Effect of Input Excitation on the Quality of Empirical Dynamic Models for Type 1 Diabetes

Daniel A. Finan, Cesar C. Palerm, Francis J. Doyle III, and Dale E. Seborg
Dept. of Chemical Engineering, University of California, Santa Barbara, CA 93106

Howard Zisser, Wendy C. Bevier, and Lois Jovanovič
Sansum Diabetes Research Institute, Santa Barbara, CA 93105

DOI 10.1002/aic.11699
Published online March 23, 2009 in Wiley InterScience (www.interscience.wiley.com).

Accurate prediction of future blood glucose trends has the potential to significantly improve glycemic regulation in type 1 diabetes patients. A model-based controller for an artificial β-cell, for example, would determine the most efficacious insulin dose for the current sampling interval given available input–output data and model predictions of the resultant glucose trajectory. The two inputs most influential to the glucose concentration are bolused insulin and meal carbohydrates, which in practice are often taken simultaneously and in a specified ratio. This linear dependence has adverse effects on the quality of linear dynamic models identified from such data. On the other hand, inputs with greater degrees of excitation may force the subject into extreme hypoglycemia or hyperglycemia, and thus may be clinically unacceptable. Inputs with good excitation that do not endanger the subject are shown to result in models that can predict glucose trends reasonably accurately, 1–2 h ahead. © 2009 American Institute of Chemical Engineers *AIChE J,* 55: 1135–1146, 2009

Keywords: type 1 diabetes, artificial pancreas, linear dynamic models, model identification

Introduction

Diabetes mellitus is a metabolic disease characterized by an inability to adequately regulate blood glucose levels (glycemia) in the body. In particular, people with type 1 diabetes produce virtually no endogenous insulin, the hormone that promotes uptake of glucose from the bloodstream into the cells where it is metabolized. Insulin is of central importance to glycemic control; too much insulin leads to low glucose levels, or hypoglycemia, whereas too little insulin results in high glucose levels, or hyperglycemia. Both hypoglycemia and hyperglycemia are associated with health risks. Acute hypoglycemia can present immediate health threats like insulin shock and coma.[1] In the absence of insulin, on the other hand, the body cannot use glucose for energy and consequently is forced to metabolize fatty acids into ketones. This process, known as diabetic ketoacidosis, causes acute hyperglycemia, which in turn can cause thirst, vomiting, and shortness of breath, to name just a few symptoms. Finally, sustained hyperglycemia greatly increases a diabetes patient's susceptibility to long-term complications such as microvascular diseases[2] (e.g., nephropathy, neuropathy, and retinopathy) and cardiovascular disease.[3] Thus, to better control their glycemia, people with type 1 diabetes rely on exogenous insulin, delivered either through multiple daily injections or a continuous subcutaneous insulin infusion (CSII) pump.

Current address for C. C. Palerm: Medtronic Diabetes, Northridge, CA.
Correspondence concerning this article should be addressed to D. A. Finan at finan@engineering.ucsb.edu or D. E. Seborg at seborg@engineering.ucsb.edu.

As of 2005, one to two million people (0.35–0.70% of the population) in the United States were estimated to have type 1 diabetes.[4] Needless to say, the disease is responsible for much pain and suffering as well as the costs associated with the consequent mitigation and treatment. Diabetes patients can significantly reduce their risk of diabetes-related complications (and associated costs) through better control of blood glucose levels. Improving one's glycosylated hemoglobin (HbA1c) level, an indication of a patient's overall glycemic control, by 10–15% reduces the risk of microvascular complications by 40%.[4]

With the development of rapid-acting insulin preparations, CSII pumps, and improved accuracy and frequency of blood glucose measurements, type 1 diabetes patients are able to obtain significantly improved glycemic control when compared with conventional standards involving infrequent finger-stick measurements and infrequent manual delivery (i.e., via a syringe) of slower-acting insulin preparations.[5–24] Despite the improved technology and the improved treatment that it has made possible, painful finger-stick measurements, constant treatment decisions, and greatly increased risks of both short-term and long-term complications (when compared with an individual without diabetes)[4] are still very much a reality for even the most diligent diabetes patients.

