Subspace-based Model Identification of Diabetic Blood Glucose Dynamics

Marzia Cescon * Fredrik Ståhl * Mona Landin-Olsson ** Rolf Johansson *

* Department of Automatic Control, LTH, Lund University, SE-221 00 Lund, Sweden (e-mail: marzia.cescon|fredrik.stahl|rolf.johansson@control.lth.se).
** Department of Endocrinology, Lund University Hospital SE-221 85 Lund, Sweden (e-mail: mona.landin-olsson@med.lu.se)

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, thus leading to serious health damages. In order to keep tight glucose control treatment happens to be based essentially on insulin delivery. In common practise, a main limiting factor in adequate choice of insulin dosing is the lack of reliable predictions of blood glucose evolution. In this work, state-space models of various orders were investigated and evaluated for their capability of description of one diabetic subject blood glucose evolution over 24 hours. It turned out that models of low complexity are capable of detecting even abrupt changes in calibration data but they are not sufficient for validation data. Blood glucose levels could be predicted fairly well up to 30 minutes ahead on validation data.

Keywords: Diabetes, mathematical model, physiological models, subspace-based system identification, glucose-insulin dynamics.

1. INTRODUCTION

Diabetes Mellitus is a syndrome of impaired carbohydrate, fat, and protein metabolism caused by either lack of insulin secretion or decreased sensitivity of the tissue to insulin. In type I diabetes, also called insulin-dependent diabetes mellitus (IDDM), there is an absolute deficiency of insulin secretion, which is due to β-cell destruction. The basic effect of insulin lack or insulin resistance to glucose metabolism is the prevention of efficient uptake and utilization of glucose by most cells of the body. As a result, blood glucose concentration increases (hyperglycemia). This has multiple effect throughout the body associated with damage, dysfunction and failure of various organs. The main complications are heart disease, retinopathy, nephropathy and neuropathy.

Type I diabetes is often controlled with insulin injections. The philosophy of insulin replacement is to mimic with injections the insulin secretion pattern in the non-diabetic person. Here, insulin is secreted at two rates: a slow basal secretion throughout the 24 hours and an augmented rate at meal times. The basal insulin concentration is sufficient to keep a constant glucose concentration during fasting conditions and the prandial insulin doses should enhance an increased glucose uptake during and after meals. If not treated, too high levels of insulin cause the blood glucose to fall to low values resulting in loss of consciousness and coma. Hence, tight glucose control (i.e., as close to normal as possible) should be maintained in order to prevent long-term complications due to hyperglycemia and short-term complications due to hypo- and hyperglycemia.

All patients must frequently adjust their insulin dosing. This represents a daily challenge, as many factors influence glycemia, such as diet, meal composition, exercise as well as stress. The task is non trivial, therefore many patients would benefit from some sort of decision support. The availability of a blood glucose predictor that would inform the patient on the near future blood glucose and offer advice on how to modulate insulin therapy in relation to food intake and insulin injections to prevent out-of-target glucose deviations would be highly valuable. Such a prediction engine of the blood glucose evolution requires a model of the dynamic interplay between previously injected insulin, food intake and current blood glucose level. This paper addresses the questions how to identify such a model and how to provide model-based short-term predictors suitable for blood glucose prediction.

Whereas there is at least thirty years of research reported e.g. in (Bergman and Cobelli 1980), (Lynch and Bequette
2002), (Hovorka et al. 2004), (Sparacino et al. 2007) there are few contributions on data-based modeling (see e.g. (Ståhl and Johansson 2008a), (Ståhl and Johansson 2008b), (Makrogliou et al. 2006), (Finan et al. 2006). To this purpose, the paper addresses mathematical modeling of glucose-insulin dynamics by means of system identification, see (Ljung 1999, Johansson 1993, Söderström and Stoica 1989).

Denoting with $u$ the exogenous variables, namely carbohydrate intake and injected insulin and with $y$ the explained variable, namely plasma glucose concentration, the focus was to fit a statistical model to the data, describing and not explaining the underlying mechanisms of data generation using subspace-based techniques applied to one patient records.

