZOH	50.6296	20.8109	−1.3747	−21.3287
104	ARMAX	76.0065	54.1321	42.0296	34.4656
	ZOH	52.1479	19.1685	−4.933	−22.7172
105	ARMAX	57.1568	27.7924	13.4519	7.264
	ZOH	47.214	14.5802	−6.1496	−18.931
106	ARMAX	54.7189	8.3002	−25.1534	−48.6235
	ZOH	44.0662	9.708	−8.2996	−15.4517
107	ARMAX	68.2954	48.3223	34.0031	26.5711
	ZOH	52.2875	16.1428	−10.6265	−29.7901
115	ARMAX	79.7602	59.3186	42.1829	32.44
	ZOH	63.6523	34.7217	12.3847	−3.4325
102	ARMAX	76.0582	49.3294	30.8174	21.5344
	ZOH	58.4849	27.0355	4.3175	−12.5778
130	ARMAX	63.114	39.9365	20.9482	9.3143
	ZOH	58.8014	29.0776	6.3448	−9.5646

Percentage FIT [%] versus Prediction Horizon [min] on validation data

Table 5 Model-based predictor performance evaluation

Patient ID	predictor	30 [min]	60 [min]	90 [min]	120 [min]
102	ARMAX	0.1788	0.6462	1.1344	1.5641
	ZOH	0.3981	1.1676	2.1253	3.1139
103	ARMAX	0.1654	0.5896	0.8474	1.0292
	ZOH	0.3323	0.8545	1.3988	2.0007
104	ARMAX	0.2635	0.9593	1.529	1.9525
	ZOH	1.0513	2.9976	5.047	6.8949
105	ARMAX	0.2842	0.8028	1.1445	1.2995
	ZOH	0.4306	1.1233	1.7245	2.1461
106	ARMAX	0.5172	2.1103	3.9098	5.4791
	ZOH	0.7973	2.0763	2.9834	3.3862
107	ARMAX	0.2433	0.6391	1.0343	1.2715
	ZOH	0.5535	1.7088	2.9718	4.0885
115	ARMAX	0.07313	0.2834	0.5559	0.7294
	ZOH	0.2423	0.7786	1.3951	1.9321
102	ARMAX	0.202	0.8952	1.6533	2.1069
	ZOH	0.6141	1.8968	3.2605	4.5099
130	ARMAX	0.7108	1.8847	3.2648	4.2967
	ZOH	0.8868	2.6279	4.5825	6.2708

Prediction Error Variance 10^3 $[(mg/dL)^2]$ versus Prediction Horizon [min] on validation data

Table 6 Model-based predictor performance evaluation

Patient ID	predictor	30 [min]	60 [min]	90 [min]	120 [min]
102	ARMAX	93.061	74.9189	55.9707	39.2934
	ZOH	84.5496	54.6813	17.5108	−20.8571
103	ARMAX	87.864	56.7519	37.8332	24.4987
	ZOH	75.6258	37.316	−2.6105	−46.7695
104	ARMAX	94.2616	79.1115	66.7055	57.4847
	ZOH	77.1072	34.7272	−9.8982	−50.1352
105	ARMAX	81.6979	48.3	26.2977	16.3122
	ZOH	72.2686	27.6622	−11.0544	−38.207
106	ARMAX	79.7178	17.2474	−53.3144	−114.8501
	ZOH	68.734	18.5836	−16.9867	−32.7826
107	ARMAX	90.0062	73.7458	57.514	47.7695
	ZOH	77.2643	29.8063	−22.0801	−67.9497
115	ARMAX	96.0871	84.8332	70.2528	60.9726
	ZOH	87.0342	58.3385	25.3581	−3.3786
120	ARMAX	94.3313	74.8741	53.5968	40.8661
	ZOH	82.7652	46.765	8.4905	−26.5775
130	ARMAX	86.3956	63.9289	37.5156	17.7668
	ZOH	83.028	49.705	12.2986	−20.0131

Percentage VAF [%] versus Prediction Horizon [min] on validation data

and qualitative assessments concerning glucose-trends detection. Indeed, in diabetes management, the perhaps most important feature for a predictor is the ability of capturing hypoglycemias and hyperglycemias, rather than being correct in the normo-glycemic range. The performances were compared to those achieved with the zero-order-hold prediction.

5 Conclusions and Future Work

5.1 Conclusions

This contribution dealt with linear modeling and short-term prediction in diabetes physiology. Specifically, data-driven techniques were investigated to the purpose of the DIAdvisor^{TM} tool application [22] and evaluated for type 1 diabetes mellitus records belonging to a population of 9 subjects in hospital conditions.

5.1.1 Modeling

An individual-specific, physiologically relevant model of the glucose-insulin interaction subsystem was identified from each of the subjects data using prediction error methods and subspace-based methods. Inputs to the models were:

- interpolated total plasma insulin concentration u_i [mIU/L] from drawn blood samples;
- plasma glucose rate of appearance $\hat{u}_g$ [mg/kg/min] after carbohydrate intestinal absorption;

and the output was:

- interpolated blood glucose y_{BG} [mg/dL] from drawn blood samples

ARMAX models of order in the range [3:6] satisfied all the criteria required, specifically:

- stability;
- uncorrelated (white) residuals;
- physiologically sensible responses to 1 [IU] of insulin and 10 [g] of carbohydrates

and were therefore selected for inclusion in the advisory tool. However, the additional requirements on model-based predictor performances were met only partially. Indeed, whereas a value of VAF > 50 % on 60 min-ahead model-based prediction on validation data was achieved by all the models except the 5th-order ARMAX model for patient 105 and the 6th-order ARMAX model for patient 106, a value of FIT > 50 % on 60 min-ahead model-based prediction on validation data was achieved by the 3rd-order ARMAX model identified from patient 104 data and the 6th-order ARMAX model identified from patient 115 data, only.

5.2 Future Work

5.2.1 Experiment Design

Poor model input excitation and input signal correlation were reported giving rise to the issue of ill-conditioning of the estimates. Whereas the importance of persistency of excitation was well recognized in the consortium, it posed problems in the ethical approval of the experimental protocol. Thus, further work is needed to investigate optimal experimental conditions and protocols in order to obtain data suitable for identification purposes without contributing to higher patient risk.

5.2.2 Modeling

Several topics call for attention:
First, in the meal simulation model, the development of new compartments accounting also for proteins, fat and fiber content of a meal is needed. A feasible alternative to that would be introducing the glycemic index G_i of each mixed meal in the already existing meal models from the literature (e.g. [16, 19]). Using the patient's diary and nutrition tables it is possible to estimate the G_i of each food

intake. Hence, considering that lower G_i nutrients prolong the glucose rate of appearance in plasma, the glucose intestinal absorption model could be modified, e.g., making the parameters *b, c,* k_{max} and k_{abs} functions of the G_i of the meal:

$$\begin{aligned} b &= 0.089 \cdot \log(\mathrm{GI}) + 0.43 \\ c &= 0.3 \cdot \log(\mathrm{GI}) - 0.61 \\ k_{max} &= 0.02 \cdot \frac{GI}{30} - 0.013 \\ k_{abs} &= 0.013 \cdot e^{0.02 \cdot GI} \end{aligned} \tag{17}$$

In addition, since patient's annotation on meal quantity and content are not always reliable, the problem of meal estimation to cope with forgotten meals and mistaken information needs to be addressed.

As far as the insulin modeling is concerned, grey-box identification of the parameters describing insulin pharmacokinetics/pharmacodynamics could be pursued, in order to fit the physiological models (e.g. [69]) to the patients data.

A first-principle-based model quantifying the impact of physical activity on blood glucose excursion deserves future studies as well.

Further improvements of the state-of-the-art in identification, e.g., the ability of handling non-uniformly sampled data, reduction of sensitivity to initial conditions and automatic selections of model parameters, would be valuable. Most importantly, new identification methods tailored to the diabetes application need to be developed: constrained optimization could be used, for instance, to obtain physiologically correct responses to inputs. In doing so, it may be helpful to relate patient data such as BMI, insulin sensitivity and/or resistance to the insulin and glucose impulse responses.

Hybrid [53] or Linear-Parameter-Varying (LPV) models [3, 66] may capture the circadian variation of glucose in a better way than Linear-Time-Invariant (LTI) models, therefore this investigation is left to future work.

