

# 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

## 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 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  $\hat{\mathbf{y}}_k$ .

$$\hat{y}_k^e = \mathbf{w}_k^T \hat{\mathbf{y}}_k \quad (1)$$

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

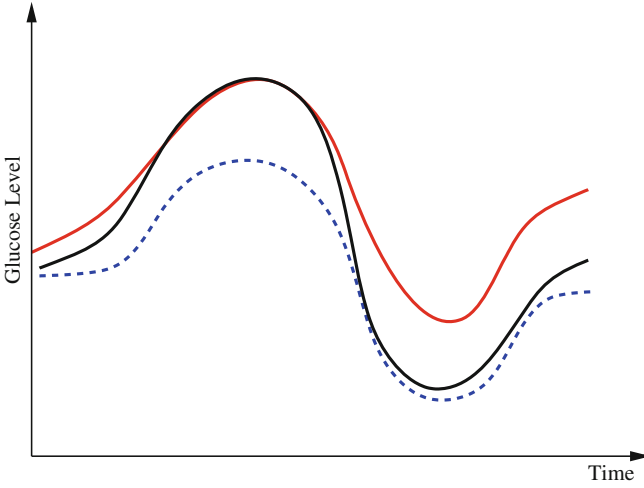
As seen from the review above, many different approaches to glucose modeling and predicting have been established. These methods may each be more suitable to specific conditions for the glucose dynamics, and improvements in robustness and prediction performance may be achieved by combining their outcomes, 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  $t_k \in \{1, 2, \dots\}$ . 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  $S_i \in S = [1, \dots, n]$ , 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,  $\hat{y}_{k+q|k}^j, j \in \{1, \dots, m\}$  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  $\hat{y}_{k+q|k}^e$  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:

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

where  $p(y_{k+q}|D_k)$  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:

$$\hat{y}_{k+q}^e = \mathbb{E}(y_{k+q}|D_k) \quad (3)$$

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

$$= \mathbf{w}_k^T \hat{\mathbf{y}}_k \quad (5)$$

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

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

where  $\hat{y}_{k+q}^e$  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  $\theta_k \in \Theta = [1, \dots, p]$ .

$$p(M_i|D_k) = \sum_j p(M_i|\theta_k = j, D_k)p(\theta_k = j|D_k) \quad (8)$$

or in matrix notation

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

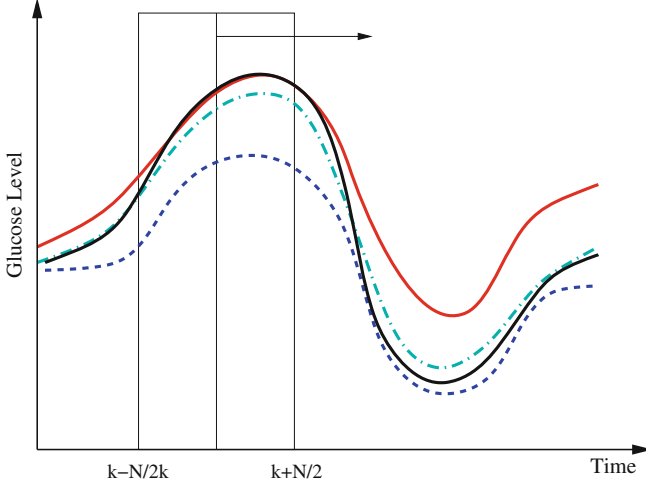
Here,  $\Theta$  represents a *predictor mode* in a similar sense to the dynamic mode  $S$ , and likewise  $\theta_k$  represents the prediction mode at time  $k$ .  $\mathbf{p}(\mathbf{w}_k|\theta_k = j)$  is a column vector of the joint prior distribution of the conditional weights of each predictor 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:

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

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  $\mathcal{L}(y, \hat{y})$  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  $\theta_{k-1}$ . For reasons soon explained, only  $\mathbf{w}_k|\theta_{k-1}$  is thus calculated:

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



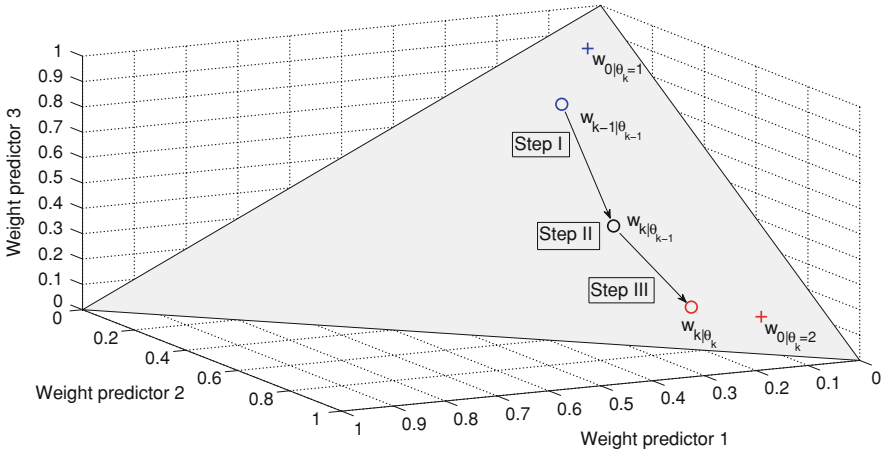
**Fig. 2** Principle of finding the predictor modes for unlabelled data over the training data set time period  $T$ . For every time point  $t_k \in T$ , the optimal  $\mathbf{w}_k$  is determined by Eq. (10), where the optimal prediction  $\mathbf{w}_k \hat{\mathbf{y}}$  (light green dash-dotted curve) 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  $\mathcal{L}$  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  $\mathbf{w}_{0|\theta=i}$ ,  $i = [1, \dots, p]$

Here,  $\mu_j$  is a forgetting factor, and  $A_{\theta_k=i}$  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  $\mathbf{p}(\mathbf{w}_{k|\theta_k})$  estimated from the data retrieved from Eq. (10) above. The  $\mathbf{p}(\mathbf{w}_{k|\theta_k})$  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  $\theta_{k-1}$  (in a sense that  $\mathbf{p}(\mathbf{w}_{k|\theta_{k-1}})$  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:

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

where

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



**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  $p(w_k|\theta_k = i)$ ,  $i = [1, \dots, p]$  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  $\lambda$  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  $w_{k|\theta_k}$  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  $\theta_k$  is active; (9) reduces to  $\mathbf{p}(w_k) = \mathbf{p}(w_{k|\theta_k})$ . 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  $\mu$ , 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.

$A_{\theta_k=i}$  should also be chosen, such that a sound balance between flexibility and robustness is found, i.e., a too small  $\|A_{\theta_k=i}\|_2$  may result in over-switching, whereas a too large  $\|A_{\theta_k=i}\|_2$  will give a stiff and inflexible predictor. Furthermore,  $A_{\theta_k=i}$  should force the weights to move within the perimeter defined by  $\mathbf{p}(w|\theta_k = i)$ . This is approximately accomplished by setting  $A_{\theta_k=i}$  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  $\mu$  can be found by evaluating different choices for some test data. However, from our experience we have seen that  $N = 10\text{--}20$  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  $\theta_k = 0 : p(\mathbf{w}_k|\theta_k = \mathbf{0}) \in N(\mathbf{1}/m, \mathbf{I})$  is added to the set of predictor modes.

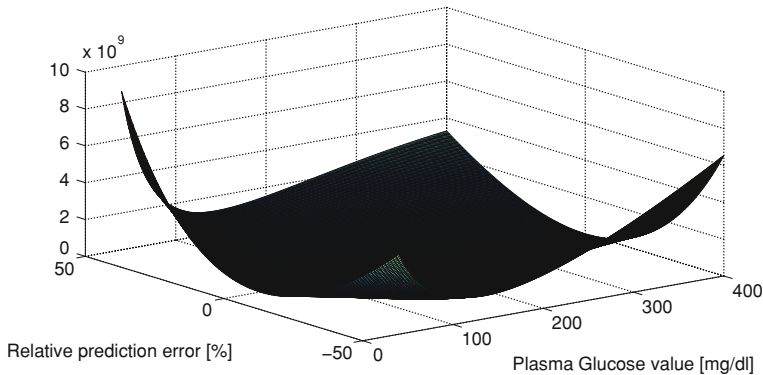
#### Summary of algorithm

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  $\{\mathbf{w}_{k|\theta=i}\}_{T_{S_i}}, \forall i \in \{1, \dots, n\}$ .
3. Classify different predictor modes, and determine density functions  $\mathbf{p}(\mathbf{w}_{k|\theta=i})$  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 $\mathcal{L}$

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.