The paper outline is as follows. In Sec. 2 the experimental conditions are introduced, the physiological models used for input estimation are shown and the identification strategy is provided. Section 3 presents the results achieved throughout the work. In Sec. 4 the work carried out is discussed. Last, Sec. 5 concludes the paper and proposes future research.

2. METHODS

2.1 Experimental Conditions & Data Acquisition

The data consisted of three days CGMS (Continuous Glucose Monitoring System) measurements collected in ambulatory conditions by one out-patient of Lund University Hospital in CSII (Continuous Subcutaneous Insulin Infusion) therapy with Medtronic MiniMed (Medtronic Inc. (2008)), that is, a portable pump is used to continuously infuse basal insulin into the subcutaneous tissue, together with a personal diary containing information on food intake and short-acting insulin dosages.

The glucose sensor was inserted into the skin and measured the glucose level in the interstitial fluid every 5 minutes, for a total of 288 readings per day. To calibrate the system a minimum of 2 fingerprick measurements were requested each day (every 12 hours). Meal time, food composition and quantity of the different nutrients were noted in the diary, as well as time and size of insulin injections. Raw data can be seen in Fig. 2 and were obtained from 13:00 on Day 1 to 12:00 on Day 3. Impulse-type inputs were filtered according to 2.2 and 2.3 in order to model the gut glucose absorption and the subcutaneous-to-intravenous insulin absorption, respectively (Fig. 3), the benefits of this procedure being shown in (Finan et al. 2006).

The data set was divided into three segments, each representing approximately one day. One segment was used for identification and the other segments for validation.

2.2 Glucose Intestinal Absorption Modeling

The rate of appearance of glucose in the blood stream following a meal is described by a three compartment model based upon Model 2 in (Dalla Man et al. 2006). The overall model represents the digestion of food in the gastro-intestinal tract and considers the following model equations:

$$
\begin{align*}
q_{sto1}(t) &= -k_{sto1} \cdot q_{sto1}(t) + D \cdot \delta(t) \\
q_{sto2}(t) &= -k_{empt} \cdot q_{sto2}(t) + k_{sto1} \cdot q_{sto1}(t) \\
q_{gut}(t) &= -k_{abs} \cdot q_{gut}(t) + k_{empt} \cdot q_{sto2}(t) \\
R_c(t) &= f \cdot k_{abs} \cdot q_{gut}(t) / m_b
\end{align*}
$$

(1)

Fig. 2. Patient raw data: carbohydrate intake and GI of the meal (top) and short-acting insulin doses (center) as noted in the diary, plasma glucose level by the CGMS vs fingerpricks (bottom)

Fig. 3. Filtered data: glucose rate of appearance in the gut (top), insulin flux from subcutaneous depots (center) and plasma glucose concentration (bottom). Notice that the sampling time is 5 min

Fig. 4. Rate of Appearance of glucose in plasma after a meal challenge consisting in 60 g CHO with different Glycemic Index (solid GI = 50, thick GI = 80, dotted GI = 140). Simulation was performed with the proposed meal model
where $q_{sto1}$ [mg] and $q_{sto2}$ [mg] are the amounts of carbohydrates in the stomach (solid and liquid phase, respectively). $D(t)$ [mg/min] is the amount of ingested carbohydrates, $q_{got}$ [mg] is the carbohydrate mass in the intestine, $k_{empt}$ is the rate of grinding, $m_b$ [Kg] is the subject’s body weight, $k_{abs}$ is the rate of absorption and $f$ the fraction of intestinal absorption that actually appears in plasma. The model incorporates a non-linear term by letting the gas-
tric emptying rate be non-linearly dependent on the amount of carbohydrates in the stomach according to the following relationship:

$$k_{empt}(q_{sto}) = k_{min} + k \cdot \tanh(\alpha(q_{sto} - b \cdot D))$$

where

$$k = \frac{k_{max} - k_{min}}{2}, \quad \alpha = \frac{5}{2 \cdot D \cdot (1 - b)}, \quad \beta = \frac{5}{2 \cdot D \cdot c}$$

In the current work the above model was modified, making the parameters $b$, $c$, $k_{max}$ and $k_{abs}$ be dependent upon the glycomic index of the meal:

$b = 0.089 \cdot \log(GI) + 0.43$
$c = 0.3 \cdot \log(GI) - 0.61$
$k_{max} = 0.02 \cdot \log D - 0.013$
$k_{abs} = 0.013 \cdot e^{0.02 GI}$

Mean population values reported in (Dalla Man et al. 2006) were instead adopted for the parameters $k_{max}$, $k_{min}$ and $f$ (Table 1). Figure 4 shows the rate of appearance of glucose in plasma following meals of the same size with different compositions simulated with the proposed model.

