Online Prediction of Subcutaneous Glucose Concentration for Type 1 Diabetes Using Empirical Models and Frequency-Band Separation

Chunhui Zhao, Eyal Dassau, Howard C. Zisser, Lois Jovanovič, Francis J. Doyle III, and Dale E. Seborg
Dept. of Chemical Engineering, University of California, Santa Barbara, CA 93106

Sansum Diabetes Research Institute, Santa Barbara, CA 93105

DOI 10.1002/aic.14288
Published online November 18, 2013 in Wiley Online Library (wileyonlinelibrary.com)

Online glucose prediction which can be used to provide important information of future glucose status is a key step to facilitate proactive management before glucose reaches undesirable concentrations. Based on frequency-band separation (FS) and empirical modeling approaches, this article considers several important aspects of on-line glucose prediction for subjects with type 1 diabetes mellitus. Three issues are of particular interest: (1) Can a global (or universal) model be developed from glucose data for a single subject and then used to make suitably accurate on-line glucose predictions for other subjects? (2) Does a new FS approach based on data filtering provide more accurate models than standard modeling methods? (3) Does a new latent variable modeling method result in more accurate models than standard modeling methods? These and related issues are investigated by developing autoregressive models and autoregressive models with exogenous inputs based on clinical data for two groups of subjects. The alternative modeling approaches are evaluated with respect to on-line short-term prediction accuracy for prediction horizons of 30 and 60 min, using independent test data. © 2013 American Institute of Chemical Engineers *AIChE J*, 60: 574–584, 2014
Keywords: global model, glucose prediction, type 1 diabetes, frequency-band separation

Introduction

Type 1 diabetes mellitus (T1DM) is a disease characterized by the inability of the body to regulate blood glucose concentration. T1DM results from autoimmune destruction of the pancreatic β-cells, which produce the hormone insulin. Without appropriate treatment with exogenous insulin, people with T1DM have difficulty maintaining their blood glucose concentration within a normal range (e.g., 70–120 mg/dL before meal and less than 160 mg/dL after meal). Consequently, they can suffer from large glycemic excursions, including episodes with very low glucose (hypoglycemia) and very high glucose (hyperglycemia); both situations are detrimental to the quality of life.[1] Generally, a careful balance is required among a person's daily activities, diet, and insulin administration to bring the blood glucose into a normal range. However, this balancing is not an easy task because large glycemic variations often go undetected, including asymptomatic hypoglycemia.[2]

Continuous glucose monitoring (CGM) devices can measure glucose time-series data on-line (i.e., every 1–5 min). The real-time data provide timely and important information about the person's current glycemic state, and also reveal its direction and rate of change. CGM devices have opened new opportunities for glycemia management of subjects with T1DM. For example, it has been reported that if the recent glucose history follows previous known patterns, future blood glucose values might be anticipated from past measurements.[3] An empirical model of glucose dynamics can facilitate glycemia management. Thus, the identification of simple, accurate glucose prediction models is a key step in the development of an effective artificial pancreas (AP).

Many empirical (or "data-driven") modeling techniques have been evaluated for both *in silico* and clinical studies.[3–14] Generally, the glucose prediction model has the form of a linear dynamic model where the future glucose concentration is predicted based on current and past glucose signals, as well as available exogenous input signals, notably insulin delivery and meal carbohydrate (CHO) estimates. Bremer and Gough[3] first suggested that glycemic time-series data had an inherent structure that could be described by a simple linear dynamic model. The linear models that have received the most attention for TIDM applications are autoregressive models (AR) and autoregressive models with exogenous inputs (ARX). For AR modeling, only CGM data are required to develop the model and to predict future glucose concentrations as a linear combination of recent measurements. The Cobelli[7,8] and Reifman research groups[9] have clinically evaluated subject-specific AR models with different model orders to improve management of glucose concentrations. Eren-Oruklu et al.[10,11] have reported subject-specific

Correspondence concerning this article should be addressed to: C. Zhao at this current address: State Key Laboratory of Industrial Control Technology, Dept. of Control Science and Engineering, Zhejiang University, Hangzhou, 310027, China; e-mail: chhzhao@zju.edu.cn; or D. E. Seborg at seborg@engineering.ucsb.edu.

recursive AR models where the model parameters were recursively updated to reflect the recent glucose history. The ARX models are an extension of AR models that include the exogenous inputs, insulin delivery, and meal CHO estimates. Finan et al.[4,12,13] have developed ARX models for *in silico* and clinical subjects. They evaluated the effects of key design issues such as the degree of input excitation, model orders, and prediction horizons. A new latent variable (LV)-based statistical modeling algorithm has recently been developed by our research group[15,16] for T1DM applications. The results for clinical and *in silico* applications have demonstrated the effectiveness of the proposed method and its improved prediction accuracy, compared with standard AR and ARX models.

In contrast to the widespread development of personalized glucose prediction models, there is a limited body of work concerning the analysis of intersubject variability and its effects on glucose prediction. In particular, Gani et al.[17] have mentioned and verified the concept of a universal model for T1DM. If a universal model is reasonably accurate, it would considerably reduce the effort associated with model development. The "universal model" of Gani et al.[17] is a 30th-order AR prediction model with constant coefficients, which was able to capture the time-series glucose correlations for different subjects. The model was based on glucose data for a single subject and then used to make short-term (30-min-ahead) glucose predictions for other subjects without any need for model customization. Their results indicated that the predictive capability of the models was not affected by intersubject differences. However, their modeling approach makes glucose predictions based on a set of time-series data that are smoothed retrospectively[17,18] using a Tikhonov regularization method.[19] Thus, the model is not suitable for on-line applications which is thus not useful for development of AP. Also, they did not check whether prediction model was universal when two exogenous inputs, insulin delivery, and meal CHO estimates, were included in the model structure.

In this article, empirical modeling approaches and frequency-band separation (FS) are investigated for on-line glucose prediction for T1DM. A variety of models are considered including the new LV models as well as standard AR and ARX models. The major issue addressed in this work is about short-term online glucose prediction. So, the key issue is whether these modeling techniques can be used to develop a "global model" that is suitable for on-line glucose prediction. Of particular interest, is whether a prediction model developed for data from one group of clinical subjects is also valid for other groups of clinical subjects without any customization in comparison with those subject-dependent (SD) standard models.

Methodology

Standard AR and ARX prediction models

In this article, AR modeling techniques based on standard least-squares (LS) algebra and a new latent-variable-based statistical analysis[15,16] are used to develop empirical prediction models from glucose time series data. Corresponding ARX models are evaluated after adding the exogenous inputs to the AR models.

The general form of the ARX model for this research is given by

$$\begin{aligned} A\left(q^{-1}\right)g(k) &= B_{\text{ins}}\left(q^{-1}\right)u_{\text{ins}}\left(k-k_{\text{ins}}\right) \\ &+ B_{\text{meal}}\left(q^{-1}\right)u_{\text{meal}}\left(k-k_{\text{meal}}\right)+\beta+\varepsilon(k) \end{aligned} \tag{1}$$

where $g(k)$ denotes the glucose concentration at sampling instant k, $u_{\text{ins}}(k-k_{\text{ins}})$, and $u_{\text{meal}}(k-k_{\text{meal}})$ are the exogenous inputs, bolus insulin, and meal carbohydrate content, at time k, β is a constant bias term, and $\varepsilon(k)$ is a zero mean, random disturbance at time k. Note that this ARX model is expressed in terms of physical variables and includes a bias term, rather than being based on deviation variables. This approach eliminates the need to specify an appropriate steady-state reference value for the glucose concentration, information that may be difficult to determine in practice due to the inherent dynamic behavior of blood glucose concentration. The input time delays, k_{ins} and k_{meal}, in Eq. 1 can be different for each input. The time delays are expressed as integer multiples of the sampling period Δt. In this article, $\Delta t = 5$ min.