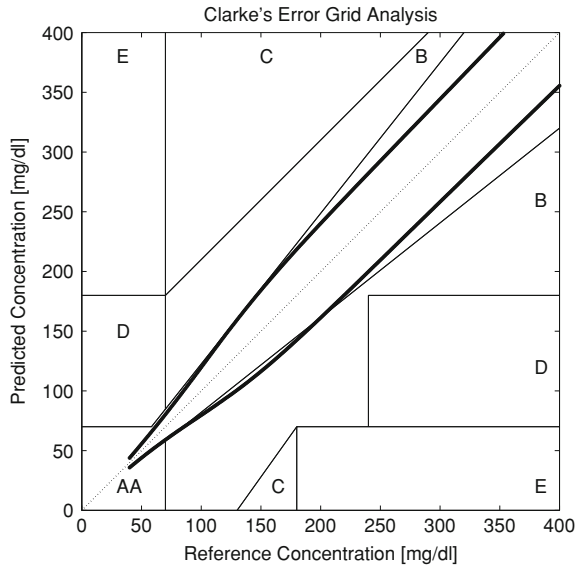
In Fig. 4 the cost function can be seen, plotted against relative prediction error and absolute blood glucose value.

#### ***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.

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

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

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

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

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

$$m_3 = \frac{H_{Eb} \cdot m_1}{1 - H_{Eb}} \quad (19)$$

$$m_4 = 0.4 \frac{C_L}{V_i \cdot M_{BW}} \quad (20)$$

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:

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

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

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

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

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

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

**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.  $k_{empr}$  is a nonlinear function of  $q_{sto}$  and  $C(t)$ :

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

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

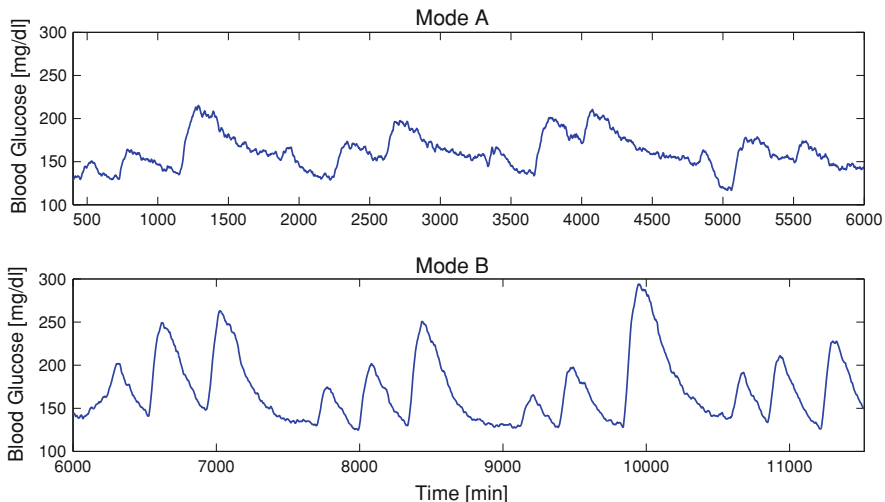
With  $k = (k_{max} - k_{min})/2$ ,  $\alpha = 5/2D(1b)$ ,  $\beta = 5/2Dc$  with parameters  $k_{max}$ ,  $k_{min}$ ,  $b$ , and  $d$

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  $t_k \in (5, 10, \dots)$  min. Short-term predictions,  $p$  steps ahead, were evaluated using the Kalman filter:

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

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

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

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

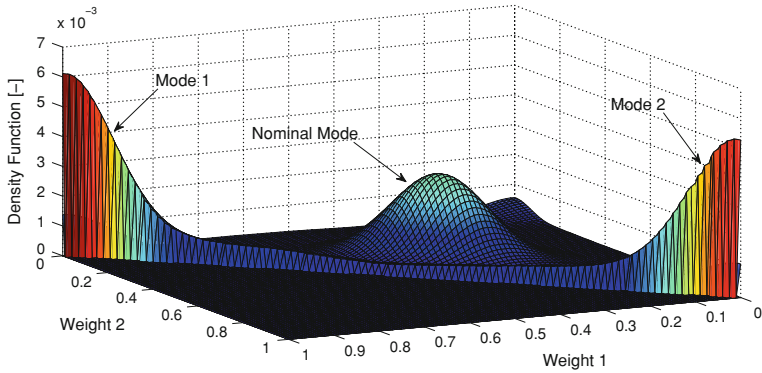
### 5.3.1 Training the Mode Switcher

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  $\mathbf{w}_{01} = [0.96, 0.03, 0.01]$  and  $\mathbf{w}_{02} = [0.03, 0.96, 0.01]$  defining the expected weights for each predictor mode. The nominal mode probability density function was set to  $N(\frac{1}{3}\frac{1}{3}, \mathbf{0.1I})$ . In Fig. 7 all density functions, including the nominal mode, projected onto the  $(w_1, w_2)$ -plane, can be seen together.