5.2.3 Prediction

Given the application, it is crucial that the predictors are correct in forecasting hypo- and hyperglycemic excursions, rather than being precise in the euglycemic range. A cost function similar to that presented in [32] could be an appropriate choice to penalize out-of-range glucose deviations.

Last, it would be interesting to challenge linear model identification methods on data collected from patients in ambulatory conditions.

Acknowledgements This research was supported by the European project DIAdvisor™, FP7 IST-216592 and the Swedish Research Council through the Linnaeus Center LCCC.

References

1. Abbott: FreeStyle Navigator™ (2013) http://www.abbottdiabetescare.co.uk. Accessed Sept 2013
2. Ackerman J, McGucking W (1964) A mathematical model of the glucose tolerance test. Phys. Medicine Biol. 9:203–213
3. Bamieh B, Giarre L (2002) Identification of linear parameter-varying models. Int J Robust Nonlinear Control 12(9):841–853
4. Bequette B (2010) Continuous glucose monitoring: real-time algorithms for calibration, filtering and alarms. J Diabetes Sci Technol 4(2):404–418
5. Bergman R, Phillips L, Cobelli C (1981) Physiologic evaluation of factors controlling glucose tolerance in man: measurement of insulin sensitivity and beta-cell sensitivity from the response to intravenous glucose. J Clin Invest 68:1456–1467
6. Bergman R, Ziya Ider Y, Bowden C, Cobelli C (1979) Quantitative estimation of insulin sensitivity. Am J Physiol 236(6):E667–E677
7. Bolie V (1961) Coefficients of normal blood glucose regulation. J Appl Physiol 16:783–788
8. Bremer T, Gough DA (1999) Is blood glucose predictable from previous values? Diabetes 48:445–451
9. Breton M, Shields D, Kovatchev B (2008) Optimum subcutaneous glucose sampling and Fourier analysis of continuous glucose monitors. J Diabetes Sci Technol 2:495–500
10. Brouns F, Bjork I, Frayn K, Gibbs A, Lang V, Slama G, Wolever T (2005) Glycaemic index methodology. Rev Nutr Res 18:145–171
11. Castillo-Estrada G, Del Re L, Renard E (2010) Nonlinear gain in online prediction of blood glucose profile in type 1 diabetic patients. In: Proceedings of 49th IEEE conference on decision and control (CDC2009), Atlanta, GA, USA, pp 1668–1673
12. Castillo-Estrada G, Kirchsteiger H, Del Re L, Renard E (2009) Model based validation of meal inputs in diabetes therapy. In: Proceedings. of 15th IFAC symposium on system identificaiton SYSID2009, Saint-Malo, France, pp 239–244
13. Chiuso A (2007) On the relation between CCA and predictor-based subspace identification. IEEE Trans. Automatic Control 52(10):1795–1812
14. Chiuso A (2007) The role of vector autoregressive modeling in subspace identification. Automatica 43(6):1034–1048
15. Cobelli C, Dalla Man C, Sparacino G, Magni L, De Nicolao G, Kovatchev B (2009) Diabetes: Models, signals and control . IEEE Rev Biomed Eng 2:54–56
16. Dalla Man C, Camilleri M, Cobelli C (2006) A system model of oral glucose absorption: validation on gold standard data. IEEE Trans Biomed Eng 53(12):2472–2477
17. Dalla Man C, Caumo A, Basu R, Rizza R, Toffolo G, Cobelli C (2004) Minimal model estimation of glucose absorption and insulin sensitivity from oral test: validation with a tracer method. Am J Physiol Endocrinol Metab 287:E637–E643
18. Dalla Man C, Caumo A, Cobelli C (2002) The oral glucose minimal model: estimation of insulin sensitivity from a meal test. IEEE Trans Biomed Eng 49(5):419–429
19. Dalla Man C, Rizza RR, Cobelli C (2007) Meal simulation model of the glucose-insulin system. IEEE Trans Biomed Eng 54(10):1740–1749
20. Dalla Man C, Yarasheski KE, Caumo A, Robertson H, Toffolo G, Polonski KS, Cobelli C (2005) Insulin sensitivity by oral glucose minimal models: validation against clamp. Am J Physiol Endocrinol Metab 289:E954–E959
21. Dassau E, Cameron F, Lee H, Bequette BW, Zisser H, Jovanovic L, Chase HP, Wilson DM, Buckingham BA, Doyle FJ (2010) Real-time hypoglycemia prediction suite using continuous glucose monitoring a safety net for the artificial pancreas. Diabetes Care 33:1249–1254
22. DIAdvisor: The DIAdvisor™ (2012) http://www.diadvisor.eu. Accessed May 2013
23. Eren-Oruklu M, Cinar A, Quinn L, Smith D (2009) Estimation of future glucose concentrations with subject-specific recursive linear models. Diabetes Techn Ther 11(4):243–253

24. Eren-Oruklu M, Cinar A, Rollins D, Quinn L (2012) Adaptive system identification for estimating future glucose concentrations and hypoglycemia alarms. Automatica 48:1892–1897
25. Facchinetti A, Sparacino G, Cobelli C (2010) Modeling the error of continuous glucose monitoring sensor data: critical aspects discussed through simulation studies. J Diabetes Sci Technol 4(1):4–14
26. Ferrannini E, Cobelli C (1987) The kinetics of insulin in man. general aspects. Diabetes Metab Rev 3:335–363
27. Finan D, Palerm C, Doyle J, Seborg D (2009) Effect of input excitation on the quality of empirical dynamic modes for type 1 diabetes. Process Syst Eng 55(5):1135–1146
28. Finan D, Zisser H, Jovanovic L, Bevier WC, Seborg DE (2006) Identification of linear dynamic models for type 1 diabetes: a simulation study. In: Proceedings of IFAC international symposium on advanced control of chemical processes (ADCHEM2006), Gramado, Brazil
29. Finan D, Zisser H, Jovanovic L, Bevier WC, Seborg DE (2007) Practical issues in the identification of empirical models from simulated type 1 diabetes data. Diabetes Technol Ther 9(5):438–450
30. Gani A, Andrei G, Srinivasan R, Kenneth W, Jaques R (2009) Predicting subcutaneous glucose concentration in humans: data-driven glucose modeling. IEEE Trans Biomed Eng 56(2):246–254
31. Gani A, Gribok A, Yinghui L, Ward W, Vigersky R, Reifman J (2010) Universal glucose models for predicting subcutaneous glucose concentration in humans. IEEE Trans Inf Technol Biomed 14(1):157–165
32. Guerra S, Sparacino G, Facchinetti A, Schiavon M, Dalla Man C, Cobelli C (2011) A dynamic risk measure from continuous glucose monitoring data. Diabetes Technol Ther (Accepted for publication)
33. Haverkamp B, Verhaegen M (1997) SMI Toolbox: State space model identification software for multivariable dynamical systems, 1.0 edn. TU Delft, Delft
34. Hemocue: HemoCue® (2013) http://www.hemocue.com . Accessed Sept 2013
35. Hovorka R, Canonico V, Chassin L, Haueter U, Massi-Benedetti M, Orsini Federici M, Pieber T, Schaller H, Schaupp L, Vering T, Wilinska M (2004) Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes. Physiol Meas 25, 905–920
36. Johansson R (1993) System Modeling and Identification. Prentice Hall, Englewood Cliffs, NJ
37. Keenan B, Mastrototaro J, Voskanyan G, Steil GM (2009) Delays in minimally invasive continuous glucose monitoring devices: a review of current technology. J Diabetes Sci Technol 3(5):1207–1214
38. Kovatchev B, Breton M, Cobelli C, Dalla Man C (2008) Method, system and computer simulation environment for testing of monitoring and control strategies in diabetes
39. Kovatchev B, Breton M, Dalla Man C, Cobelli C (2008) In silico model and computer simulation environment approximating the human glucose/insulin utilization. Master file MAF-1521, Food and drug administration (FDA), Silver Spring
40. Lehmann E, Deutch T (1992) A physiological model of glucose-insulin interaction in type 1 diabetes mellitus. J Biomed Eng 14:235–242
41. Lehmann E, Hermanyi I, Deutch T (1994) Retrospective validation of a physiological model of glucose-insulin interaction in type 1 diabetes mellitus. Med Eng Phys 16(4):351–352
42. Ljung L (1999) System identification: theory for the user. Prentice-Hall, Upper Saddle River
43. Lynch S, Bequette W (2002) Model predictive control of blood glucose in type I diabetics using subcutaneous glucose measurements. In: Proceedings of the American Control Conference (ACC2002), Anchorage, pp 2714–2719
44. Maciejovski J (2002) Predictive control with constraints. Prentice Hall, London
45. Makroglou A, Li J, Kuang Y (2006) Mathematical models and software tools for the glucose—insulin regulatory system and diabetes: an overview. Appl Numer Math 56:559–573
46. MathWorks: MathWorks (2013) http://www.mathworks.com/products/matlab/. Accessed May 2013