Table 1. Mean population parameters values adopted in the glucose intestinal absorption model and in the insulin kinetics model, respectively

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{max}$</td>
<td>0.03</td>
</tr>
<tr>
<td>$k_{min}$</td>
<td>0.006</td>
</tr>
<tr>
<td>$f$</td>
<td>0.86</td>
</tr>
<tr>
<td>$k_{01}$</td>
<td>1.12</td>
</tr>
<tr>
<td>$k_{02}$</td>
<td>2.1</td>
</tr>
<tr>
<td>$k_c$</td>
<td>1.09</td>
</tr>
<tr>
<td>$k$</td>
<td>0.67</td>
</tr>
<tr>
<td>$B$</td>
<td>1.55</td>
</tr>
</tbody>
</table>

### 2.3 Insulin Kinetics

Insulin kinetics from subcutaneous depots to blood was represented by the four-compartment in (Wilinski et al. 2005; Model 9). Two channels were used to describe insulin in the s.c. tissue: one to cope with the delay in insulin absorption and one to represent faster absorption. The reader is referred to the paper for the model equations. Nominal parameter values for $k_{a1}$, $k_{a2}$, $k_c$, $k$, $B$ are reported in Table 1.

### 2.4 Subspace-based model identification

As already mentioned, the approach considered for modeling is subspace-based system identification of linear state space models. The dynamic of the glucose-insulin subsystem of each day was modeled as a discrete-time, time invariant linear model in state space form with exogenous inputs. The inputs $u(t) \in \mathbb{R}^2$ of the model are the glucose flux from the gut and the insulin flux from the subcutaneous depots, and the output $y(t) \in \mathbb{R}$ is the plasma glucose concentration, all obtained according to the procedure outlined in Sec. 2.

Assuming that data $\{y_k, u_k\}, k = 1 \ldots N$ are available it is possible to represents the dynamical model in the following innovations form:

$$\begin{align*}
x_{k+1} &= Ax_k + Bu_k + Kw_k \\
y_k &= Cx_k + Du_k + w_k
\end{align*}$$

reducing with $n$ the dimension of the state-space, $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times 2}$, $C \in \mathbb{R}^{1 \times n}$, $D \in \mathbb{R}^{1 \times 2}$, $K \in \mathbb{R}^{n \times 1}$ and $\{w_k\}$ is the innovations process—i.e., the one-step-ahead prediction error, which is a zero mean white noise. In particular two toolboxes were used: the Matlab® System Identification Toolbox n4sid.m routine (Van Overschee and De Moor 1994) and the SM1 Toolbox (Haverkamp and Verhaegen 1997). The general methodology consisted of identifying state space models of various orders, namely 3, 4, 5, 6 and then quantifying the ability of the models to capture the circadian behaviour of the glucose-insulin system for the calibration data with respect to some sort of metrics. There exists quite a big wealth of statistical indices in the literature. The metrics chosen were the variance of the simulation error, that is $\mathbb{E}[(y_k - \hat{y}_k)^2(t)]$, $\hat{y}_k = y_k - \hat{y}_k$, where the subscript index stands for the simulated signals using the estimated initial conditions (initial conditions were estimated ex-novo for each set of data), and the variance-accounted-for, that is

$$\text{VAF} = 1 - \frac{\mathbb{E}[(y_k - \hat{y}_k)(y_k - \hat{y}_k)^T]}{\mathbb{E}[y_k^T y_k]}$$

However, mathematical metrics may sound promising but not necessarily meaningful to a clinician. Indeed, for this particular application there is a need of quantifying the performance in terms of number and severity of hypo- and hyperglycemic events that can be detected. Thus, in addition to numeric coefficients also qualitative assessments were applied. Pole-zero maps of the estimated systems were drawn and the cancelations were penalized in order to avoid overparametrization. Stability and controllability of each of the previously estimated models were guaranteed.