In Eq. 1, $A(q^{-1})$, $B_{\text{ini}}(q^{-1})$, and $B_{\text{meal}}(q^{-1})$ denote polynomials in q^{-1}, where q^{-1} is the backward shift operator, that is, $q^{-1}g(k) \equiv g(k-1)$. For example

$$A\left(q^{-1}\right)=a_0+a_1q^{-1}+a_2q^{-2}+\cdots+a_{n_A}q^{-n_A} \tag{2}$$

where n_A is the order of the $A(q^{-1})$ polynomial. It determines the number of previous glucose measurements that are relevant for prediction. When polynomials $B_{\text{ini}} = B_{\text{meal}} = 0$, the ARX model in Eq. 1 reduces to an AR model. The identification of an AR or ARX model with specified model orders can be performed analytically using standard LS regression[20] to estimate the model coefficients. Alternatively, the new LV approach of the next section can be employed.

LV models

An LV-based glucose prediction method that has been recently reported[16] is also considered in this article. The resulting models (LVX/LV) have the same structures as standard ARX/AR models. However, the model parameters are calculated using different principles and algorithms. The LV-based models are developed from two sets of data:[16] predictor variable data $\mathbf{X}(N\times J)$ and output variable data $\mathbf{y}(N\times 1)$ where N is the number of observations and J is the number of predictor variables. The predictor data matrix $\mathbf{X}$ consists of past and current glucose concentrations and the two exogenous inputs. When no exogenous inputs are included, the LVX model reduces to an LV model. The single-output variable is the future glucose concentration. For T1DM, the time-series glucose measurements contain the autocorrelation structure relating past and future glucose concentration values.

In the LV-based modeling method, a few LVs are first calculated which capture the important glucose dynamics and are linear combinations of the available measurements. In the second step, the quantitative relationship between the LVs and future glucose concentration is determined. A variety of LV-based regression methods have been developed with the chief difference being how the LVs are calculated. The details of the LV-based methods for this research are summarized in Appendix A. The LV-based glucose prediction model is briefly described in Appendix B.

Frequency-band separation

Rahaghi and Gough[21] have recently suggested that the glucose dynamics for subjects without diabetes can be

Table 1. Clinical Details for the Two Groups of Subjects

Group #	Number of Subjects	Age (Years)	Weight (kg)	Height (cm)	CGM Device
1	7	48 ± 15	87 ± 25	174 ± 11	DexCom 7 Plus™
2	10	49 ± 10	79 ± 10	173 ± 9	DexCom 7™

divided into four distinct frequency ranges with different periods:

- *Band I*: 5–15 min
- *Band II*: 60–120 min
- *Band III*: 150–500 min
- *Band IV*: $\geq$ 700 min

Each band is driven by different physiological mechanisms and intrinsic blood glucose dynamics. Lu et al.[22] considered the relative importance and predictive power of these four frequency bands using SD AR models for short-term glucose prediction. Their results indicated that a combination of the datasets for Band II, and either Band III or IV, represent the underlying glucose dynamics necessary for model development. These papers provide meaningful insights into the underlying glucose dynamics from a frequency domain viewpoint. However, there is very little literature that considers an important aspect of glucose prediction: the possibility that for a given frequency band, the glucose dynamics may vary from subject to subject (i.e., intersubject variability). Consequently, in contrast to previous T1DM papers on frequency domain analysis,[21,23] in this article, glucose dynamics are investigated from an intersubject perspective. The objective is to extract the common glucose dynamics for global model development when only two frequency bands are considered.

For frequency band separation, a first-order, low-pass Butterworth filter is employed after specifying its threshold period P. The filter has the form

$$\tilde{x}(k)=\beta_1 x(k)+\beta_2 x(k-1)\text{-}\alpha\tilde{x}(k-1) \tag{3}$$

where, x is the glucose measurement, $\tilde{x}$ is the filtered value, and the constant filter parameters are α, β_1, and β_2. Thus, filter output $\tilde{x}$ is a linear combination of the previous filtered value, and the previous and current measurements.

In the frequency band separation approach, the glucose data are first filtered to divide them into two frequency bands, a low-frequency band and a high-frequency band. For subject i, the low-frequency band data are denoted by $\mathbf{x}_i^L(K\times 1)$ (where K is the number of observations). These data are generated by passing the original CGM glucose data through the low-pass filter in Eq. 3. The high-frequency band data for subject i, $\mathbf{x}_i^H(K\times 1)$, are then obtained by subtracting the low-frequency band data from the original data

$$\mathbf{x}_i^H = \mathbf{x}_i - \mathbf{x}_i^L \tag{4}$$

The empirical prediction models can then be developed based on data for either a single band or both bands. To evaluate the feasibility of developing a global model, four modeling approaches are considered:

1. Global low-frequency model (GL): The global low-frequency model is identified based on the low-frequency glucose data for a single subject. This model is then used to make glucose predictions for the other subjects.

2. Subject-dependent low-frequency model (SL): A separate low-frequency model is identified for each subject.

3. Subject-dependent two-frequency models (ST): Individual low-frequency and high-frequency models are designed for each subject. Then, glucose predictions are made by adding the individual model predictions together.

4. Standard subject-dependent model (SD): A separate empirical model is designed for each subject; no frequency band filtering is employed.

For ARX/LVX modeling, the exogenous inputs are included and transformed using two different second-order transfer functions,[22] producing time-smoothed inputs. In this way, the effects of the two inputs can be separated, which facilitates the model identification.

Performance metrics

To characterize the glucose prediction performance, a variety of metrics are available. In this article, two metrics are used:

1. Root mean-square error [RMSE (mg/dL)][17,18]

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i\in N}\left(y(i)-\hat{y}(i)\right)^2} \tag{5}$$

where, $\hat{y}(i)$ is the predicted value, $y(i)$ is the measurement; and N is the number of samples.

1. Mean absolute relative deviation [MARD (%)][24,25]

$$\text{MARD}\,(\%) = \frac{1}{N}\sum_{i\in N}\left|\frac{y(i)-\hat{y}(i)}{y(i)}\right|\times 100\% \tag{6}$$

where $|\bullet|$ denotes the absolute value.

Results and Discussion

Study subjects

Two groups of de-identified ambulatory clinical data from subjects with T1DM were used in this investigation. The subjects gave their voluntary and written informed consent to participate. The subcutaneous glucose data were collected using two types of CGM devices and a 5-min sampling interval. The clinical information for the two groups is listed in Table 1 where the mean value and standard deviation were calculated across the biometric information of subjects in each group.

For Group 1, meal CHO estimate and bolus insulin delivery information were available which facilitated the identification of ARX/LVX models. Data (2 or 3 days) were collected for each subject. Data for the first day were used for model identification and the remaining data was used for model testing. The latter will be referred to as the test data. For Group 2, only CGM data for a period of 4–7 days were collected for each subject. Because more data were available for Group 2, compared with Group 1, the first 2 days of data were used for model identification, whereas the remaining data were used as test data. However, the prediction results were not statistically different when only 1 day of data was used for model identification based on a paired t-test ($\alpha = 0.05$).

AR and LV empirical models

The AR modeling technique was evaluated as a candidate for developing a global model, in the following manner.

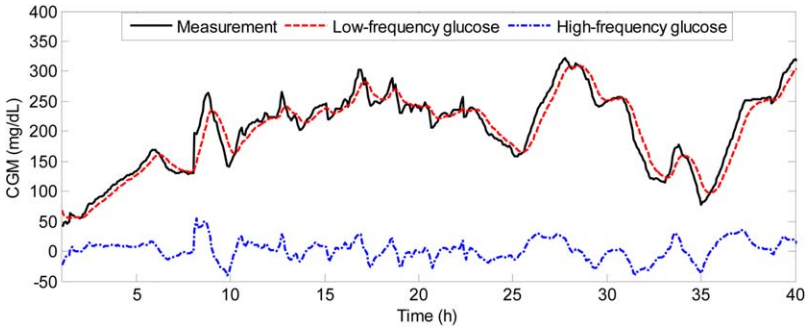

Figure 1. Comparison of subcutaneous glucose concentration provided by CGM and the separated profiles in two different bands for Subject #1 in Group1.

