LUND UNIVERSITY

Bayesian Combination of Multiple Plasma Glucose Predictors

Ståhl, Fredrik; Johansson, Rolf; Renard, Eric

[Link to publication](#)

Bayesian Combination of Multiple Plasma Glucose Predictors

F. Ståhl, R. Johansson and Eric Renard

Abstract—**This paper presents a novel on-line approach of merging multiple different predictors of plasma glucose into a single optimized prediction. Various different predictors are merged by recursive weighting into a single prediction using regularized optimization. The approach is evaluated on 12 data sets of type I diabetes data, using three parallel predictors. The performance of the combined prediction is better, or in par, with the best predictor for each evaluated data set. The results suggest that the outlined method could be a suitable way to improve prediction performance when using multiple predictors, or as a means to reduce the risk associated with definite a priori model selection.**

I. INTRODUCTION

Type I Diabetes Mellitus (IDDM) is a chronic metabolic disease characterized by impaired plasma glucose regulation. Maintaining normoglycemia is crucial in order to avoid both immediate and long-term complications, and to this end, several insulin injection are taken daily by pen injection devices or using an insulin pump. Therapy decision are determined based on-the-spot information provided by infrequent blood (plasma) glucose (BG) measurements, and more recently also by the so-called Continuous Glucose Measurement Systems (CGM) providing interstitial glucose values every 5 or 10 min, together with the patient's and/or the care provider's knowledge and understanding of the patient-specific glucose dynamics.

To further facilitate the therapy outcome, numerous predictive models of plasma glucose dynamics have been suggested in the literature [19], [3], both for the purpose of closed-loop insulin pump based control in a MPC framework [6], and as a means to decision support in itself [24]. These predictors are based on different methods and data; purely empirical [14], derived from physiological models [9] or a combination thereof [7], [10], and designed for, and validated on different usage scenarios. Selecting among the different predictors a priori is a challenging task, and considering the complex dynamics of plasma glucose metabolism, there is good reason to believe that different predictors may be experts in specific conditions, and that no single model will be able to fully capture the dynamics alone. To overcome the problem of a priori model selection, and to take advantage of the different levels of expertise, merging of multiple predictors on an on-line basis is an interesting option.

F. Ståhl and R. Johansson are with Dept. Automatic Control, Lund University, PO Box 118, SE22100 Lund Sweden, Fredrik.Stahl@control.lth.se, Phone +4646 222 87 84, Fax +4646 138118.

Eric Renard is with Department of Endocrinology, Diabetes and Nutrition,and INSERM CIC 1001, Montpellier University Hospital, Montpellier, France

Fusion of models for the purpose of prediction was developed in different research communities, such as in the meteorological and econometric communities where regression-oriented ensemble prediction has been a vivid research area since the late '60s, see e.g. [25] and [11].

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 developed such as the bagging, boosting [5] and weighted majority [20] algorithms, and on-line versions of these [23], [17].

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

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

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 [12].

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. [15], [22]. 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 [1].

Combinations of specifically adaptive filters has also stirred some interest in the signal processing community. Typically, filters with different update rate are merged to benefit from each filter's specific change responsiveness respectively steady state behavior [2].

Finally, in fuzzy modeling, soft switching between multiple models is offered using fuzzy membership rules in the Takagi-Sugeno systems [27].

This paper presents a novel approach combining elements from both the switching and averaging techniques above, forming a 'soft' switcher in a Bayesian framework. The paper is organized as follows; in Sec. II the problem formulation is presented, Sec. III describes the algorithm, Sec. IV discusses

suitable cost functions, Sec. V presents the data obtained in the DIAdvisor project and the evaluation metrics, Sec. VI gives the results, which are discussed in Sec. VII, and Sec. VIII concludes the paper.

II. PROBLEM FORMULATION

CGM measurements $y_k \in \mathbb{R}$, used as a proxy for plasma glucose, and additional data u_k are sampled at a fixed rate, together forming the data z_k, at time $t_k \in \{1, 2, ...\}$. The data u_k provide information about meal and insulin intake.