Short of a cure, the most effective treatment of type 1 diabetes would be achieved using an "artificial β-cell," a medical device with three components: a continuous glucose sensor, a controller, and an insulin infusion pump. The controller of this device would automatically regulate a subject's glucose level by deciding the best insulin infusion rate for the current time based on the continuous glucose measurements, previously delivered insulin dosage information, and possibly meal information. A model-based controller would also take into account predictions of future glucose trends obtained from a reasonably accurate mathematical model of the subject's glucose-insulin dynamics. Thus, the development of accurate and (ideally) simple models is an important step in the development of an artificial β-cell.

The quality of the data from which a model is identified directly influences the predictive capability of that model.[25] In practical diabetes therapy, it is not uncommon to deliver the two most important inputs to the system–insulin boluses and meals–simultaneously and in the same ratio, known as the insulin-to-carbohydrate ratio (ICR). This delivery pattern makes the two inputs linearly dependent, greatly confounding a linear dynamic model's ability to distinguish the two very different effects of each input on the resultant glucose levels.[26] In many other applications of system identification, this problem is avoided by using highly excited inputs to the system, such as pseudo-random binary sequences (PRBS), and by making sure that each input is independent.[27] Unfortunately, these types of input excitation are infeasible when the "plant" to be identified is a person; highly excited input signals like PRBS are difficult to implement through an insulin pump and infeasible for meals. More importantly, the subjects' safety is of utmost importance and thus input signals that are too drastic are unacceptable if they force the subject into severe hypoglycemia or hyperglycemia. Therefore, the appropriate compromise is to increase the excitation quality of the input data while keeping the subject at acceptable glucose levels. This identification strategy is investigated *in silico* in the current article.

Empirical Models

Empirical, or data-driven, models of type 1 diabetes have been far less prevalent in the literature than physiological, first-principles-based models. Empirical models are less attractive than physiological models in the sense that the former often contain less physical meaning than the latter. The structures of empirical models are not derived from the underlying physiology, and their parameters (usually) do not correspond directly to physiological parameters (e.g., "insulin sensitivity" or "glucose distribution volume"). However, empirical models are more attractive than physiological models in the sense that they are often much simpler and their parameters readily estimated. Types of empirical models include dynamic, input–output models and neural networks.

The family of linear dynamic, input–output models includes autoregressive (AR), impulse-response (IR), autoregressive exogenous input (ARX), and autoregressive moving average exogenous input (ARMAX) models. The AR models[28,29] provide short-term predictions of future outputs (glucose concentrations) based on linear combinations of previous outputs. Because the exogenous inputs (meals and exogenous insulin) are omitted, only time-series glucose data are needed to estimate the parameters of AR models. These models can be easily identified recursively, wherein the model parameters are updated with each new sample to better capture the changing process dynamics. In standard least-squares recursive identification, a "forgetting factor" can be used to weight the most recent glucose measurements more heavily than previous measurements. Sparacino et al.[29] have investigated the effect of the forgetting factor on the ability of recursively identified AR models to accurately predict hypo- and hyperglycemic events 30 and 45 min into the future. In a sense, IR models are the opposite of AR models; IR models predict future glucose trends based only on linear combinations of the previous inputs. Parker et al.[30,31] have used these models in a simulated model predictive control (MPC) application.

In general, ARX models are widely used due to the incorporation of both the effects of exogenous inputs and the system dynamics. The recursive identification of ARX models is straightforward and has been performed with both diabetic human[32,33] and canine glucose and insulin data, but with infrequent measurements. ARMAX models, which are similar to ARX models but describe the prediction errors as moving averages of noise, have also been investigated for diabetic swine.[34]

Among empirical modeling techniques, neural networks (NNs) for type 1 diabetes have received the greatest degree of research attention. In simulation studies, Trajanoski et al.[35–37] have used NNs to identify nonlinear ARX (NARX) models. A few other studies[38–46] present novel contributions to the identification of NNs from diabetes data, such as using measurements of counterregulatory hormones to predict the future glucose trends[39] or incorporating fuzzy logic into the NNs.[45]

Other types of empirical models applied to diabetes data include qualitative, or fuzzy, logic systems,[47] Volterra series

models,[48] and response surface methodology and data mining.[49]

Physiological Model

The prominent physiological models of type 1 diabetes in the current literature range widely in linearity and complexity, from the aptly named three-state minimal model[50,51] to more comprehensive models with 20 or more state variables.[52,53] The model selected for the current research is the model of Hovorka et al.,[54–56] which presents a tradeoff between simplicity and physiological phenomena. The full nonlinear model contains 11 state variables and describes subcutaneous-to-intravenous insulin absorption as well as glucose absorption from a carbohydrate (CHO) meal.