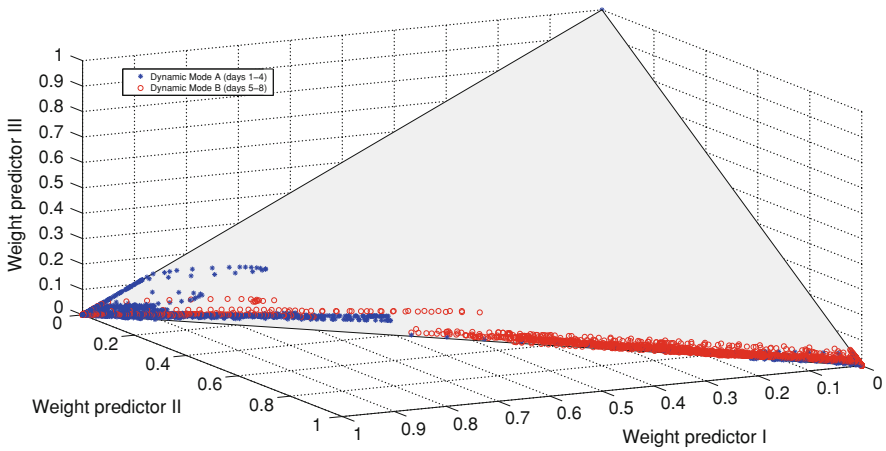
### 5.3.2 Ensemble Prediction Versus Individual Predictions

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.

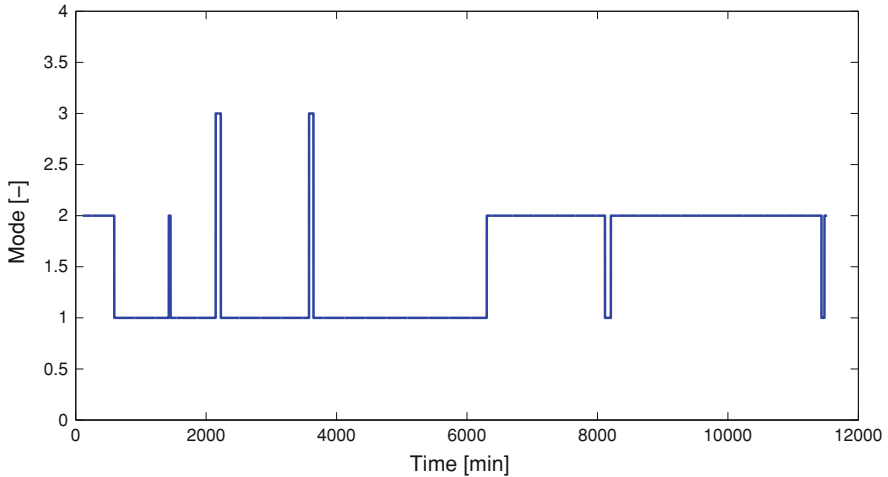


**Fig. 7** Estimated probability density functions for the weights in the training data, including nominal mode. Example I: UVa/Padova model



**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.

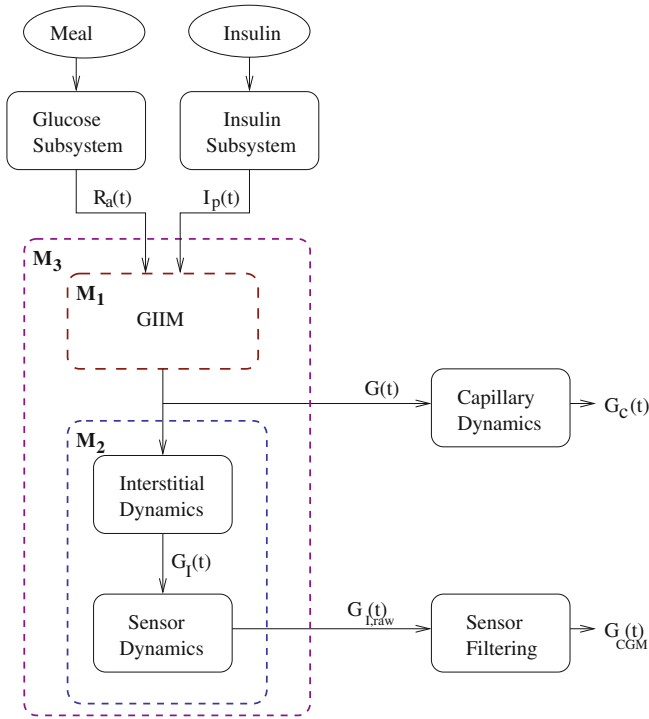
The first trial data (DAQ) were used to train the individual predictor models. The second and third trial data (DIAdvisor I.B and C) were used to train and cross-validate the SW-BMA, i.e., the SW-BMA was trained on B data and validated on C data, and vice versa.