Given m number of expert p-steps-ahead predictions at time t_k, $\hat{y}_{k+p|k}^j, j \in \{1, ..m\}$, each utilizing different methods, and/or different training sets; how is a Clarke Error [8] optimal p-steps-ahead prediction $\hat{y}_{k+p|k}^e$ of the plasma glucose y_k determined?

III. 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 [27], [2] to theory based approaches, e.g. [16]. Specifically, in a Bayesian Model Averaging framework [16], which will be adopted in this paper, the weights are interpreted as partial beliefs in each predictor model M_i, and the merging is formulated as:

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

and, if only point-estimates are available one can e.g. use:

$$\hat{y}_{k+p|k}^e = \mathbb{E}(y_{k+p}|D_k) \quad (3)$$
$$= \sum_i \mathbb{E}(M_i|D_k)\mathbb{E}(y_{k+p}|M_i, D_k) \quad (4)$$
$$= \mathbf{w}_k^T \hat{\mathbf{y}}_k \quad (5)$$
$$\mathbf{w}_k^{(i)} = \mathbb{E}(M_i|D_k) \quad (6)$$
$$p(\mathbf{w}_k^{(i)}) = p(M_i|D_k) \quad (7)$$

Where $\hat{y}_{k+p|k}^e$ is the combined prediction of y_{k+p} using information available at time k, $D_k : \{z_{1:k}\}$ is the data received up until time k, $\mathbf{w}_k^{(i)}$ indicates position i in the weight vector. The conditional probability of predictor M_i can be further expanded by introducing the latent variable Θ_j.

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

or in matrix notation

$$p(\mathbf{w}_k) = \left[p(\mathbf{w}_{k|\theta_k=\Theta_1}) \cdots p(\mathbf{w}_{k|\theta_k=\Theta_M}) \right] \mathbf{p}(\Theta|D_k) \quad (9)$$

Here Θ_j represents a *predictor mode*, and likewise θ_k the prediction mode at time k, $\mathbf{p}(\Theta|D_k)$ is a column vector of $p(\Theta_j|D_k), j = \{1 \ldots M\}$ and $p(\mathbf{w}_{k|\Theta_i})$ is the joint prior distribution of the conditional weights of each predictor model given the predictor mode Θ_i.

Data for estimating the distribution for $p(\mathbf{w}_{k|\Theta_i})$ is given by the following constrained optimization.

$$\{\mathbf{w}_k\}_T = \arg\min \sum_{i=k-N/2}^{k+N/2} \mathscr{L}(y(t_i), \mathbf{w}_k^T \hat{\mathbf{y}}_i), \quad k \in T \quad (10)$$

$$\text{s.t.} \sum_j w_k^{(j)} = 1 \quad (11)$$

Where T represents the time points in the training data, N is the size of the evaluation window and $\mathscr{L}(y, \hat{y})$ is a cost function. Next, cluster analysis is attempted by e.g., using a Gaussian Mixture Model (GMM) or the k-means algorithm, giving M different *predictor mode* clusters. Unless given by the cluster identification, prior distributions can be estimated by the Parzen window method [4], giving mean $\mathbf{w}_{0|\Theta_i} = \mathbb{E}(\mathbf{w}_{k|\Theta_i})$ from these cluster data sets. An alternative to the Parzen approximation is of course to estimate a more parsimoniously parametrized pdf (e.g., Dirichlet) for the extracted data points. Now, in each time step k the $\mathbf{w}_{k|\theta_{k-1}}$ is determined from the sliding window optimization below by taking advantage of the information provided in the cluster pdfs and the posterior $p(M|D_k)$, if provided. First, determine the optimal weight given the N most recent predictions and the predictor mode θ_{k-1}.

$$\mathbf{w}_{k|\theta_{k-1}} = \arg\min \sum_{j=k-N}^{k-1} \mu^{k-j} \mathscr{L}(y_j, \mathbf{w}_{k|\theta_{k-1}}^T \hat{\mathbf{y}}_j) \quad (12)$$

$$+ (\mathbf{w}_{k|\theta_{k-1}} - \mathbf{w}_{0|\theta_{k-1}})\Lambda(\mathbf{w}_{k|\theta_{k-1}} - \mathbf{w}_{0|\theta_{k-1}})^T \quad (13)$$

$$\text{s.t.} \sum_j w_{k|\theta_{k-1}}^{(j)} = 1 \quad (14)$$

Here, μ is a forgetting factor, $\mathbf{w}_{0|\theta_{k-1}}$ is the mode center for the active mode and Λ is a regularization parameter.