Sources of nonlinearity in the model include "insulin actions" which describe the effects of insulin on glucose distribution/transport, disposal, and endogenous production. Other sources of nonlinearity include physiologically-based saturation effects, which switch "on" or "off" at certain glucose concentrations. See the Appendix for the model equations and a brief description.

Linear Dynamic Models

Model structures

For this diabetes application, the empirical models are linear dynamic input–output polynomial models that have two inputs, the insulin bolus amount u_{bol} and meal CHO amount u_{meal}, and one output, the glucose concentration G. The general model form is given by Eq. 1.[25]

$$A\left(q^{-1}\right)G(t) = \frac{B_1(q^{-1})}{F_1(q^{-1})}u_{\text{bol}}(t) + \frac{B_2(q^{-1})}{F_2(q^{-1})}u_{\text{meal}}(t) + \frac{C(q^{-1})}{D(q^{-1})}\varepsilon(t) \tag{1}$$

The numerators and denominators of the discrete-time transfer functions, A^{-1}, B_1/F_1, B_2/F_2, and C/D, are polynomials in q^{-1}, where q^{-1} is the backward shift operator, that is, $q^{-1}x(t) \equiv x(t-1)$, and t is the current sample time. The disturbance term $\varepsilon(t)$ is assumed to be zero-mean white Gaussian noise. The identification of these linear dynamic models involves specifying the model orders and estimating the coefficients of q^{-1} in these transfer functions.

Special cases of this general model can be obtained by simplifying one or more of the transfer functions. For example, an autoregressive exogenous input (ARX) model[32,33] is obtained for $C = D = F_1 = F_2 = 1$:

$$A\left(q^{-1}\right)G(t) = B_1\left(q^{-1}\right)u_{\text{bol}}(t) + B_2\left(q^{-1}\right)u_{\text{meal}}(t) + \varepsilon(t) \tag{2}$$

The ARX model is possibly the most widely used linear dynamic model because its parameters can be estimated via straightforward linear regression. However, because the denominators $D = F_1 = F_2 = 1$, both inputs and the disturbance term have the same denominator dynamics A.

Similarly, an autoregressive moving average exogenous input (ARMAX) model is obtained for $D = F_1 = F_2 = 1$:

$$A\left(q^{-1}\right)G(t) = B_1\left(q^{-1}\right)u_{\text{bol}}(t) + B_2\left(q^{-1}\right)u_{\text{meal}}(t) + C\left(q^{-1}\right)\varepsilon(t) \tag{3}$$

The ARMAX model allows the disturbance term to have numerator dynamics C which results in a moving average of recent values of $\varepsilon(t)$. The model parameters must be estimated from a prediction error method, and thus are susceptible to convergence to local minima in the parameter space.[25]

Finally, a Box-Jenkins (BJ) model is obtained by setting $A = 1$:

$$G(t) = \frac{B_1(q^{-1})}{F_1(q^{-1})}u_{\text{bol}}(t) + \frac{B_2(q^{-1})}{F_2(q^{-1})}u_{\text{meal}}(t) + \frac{C(q^{-1})}{D(q^{-1})}\varepsilon(t) \tag{4}$$

The BJ model allows both inputs and the disturbance term to have independent numerator and denominator dynamics. The parameters of this model also must be estimated using a prediction error method.