The threshold period is 80 min. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

First, a threshold filter value P was selected. Then, the low-pass filter in Eq. 3 was used to divide the original CGM signal into two parts. Figure 1 illustrates the data filtering for a typical type 1 diabetes subject in Group 1 and a threshold period of $P = 80$ min. In general, the low-frequency glucose signals capture the overall glucose trends, whereas the high-frequency signals are more random with a mean value of zero. In a preliminary investigation of SD standard AR/LV model development, it was determined that as the number of glucose variables (i.e., the predictor length PL) increases, glucose prediction accuracy did not improve for $PL > 7$. Therefore, a value of $PL = 7$ was used for global model identification. Then, for each of the 17 subjects in Groups 1 and 2, training data for a single subject were used to develop a low-frequency AR model. The prediction accuracy of the model was evaluated for the other 16 subjects. Thus for each test subject, 16 sets of glucose predictions were obtained from the individual models.

Let $y_{i,j}$ denotes the RMSE value of glucose prediction for Subject i using the low-frequency AR model developed based on training data for subject j. An average RMSE value

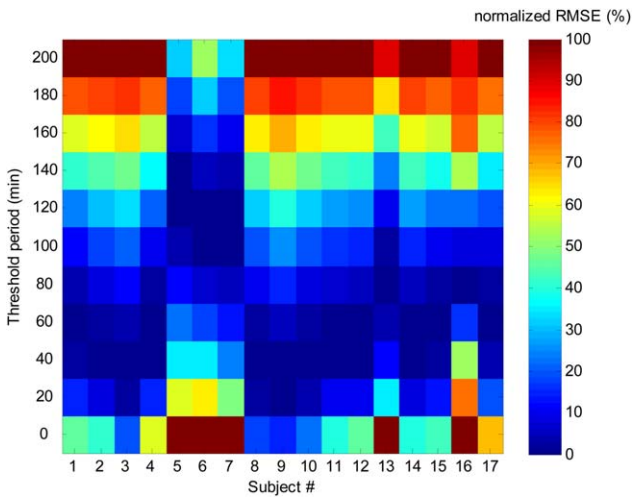

Figure 2. Effects of threshold period (min) for frequency-band separation on 30-min prediction performance [evaluated by normalized RMSE (%)] for 17 subjects using low-frequency AR models.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

y_i for subject i was calculated by averaging the results for the 16 sets of test data

$$y_i = \frac{1}{16}\sum_{j=1}^{16} y_{i,j} \tag{7}$$

These calculations were repeated for different values of threshold period P. The y_i values for test subject i and the threshold periods can be arranged in a vector, $\mathbf{y}_i = [y_{i,1}, y_{i,2}, \ldots, y_{i,m}, \ldots, y_{i,M}]$, where M is the number of threshold periods. Because the RMSE values have different ranges for different subjects, the $\mathbf{y}_i$ values were normalized

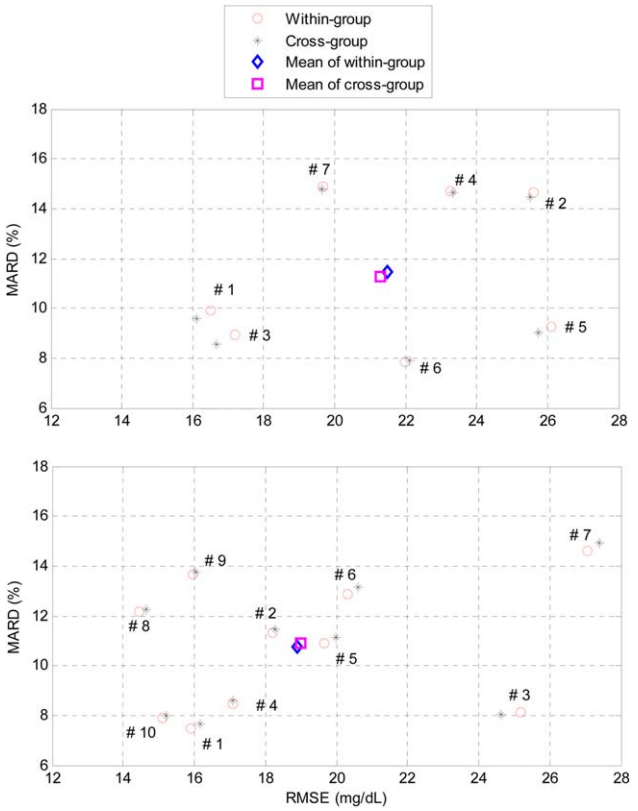

Figure 3. GL AR model based 30-min prediction results for seven subjects in Group 1 (Top) and 10 subjects in Group 2 (bottom) with respect to within-group and cross-group analyses.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

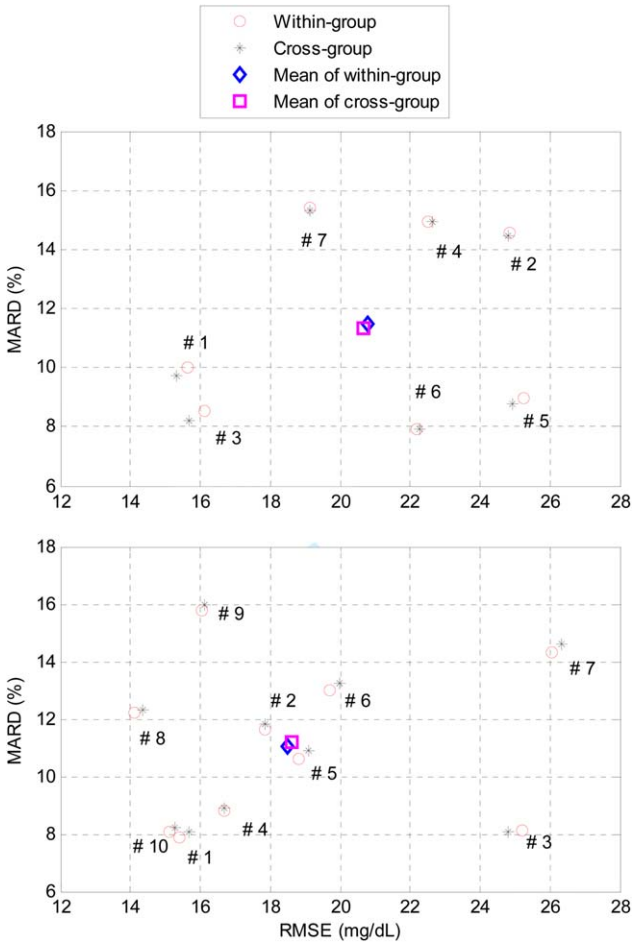

Figure 4. GL LV model based 30-min prediction results for seven subjects in Group 1 (Top) and 10 subjects in Group 2 (bottom) with respect to within-group and cross-group analyses.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

and transformed into a percentage between 0 and 100%, which was more convenient for plotting. For subject i, the transformed $\mathbf{y}_i$ value, $\overline{\mathbf{y}}_i$, is defined as

$$\overline{\mathbf{y}}_i \triangleq \frac{\mathbf{y}_i - \mathbf{y}_{i,\min}}{\max(\mathbf{y}_i) - \min(\mathbf{y}_i)} \times 100\% \tag{8}$$

where $\max(\mathbf{y}_i)$ denotes the maximum RMSE value for the different threshold periods and$\min(\mathbf{y}_i)$ denotes the corresponding minimum value. Their difference denotes the glucose prediction range. In Eq. 8, $\mathbf{y}_{i,\min}$ is an M-dimensional vector where each element is $\min(\mathbf{y}_i)$.

Figure 2 shows the effects of threshold period P on the prediction accuracy of the average AR models. The 17 subjects of Groups 1 and 2 served as the test subjects for the low-frequency AR models. Note that a value of $P = 0$ means that no FS was used and thus a standard AR model was obtained. Clearly, the large average RMSE values indicate that the standard AR model is not global.

Based on the results in Figure 2, a value of $P = 80$ min was chosen for subsequent frequency band separation. For the range, 40 min $< P <$ 80 min, the accuracy of the low-frequency AR model is not statistically superior for any specific value of P, based on a paired t-test ($\alpha = 0.05$). Consequently, a value of $P = 80$ min was selected arbitrarily for frequency band separation.