The $p(\mathbf{w}_{k|\theta_k})$ 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 θ_{k-1} (in a sense that $p(\mathbf{w}_{k|\theta_{k-1}})$ is very low) it can thus be seen as an indication of that a mode switch has taken place.

Next, a logical test is used to determine if a mode switch has occurred. The predictor mode is switched to mode Θ_i, if:

$$\begin{cases} p(\Theta_i|\mathbf{w}_k, D_k) > \lambda, \text{ and} \\ p(\mathbf{w}_k|\Theta_i, D_k) > \delta \end{cases} \quad (15)$$

where

$$p(\Theta_i|\mathbf{w}_k, D_k) = \frac{p(\mathbf{w}_k|\Theta_i, D_k)p(\Theta_i|D_k)}{\sum_j p(\mathbf{w}_k|\Theta_j, D_k)p(\Theta_j|D_k)} \quad (16)$$

Where a λ somewhat larger than 0.5 gives a hysteresis effect to avoid chattering between modes, and δ assures that non-conclusive situations, evaluated on the outskirts of the probability functions, don't result in switching. Unless otherwise estimated from data, the conditional probability of each prediction mode $p(\Theta_i|D_k)$ is set equal for all possible modes, and thus cancels in (16). The logical test is evaluated using the priors received from the pdf estimate and the $\mathbf{w}_{k|\theta_k}$ received from (14).

Finally, the weights are determined by rerunning (14) with the new mode.

Now, since only one prediction mode θ_k is active; (9) reduces to $p(\mathbf{w}_k) = p(\mathbf{w}_{k|\theta_k})$.

The length N of the evaluation period is, together with the forgetting factor μ, a crucial factor 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.

Λ should also be chosen such that a sound balance between flexibility and robustness is found, i.e., a too small Λ may result in over-switching, whereas a too large Λ will give a stiff and inflexible predictor. Furthermore, Λ should force the weights to move within the perimeter defined by $p(\mathbf{w}|\Theta_i)$. This is approximately accomplished by setting Λ equal to the inverse of the covariance matrix $\mathbf{R}_{\theta_k}$ of an approximative Gaussian distribution of the active cluster distribution.

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

IV. CHOICE OF $\mathcal{L}$

A suitable cost function for determining appropriate weights should take into account that the consequences of acting on too high glucose predictions in the lower BG 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 non-diabetics 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 [18] the selected cost 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. 1 the cost function can be seen plotted against relative prediction error and absolute blood glucose value.

A. 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 [8]. This metric meets the minimum criteria raised earlier. However, other aspects makes it less suitable; no distinction between prediction errors within error zones, instantaneous switches in evaluation score, etc.

In Fig. 2 the isometric cost of the chose cost function for different prediction errors at different BG values has been plotted together with the Clarke Grid Plot. The boundaries
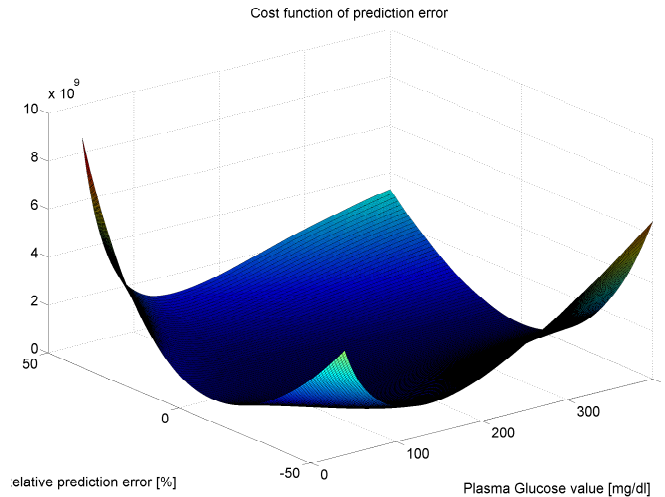