The order of each transfer function determines how many previous samples the current prediction depends on. For example, a first-order ARX model implies that $G(t)$ is a function of the input and output variables at time $t - 1$:

$$\begin{aligned} G(t) &= -a_1 q^{-1}G(t) + b_1 q^{-1}u_{\text{bol}}(t) + b_2 q^{-1}u_{\text{meal}}(t) + \varepsilon(t) \\ &= -a_1 G(t-1) + b_1 u_{\text{bol}}(t-1) + b_2 u_{\text{meal}}(t-1) + \varepsilon(t) \end{aligned} \tag{5}$$

In this article, the inputs $u_{\text{bol}}(t)$ and $u_{\text{meal}}(t)$ are modeled as impulses.

Prediction horizons

To investigate the relationship between model predictions and the prediction horizon, a range of prediction horizons was investigated in this research, including the following two limits.

- One-step-ahead (5-min-ahead) predictions use only the available input and output data to generate a prediction at the next sample time. The disturbance term $\varepsilon(t)$, then, represents the prediction errors.
- Infinite-step-ahead predictions use only the available input data and previous model predictions to generate future output predictions. Thus, since the model is given no information about the actual process output data, infinite-step-ahead predictions are noise-free simulations and $\varepsilon(t)$ is ignored.

Because of their potential relevance to model-based control for this application, 12-step and 24-step-ahead (1-h and 2-h-ahead) prediction horizons are also investigated.

Methods

Simulation conditions

First, a nominal dataset was generated in which three simultaneous boluses and meals in the same ratio, ICR_{N}, were administered to the virtual subject to simulate a 20-g (CHO) breakfast, 40-g lunch, and 60-g dinner over a period of 24 h. All datasets simulated 24 h of 5-min data with breakfast, lunch, and dinner administered at 8:00 AM, 12:00 PM, and 6:00 PM, respectively. Four additional datasets were generated in which different levels of input excitation were achieved by varying the ICR for the meals. These first five datasets are labeled as 1a–5a. Five more datasets were then simulated which were identical to 1a–5a except that the

Table 1. Datasets for Identification

Characteristic		Dataset									
		1a	2a	3a	4a	5a	1b	2b	3b	4b	5b
ICR_r^*	B	1	1	1	0.5	1	1	1	1	0.5	1
	L	1	1	0.5	1	0.5	0.5	0.5	0.25	0.5	0.25
	D	1	0.67	1	1	1.25	1	0.67	1	1	1.25
$(u_b)_r^\dagger$		–	–	–	–	–	0.5	0.5	0.25	0.5	0.25
κ		∞	14.4	10.4	19.1	7.5	6.8	7.5	6.3	6.5	5.3

*B, breakfast; L, lunch; and D, dinner. The ICR is listed relative to the respective nominal ICR, that is, $ICR_r = ICR/ICR_N$.
†For datasets 1b–5b, the correction bolus is listed relative to the nominal lunch bolus, that is, $(u_b)_r = u_b/u_N$.

lunch bolus was halved. Thus, the subject underbolused for lunch and became mildly hyperglycemic. A correction bolus was then taken 90 min after lunch in order to return the subject to normoglycemia before dinner. These latter datasets are labeled as 1b–5b. In all datasets, the virtual subject's glucose levels were kept between 65 and 200 mg/dL, a clinically acceptable glucose range. Gaussian white noise (σ = 3.3 mg/dL) was added to the simulated glucose values for all datasets. This noise level is consistent with previous studies[26,48] and is based on observations and analyses of subject data. Table 1 describes all 10 datasets in terms of the ICR used for each meal relative to the nominal case (i.e., dataset 1a) and, if applicable, the correction bolus relative to the nominal lunch bolus.

The level of input excitation of the 10 calibration datasets described in Table 1 is characterized by the condition number of the input matrix, $[u_{bol}(t)\ u_{meal}(t)]$, and is denoted by κ. The condition number of this matrix is a measure of the degree of linear dependence between the two inputs. It is the ratio of the largest singular value of the matrix to the smallest[57]; thus, a small condition number implies a small degree of linear dependence (high level of excitation). The limiting cases are $\kappa = 1$, implying that the inputs are orthogonal, and $\kappa = \infty$, implying that the inputs are linearly dependent. The latter case applies to dataset 1a.