47. Miller M, Strange P (2007) Use of Fourier models for analysis and interpretation of continuous glucose monitoring glucose profiles. J Diabetes Sci Technol 1:630–638
48. Naumova V, Pereverzyev S, Sivananthan S (2012) A meta-learning approach to the regularized learning-case study: blood glucose prediction. Neural Networks 33(9):181–193
49. Nucci G, Cobelli C (2000) Models of subcutaneous insulin kinetics. A critical review. Comput Methods Programs Biomed 62:249–257
50. Palerm C, Bequette B (2007) Hypoglycemia detection and prediction using continuous glucose monitoring - a study on hypoglycemic clamp data. J Diabetes Sci Technol 1(5):624–629
51. Palerm C, Willis J, Desemone J, Bequette B (2005) Hypoglycemia prediction and detection using optimal estimation. Diabetes Technol Ther 7(1):3–14
52. Palerm CC, Rodriguez-Fernandez M, Bevier WC, Zisser H, Banga JR, Jovanovic L, Doyle FJ (2006) Robust parameter estimation in a model for glucose kinetics in type 1 diabetes subjects. In: Proceedings of 28th international conference of the IEEE engineering in medicine and biology society (EMBS2006), New York City, USA, pp 319–322
53. Paoletti S, Juloski A, Ferrari-Trecate G, Vidal R (2007) Identification of hybrid systems: a tutorial. Eur J Control 13:242–260
54. Pappada S, Cameron B, Rosman P, Bourey R, Papadimos T, Olorunto W, Borst M (2011) Neural network-based real-time prediction of blood glucose in patients with insulin-dependent diabetes. Diabetes Technol Ther 13(2):135–141
55. Reifman J, Rajaraman S, Gribok A, Ward W (2007) Predictive monitoring for improved management of glucose levels. J Diabetes Sci Technol 1(4):478–486
56. Roy A, Parker RS (2006) Dynamic modeling of free fatty acid, glucose, and insulin: an extended minimal model. Diabetes Technol Ther 8(6):617–626
57. Sorensen J (1985) A physiologic model of glucose metabolism in man and its use to design and assess improved insulin therapies for diabetes. Ph.D. thesis, Massachussetts institute of technology, Department of chemical engineering (1985)
58. Sparacino G, Facchinetti A, Maran A, Cobelli C (2008) Continuous glucose monitoring time series and hypo/ hyper glycemia prevention: requirements, methods, open problems. Curr Diab Rev 4(3):181–192
59. Sparacino G, Zanderigo F, Maran A, Facchinetti A, Cobelli C (2007) Glucose concentration can be predicted ahead in time from continuous glucose monitoring sensor time-series. IEEE Trans Biomed Eng 54(5):931–937
60. Spurr G, Prentice A, Murgatroyd P, Goldberg G, Reina J, Christman N (1988) Energy expenditure from minute-by-minute heart-rate recording: a comparison with indirect calorimetry. Am J Clin Nutr 48(3):552–559
61. Stahl F, Johansson R (2008) Short-term diabetes blood glucose prediction based on blood glucose measurements. In: Proceedings of 30th international conference of the IEEE engineering in medicine and biology society (EMBC2008), Vancouver, British Columbia, Canada, pp 291–294
62. Stahl F, Johansson R (2009) Diabetes mellitus modeling and short-term prediction based on blood glucose measurements. Math Biosci 217:101–117
63. Steil G, Rebrin K, Hariri F, Jinagonda S, Tadros S, Darwin C, Saad M (2005) Interstitial fluid glucose dynamics during insulin-induced hypoglycemia. Diabetologia 48:1833–1840
64. Association The American Diabetes (2010) Diagnosis and classification of diabetes mellitus. Diabetes Care 33(Supplement 1):S11–S66
65. Van Overschee P, De Moor B (1994) N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems. Automatica 30:75–93
66. Verdult V, Verhaegen M (2002) Subspace identification of multivariable linear parameter-varying systems. Automatica 38:805–814
67. Verhaegen M (1994) Identification of the deterministic part of MIMO state space models given in innovations form from input-output data. Automatica 30(1):61–74

68. Wilinska M, Chassin L, Acerini C, Allen J, Dunger D, Hovorka R (2010) Simulation environment to evaluate closed-loop insulin delivery systems in type 1 diabetes. J Diabetes Sci Technol 4(1):132–144
69. Wilinska M, Chassin L, Schaller H, Schaupp L, Pieber T, Hovorka R (2005) Insulin kinetics in type-1 diabetes: continuous and bolus delivery of rapid acting insulin. IEEE Trans Biomed Eng 52(1):3–12
70. Williams G, Pickup J (1992) Handbook of Diabetes, 2nd edn. Blackwell Science, Oxford
71. Zecchin C, Facchinetti A, Sparacino G, Cobelli C (2013) Reduction of number and duration of hypoglycemic events by glucose prediction methods: a proof-of-concept in silico study. Diabetes Technol Ther 15(1):66–77
72. Zecchin C, Facchinetti A, Sparacino G, De Nicolao G, Cobelli C (2011) A new neural network approach for short-term glucose prediction using continuous glucose monitoring time-series and meal information. In: Proceedings of 33rd International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC2011). Boston, Massachussets, USA
73. Zecchin C, Facchinetti A, Sparacino G, De Nicolao G, Cobelli C (2012) Neural network incorporating meal information improves accuracy of short-time prediction of glucose concentration. IEEE Trans Biomed Eng 59(6):1550–1560
74. Zhao C, Dassau E, Jovanovic L, Zisser H, Doyle F, Seborg D (2013) Predicting subcutaneous glucose concentration using a latent-variable-based statistical method for type 1 diabetes mellitus. J Diabetes Sci and Technol 6(3):617–633

Nonlinear Modeling of the Dynamic Effects of Free Fatty Acids on Insulin Sensitivity

Vasilis Z. Marmarelis, Dae C. Shin and Georgios D. Mitsis

Abstract This chapter presents a nonlinear model of the combined dynamic effects of spontaneous variations of plasma insulin and free fatty acids on glucose concentration in a fasting dog. The model is based on the general nonparametric modeling methodology that employs the concept of Principal Dynamic Modes (PDMs) to obtain a Volterra-equivalent nonlinear dynamic model with two inputs (insulin and free fatty acids) and one output (glucose) that are measured experimentally every 3 min in a fasting dog as time-series data over 10 hr. This model is deemed valid and predictive for all input waveforms within the experimental dynamic range. The obtained model is composed of two PDMs for each input and cubic Associated Nonlinear Functions (ANFs), in addition to two cross-terms. The waveform of the obtained PDMs offers potentially valuable interpretation of the implicated physiological mechanisms. The system nonlinearities are described, in turn, by the obtained ANFs. The evaluation of the overall model performance is facilitated by the use of specialized inputs, such as pulses or impulses. For instance, the use of insulin input pulses can yield estimates of "dynamic insulin sensitivity" (as the ratio of the steady-state glucose response to the input pulse amplitude) for various levels of free fatty acids. The obtained result indicates (in a quantitative manner) the widely held view that insulin sensitivity decreases with rising levels of free fatty acids. Furthermore, it indicates that this effect depends on the input insulin strength (dose-dependent insulin sensitivity). Drastic reduction of

V. Z. Marmarelis (✉) · D. C. Shin
Department of Biomedical Engineering, University of Southern California, Los Angeles, CA, USA
e-mail: vzm@usc.edu

D. C. Shin
e-mail: shind@usc.edu

G. D. Mitsis
Department of Electrical and Computer Engineering, University of Cyprus, Nicosia, Cyprus
e-mail: mitsis.georgios@ucy.ac.cy

V. Marmarelis and G. Mitsis (eds.), *Data-driven Modeling for Diabetes*, Lecture Notes in Bioengineering, DOI: 10.1007/978-3-642-54464-4_10,

insulin sensitivity is predicted by the model above a critical level of free fatty acids for low-to-moderate values of plasma insulin. This result demonstrates the potential utility of the proposed modeling approach for advancing our quantitative understanding of the processes underpinning obesity and Type II diabetes.