Fig. 1.   Cost function of relative prediction error.

of the A/B/C/D/E areas of the Clarke Grid can be regarded lines of isometric cost according to the Clarke metric. In the figure the isometric cost of the cost function has been chosen to correspond to the lower edge defined by the intersection of the A and B Clarke areas. Thus, the area enveloped by the isometric cost can be regarded as the corresponding A area of this cost function. Apparently it puts much tougher demands both in the lower and upper BG regions in comparison to the Clarke Plot.
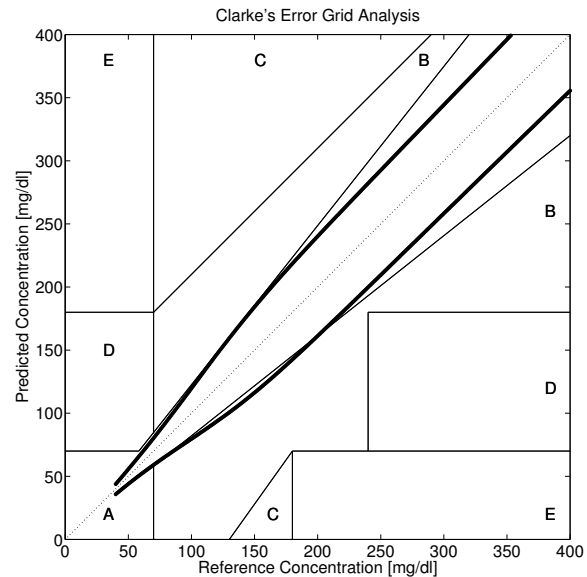

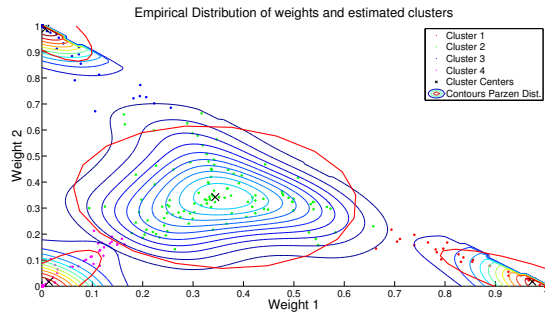Fig. 2.   Isometric cost in comparison to the Clarke Grid.

Fig. 3. Example of distribution of weights in the training data by (11) and clusters given by the k-means algorithm. The red ellipses represent the fitted Gaussian covariances of each cluster. (Patient 0103, Trial B)

Fig. 4. Example of estimated probability density functions for the different predictor mode clusters in the training data. (Patient 0103, Trial B)

V. DATA AND EVALUATION CRITERIA

A. Predictors

Three different predictors of different structure were used in this study; a state-space-based model (SS) (model M_1 in [26]), a recursive ARX model [13] and a kernel-based predictor [21]. The SS and ARX models furthermore utilized inputs generated by population parametrized sub models describing the flux and digestion of insulin and glucose following an insulin injection or meal intake [9]. For further information regarding these predictors and the underlying models the reader is referred to [26], [13], [21].

B. Training and Test data

Data records from three different in-hospital trials conducted at the Montpellier University Hospital have been used. Each trial ran over three days with standardized meals at scheduled hours. No specific intervention on the usual diabetes treatment was undertaken during the studies since a trueful picture of normal blood glucose fluctuation and insulin-glucose interaction was pursued. Apart from notations on meal and insulin administration, glucose was monitored by CGM and frequent plasma glucose measurements (> 40 daily). The BG data was sufficiently rich to allow interpolation (cubic), and was used for evaluation purposes.

A number of patients participated in all three trials. Based on data completeness, six of these have been selected for this study.

The first trial data (DAQ) was 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.

C. Evaluation Criteria

The prediction results were compared to the interpolated BG in terms of Clarke Grid Analysis and the complementary Root Mean Square Error (RMSE).