Model identification and validation

For each of the 10 datasets, third-order through fifth-order ARX, ARMAX, and BJ models were identified using the MATLAB System Identification Toolbox.[25] To reduce the effect of a particular noise sequence on an identified model, models were identified from 10 different noise sequences. Thus, for each of the 10 datasets, 30 ARX models, 30 ARMAX models, and 30 BJ models were identified. The models were then validated on the other nine datasets. The datasets were evaluated based on the quality of the identified models, that is, the ability of the models to predict the other nine validation datasets. The quality of the models was assessed based on their ability to accurately predict validation data for the range of prediction horizons mentioned in the Prediction Horizons subsection.

Quantification of model prediction accuracy

An inherent difficulty in quantifying a model prediction is selecting suitable metrics that most closely characterize the "goodness" of the prediction. In this research, three metrics are used to quantify the accuracy of the model predictions: the FIT (closely related to the coefficient of determination R^2), the median relative absolute difference (RAD), and the Clarke Error Grid (CEG). The latter two metrics have often served different purposes than in the current article. Median and mean RAD values have quantified the agreement between a continuous glucose sensor and a glucose meter or laboratory reference.[58,59] The CEG, on the other hand, was initially developed to quantify the agreement between subject estimates of their glucose concentration and the meter values.[60] Later it was used to ascertain the accuracy of meter readings when compared with a reference value.[61] Recently, the CEG and a subsequent augmentation, the continuous glucose-error grid analysis (CG-EGA), were used to assess the performance of continuous glucose sensors when compared with reference values[5,8,59,62–66] (See Clarke et al.[60] for a history of the CEG.). Regardless of the specific application, the CEG compares an "estimated" glucose reading to a "reference" reading, and exploits properties specific to diabetes. Hovorka et al.[55] analogously applied the CEG to determine the accuracy of model predictions (the "estimate") when compared with frequent glucose readings (the "reference"). This use of the CEG is appropriate and adopted in this article.

The FIT of a model prediction is a statistical metric that quantifies how much of the variation in the data is explained by the model prediction[25]:

$$\text{FIT} = \left(1 - \frac{\left\|G - \hat{G}\right\|}{\left\|G - \overline{G}\right\|}\right) \times 100\% \tag{6}$$

where G is the vector of measured (or in this case simulated via the physiological model) outputs, $\hat{G}$ is the vector of model predictions, $\overline{G}$ is the mean of the measured output, and the norms are Euclidean. Thus, FIT = 100% is obtained for a perfect prediction, FIT = 0 can be obtained by predicting the mean of the measured output at every sample, and FIT $<$ 0 is obtained for very poor model predictions. The FIT is very closely related to the coefficient of determination, R^2:

$$R^2 = \left(1 - \frac{\left\|G - \hat{G}\right\|^2}{\left\|G - \overline{G}\right\|^2}\right) \times 100\% \tag{7}$$

Thus, when the model predictions are accurate, the second term of Eq. 6 is small (i.e., much less than one), and the second term of Eq. 7 is very small. This makes FIT a more sensitive metric when the model predictions are good, and is the reason FIT is used to quantify model predictions in this article.

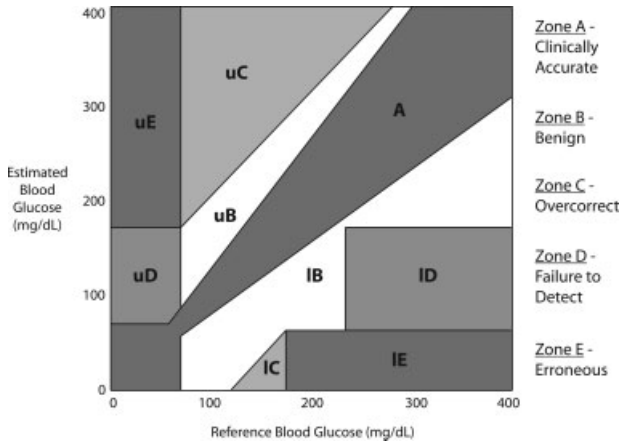

Figure 1. The Clarke Error Grid (CEG).[60]

Upper (u) and lower (l) regions for each Zone A–E. The zones are explained further in the text.