1 Introduction

The multiple effects (direct and indirect) of insulin on blood glucose have been studied extensively in the context of diabetes mellitus, motivated by the need for improved diagnostic procedures and effective treatment of diabetic patients. Parametric/compartmental models (assuming the form of sets of differential and algebraic equations) have been developed for this purpose and seek to describe the causal effect of infused insulin on blood glucose concentration—typically in connection with specific testing protocols, such as the Glucose Tolerance Test [1, 2]. More complex models that take into account the effects of glucagon and free-fatty acids have also been proposed [3, 4]. Nonparametric data-based modeling has also been suggested for this purpose using variants of the Volterra-Wiener approach to input-output system modeling [5, 6]. Our group has pioneered the Laguerre-Volterra network [5] and the Principal Dynamic Modes (PDMs) approaches [6], which both utilize Laguerre expansions of the system kernels. The nonparametric approach has the advantage of being true to the data and not requiring a priori postulation of a specific model form (e.g. differential equations). The PDM-based approach is used in the present study, which seeks to elucidate the dynamic effects of free fatty acids (FFAs) on insulin-glucose interactions. This subject is attracting increasing attention in the context of the relation between obesity and diabetes [7, 8]. Elevated FFAs have been shown to increase plasma glucose and hepatic glucose output, as well as increase peripheral and hepatic insulin resistance in a dose-dependent manner [9–11].

Thus, FFAs are viewed as a major link between obesity and insulin resistance or Type 2 diabetes. In the liver, FFAs cause insulin resistance by inhibiting insulin suppression of glycogenolysis. FFAs also promote glucose-stimulated insulin secretion by the pancreatic beta cells. The latter is viewed as a possible reason preventing the development of Type 2 diabetes in most obese insulin-resistant people. FFAs have also been shown to facilitate inflammatory processes and, therefore, may contribute to the pathogenesis of coronary artery disease [8, 9]. Dysregulation of FFA metabolism may cause insulin resistance because of preferential oxidation of FFA over glucose [10].

In the present study, we examine how spontaneous variations in plasma insulin and FFA (viewed as the "input" variables) jointly affect the level of blood glucose (viewed as the "output" variable) under fasting conditions. To achieve this goal, we quantify the dynamic effects of changes in the two input variables upon the output variable via the Volterra-equivalent PDM-based model estimated from

actual experimental time-series data collected in a fasting dog. The obtained model has predictive capability for arbitrary inputs that remain within the experimental dynamic range. The model is *dynamic* (i.e. it predicts the present value of output glucose concentration on the basis of the entire epoch of insulin and FFA input values) and *nonlinear* (i.e. the effects of input changes upon the output are not additive and do not scale proportionally). The employed modeling methodology is generally applicable to nonlinear dynamic systems and robust in the presence of measurement noise or systemic interference. In addition, this methodology is applicable to short data-records, thereby rendering feasible the estimation of the model with limited number of experimental measurements [12].

The objective of this chapter is to demonstrate the efficacy of the proposed approach with experimental data from a fasting dog and to provide some novel (albeit preliminary) physiological insight into the joint causal effects of plasma insulin and FFA variations on plasma glucose. This insight is offered in the quantitative form of a predictive model that may be used to test rigorously postulated hypotheses within the dynamic range of the available data. It is hoped that the efficacy of this approach will enable fruitful applications in this physiological domain.

2 Methods

Experimental time-series data of plasma glucose, insulin and free fatty acids (FFAs) were collected in a healthy male mongrel dog every 3 minutes over a 10-hour period (200 time-series samples). The data were collected under fasting conditions of spontaneous activity and the animal was judged to be in good health. The University of Southern California Institutional Animal Care Committee approved all surgical and experimental procedures. Details of the experimental procedures can be found in [6].

The collected time-series datasets are shown in the left panel of Fig. 1 and their mean (standard deviation) values are: 81.14 (2.39) mg/dl for glucose; 44.85 (13.81) pM for insulin; and 0.52 (0.07) mM for FFA. Since we are interested in studying the dynamics of this system over time horizons longer than 15 min, we perform moving averaging with a five-point Hanning window (equivalent to low-pass filtering below 0.1 cycles/min) that removes the very rapid variations of the data. We analyze the de-meaned filtered data that have standard deviations (SDs) of 1.84 mg/dl for glucose; 9.68 pM for insulin; and 0.06 mM for FFA, and are shown in the right panel of Fig. 1. The SDs of the filtered data are smaller because sharp peaks are smoothed, especially in the insulin data. We deem this acceptable because we focus on capturing the system dynamics over cyclical changes with periods longer than 10 min per cycle. The random and broadband nature of these datasets is evident. It is difficult to discern visually any consistent correlation between the fluctuations of these signals—a fact that motivates the use of dynamic modeling of the data using our nonlinear methodology. The latter employs the

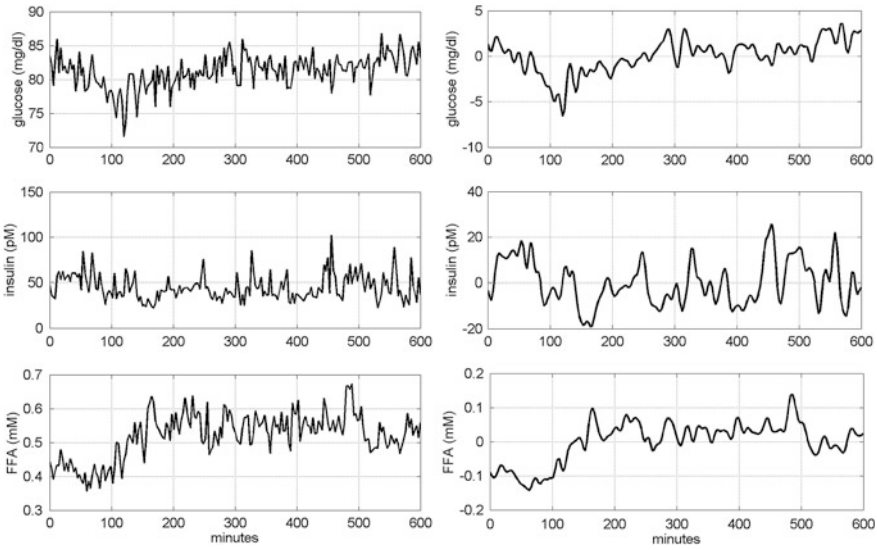


Fig. 1 *Left* the collected experimental time-series data of plasma glucose (*top panel*), insulin (*middle panel*) and free fatty acids (FFAs) (*bottom panel*) from a fasting dog at rest over 10 hours (sampling every 3 min). *Right* the time-series data after mean subtraction and low-pass filtering

concept of Principal Dynamic Modes (PDMs) to obtain a model with two inputs (insulin and FFA) and one output (glucose). The obtained Volterra-equivalent PDM-based nonlinear dynamic model of the input-output relationship does not require a priori model postulates and yields data-based models that have general predictive capability for arbitrary input waveforms within the experimental dynamic range. The employed modeling methodology is summarized in Appendix I. Details can be found in [12].