VI. RESULTS

A. Training the mode switcher

1) Cluster Analysis - Finding the Modes: The three predictors were used to create 40 minute ahead predictions for both
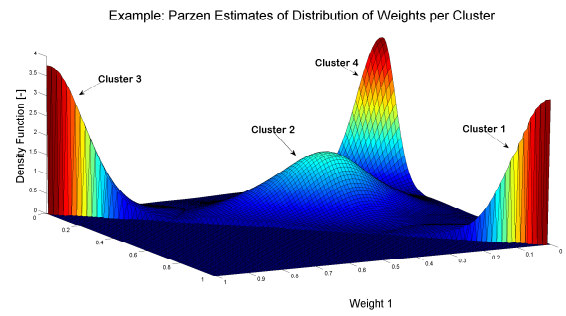
training data sets $D_{T_{B(C)}}$. Using (11) with $N = 20$ the weights $\{\mathbf{w}_k\}_{T_{B(C)}}$ were obtained; example depicted in the (w_1, w_2) plane in Fig. 3. The weights received from the training are easily visually recognized as belonging to different groups (true for all patients, not shown). Attempts were made to find clusters using a 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 $\sum w = 1$. 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 $w_{0|\Theta_i}$.

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. 4. Gaussian distributions were fitted to give the covariance matrices R_{Θ_i} used in (14).

2) Feature selection: The posterior mode probability $p(\Theta|D_k)$ 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. 5). The binary features in Table I were selected.

Meals were considered to be announced 30 minutes before the meal.

From the training data the posterior mode probabilities $p(\Theta_i|f_j)$ 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 $p(\Theta_i)$ was determined by the total ratio of active time per cluster over the entire test period.

The different features are overlapping, and to resolve this issue they were given different priority—only allowing the feature of highest priority, f_k^* to be present at each time step t_k. Thereafter, $p(\Theta|D_k) = p(\Theta|f_k^*)$ is determined. If no feature is active the $p(\Theta|D_k)$ is approximated by the $p(\Theta_i)$ estimate.
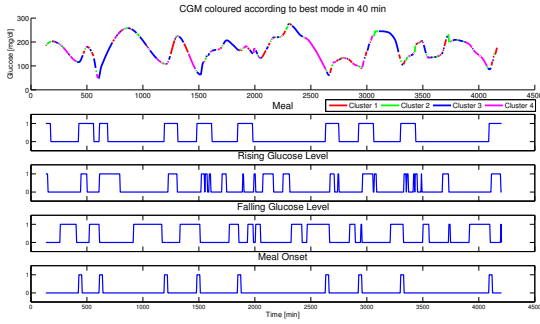
Fig. 5. Example of CGM coloured according to best predictor mode in 40 minutes together with active features. (patient 0103, Trial B)

TABLE I
SELECTED FEATURES. ε CORRESPONDS TO THE MAXIMUM AMPLITUDE OF GLUCOSE RATE-OF-APPEARANCE, Ra AFTER DIGESTING 10 G CHO, AND $\Delta BG = BG_k - BG_{k-5}$

Feature	Threshold	Priority
Meal	$\max(Ra_k, .., Ra_{k+30}) > \varepsilon$	1
Rising BG	mean $(\Delta BG_{k-10}, \dots, \Delta BG_k)$ > 30 mg/(dl· h)	2
Falling BG	mean $(\Delta BG_{k-10}, \dots, \Delta BG_k)$ < -18 mg/(dl· h)	3
Meal and rising BG	See above.	4
Meal Onset	$\max Ra(k-30, ..., k) < \varepsilon$ and $\max Ra(k, ..., k+30) > \varepsilon$	5

B. 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 μ was set to 0.8 and N to 20 minutes. An example of the distribution of the weights $\mathbf{w}_k$ for the three predictors can be seen in Fig. 6.