The hypothesis of a global low-frequency (GL) prediction model was also tested, based on AR models for individual subjects. Each GL AR model was used for glucose prediction for the test set of 17 subjects. A paired t-test ($\alpha = 0.05$) was again used to assess the differences in prediction accuracy between these low-frequency AR models. The results (not shown here) indicated that the choice of the subject used to generate the GL AR model did not have a significant effect on model accuracy. Thus, training data for any subject could be used to develop a GL AR model. For the rest of

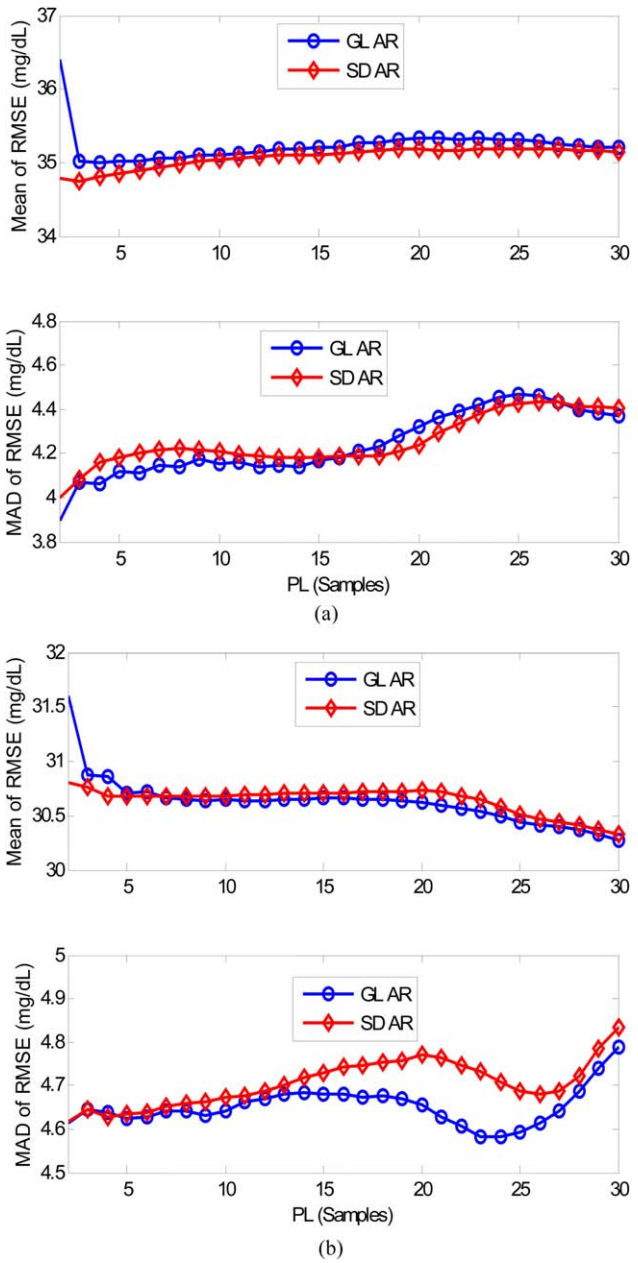

Figure 5. Effect of PL on 60-min prediction performance using GL AR and SD AR models [RMSE (mg/dL)] for (a) seven subjects in Group 1 and (b) 10 subjects in Group 2.

Top: Mean of RMSE values; Bottom: MAD of RMSE values. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

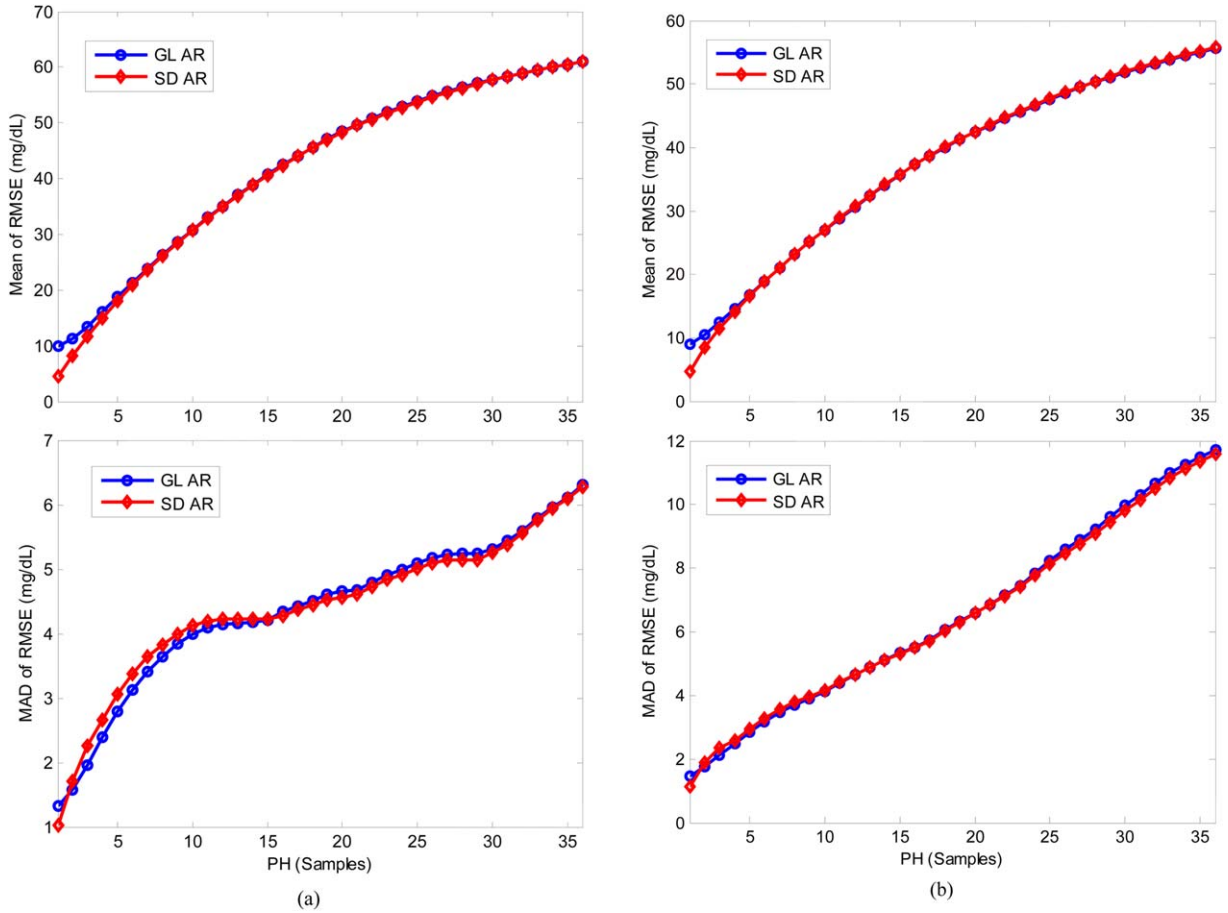

Figure 6. Effect of *PH* on prediction performance using GL AR and SD AR models [RMSE (mg/dL)] for (a) seven subjects in Group 1 and (b) 10 subjects in Group 2.

Top: Mean of RMSE values; Bottom: MAD of RMSE values. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

the article, the GL AR model was obtained by averaging individual GL AR models from different subjects. The average model is considered to be more representative than an individual model.

Next, the prediction accuracy of the GL AR model will be assessed by both within-group and cross-group analysis. For the within-group analysis, the GL AR model for the group was obtained by averaging the individual GL models for one group and then evaluating the on-line glucose prediction accuracy for test data for the same group. For the cross-group analysis, the GL AR model for one group was applied to the test data for the subjects in the second group.