3 Results

Using the PDM-based modeling methodology outlined in Appendix I with three Laguerre basis functions having alpha parameter 0.4 for the insulin input and 0.8 for the FFA input (determined via a search procedure minimizing the prediction error), we obtain the model shown schematically in Fig. 2 that has two PDMs for each input and two cross-terms (one between the 1st PDM of insulin and the 2nd PDM of FFA, and the other between the 1st and 2nd PDM of FFA). The computed PDMs are shown in Fig. 3 in the time-domain. Each PDM can be viewed as the impulse response function of a linear filter that transforms the respective input epoch into a state variable (the PDM output) that has a contemporaneous nonlinear relationship with the output variable described by the respective ANF [12]. The computed ANFs are shown in Fig. 4, plotted over the abscissa range of ± 1 SD of

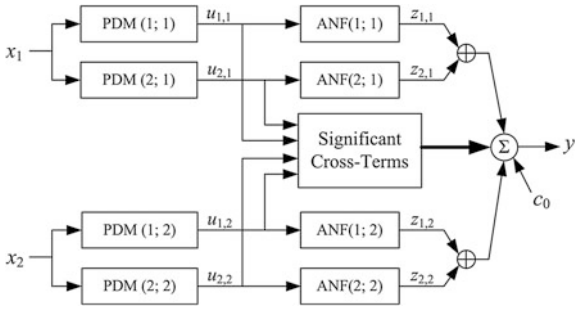


Fig. 2 Schematic of the PDM-based model, which is composed of two PDMs for each input (insulin and FFA deviations from the respective mean values) that can be viewed as impulse response functions of two linear filters receiving the respective input. Each PDM is followed by a static nonlinearity, termed Associated Nonlinear Function (ANF), which transforms the PDM output into an additive component of the model-predicted glucose output (deviation from its mean value). The model also includes the significant cross-terms that are selected on the basis of the statistical significance of their correlation with the output (see Appendix I). Two cross-terms were found significant in this system: one between PDMs of the two inputs (1st of insulin and 2nd of FFA) and the other between the two PDMs of the FFA input. The ordinate of the insulin PDMs is in mg/dl per pM, and in mg/dl per mM for the FFA PDMs. The abscissa of the PDMs is in min. The abscissa and ordinate of the ANFs are both in mg/dl

the PDM output that corresponds to the experimental data. The abscissa of the ANF plot is the PDM output and the ordinate is the corresponding component of the model output prediction (glucose) at the same time instant. Figure 5 shows the PDM-based model prediction (after the initial transient of 120 min that is roughly equal to the system memory), along with the actual glucose-output. The normalized mean-square error of this model prediction is 21 %. We view this low prediction error (relative to what can be achieved with other methods) as validating the obtained model. The specific form of the obtained PDMs and ANFs is discussed in the following section with regard to plausible physiological interpretations.

The output equation of this two-input PDM-based model is:

$$G(t) = G_0 + f_{I,1}\left[P_{I,1} * I(t)\right] + f_{I,2}\left[P_{I,2} * I(t)\right] + f_{F,1}\left[P_{F,1} * F(t)\right] + f_{F,2}\left[P_{F,2} * F(t)\right] + \sum_{j,k} C_{j,k}(t)$$

where * denotes the convolution operation, $G(t)$, $I(t)$ and $F(t)$ denote the demeaned time-series data of glucose, insulin and FFA respectively, f denotes the cubic ANF and P denotes the PDM for the input and rank indicated in the respective subscript, G_0 is the model constant, and $C_{j,k}$ (t) denotes the significant cross-terms of the model (products of PDM outputs that have statistically significant correlation with the glucose-output signal)—in this model, the following cross-term combinations were found to be significant: ($P_{I,1}$, $P_{F,2}$) and ($P_{F,1}$, $P_{F,2}$). This model has 15 free parameters (12 ANF coefficients, 2 cross-term coefficients

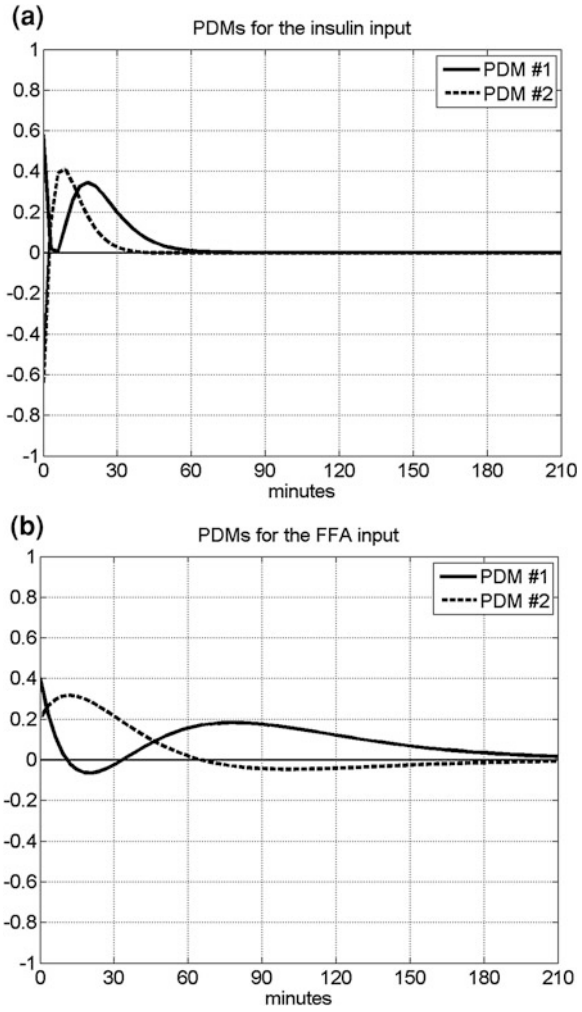


Fig. 3 The computed two PDMs for the insulin input (*left*) and the FFA input (*right*)

and the constant G_0). The PDMs are estimated separately from the kernel estimates via the Laguerre expansion technique (see Appendix I) that requires in this case the least-squares estimation of 28 free parameters (9 for the self-kernels of each of the two inputs, 9 for the cross-kernel and the constant). The available data (200 samples) are deemed adequate for the estimation of these parameters without risk of over-fitting.

Having obtained the PDM-based model, we can use it to advance our understanding of the dynamics of this system. Some discussion on plausible interpretations of the specific form of the obtained PDMs and the respective ANFs are provided in the following section. In this section, we illustrate the model prediction for specialized inputs (i.e. insulin impulse and FFA pulse) that can be evaluated against known physiology. We also explore one important aspect of the system

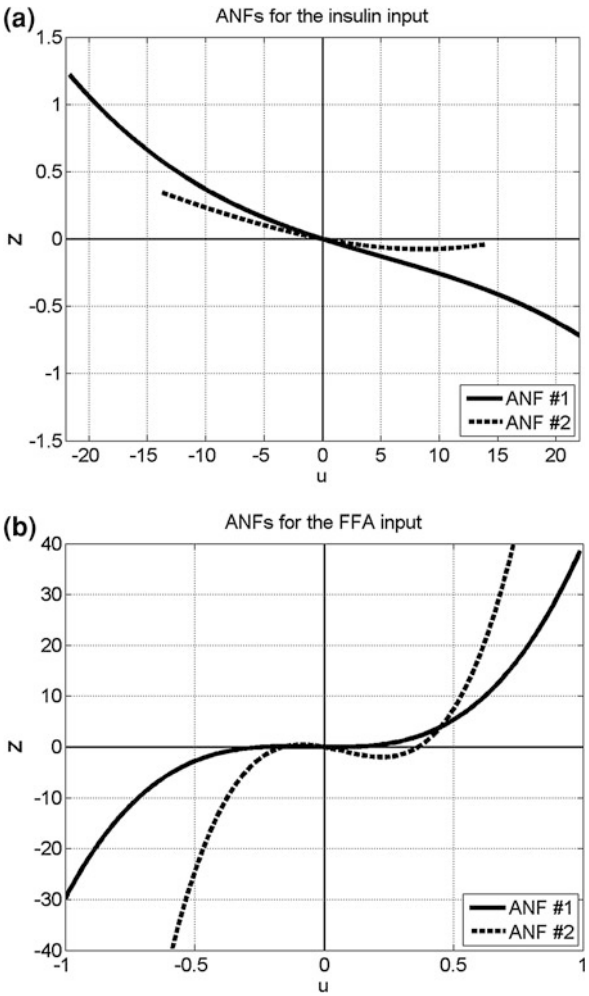


Fig. 4 The computed ANFs for the insulin input (*left*) and the FFA input (*right*)

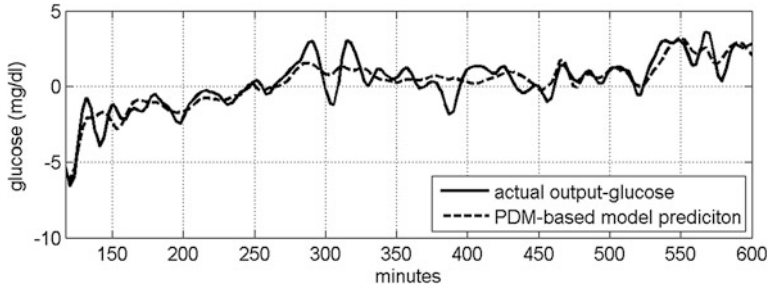


Fig. 5 The PDM-based model prediction of the output (*dashed line*) and the actual output-glucose data (*solid line*) after de-meaning and low-pass filtering. The resulting Normalized Mean-Square Error (NMSE) of this prediction is 21 %