Table II summarizes a comparison of predictive performance over the different patient test data sets for the RMSE evaluation criteria, and in Table III the evaluation in terms of Clarke Grid Analysis is given. The optimal switching approach, here defined as using the non-causal fitting by Eq.
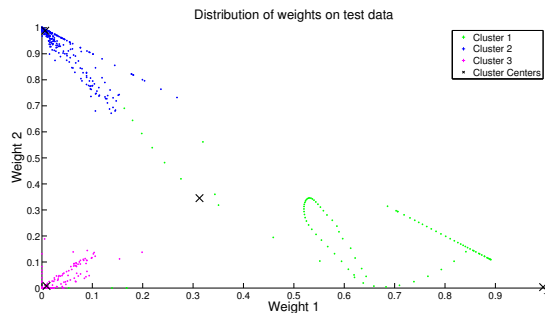

Fig. 6. Example of the distribution of weights in the test data using the estimated clusters and feature correlations. (Patient 0108, Trial B)

(11), is used as a measure of optimal performance of a linear combination of the different predictors.

TABLE II
PERFORMANCE EVALUATION FOR THE 40 MINUTE SW-BMA PREDICTION COMPARED TO THE OPTIMAL SWITCHING AND THE INDIVIDUAL PREDICTORS. THE METRIC IS THE ROOT MEAN SQUARE ERROR (RMSE), NORMALIZED AGAINST THE BEST INDIVIDUAL PREDICTOR $M_1 - M_3$ FOR EACH PATIENT.

Merging Strategy	median $\frac{RMSE}{RMSE_{best}}$ [min-max]	
	Trial B	Trial C
SW-BMA	1.03 [0.75-1.04]	1.03 [0.94-1.05]
Optimal switching	0.97 [0.54-1.0]	0.94 [0.73-1.0]
2:nd 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]

TABLE III
PERFORMANCE EVALUATION FOR THE 40 MINUTE SW-BMA PREDICTION COMPARED TO THE OPTIMAL SWITCHING AND THE BEST INDIVIDUAL PREDICTOR BY THE AMOUNT OF DATA (%) IN THE ACCEPTABLE A/B ZONES VS 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

VII. DISCUSSION

Compared to the individual predictors 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 ($\frac{RMSE}{RMSE_{best}} = 0.75$). However, comparison to the optimal switcher indicates that there is still further room for improvement. To fill this gap, timely switching is most important. A crucial part of the algorithm is thus to select important features with significant correlations to mode switching, in order to improve the likelihood that the best predictor mode is used at each time. Further research is needed to improve this aspect.

VIII. CONCLUSIONS

In this paper a novel merging mechanisms for multiple glucose predictor has been proposed and evaluated on 12 data sets from a clinical trial. The results show that the merged prediction has a predictive performance in comparison with the best individual predictor in each case.

This early assessment indicates that the concept may prove useful when dealing with several individual glucose predictors of uncertain reliability, or as a means to improve predictive quality if the predictions are diverse enough.

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

References

[1] A. Alessandri, M. Baglietto, and G. Battistelli. Receding-horizon estimation for switching discrete-time linear systems. *Automatic Control, IEEE Trans. on*, 50(11):1736 – 1748, Nov. 2005.

[2] J. Arenas-Garcia, M. Martinez-Ramòn, A. Navia-Vazquez, and A. R. Figueiras-Vidal. Plant identification via adaptive combination of transversal filters. *Signal Processing*, 86(9):2430 – 2438, 2006. Special Section: Signal Processing in UWB Communications.

[3] N. P. Balakrishnan, G. P. Rangaiah, and L. Samavedham. Review and analysis of blood glucose (bg) models for type 1 diabetic patients. *Industrial Engineering Chemistry Research*, 50(21):12041–12066, 2011.

[4] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, Secaucus, NJ, USA, 2006.

[5] L. Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996.

[6] D. Bruttomesso and et al. Closed-loop artificial pancreas using subcutaneous glucose sensing and insulin delivery and a model predictive control algorithm: Preliminary studies in padova and montpellier. *J Diabetes Sci Technol.*, 5(3):1014–1021, Sep. 2009.

[7] M. Cescon and E. Renard. Adaptive subspace-based prediction of t1dm glycemia. In *50th IEEE Conference on Decision and Control and European Control Conference (CDC2011)*, Orlando, Florida, USA, December 2011.