After the calibration procedure of obtaining the optimal model structure and model parameters for each of the day, the best configuration (in bold cases in Tables 2, 3, 4) was validated against all other days. Infinite-step-ahead model predictions—i.e., the future output is calculated based upon the input signal only—were evaluated.

### 2.5 Look-ahead Predictors

In order to provide short-term predictions of blood glucose evolution two strategies were followed: Firstly, short-term predictors—e.g., Kalman estimators and ARX predictors—based on the identified models and direct identification were designed.
\[ x_{k+1} = (A - KC)x_k + Bu_k + Ky_k, \quad x_k \in \mathbb{R}^n \quad (6) \]
\[ \hat{y}_{k+d|k} = C_d x_k + D_d u_k + D_k y_k \quad (7) \]
\[ \varepsilon_{k|k-d} = y_k - \hat{y}_{k|k-d} \quad (8) \]

Predictions of the blood glucose level up to 30 minutes were made on validation data. Secondly, predictors were found by means of direct least-squares predictor identification of Eqs. (4-5) and subspace-based identification of Eqs. (4-8).

3. RESULTS

Modeling results are listed in Tables 2-4. The variance of the simulation error was smaller for the models obtained with the PO – MOESP-type of algorithms than the ones obtained with the n4sid in all the examined cases but two, in bold cases in Tables 3 and 4, respectively. No intuitive explanation may be attributed to this fact. The correctness of the models was further verified by comparing the real output with the estimated output from the model, quantified by the VAF metric. According to it, improvements were registered with SMI models. Figures 6 and 7 depict the simulation results for each of the days considered on calibration data with the previously selected models.

Cross validation was also made, testing the models on consecutive days but with the same initial conditions applied before. It turned out that not every model was capable to capture the evolution of the plasma glucose concentration over the 24 hours. The best result achieved with cross validation can be seen in Figs. 8-9.

Short-horizon look ahead predictors were tested on calibration data (Fig. 5) as well as on validation data (Figs. 8-9).

![Fig. 5. Day 1 CGMS evolution: real (thick line) vs 30 min ahead prediction using the LS approach (dotted red line) and the 3rd-order model obtained with n4sid (solid black line), respectively. Notice that only calibration data were used.](image)

![Fig. 6. Day 2 CGMS profile: real (thick line) vs simulated using the 4th-order model obtained with n4sid (solid thin line) and 4th-order model obtained with SMI (dotted line), respectively](image)

Table 2. Variance of the simulation error and variance-accounted-for for different estimated models of Day 1

<table>
<thead>
<tr>
<th>Order</th>
<th>Metrics</th>
<th>n4sid</th>
<th>SMI</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>(\sigma^2(e_{\text{sim}}))</td>
<td>0.9445</td>
<td>0.6556</td>
</tr>
<tr>
<td></td>
<td>VAF</td>
<td>93.2582</td>
<td>95.3201</td>
</tr>
<tr>
<td>4</td>
<td>(\sigma^2(e_{\text{sim}}))</td>
<td>4.8229</td>
<td>0.6762</td>
</tr>
<tr>
<td></td>
<td>VAF</td>
<td>65.5737</td>
<td>95.1732</td>
</tr>
<tr>
<td>5</td>
<td>(\sigma^2(e_{\text{sim}}))</td>
<td>1.0181</td>
<td>0.4202</td>
</tr>
<tr>
<td></td>
<td>VAF</td>
<td>92.7331</td>
<td>97.007</td>
</tr>
<tr>
<td>6</td>
<td>(\sigma^2(e_{\text{sim}}))</td>
<td>0.8862</td>
<td>0.6097</td>
</tr>
<tr>
<td></td>
<td>VAF</td>
<td>93.6742</td>
<td>95.6477</td>
</tr>
</tbody>
</table>