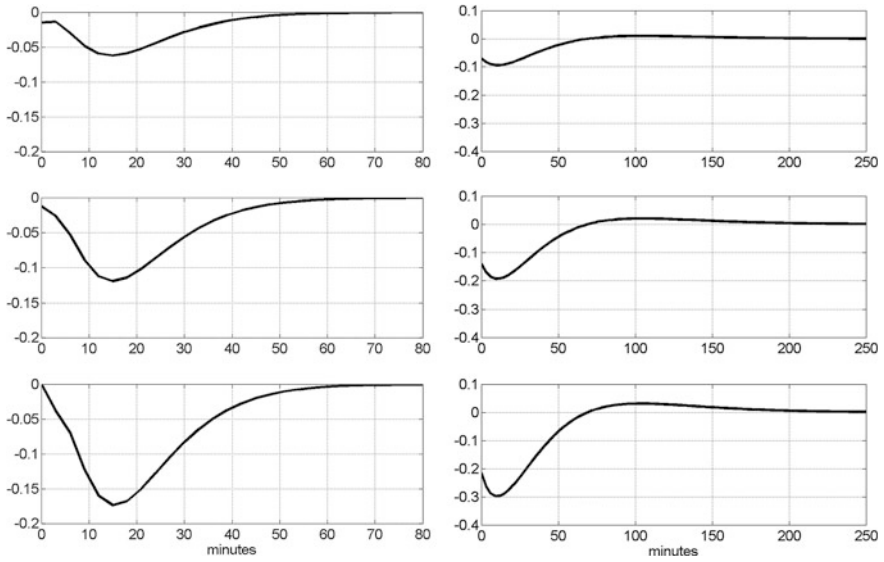


Fig. 6 *Left panels*, *top* to *bottom* the model-predicted glucose responses for three magnitudes of insulin impulses (0.5, 1 and 1.5 SD of the recorded insulin data). *Right panels*, *top* to *bottom* the model-predicted glucose responses for three levels of FFA impulses (0.5, 1 and 1.5 SD of the recorded FFA data). The units of glucose are mg/dl. Evidently, the FFA effect is much larger

dynamics that pertains to the notion of "dynamic insulin sensitivity" (DIS) by utilizing the predictive capability of the PDM-based model to compute the predicted glucose response to an insulin pulse input for a given level of FFA. The insulin and FFA values are deviations from the respective mean values of the experimental data (i.e. relative to the observed operating point of the system). The model predictions for three magnitudes of insulin impulses (0.5, 1 and 1.5 SD of the recorded insulin data) and three levels of FFA impulses (0.5, 1 and 1.5 SD of the recorded FFA data) are shown in Fig. 6 and suggest that glucose is reduced in a sub-linear manner in response to an increase of insulin (consistent with current view), and the glucose response to an FFA increase is biphasic exhibiting an early sub-linear reduction and a later phase of counter-regulation (after about 2 hr) where glucose is increased. The early phase of the glucose response to an FFA increase may be due to the reported facilitation of pancreatic beta-cell secretion (leading to an increase of insulin and subsequent reduction of glucose), whereas the late-phase counter-regulation may be due to the known inhibition of FFA to the insulin facilitation of glucose uptake by tissues and to the insulin inhibition of the processes of glycogenolysis and gluconeogenesis (both leading to glucose increase). It is posited from the different dynamic time-scales that the latter effect takes longer than the former.

As an illustration of the nonlinear interaction between the two inputs as they affect the glucose output, we show in Fig. 7 the model-predicted glucose responses to an insulin pulse input with amplitude equal to 1 SD of the experimental insulin data for four different levels of FFA equal to 0, 0.5, 1 and 1.5 SD of the

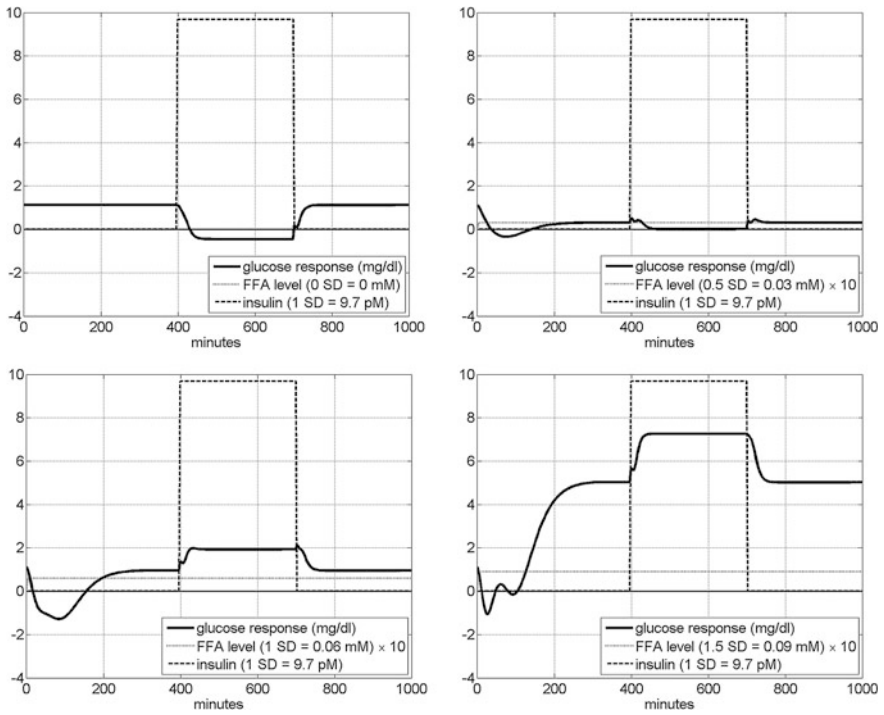


Fig. 7 The model-predicted glucose response in mg/dl to an insulin pulse input equal to 1 SD of the experimental insulin data (9.7 pM) for four different levels of FFA: 0 (*top left*), 0.5 (*top right*), 1 (*bottom left*) and 1.5 (*bottom right*) SD of the experimental FFA data. We observe that the glucose response to insulin is reversed from the normal (increase instead of decrease) for high levels of FFA. The early transient glucose response (reduction) to a non-zero FFA step input (prior to the application of the insulin pulse) is consistent with the response profile presented in Fig. 6

experimental FFA data. It is evident that the steady-state value of the glucose output (in response to the given insulin pulse input) is smaller for higher level of FFA and, surprisingly, it is reversed from the normal (increase instead of decrease) for high levels of FFA. This surprising result must be examined with controlled experiments in the future. The observed early transient glucose response (reduction) to a non-zero FFA step input is consistent with the response profile presented and discussed in Fig. 6.

The Dynamic Insulin Sensitivity (DIS) is defined as the ratio of the predicted steady-state glucose response to the corresponding amplitude of an insulin pulse input (for a given FFA level). The sign is inverted, so that normal DIS (i.e. reduction of glucose for raised insulin) takes positive values. Since the model is nonlinear (3rd order), the computed DIS values will generally follow a nonlinear (cubic) function of the insulin pulse amplitude and of the respective FFA level. These DIS nonlinear functions will be generally different for different operating points of the system.