The relative absolute difference, RAD, of a prediction scales the absolute error of a prediction by the measured value at that sample time:

$$\mathrm{RAD}(t) = \frac{\left|G(t) - \hat{G}(t)\right|}{G(t)} \times 100\% \tag{8}$$

Thus, a 30 mg/dL error when the subject is at 300 mg/dL is weighted the same as a 10 mg/dL error when the subject is at 100 mg/dL (i.e., for both, RAD = 10%). Because, in general, the median of a dataset is less sensitive to outliers than the mean, we will consider the median RAD in this article.

Finally, the CEG takes into account the absolute values of the prediction and measured value at that sample, and the effects that the corresponding error might have on an insulin dosage decision. Each prediction-measured value pair falls into one of five "zones," labeled A–E (See Figure 1.). Data that fall into Zone A are considered clinically accurate. Data in Zone B are benignly erroneous (i.e., having little or no effect on a treatment decision). Data in Zone C may prompt an overcorrected treatment decision (i.e., administering too much insulin or ingesting too much CHO). Conversely, data in Zone D imply that a severe hypoglycemic or hyperglycemic episode has gone undetected, effectively resulting in an undercorrected treatment decision. Lastly, data in Zone E are highly erroneous. The metric used to quantify model accuracy is the percentage of model predictions in Zone A.

Results

Tables 2–4 display the identification results based on FIT values, median RAD values, and CEG Zone A values, respectively. The values represent medians over all models and validation datasets. For example, the suite of models identified from dataset 1a has been validated on datasets 2a–5a and 1b–5b. The average of these validation metrics is the value corresponding to "Calibration Dataset 1a" in the tables.

Comparisons among model types

Based on overall median values, the ARX and ARMAX models predict the validation data more accurately than the BJ models for a 1-step (5-min) prediction horizon. However, given the substantially larger time constants associated with the glycemic response to insulin delivery and CHO absorption from a meal, 1-step-ahead predictions-even the most accurate ones–are of limited utility in a closed-loop, model-based control scheme.

For the longer prediction horizons (1-h, 2-h, and infinite), regardless of the prediction metric, the ARMAX models show the best prediction capabilities, followed by the BJ models, and then the ARX models. For a 1-h prediction horizon, for example, the overall median FIT values for ARMAX, BJ, and ARX models were 77%, 73%, and 51%, respectively (the corresponding median RAD values were 4.7, 5.2, and 7.6%, whereas the corresponding frequencies of predictions in Zone A of the CEG were 99%, 97%, and 89%). Using a 2-h prediction horizon, the overall median FIT values for ARMAX, BJ, and ARX models were 65%, 59%, and 25%, respectively (the corresponding median RAD values were 6.3, 7.4, and 12.9%, whereas the corresponding

Table 2. Average Validation Results Based on FIT Values (%) for All Prediction Horizons, Model Types, and Calibration Datasets*

		Calibration Dataset										
Prediction Horizon	Model	1a	2a	3a	4a	5a	1b	2b	3b	4b	5b	Median
5 min	ARX	87	86	86	87	86	87	86	86	86	87	86
	ARMAX	84	83	86	77	86	86	86	86	83	85	85
	BJ	82	78	**86**	82	85	85	84	**86**	80	84	84
1 h	ARX	51	51	50	50	50	52	**55**	52	53	51	51
	ARMAX	67	74	78	70	**79**	77	**79**	**79**	78	76	77
	BJ	44	63	75	59	77	75	70	**78**	71	75	73
2 h	ARX	25	25	22	23	23	27	**32**	26	28	25	25
	ARMAX	48	61	68	58	**69**	60	67	67	66	64	65
	BJ	20	47	59	47	63	62	52	**66**	59	60	59
Infinite	ARX	−3	−2	−2	−4	−2	−4	**3**	1	2	−1	−2
	ARMAX	27	42	54	45	**55**	47	49	49	54	53	49
	BJ	−4	36	36	32	38	35	42	50	**51**	35	36
Median		46	56	64	54	66	61	61	**67**	63	62	

*Shaded values represent the worst validation results for a given model type and prediction horizon; boldfaced values represent the best.