Table 3. Variance of the simulation error and variance-accounted-for for different estimated models of Day 2

<table>
<thead>
<tr>
<th>Order</th>
<th>Metrics</th>
<th>n4sid</th>
<th>SMI</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>(\sigma^2(e_{\text{sim}}))</td>
<td>1.6753</td>
<td>0.9299</td>
</tr>
<tr>
<td></td>
<td>VAF</td>
<td>84.9467</td>
<td>91.6446</td>
</tr>
<tr>
<td>4</td>
<td>(\sigma^2(e_{\text{sim}}))</td>
<td>0.1149</td>
<td>0.1685</td>
</tr>
<tr>
<td></td>
<td>VAF</td>
<td>98.9675</td>
<td>98.4857</td>
</tr>
<tr>
<td>5</td>
<td>(\sigma^2(e_{\text{sim}}))</td>
<td>1.1914</td>
<td>1.0583</td>
</tr>
<tr>
<td></td>
<td>VAF</td>
<td>89.2942</td>
<td>90.4905</td>
</tr>
<tr>
<td>6</td>
<td>(\sigma^2(e_{\text{sim}}))</td>
<td>2.8085</td>
<td>0.3017</td>
</tr>
<tr>
<td></td>
<td>VAF</td>
<td>74.7600</td>
<td>97.2888</td>
</tr>
</tbody>
</table>

Table 4. Variance of the simulation error and variance-accounted-for for different estimated models of Day 3

<table>
<thead>
<tr>
<th>Order</th>
<th>Metrics</th>
<th>n4sid</th>
<th>SMI</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>(\sigma^2(e_{\text{sim}}))</td>
<td>0.7578</td>
<td>0.5071</td>
</tr>
<tr>
<td></td>
<td>VAF</td>
<td>83.007</td>
<td>88.6252</td>
</tr>
<tr>
<td>4</td>
<td>(\sigma^2(e_{\text{sim}}))</td>
<td>0.1423</td>
<td>0.1738</td>
</tr>
<tr>
<td></td>
<td>VAF</td>
<td>96.8068</td>
<td>96.1020</td>
</tr>
<tr>
<td>5</td>
<td>(\sigma^2(e_{\text{sim}}))</td>
<td>0.1288</td>
<td>0.4549</td>
</tr>
<tr>
<td></td>
<td>VAF</td>
<td>97.113</td>
<td>89.7961</td>
</tr>
<tr>
<td>6</td>
<td>(\sigma^2(e_{\text{sim}}))</td>
<td>0.1022</td>
<td>0.1539</td>
</tr>
<tr>
<td></td>
<td>VAF</td>
<td>97.7084</td>
<td>96.5692</td>
</tr>
</tbody>
</table>

4. DISCUSSION

The present paper focused on the problem of modeling blood glucose-insulin dynamics in diabetes. Subspace-based identification techniques were used, the reason being that such methods are based on robust, non-iterative and numerically efficient linear algebra tools which, contrary to other methods based on the optimization of some cost function (e.g., prediction-error methods in (Johansson 1993, Ljung 1999, Söderström and Stoica 1989) do not require performing costly iterative minimization thus also avoiding the risk of getting stuck in local minima—e.g., (Haverkamp 2000), (Johansson 1993). Many difficulties were encountered, mainly involving data both quantitatively and qualitatively. Indeed the records were collected by one single subject during three sample days giving rise to the question of validity of results obtained. The patient chosen was selected out of a database containing 50 patient records based on the following criteria:

- Linear regression
- No severe hypoglycemia
- Adequate glycemic control
• Insulin pump patient: By choosing a pump patient the risk of interference of the slow acting insulin profile in the estimate of the fast acting insulin is eliminated.

• Well documented food intake estimates: Quantifying the amount of ingested carbohydrates in a meal is of key importance for correct meal modeling.

• Some meals and insulin injections were separated in the data, i.e., meals taken without insulin and conversely: Since the effect of the meal intake and the insulin injec-

• No major flaws in the CGMS data record, such as missing data segments or major outliers due to poor calibration.