The prediction accuracies for the within-group and cross-group analyses for the GL AR models are shown in Figure 3, for the MARD and RMSE indices. These results are displayed for individual subjects. The best prediction performance occurs when the results are closest to the left bottom corner of each subfigure. The mean values for all of the subjects are also included in Figure 3. It is noteworthy that the cross-group prediction results almost overlap the within-group prediction results for both groups, which suggests that the accuracy of global model is not dependent on the specific CGM sensor or clinical trial.

Figure 4 presents a similar analysis performed for the LV modeling method. These results are very similar to those in Figure 3 for the AR method. Based on a paired t-test ($\alpha = 0.05$), the accuracy of the GL AR model is not statistically different from that of the GL LV model. The results in Figures 3 and 4 show that both the AR and LV methods can be used to identify global prediction models that capture the common glucose dynamics. However, the prediction accuracy of the global model does vary significantly for individual test subjects due to individual differences.

The effect of key design parameters on the prediction accuracy of the average GL AR model is also checked. First, by setting different values of design parameters, different low-frequency AR models were identified based on training data; then, the model averaged across subjects was used as the representative GL AR model with certain parameter values and applied to testing data of different subjects in each group for online glucose prediction. The mean value and the median absolute deviation (MAD) for the RMSE index were averaged across the subjects in each group. The MAD statistic characterizes the intersubject variability for the average GL AR model. For purposes of comparison, the analogous results for standard SD AR models were also calculated.

First, the effect of predictor length *PL* on the accuracy of 30-min predictions was assessed for all 17 subjects. When choosing different *PL* values, different models are developed based on training data and the corresponding models are then applied to testing data for glucose prediction. As shown in Figure 5, as *PL* increases, the trends for the GL and SD AR models are quite similar except when *PL* is very small, a situation that is not meaningful for on-line glucose

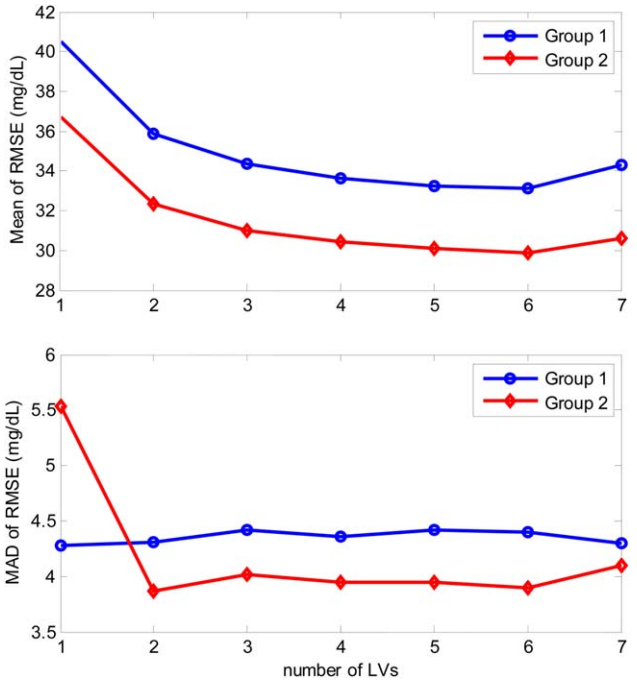

Figure 7. Effect of the number of retained LVs on 60-min prediction performance using GL AR models [RMSE (mg/dL)] for seven subjects in Group 1 and 10 subjects in Group 2.

Top: Mean of RMSE values; Bottom: MAD of RMSE values. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

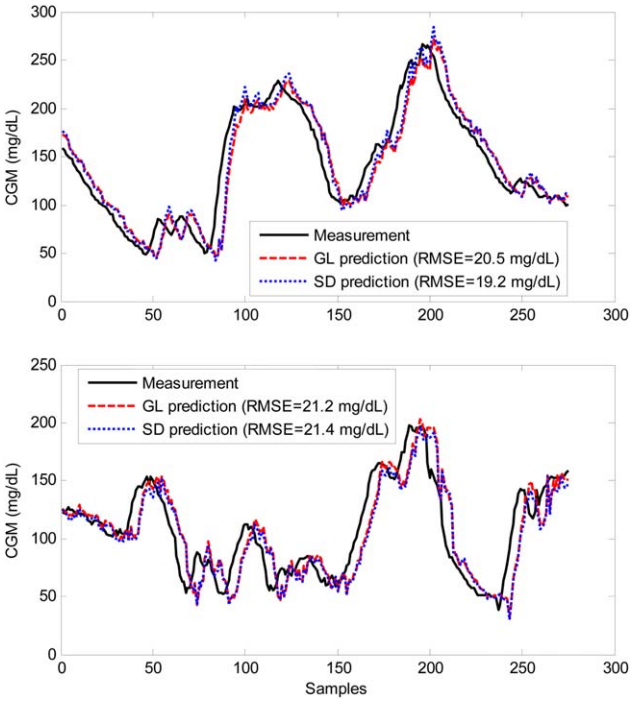

Figure 8. Comparison of representative measured and 30-min predicted glucose profiles for Subjects #1 (top) and #7 (bottom) in Group 1 based on the global low-frequency (GL) and standard subject-dependent (SD) AR models.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

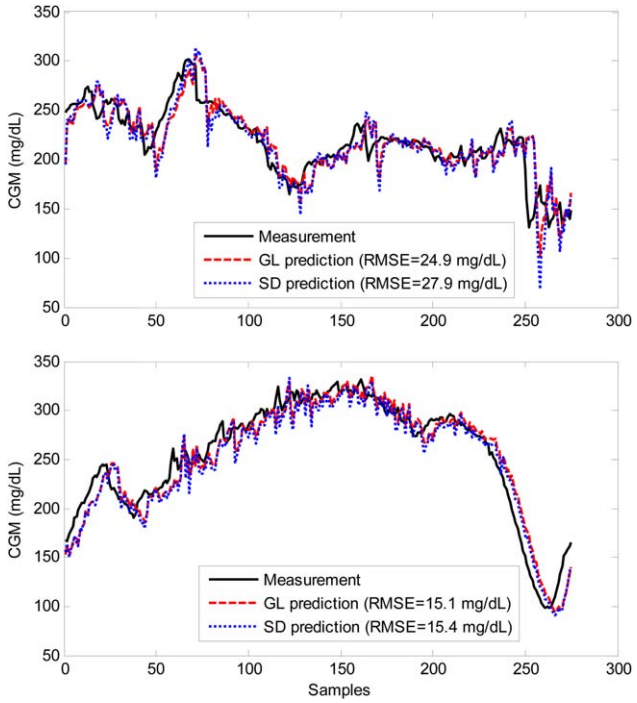

Figure 9. Comparison of representative measured and 30-min predicted glucose profiles for Subjects #3 (top) and #10 (bottom) in Group 2 based on the global low-frequency (GL) and standard subject-dependent (SD) AR models.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

prediction. For both models, the prediction performance is not very sensitive to the choice of the *PL* value (note expanded ordinate scale). For the subsequent AR model comparisons, a value of $PL = 7$ was used. Figure 6 presents an analogous evaluation for a second important design parameter, the prediction horizon, *PH*. As *PH* increases, the prediction accuracy of both types of models decreases, as expected, revealing that some significant unknown extraneous variables may need to be considered. Furthermore, the prediction accuracies of the GL AR and SD AR models are very similar except for small *PH* values. In particular, the differences are not statistically significant, based on a paired *t*-test ($\alpha = 0.05$). The comparisons in Figures 5 and 6 indicate that the global AR model is as accurate as the SD AR model, an important advantage for glucose prediction applications that involve large numbers of subjects. Another important design parameter is the number of retained LVs when using LV method to develop the prediction model. Figure 7 shows the results to evaluate the effect of the number of retained LVs on global AR model where a value of $PL = 7$ was used as mentioned before for both groups of subjects. As the number of LVs increases, the prediction accuracy of both groups first increases and then decreases after the number is larger than six. That is, when all LVs are retained, the prediction accuracy is not necessary better than the other cases which may result from the overfitting problem.