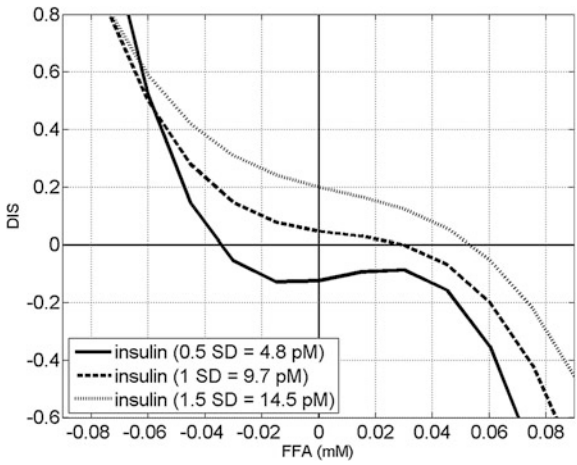


Fig. 8 The computed DIS curves as a function of FFA level for three amplitudes of the insulin input pulse (0.5, 1 and 1.5 SD of the experimental insulin data)

The obtained results of computed DIS values are shown in Fig. 8 for three amplitudes of the insulin input pulse (0.5, 1 and 1.5 SD of the experimental insulin data), plotted over constant FFA levels ranging from −1.5 to +1.5 SD of the actual FFA experimental data. It is evident in Fig. 8 that the DIS is reduced significantly for high FFA levels, with a critical soft-threshold value seen around 0.05 mM. This effect is sharper for lower insulin values. Additional discussion on this result and the relation of DIS to the widely used "insulin sensitivity" (S_I) parameter associated with the "Minimal Model" of the Glucose Tolerance Test is provided in the following section.

Using the model output equation, we can remove, in principle, the dependence of the DIS estimate from the specific operating point of the experimental data (as defined by the mean values) by replacing the deviations from the mean with the data value prior to de-meaning. It can be shown that the 3rd-degree coefficients are not affected, although the ANF coefficients of degree lower than 3rd and the constant term are affected by the mean values. The PDMs remain the same.

As a final illustration of the capability of the predictive model in enhancing our understanding of the system functional characteristics, we plot in Fig. 9 the steady-state value of the glucose output in response to an insulin input pulse for various amplitudes ranging from −2 to +2 SD of the experimental insulin data. This steady-state insulin-glucose relationship is nonlinear and varies for different levels of FFA, as demonstrated in Fig. 9 for five different levels of FFA. It is seen that the insulin-glucose relation exhibits only small changes when the FFA levels remain close to the baseline value (e.g. from −0.05 to 0.05 mM), but changes drastically when the FFA levels take larger values (e.g. ±0.1 mM). In this regard—i.e. high levels of FFA result in drastic reduction of insulin sensitivity (rise of insulin resistance), as indicated by the slope of these curves. It is also seen that increased levels of insulin generally improve the insulin sensitivity (by making the slope of

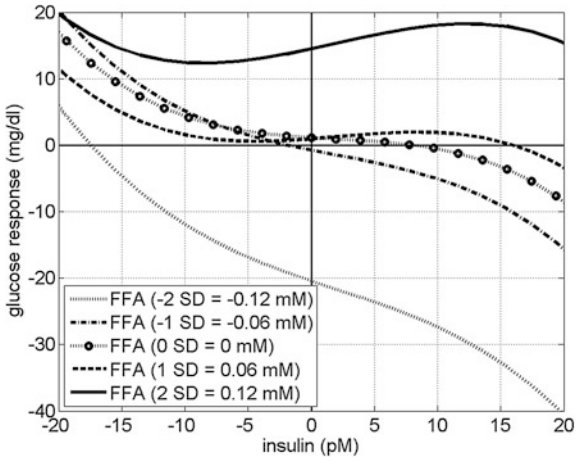


Fig. 9 The computed steady-state insulin-glucose curves (over ±2 SD of the experimental insulin data) for various FFA levels. Insulin resistance is highest for the highest FFA level and is generally reduced as the insulin values increase

these curves more negative) for all FFA levels. A critical value of insulin around 15 pM (i.e. about 1.5 SD of the experimental insulin data) emerges from the results in this regard (see Fig. 9).

4 Discussion

The PDM-based model of Fig. 2 provides insights into the dynamic interrelationships between insulin, FFA and glucose variations during spontaneous fasting conditions in a dog. If analogies hold in human subjects, these insights may have important implications for the in-vivo regulation of these variables in diabetics.

The waveforms of the obtained two PDMs and ANFs for the insulin input indicate a "glucoleptic" effect (i.e. glucose reduction for insulin increase) for both PDMs—with the 1st PDM exhibiting faster dynamics. The 1st ANF (see Fig. 4) exhibits a symmetric and mildly supralinear response characteristic, but the 2nd ANF is asymmetric indicating larger effect for negative PDM output. These findings are consistent with existing qualitative physiological knowledge regarding the process of insulin-facilitated glucose uptake by muscle, adipose and organ tissues, as well as inhibition of the processes of glycogenolysis and lipolysis, all of them leading to reduction in plasma glucose, as described quantitatively by the two glucoleptic PDM-ANF cascades. However, it is not currently known which set of those glucoleptic processes is faster and thus corresponds to the 1st PDM. This question should be resolved with specialized experiments in the future.

The waveforms of the obtained two PDMs for the FFA input indicate mainly glucogenic characteristics (i.e. positive PDM values), probably related to the known inhibitory effects of FFA on insulin facilitation of glucose uptake and the process of glycogenesis. Specifically, the 1st FFA PDM exhibits early positive values indicating immediate rise of glucose in response to an impulsive rise of

FFA (as long as it is of adequate size to exceed the observed dead-zone—which is not the case in the illustration of Fig. 6). The positive values of this PDM rapidly diminish and become slightly negative between 10 and 30 min after input onset. Subsequently, the values of this PDM rise to reach a peak in ∼70 min and remain positive for over three hours. This late positive "hump" is responsible for the late-phase counter-regulation observed in the illustration of Fig. 6 for large FFA pulse input and explains the surprising result of reversed glucose response to insulin (increase instead of decrease) for high levels of FFA that is shown in Fig. 7. This intriguing finding ought to be examined with specialized experiments in the future because, if confirmed, bears huge implications for the understanding of Type 2 diabetes. The 2nd FFA PDM starts positive and reaches a peak within ∼10 min (although its respective ANF exhibits small negative values near the origin, which accounts for the early negative glucose response to FFA pulse input in the illustration of Fig. 6), diminishing afterwards and reaching near-zero levels around 60 min (see Fig. 3). The respective ANFs (see Fig. 4) have nearly symmetric response characteristics with a central dead-zone (i.e. supralinear increase/decrease of glucose for increase/decrease of FFA, as long as the latter has adequate size to exceed the dead-zone). These findings are consistent with existing qualitative physiological knowledge regarding the inhibitory effects of FFA on insulin facilitation of glucose uptake and on the process of glycogenesis. The precise time-constants or the relative magnitudes of these physiological effects are not currently known and, therefore, confirmation of the validity of the PDM analysis of this system with specialized experiments in the future may provide valuable new insights.

One interesting finding that was presented in the previous section is the relation of "dynamic insulin sensitivity" (DIS) to the widely used "insulin sensitivity" (S_I) associated with the "Minimal Model" of the Glucose Tolerance Test [1], which is defined as the ratio of two parameters of that model: $S_I = p_3/p_2$. The computed DIS values using the PDM-based model (see Fig. 8) suggest that DIS is reduced for increasing levels of FFA—a finding consistent with previous observations regarding the effect of elevated FFA on plasma glucose and insulin sensitivity [7]—and this reduction becomes drastic when the FFA level exceeds a critical value (∼0.05 mM). Our findings also suggest that the induction of insulin resistance by elevated FFA is insulin-dose dependent. When our definition of the DIS is applied to the Minimal Model, the following expression is derived between DIS and S_I: $\mathbf{DIS = (G_b{*}S_I) / (p_1 + A\ S_I)}$, where G_b denotes the basal glucose value, p_1 is the glucose disappearance parameter of the Minimal Model and A is the amplitude of the insulin input pulse. It is evident from this expression that the DIS is dependent on the insulin dose A and the operating point (defined by G_b). We note that the Minimal Model does not take into account the effects of FFA.

The presented modeling approach is data-driven and offers quantitative insights into the causal inter-relationships between insulin, glucose and FFA. These insights may prove useful in understanding these critical processes in patho-physiological conditions such as obesity and Type 2 diabetes (with the eventual promise of improved diagnosis and disease management). The presented results

are preliminary and, since they are based on only one dataset, they are simply illustrative of the PDM-based modeling approach and its potential in this field of physiology. The obtained model is not proposed as the definitive dynamic relation among these variables. Nonetheless, the presented results demonstrate the potential of the PDM-based modeling approach to advance our quantitative understanding of the subject system.