[8] W. L. Clarke, D.l Cox, L. A. Gonder-Frederick, W. Carter, and S. L. Pohl. Evaluating clinical accuracy of systems for self-monitoring of blood glucose. *Diabetes Care*, 10:622–628, 1987.

[9] C. Dalla Man, R. A. Rizza, and C. Cobelli. Meal simulation model of the glucose-insulin system. *IEEE Trans. Biomedical Engineering*, 54(10):1740–1749, 2007.

[10] V. C. Protopappas E. I. Georga and D. I. Fotiadis. *Knowledge-Oriented Applications in Data Mining*, chapter Glucose Prediction in Type 1 and Type 2 Diabetic Patients Using Data Driven Techniques. InTech, 2011.

[11] G. Elliott, C. W.J. Granger, and A. Timmermann, editors. *Handbook of Economic Forecasting*, chapter 10. Forecast Combinations. Elsevier, 2006.

[12] E. J. Elton, M. J. Gruber, and M. W. Padberg. Simple criteria for optimal portfolio selection. *The Journal of Finance*, 31(5):1341–1357, Dec. 1976.

[13] G.C. Estrada, H. Kirchsteiger, L. del Re, and E. Renard. Innovative approach for online prediction of blood glucose profile in type 1 diabetes patients. In *American Control Conference (ACC), 2010*, pages 2015 – 2020, 30 2010-july 2 2010.

[14] A. Gani, A. V. Gribok, Y. Lu, W. K. Ward, R. A. Vigersky, and J. Reifman. Universal glucose models for predicting subcutaneous glucose concentration in humans. *Trans. Info. Tech. Biomed.*, 14(1):157–165, January 2010.

[15] F. Gustafsson. *Adaptive Filtering and Change Detection*. John Wiley & Sons, Hoboken, New Jersey, USA, 2000.

[16] J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky. Bayesian model averging: A tutorial. *Statistical Science*, 14(4):382–417, 1999.

[17] J. Z. Kolter and M. A. Maloof. Dynamic weighted majority: A new ensemble method for tracking concept drift. *Data Mining, IEEE International Conference on*, 0:123–130, 2003.

[18] B.P. Kovatchev, M. Straume, D.J. Cox, and L.S. Farhy. Risk analysis of blood glucose data: A quantitative approach to optimizing the control of insulin dependent diabetes. *J. of Theoretical Medicine*, 3:1–10, 2000.

[19] A. Makroglou, J. Li, and Y. Kuang. Mathematical models and software tools for the glucose-insulin regulatory system and diabetes: an overview. *Applied Numerical Mathematics*, pages 559–573, 2006.

[20] Littlestone N. and Warmuth M. K. The weighted majority algorithm. *Information and Computation*, 108(2):212–261, feb. 1994.

[21] S.; Sampath S. Naumova, V.; Pereverzyev. 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, 2011.

[22] H Ohlsson, L. Ljung, and S. Boyd. Segmentation of ARX-models using sum-of-norms regularization. *Automatica*, 46(6):1107–1111, 2010.

[23] N.C. Oza. Online bagging and boosting. In *Systems, Man and Cybernetics, 2005 IEEE International Conference on*, volume 3, pages 2340 – 2345 Vol. 3, Oct. 2005.

[24] J.U. Poulsson and et al. A diabetes management system empowering patients to reach optimised glucose control: From monitor to advisor. In *Proc. IEEE EMBS 32nd Annual International Conference (EMBC2010)*, pages 5270–1, Sep. 2010.

[25] A. E. Raftery, T. Gneiting, F. Balabdaoui, and M.l Pololakowski. Using bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review*, 133:1155–1174, May 2005.

[26] F. Ståhl and R. Johansson. Observer based plasma glucose prediction in type i diabetes. page 1620.

[27] T. Takagi and M. Sugeno. Fuzzy identification of system and its applications to modelling and control. *IEEE Tran. on Systems, Man, and Cybernetics.*, SMC-15:116–132, Jan-Feb 1985.