Representative comparisons of the measured and predicted glucose profiles for GL AR and SD AR models are shown in Figure 8. These results are for Subjects #1 and #7 in Group 1

Table 2. RMSE (mg/dL) (Mean ± MAD) for Glucose Prediction and Seven Subjects in Group 1 with Respect to Different Types of AR/LV Models

	$PH = 6$		$PH = 12$	
Type of Model	AR	LV	AR	LV
Global low-frequency model	21.4 ± 3.1	20.7 ± 3.1	35.1 ± 4.1	33.1 ± 4.4
Global standard model	23.7 ± 5.6	22.4 ± 3.9	36.9 ± 4.9	35.1 ± 5.6
Subject-dependent low-frequency model	21.3 ± 3.3	20.4 ± 3.2	34.9 ± 4.2	31.0 ± 4.0
Subject-dependent two-frequency model	20.9 ± 3.5	19.9 ± 3.5	34.9 ± 4.3	31.0 ± 4.0
Subject-dependent standard model	20.8 ± 3.5	19.5 ± 3.3	34.9 ± 4.2	30.6 ± 4.0

Table 3. RMSE (mg/dL) (Mean ± MAD) for Glucose Prediction and 10 Subjects in Group 2 with Respect to Different Types of AR/LV Models

	$PH = 6$		$PH = 12$	
Type of Model	AR	LV	AR	LV
Global low-frequency model	18.9 ± 3.2	18.5 ± 3.0	30.7 ± 4.6	29.9 ± 3.9
Global standard model	20.6 ± 4.1	20.3 ± 4.0	33.4 ± 5.6	33.2 ± 5.5
Subject-dependent low-frequency model	19.0 ± 3.2	18.5 ± 3.1	31.3 ± 5.2	29.4 ± 4.8
Subject-dependent two-frequency model	19.3 ± 3.5	18.8 ± 3.4	31.6 ± 5.3	29.7 ± 4.9
Subject-dependent standard model	19.3 ± 3.5	18.7 ± 3.4	30.7 ± 4.7	29.7 ± 4.8

and a representative day of test data. In general, the evolving glucose trends were captured by both methods and the difference in prediction accuracy is not statistically significant based on a paired t-test ($\alpha = 0.05$). Figure 9 presents similar results for two representative subjects in Group 2.

In Tables 2 and 3, the prediction accuracy of the different types of AR models is compared for prediction horizons of 30 and 60 min. The merits of the frequency separation approach are apparent from a comparison of the two global models in the first two rows in each table. When frequency separation is used, the resulting "global low frequency model" is more accurate than the "global standard model" for both Groups 1 and 2. Furthermore, the global low frequency model exhibits about the same degree of accuracy as the SD models. The results in Tables 2 and 3 and Figures 5 and 6 support the hypothesis that a global AR or LV model is feasible for on-line glucose prediction. In general, the LV-based AR models in Tables 2 and 3 are more accurate than the LS-based AR models. However, the comparisons of model prediction accuracy in Tables 2 and 3 are not statistically significant for a paired t-test ($\alpha = 0.05$), except for the global standard model which is statistically less accurate.

ARX and LVX empirical models

In this section, ARX and LVX models are developed to assess whether or not they can provide a global model for on-line glucose prediction when the model contains the two exogenous inputs, insulin delivery, and estimated meal CHO. A preliminary investigation was made for SD ARX and LVX models developed from pure measurement data, the model orders, and time delays for two different inputs can be determined to achieve best accuracy. Based on the results, for most subjects, the model orders for glucose, insulin, and meals were set equal to 7, 1, and 1 sample, respectively, and the time delays for the insulin delivery and meals were 5 and 6 samples, respectively. The above setting can get the best prediction accuracy for SD models, which are thus used here to evaluate whether subject-independent ARX and LVX model can be acceptable. The same threshold period ($P = 80$ min) was chosen for the FS modeling.

The development of ARX and LVX models was analogous to the development of AR and LV models in the previous section. The model details and calculations are summarized in Appendices A and B. Thus, different low-frequency models were identified for each subject, and their glucose prediction accuracy was evaluated using test data for the other subjects. For a prediction horizon of 30 min ($PH = 6$ samples), the prediction accuracy was similar, regardless of which individual model was used as the global model; the differences between the models was not

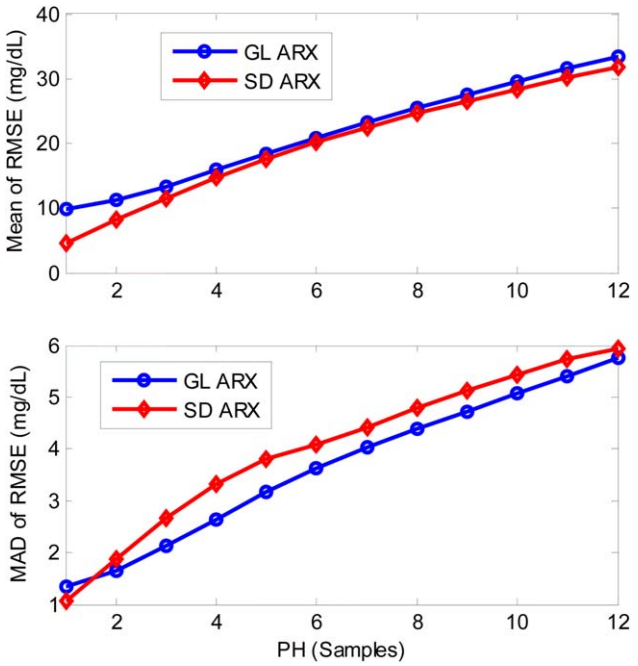

Figure 10. Effect of *PH* on prediction performance [RMSE (mg/dL)] for seven subjects based on GL and SD ARX models.

Top: Mean of RMSE values; Bottom: MAD of RMSE values. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Table 4. RMSE (mg/dL) (Mean ± MAD) for Glucose Prediction and Seven Subjects with Respect to Different Types of ARX/LVX Models

	$PH = 6$		$PH = 12$	
Type of Model	ARX	LVX	ARX	LVX
Global low-frequency model	20.9 ± 3.6	20.8 ± 3.6	34.0 ± 5.8	33.4 ± 5.6
Global standard model	23.9 ± 5.9	22.4 ± 4.3	35.5 ± 6.3	35.3 ± 6.3
Subject-dependent low-frequency model	20.6 ± 3.8	19.8 ± 3.4	31.8 ± 5.6	29.6 ± 5.1
Subject-dependent two-frequency model	20.5 ± 4.3	19.1 ± 3.9	32.2 ± 6.0	29.6 ± 5.6
Subject-dependent standard model	20.2 ± 4.1	18.7 ± 3.7	31.9 ± 5.9	29.2 ± 5.5

statistically significant for $\alpha = 0.05$. However, for 60-min predictions ($PH = 12$), the differences were more significant (these results are not shown).

To show the influence of PH on the prediction accuracy of global ARX models, prediction accuracy results for an average global model were calculated as in the previous section. Figure 10 shows the mean and MAD values of the RMSE metric predictions as a function of PH for GL and SD ARX models. As expected, the prediction accuracy deteriorates for both models as PH increases. The SD ARX models are statistically more accurate than the GL ARX models for all PH values except $PH = 6$ and 7, based on a paired t-test ($\alpha = 0.05$). The results reveal that a global ARX model is not acceptable for online glucose prediction.

The LV modeling technique was also evaluated for suitability in developing a global LVX model by considering the seven subjects of Group 1. The prediction accuracy results were quite similar to those for the ARX model, as shown in Table 4. The RMSE values in Table 4 indicate that the SD models are more accurate than the global models, especially for the larger PH value of 12. Thus, although the global modeling approach is promising for the AR/LV models of the previous section, it is not promising for the ARX/LVX models of this section.

A comparison of the SD models in Table 4 indicates that frequency band separation does not provide superior ARX models. For SD models, the new LV approach results in more accurate ARX models than the standard LS approach, which is consistent with the AR model results of Tables 2 and 3.