Acknowledgments This work was supported in part by the NIH/NIBIB center grant No. P41-EB001978 to the Biomedical Simulations Resource at the University of Southern California and by the University of Cyprus Internal Research Grant "Nonlinear, data-driven modeling and model-based control of blood glucose". The authors acknowledge thankfully Dr. R. Bergman and Dr. K. Hücking for making available the experimental data.

Appendix I: Summary of PDM-Based Modeling Methodology

We follow the general Volterra modeling approach which is applicable to all finite-memory stationary dynamic nonlinear systems [12]. For the two-input system of this study, we begin with the estimation of a 2nd order Volterra model using Laguerre expansions of the kernels [12]:

$$\begin{aligned} G(t) = k_0 &+ \int_0^\infty k_I(\tau)\, I(t-\tau)\, d\tau \;+\; \int_0^\infty k_F(\tau)\, F(t-\tau)\, d\tau \\ &+ \int\int_0^\infty k_{II}(\tau_1,\tau_2)\, I(t-\tau_1) I(t-\tau_2)\, d\tau_1 d\tau_2 \\ &+ \int\int_0^\infty k_{FF}(\tau_1,\tau_2)\, F(t-\tau_1) F(t-\tau_2)\, d\tau_1 d\tau_2 \\ &+ \int\int_0^\infty k_{IF}(\tau_1,\tau_2)\, I(t-\tau_1) F(t-\tau_2)\, d\tau_1 d\tau_2 \;+\; \varepsilon(t) \end{aligned} \tag{A1}$$

where, $G(t)$ denotes the glucose output, $I(t)$ denotes the insulin input, $F(t)$ denotes the FFA input and $\varepsilon(t)$ denotes the model prediction errors. The dynamic characteristics of this system/model are described by the kernels: k_I, k_F, k_{II}, k_{FF}, k_{IF}, which are estimated from given input-output data by means of Laguerre expansions and least-squares fitting as described below. A key step in the use of the Laguerre expansion technique is the proper selection of the Laguerre parameter "alpha", which is accomplished through a search procedure minimizing the mean-square prediction error. The Laguerre expansions of the 1st-order kernels using L Laguerre basis functions $\{b_j\}$ are given by:

$$k_i(\tau) = \sum_j a_i(j)\, b_j(\tau) \tag{A2}$$

where the subscript i denotes the input (I or F), a_j are the Laguerre expansion coefficients and the summation is taken over $j = 1,\ldots, L$. The Laguerre expansions of the 2nd-order kernels are given by:

$$k_{i_1,i_2}(\tau_1,\tau_2) = \sum_{j_1}\sum_{j_2} a_{i_1,i_2}(j_1,j_2)\; b_{j_1}(\tau_1)\; b_{j_2}(\tau_2) \tag{A3}$$

where the subscripts i_1, i_2 denote the inputs (I or F), a_{i_1,i_2} are the Laguerre expansion coefficients and the double summation is taken over j_1 and j_2 from 1 to L. The self-kernels correspond to the case when i_1 and i_2 denote the same input, and the cross-kernel when i_1 is different from i_2. Then, we have the following input-output relation which involves *linearly* the Laguerre coefficients:

$$y(t) = a_0 + \sum_i \sum_j a_i(j)\; v_{j,i}(\tau) \; + \sum_{i_1,i_2}\sum_{j_1}\sum_{j_2} a_{i_1,i_2}(j_1,j_2)\; v_{j_1,i_1}(\tau_1)\; v_{j_2,i_2}(\tau_2) \tag{A4}$$

where a_0 is a baseline constant, and the signals $v_{j,i}(t)$ are the convolutions of the Laguerre basis function b_j with the respective input i. The fact that the Laguerre expansion coefficients enter linearly in the nonlinear input-output model of Eq. (A4) allows their estimation via least-squares fitting. Following estimation of the Laguerre expansion coefficients, we can construct the Volterra kernel estimates using Eqs. (A2) and (A3), which allows computation of the model prediction for *any* given input using Eq. (A1) or (A4).

The concept of Principal Dynamic Modes (PDMs) aims at identifying an efficient basis of functions distinct and characteristic for each system, which is capable of representing adequately the system kernels. The computation of the PDMs for each input is based on Singular Value Decomposition (SVD) of a rectangular matrix composed of the 1st order kernel estimate (as a column vector) and the 2nd order self-kernel estimate (as a block matrix) weighted by the standard deviation of the respective input. The PDMs are defined for the significant singular values (two for each input in this application). The resulting PDMs form a filter-bank that receives the respective input signal and generates (via convolution) signals that are subsequently transformed by the "Associated Nonlinear Function" (ANF), which represents the nonlinear characteristics of the system for the respective PDM dynamics, to form additively the system output, as depicted schematically in Fig. 2. Thus, the PDM-based model separates the dynamics (PDMs) from the nonlinearities (ANFs). Since the "separability" of the system nonlinearity cannot be generally assumed, we include "cross-terms" in the PDM-based model that are properly selected on the basis of a statistical significance test on the computed correlation coefficient between each cross-term (i.e. the pair product of PDM outputs) and the output signal, using the w-statistic. Two cross-terms were found to be adequate in this application.

The structure of the PDM-based model of the two-input/one-output system is shown in Fig. 2. The employed PDMs represent a common "functional basis" for

efficient representation of *all* kernels of the system. Upon selection of the PDMs and the significant cross-terms, we estimate the Associated Nonlinear Functions (ANFs), which are polynomial functions (cubic in this case) transforming the output of the respective PDM into a variable that represents an additive component of the model output, along with the other ANF outputs and the cross-terms:

$$y(t) = c_0 + \sum f_I[u_I(t)] + \sum f_F[u_F(t)] + \text{Cross-Terms} + \varepsilon(t) \qquad \text{(A5)}$$

where $\{u_I\}$ and $\{u_F\}$ are the PDM outputs for the inputs I and F respectively (i.e. convolutions of each input with the respective PDM), and $\{f_I\}$ and $\{f_F\}$ are the ANFs associated with each PDM. The "Cross-Terms" in Eq. (A5) are pair products of $\{u_I\}$ and $\{u_F\}$ that have significant correlation with the output. The coefficients of the cubic ANFs and of the selected Cross-Terms (pair products of PDM outputs with significant correlation with the system output) are estimated, along with baseline constant c_0, via least-squares regression of the output Eq. (A5).

References

1. Bergman RN, Lovejoy JC (1997) The minimal model approach and determinants of glucose tolerance. Louisiana State University Press, Baton Rouge
2. Vicini P, Caumo A, Cobelli C (1997) The hot IVGTT two-compartment minimal model: indexes of glucose effectiveness and insulin sensitivity. Am J Physiol 273:E1024–E1032
3. Cobelli C, Mari A (1983) Validation of mathematical models of complex endocrine-metabolic systems. A case study on a model of glucose regulation. Med Biol Eng Compu 21:390–399
4. Roy A, Parker RS (2006) Dynamic modeling of free fatty acid, glucose, and insulin: an extended "minimal model". Diab Technol Ther 8:617–626
5. Mitsis GD, Markakis MG, Marmarelis VZ (2009) Nonlinear modeling of the dynamic effects of infused insulin on glucose: comparison of compartmental with Volterra models. IEEE Trans Biomed Eng 56(10):2347–2358
6. Marmarelis VZ, Mitsis GD, Huecking K, Bergman RN (2002) Nonlinear modeling of the insulin-glucose dynamic relationship in dogs. In: Proceedings of the second joint EMBS/BMES conference, Houston, TX, pp 224–225
7. Rebrin K, Steil GM, Getty L, Bergman RN (1995) Free fatty acid as a link in the regulation of hepatic glucose output by peripheral insulin. Diabetes 44:1038–1045
8. Boden G (2002) Interaction between free fatty acids and glucose metabolism. Curr Opin Clin Nutr Metab Care 5(5):545–549
9. Boden G (2003) Effects of free fatty acids (FFA) on glucose metabolism: significance for insulin resistance and type 2 diabetes. Exp Clin Endocrinol Diab 111(3):121–124
10. Delarue J, Magnan C (2007) Free fatty acids and insulin resistance. Curr Opin Clin Nutr Metab Care 10(2):142–148
11. Kraegen EW, Cooney GJ (2008) Free fatty acids and skeletal muscle insulin resistance. Curr Opin Lipidol 19(3):235–241
12. Marmarelis VZ (2004) Nonlinear dynamic modeling of physiological systems. IEEE-Wiley, Piscataway