As predictable, the data collected offered poor model input excitation despite the careful selection of the subject, because of the correlation between food intake and consequent insulin injection, the consequence being that estimated models have too low complexity and are inadequate for cross validation in all the possible cases. It is still an open issue in this field 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 clinic events.

Interstitium-to-plasma glucose dynamics was disregarded, the reason being that to the best of the authors knowledge such a problem has not been solved yet, deserving future attention. Hence, CGMS measurements were considered to represent plasma glucose level.

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. Using the diary and nutrition tables the glycemic index (GI) of each meal (see e.g. (Brouns et al. 2005)) was estimated. Hence, considering that lower GI nutrients prolong the rate

Fig. 7. Day 3 CGMS profile: real (thick line) vs simulated using the 6th-order model obtained with n4sid (solid thin line) and 6th-order model obtained with SMI (dotted line), respectively.

Fig. 8. CGMS evolution in Day 2 (thick line) vs. infinite-step-ahead prediction based upon the model obtained in Day 1 with n4sid (solid thin line) and SMI (dotted line), respectively.

Fig. 9. CGMS evolution in Day 1 (thick line) vs. infinite-step-ahead prediction based upon the model obtained in Day 2 with n4sid (solid thin line) and SMI (dotted line), respectively.

Fig. 10. Day 2 CGMS evolution (thick line) vs. 30 min ahead prediction based upon the model of Day 1 obtained with n4sid (magenta solid line) and SMI (magenta dotted line). Confidence interval ±σ(ε_{pred}) (black dashed line).

Fig. 11. Day 3 CGMS evolution (thick line) vs. 30 min ahead prediction based upon the model of Day 2 obtained with n4sid (magenta solid line) and SMI (magenta dotted line). Confidence interval ±σ(ε_{pred}) (black dashed line).
of appearance, the physiological model in (Dalla Man et al. 2006) was modified, letting the parameters $b$, $c$, $k_{max}$ and $k_{abs}$ be subject to the GI of the meal.

Population mean values reported in the literature were used for the parameters appearing in the meal model and in the insulin kinetics model. Given the inter-personal variability, adaptation of the parameters to the subject would be more effective.

It is well known by clinicians that under the same experimental conditions it is possible to recognize the same pattern on a subject among consecutive days, which may lead to the conclusion of time-invariant dynamics over the days. This was tested by performing infinite-step-ahead model output prediction. However, the result is still unacceptable. One hypothesis is that the limited accuracy of the infinite-step-ahead prediction can be understood as effect of input correlation and unaccounted input, e.g., physical activity.

As for the identified models, it turned out that the n4sid technique resulted in models sensitive to the model order and error-prone in cross validation of infinite-step-ahead predictions. Although poor in cross validation, the behaviour of the SMI models may indicate the superiority of this identification method compared to the n4sid.

5. CONCLUSIONS AND FUTURE WORK

This paper dealt with modeling and short-time prediction of diabetic blood glucose measurements using linear models based on subspace-based system identification. No attention was paid to the physiology of the glucose-insulin interaction and it is therefore postponed to future work. Models of low complexity were capable of describing blood glucose evolution over 24 hours, detecting even abrupt changes, in calibration data but they were not sufficient for validation data. Blood glucose levels could be predicted with a confidence interval of ±1 standard deviation of the prediction error up to 30 minutes ahead on validation data.

Records of one patient only were used: future investigations will be made on a larger population, including non-pump therapy subjects, clarifying whether the identification methodology is generally applicable or not.

Poor model input excitation and input signal correlation were reported giving rise to the issue of ill-conditioning of the estimates. Further work is, thus, needed to investigate optimal experimental conditions and protocols in order to obtain data suitable for identification purposes without contributing to higher patient risk.

Regarding meal modeling, only the carbohydrate intake was considered. Hence, the development of new compartment model accounting also for proteins, fat and fiber content of a meal is needed. 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 in the future.

ACKNOWLEDGEMENTS

As a continuation, the DIAdvisor™ project within the European FP7-ICT program will pursue research on prediction and predictive control of blood glucose concentration.

REFERENCES