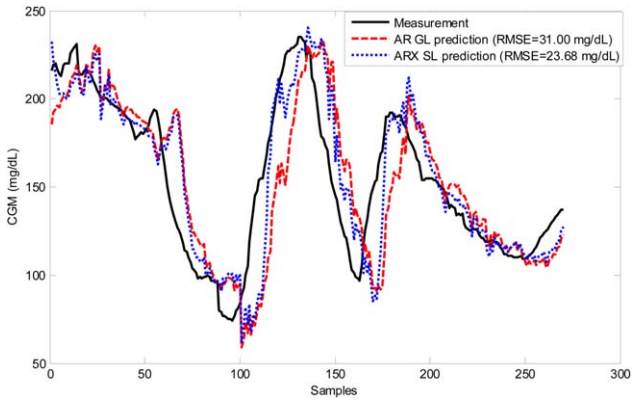

Figure 11. Comparison of representative measured and 60-min predicted glucose profiles for Subjects #3 in Group 1 based on the global low-frequency (GL) AR model and subject-dependent low-frequency (SL) ARX model.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

It is also interesting to compare the relative prediction accuracy of the AR and ARX models for the Group 1 subjects, where insulin delivery data and meal CHO estimates were available. A comparison of the prediction accuracy results in Tables 2 and 4 indicate that in general the global ARX models are not acceptable. However, a comparison of SD models indicates that the ARX models tend to be more accurate than the AR models, especially for the larger prediction horizon ($PH = 12$). Representative comparisons of the measured and predicted glucose profiles for GL AR and SD low-frequency (SL) ARX models are shown in Figure 11 for 60-min predictions. These results are for Subject #3 in Group 1 and a representative day of test data. The comparison indicates that ARX model may have smaller time lag than AR model for online prediction, revealing that the consideration of exogenous inputs may improve the prediction time lag more or less.

Conclusions

In this article, a combination of FS and empirical models has been experimentally evaluated in two clinical studies concerning online short-term ahead glucose prediction in T1DM. It is demonstrated that a global AR model based on the frequency separation approach and training data for a single subject could be used to predict future glucose concentrations for other subjects without adjusting model parameters. For prediction horizons of 30 and 60 min, the global model exhibited the same degree of prediction accuracy as models developed for individual subjects. The small differences in prediction accuracy were not statistically significant based on a paired-t test ($\alpha = 0.05$). Thus, these clinical results suggest that it is feasible to develop a global AR model based on one subject and then apply the models to other subjects.

However, global ARX/LVX models were less accurate than the SD ARX/LVX models as different subjects may give different responses to exogenous inputs. These results demonstrate that a global ARX/LVX model is not acceptable, which, thus, cannot be directly used for glucose control. These promising analyses results encourage extensions of this research methodology. For example, a critical problem concerns glucose controller based on on-line prediction of future glucose. How to analyze the influences of exogenous inputs on glucose values and develop the control model by properly adjusting model parameters of exogenous inputs across different subjects are meaningful issues which deserves further efforts.

Acknowledgments

This research work was supported by the Juvenile Diabetes Research Foundation (grants 22-2009-797, 22-2009-76), the National Natural Science Foundation of China (No.

61273166), and Specialized Research Fund for the Doctoral Program of Higher Education of China (20120101120182).

Literature Cited

1. Rubin RR, Peyrot M. Quality of life and diabetes. *Diabetes Metab Res Rev*. 1999;15:205–218.
2. Rabasa-Lhoret R, Garon J, Langelier H. Effects of meal carbohydrate content on insulin requirements in type 1 diabetic patients treated intensively with the basal-bolus (ultralente-regular) insulin regimen. *Diabetes Care*. 1999;22:667–673.
3. Bremer T, Gough DA. Is blood glucose predictable from previous values? A solicitation for data. *Diabetes*. 1999;48:445–451.
4. Finan DA, Palerm CC. Doyle III FJ, Effect of input excitation on the quality of empirical dynamic models for type 1 diabetes. *AIChE J*. 2009;55:1135–1146.
5. Trajanoski Z, Regittnig W, Wach P. Simulation studies on neural predictive control of glucose using the subcutaneous route. *Comput Meth Prog Bio*. 1998;56:133–139.
6. Dua P, Doyle FJ III, Pistikopoulos EN. Model-based blood glucose control for type 1 diabetes via parametric programming. *IEEE Trans Biomed Eng*. 2006;53:1478–1491.
7. Sparacino G, Zanderigo F, Corazza S, Maran A, Facchinetti A, Cobelli C. Glucose concentration can be predicted ahead in time from continuous glucose monitoring sensor time-series. *IEEE Trans Biomed Eng*. 2007;54:931–937.
8. Zanderigo F, Sparacino G, Kovatchev B, Cobelli C. Glucose prediction algorithms from continuous monitoring data: assessment of accuracy via continuous glucose error-grid analysis. *J Diabetes Sci Technol*. 2007;1:645–651.
9. Reifman J, Rajaraman S, Gribok A. Predictive monitoring for improved management of glucose levels. *J Diabetes Sci Technol*. 2007;1:478–486.
10. Eren-Oruklu M, Cinar A, Quinn L. Estimation of future glucose concentrations with subject-specific recursive linear models. *Diabetes Technol Ther*. 2009;9:438–450.
11. Eren-Oruklu M, Cinar A, Quinn L. Hypoglycemia prediction with subject-specific recursive time-series models. *J Diabetes Sci Technol*. 2010;4:25–33.
12. Finan DA, Zisser H, Jovanovič L. Practical issues in the identification of empirical models from simulated type 1 diabetes data. *Diabetes Technol Ther*. 2007;9:438–450.
13. Finan DA, Palerm CC, Doyle FJ III. Identification of empirical dynamic models from type 1 diabetes subject data. *In: Proc. Am Control Conf*. Seattle, WA, 2008;2099–2104.
14. Palerm CC, Willis JP, Desemone J. Hypoglycemia prediction and detection using optimal estimation. *Diabetes Technol Ther*. 2005;7:3–14.
15. Zhao CH, Dassau E, Harvey RA. Predictive glucose monitoring for type 1 diabetes using latent variable-based multivariate statistical analysis. *In: Proc IFAC*, Milano, Italy, Vol. 18. 2011;7012–7017.
16. Zhao CH, Dassau E, Jovanovič L. Predicting subcutaneous glucose concentration using latent variable (LV)-based statistical analysis method for type 1 diabetes mellitus. *J Diabetes Sci Technol*. 2012;6:617–633.
17. Gani A, Gribok AV, Lu YH. Universal glucose models for predicting subcutaneous glucose concentration in Humans. *IEEE Trans Inf Technol Biomed*. 2010;14:157–165.
18. Gani A, Gribok AV, Rajaraman S. Predicting subcutaneous glucose concentration in humans: data-driven glucose modeling. *IEEE Trans Biomed Eng*. 2009;56:246–254.
19. Tikhonov AN, Arsenin VY. *Solutions of Ill-Posed Problems*. Washington DC: Winston, 1977.
20. Ljung L. *System Identification: Theory for the User*. Upper Saddle River, NJ: Prentice-Hall, 1999.
21. Rahaghi FN, Gough DA. Blood glucose dynamics. *Diabetes Technol Ther*. 2008;10:81–94.
22. Grosman B, Dassau E, Zisser H. Zone model predictive control: a strategy to minimize hyper- and hypoglycaemic events. *J Diabetes Sci Technol*. 2010;4:961–975.
23. Lu YH, Gribok AV, Ward WK. The importance of different frequency bands in predicting subcutaneous glucose concentration in Type 1 diabetic patients. *IEEE Trans Biomed Eng*. 2010;57:1839–1846.
24. Noujaim SE, Horwitz D, Sharma M. Accuracy requirements for a hypoglycemia detector: an analytical model to evaluate the effects of bias, precision, and rate of glucose change. *J Diabetes Sci Technol*. 2007;1:652–668.
25. Stahl F, Johansson R. Diabetes mellitus modeling and short-term prediction based on blood glucose measurements. *Math Biosci Eng*. 2009;217:101–117.
26. Yu H, MacGregor JF. Post processing methods (PLS-CCA): simple alternatives to preprocessing methods (OSC-PLS). *Chemometr Intell Lab Syst*. 2004;73:199–205.
27. Cserhati T, Kosa A, Balogh S. Comparison of partial least-square method and canonical correlation analysis in a quantitative structure-retention relationship study. *J Biochem Bioph Methods*. 1998;36:131–141.
28. Anderson TW. Canonical correlation analysis and reduced rank regression in autoregressive models. *Ann Stat*. 2002;30:1134–1154.
29. Hardoon DR, Szedmak S, Shawe-Taylor J. Canonical correlation analysis: an overview with application to learning methods. *Neural Comp*. 2004;16:2639–2664.
30. Burnham AJ, Viveros R, MacGregor JF. Frameworks for latent variable multivariate regression. *J Chemometr*. 1996;10:31–45.
31. Dayal B, Macgregor JF. Improved PLS algorithms. *J Chemometr*. 1997;11:73–85.
32. Lindgren F, Geladi P, Wold S. The kernel algorithm for PLS. *J Chemometr*. 1993;7:45–59.