## 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  $G_{I,raw}(t)$  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).

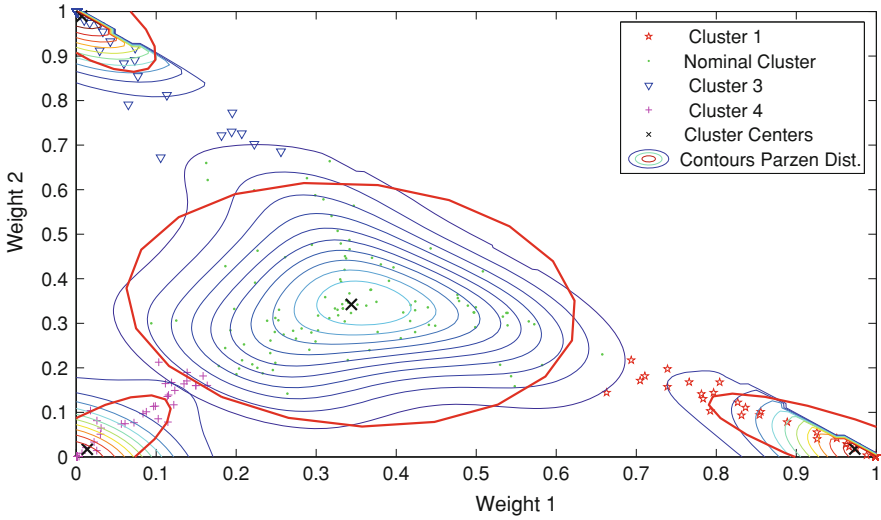
#### 6.5.2 Feature Selection

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.

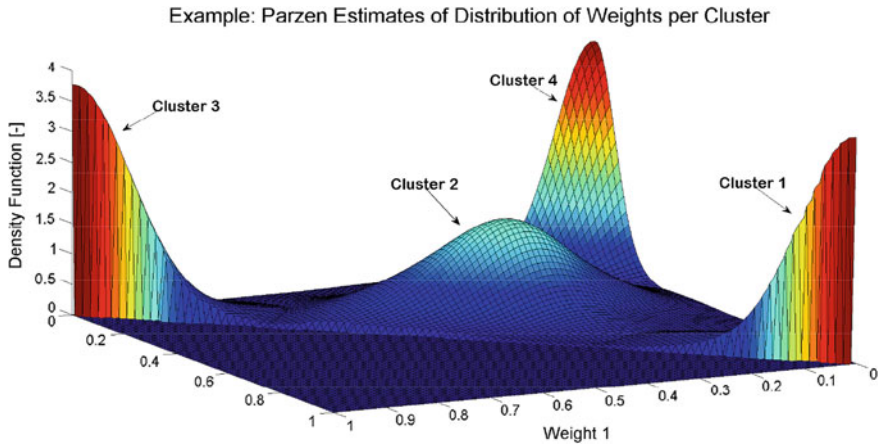
When extracting the features, meal timing and content were considered to be known 30 min before the meal.