Appendix A: LV-Based Modeling Methods

In this research, the LV-based modeling method is a partial LS-canonical correlation analysis (PLS-CCA) approach,[26] where CCA[27–29] provides post processing of PLS[30–32] modeling results. PLS is a common statistical regression tool; it calculates an LV estimate as a linear combination $\mathbf{w}$ of the predictors, which has the maximal covariance with the response variable. The objective function for the first weight vector $\mathbf{w}_1$ can be expressed as follows

$$\begin{aligned} &\max \left(\mathbf{w}_1{}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{y}\mathbf{y}^{\mathrm{T}}\mathbf{X}\mathbf{w}_1\right) \\ &\text{subject to } \mathbf{w}_1{}^{\mathrm{T}}\mathbf{w}_1=1 \end{aligned} \tag{A1}$$

The maximum for (A1) is obtained at $\mathbf{w}_1$, the largest eigenvector of the matrix $\mathbf{X}^{\mathrm{T}}\mathbf{y}\mathbf{y}^{\mathrm{T}}\mathbf{X}$, resulting in the first LV, $\mathbf{t}_1=\mathbf{X}\mathbf{w}_1$. To obtain further weight vectors, the algorithm is repeated with $\mathbf{X}$ and $\mathbf{y}$ deflated by each LV. The classical PLS calculation procedure described by Lindgren et al.[32] was used.

One problem associated with PLS method requires special attention. Its objective is to model the variations in $\mathbf{X}$ and maximize their covariance with $\boldsymbol{y}$. Large covariance does not necessarily mean strong correlation. When the predictor matrix contains a considerable amount of process variations uncorrelated with the response, it is possible that the PLS LV model may capture the major systematic variations in the predictors $\mathbf{X}$ but have relatively weak correlation with the response $\mathbf{y}$. This situation leads to a complex model structure and an overfitting problem.

Unlike PLS, CCA inherently ignores the variations in $\mathbf{X}$ that are uncorrelated with $\mathbf{y}$ and directly maximizes the correlation of $\mathbf{X}$and $\boldsymbol{y}$. It uses a different objective function

$$\begin{aligned} &\max \left(\mathbf{w}_1{}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{y}\right) \\ &\text{subject to } \mathbf{w}_1{}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{w}_1=1 \end{aligned} \tag{A2}$$

The first weight $\mathbf{w}_1$ is the largest eigenvector of the matrix $\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}(\mathbf{y}^{\mathrm{T}}\mathbf{y})^{-1}\mathbf{y}^{\mathrm{T}}\mathbf{X}$. The maximum number of CCA components L_{cca} is min (L_x, J_y), where L_x and J_y are the number of predictor and response variables, respectively. In this article, a single response variable is considered and thus $L_{\mathrm{cca}}=1$. Compared with Eq. A1, the objective of CCA is to maximize correlation instead of covariance.

However, directly applying CCA to the raw measurements will lead to the common ill-conditioned problem resulting from the inclusion of $\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}$ in the calculations. Yu and

MacGregor[26] have suggested a two-step LV-based modeling algorithm
(PLS-CCA), where CCA is used as a post processing technique to further improve the PLS LVs. In this way, a parsimonious regression model with the same prediction ability as the standard PLS model can be obtained. Based on these considerations, the PLS-CCA algorithm is employed in this article to develop the regression model for glucose concentration prediction. Compared with the conventional AR and ARX modeling methods, the **y**-related variability in the glucose and exogenous input history is modeled by only a few LVs, which are calculated in order based on their relationship with future glucose values.

Appendix B: LV/LVX Prediction Model

By applying the PLS-CCA algorithm to the modeling data pair, $\{\mathbf{X}, \mathbf{y}\}$, the basic LV-based regression model can be readily calculated

$$\mathbf{t} = \mathbf{X}\mathbf{w} \tag{B1}$$

$$\mathbf{p}^{\mathrm{T}} = \left(\mathbf{t}^{\mathrm{T}}\mathbf{t}\right)^{-1}\mathbf{t}^{\mathrm{T}}\mathbf{X} \tag{B2}$$

$$\mathrm{X} = \hat{\mathbf{X}} + \mathbf{E} = \mathbf{t}\mathbf{p}^{\mathrm{T}} + \mathbf{E} \tag{B3}$$

$$q = \left(\mathbf{t}^{\mathrm{T}}\mathbf{t}\right)\mathbf{t}^{\mathrm{T}}\mathbf{y} \tag{B4}$$

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{f} = \mathbf{t}q + \mathbf{f} \tag{B5}$$

where, $\mathbf{w}(Z\times 1)$ is the PLS-CCA weight and $\mathbf{t}(N\times 1)$ is the calculated LV. Vector $\mathbf{t}$ in Eq. B1 is a linear combination of predictor variables via regression weights $\mathbf{w}(J_x\times 1)$, indicating the systematic variations in predictor variables that are closely related to the response variable. In the two-step modeling algorithm, PLS LVs are first modeled to generate a preliminary predictor LV set and then CCA is used to further process them. The number of retained PLS LVs can be as large as J_x. In this article, only one PLS-CCA LV is retained due to the single response variable **y**, no matter how many PLS LVs are calculated. In this way, the underlying systematic glycemic variability that is closely related with the future glucose concentration (response) is characterized by one LV and modeled from the historical data, the available glucose measurements, insulin infusion and CHO meal estimate.

In (B2), Vector $\mathbf{p}(Z\times 1)$ is the loading vector for the predictors and q is the loading coefficient scalar of the response obtained by regressing **y** on **t**; they indicate how much **t** contributes to the predictor variables and response, respectively. Based on the extracted LV, the future glucose prediction is calculated as $\hat{\mathbf{y}} = \mathbf{t}q$ and the predictor variations are modeled as $\hat{\mathbf{X}}$. $\mathbf{E}(N\times Z)$ and $\mathbf{f}(N\times 1)$ are the predictor and response residuals, respectively.

During the on-line application, the newly available predictor vector $\mathbf{x}_{\mathrm{new}}{}^{\mathrm{T}}(1\times Z)$ at each sampling instant can be expressed as $\left[\mathbf{g}_{\mathrm{new}}{}^{\mathrm{T}}(1\times L_{\mathrm{G}}), \mathbf{u}_{\mathrm{I,new}}{}^{\mathrm{T}}(1\times L_{\mathrm{I}}), \mathbf{u}_{\mathrm{M,new}}{}^{\mathrm{T}}(1\times L_{\mathrm{M}})\right]$ by including their antecedent information. The prediction $\hat{y}_{\mathrm{new}}$ is then made:

$$t_{\mathrm{new}} = \mathbf{X}_{\mathrm{new}}\mathbf{T}_{\mathbf{w}} \tag{B6}$$

$$\hat{y}_{\mathrm{new}} = t_{\mathrm{new}}\,q \tag{B7}$$

$$f_{\mathrm{new}} = y_{\mathrm{new}} - \hat{y}_{\mathrm{new}} \tag{B8}$$

Manuscript received Jun. 4, 2013, and revision received Sept. 10, 2013.