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

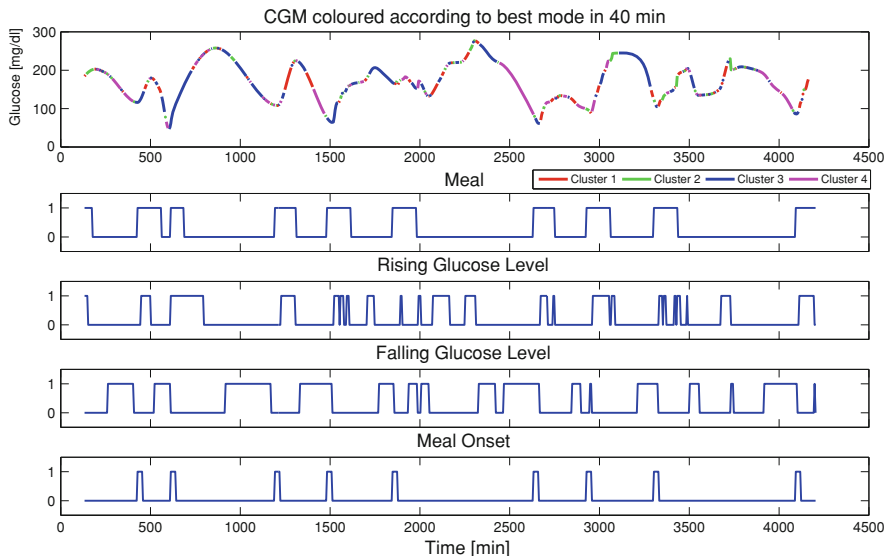


**Fig. 11** Example of distribution of weights in the training data by (10) and clusters given by the k-means algorithm. The red ellipses represent the fitted Gaussian covariances of each cluster (patient 0103, Trial B). Example II: DIAdvisor Data



**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,  $p(\theta_k = i | D_k) = p(\theta_k = i | f_k^*)$  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

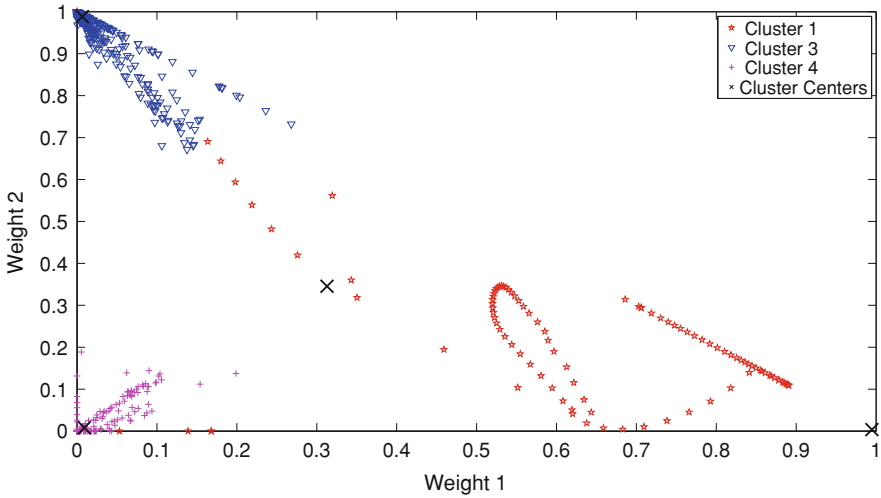
Feature	Threshold	Priority
Meal	$\max (Ra_k, \dots, Ra_{k+30}) > \varepsilon$	1
Rising G	$\text{mean} (\Delta G_{k-10}, \dots, \Delta G_k) > 30 \text{ mg}/(\text{dl}\cdot\text{h})$	2
Falling G	$\text{mean} (\Delta G_{k-10}, \dots, \Delta G_k) < -18 \text{ mg}/(\text{dl}\cdot\text{h})$	3
Meal and rising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

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

### 6.6 Prediction Performance on Test Data

Using the estimated mode clusters  $\{w_{0|i}, R_{0|i}\}$ ,  $i = [1, \dots, M]$ , 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  $\mu$  was set to 0.8 and  $N$  to 20 min. An example of the distribution of the weights  $w_k$  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: DIAadvisor 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 <sub>best</sub> [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 ( $\text{RMSE}/\text{RMSE}_{\text{best}} = 0.75$ ). 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  $\lambda$  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-García J, Martínez-Ramón M, Navia-Vázquez 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](https://doi.org/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](https://doi.org/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](https://doi.org/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, Jovanović 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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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, Jovanović 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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1172/JCI118441)
84. Puckett WR (1992) Dynamic modeling of diabetes mellitus. PhD thesis. University of Wisconsin- Madison
85. Raftery AE, Gneiting T, Balabdaoui F, Pololowski M (2005) Using Bayesian model averaging to calibrate forecast ensembles. *Mon Weather Rev* 133:1155–1174
86. Raftery AE, Kárný M, Ettl 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](https://doi.org/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 absorption in the gut. *Med Inform* 22(1):35–45
106. Zecchin C, Fachinetti 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](https://doi.org/10.1109/IEMBS.2011.6091368)