Table 3. Average Validation Results Based on Median RAD Values (%) for All Prediction Horizons, Model Types, and Calibration Datasets*

		Calibration Dataset										
Prediction Horizon	Model	1a	2a	3a	4a	5a	1b	2b	3b	4b	5b	Median
5 min	ARX	2.8	2.9	2.9	2.9	2.8	2.8	2.9	2.9	2.8	2.8	2.8
	ARMAX	3.2	3.3	2.6	2.8	**2.5**	**2.5**	2.6	**2.5**	3.2	2.9	2.7
	BJ	3.2	4.2	2.7	2.9	**2.6**	2.8	3.1	**2.6**	4.1	3.0	3.0
1 h	ARX	7.2	7.8	8.0	7.5	7.8	**7.1**	7.3	7.7	7.3	7.7	7.6
	ARMAX	5.7	5.2	4.3	5.0	**4.1**	4.9	4.2	**4.1**	4.6	4.8	4.7
	BJ	6.4	7.1	4.7	7.4	**4.4**	5.1	6.2	4.5	5.2	5.0	5.2
2 h	ARX	12.1	13.3	13.8	13.0	13.6	**11.6**	11.7	12.8	11.8	13.0	12.9
	ARMAX	8.4	7.1	**5.7**	7.0	5.8	7.1	6.0	5.8	6.0	6.6	6.3
	BJ	10.2	9.1	7.0	10.3	**6.3**	7.5	9.3	6.4	7.2	6.9	7.4
Infinite	ARX	21.9	24.7	25.3	23.3	24.0	**21.8**	23.7	24.9	23.0	24.0	23.8
	ARMAX	10.5	11.7	9.0	10.2	**7.6**	9.4	9.5	8.9	7.9	8.1	9.2
	BJ	16.6	12.1	11.2	12.8	12.6	10.9	10.9	9.9	**9.5**	11.7	11.4
Median		7.8	7.5	6.4	7.5	**6.1**	7.1	6.8	**6.1**	6.6	6.8	

*Shaded values represent the worst validation results for a given model type and prediction horizon; boldfaced values represent the best.

frequencies of predictions in Zone A of the CEG were 94%, 92%, and 70%). Again, the results for the 1-h and 2-h prediction horizons are emphasized because they are appropriate for a model-based controller for an artificial β-cell.

Comparisons among calibration datasets

Tables 2–4 indicate that the dataset 1a often results in models whose predictive capabilities are significantly worse relative to models identified from the other nine datasets. For example, using ARMAX models and a 2-h prediction horizon, the models identified from dataset 1a had an average FIT value of 48%, whereas the models identified from the other datasets averaged 58–69% (the corresponding median RAD values were 8.4% for dataset 1a and 5.7–7.1% for the other nine datasets, whereas the corresponding frequencies of predictions in Zone A of the CEG were 87% for dataset 1a and 92–98% for the other nine datasets).

Conversely, some calibration datasets often resulted in the best models. Depending on the metric used to quantify the predictions, datasets 5a and 1b–4b produced consistently accurate models. One of these datasets used three different ICR values for the three meals (5a), three used two different ICR values and a correction bolus (1b, 3b, and 4b), and one used three different ICR values and a correction bolus (2b). Thus, these datasets had small condition numbers, ranging from $\kappa = 6.3$ to $\kappa = 7.5$, relative to datasets 1a–4a. In general, these latter datasets with lower levels of excitation (large condition numbers) produced less accurate models. Even with the highest degree of excitation (smallest condition number), dataset 5b often did not produce the best models; it never, however, produced the worst.

To further illustrate the effect of the condition number κ, a property of the calibration dataset, on the overall performance of the identified models, Figures 2 and 3 display the overall median 1-h and 2-h FIT values for validation data, respectively. The overall median FIT values for each type of model are plotted as a function of κ. Also shown are the least-squares trend lines and the associated Pearson correlation coefficients r_{xy}, a statistical metric which ranges from -1 (perfectly negatively correlated) to 0 (not correlated) to $+1$ (perfectly positively correlated). Figure 2 illustrates that for 1-h predictions, the performance of the ARMAX and BJ models is strongly correlated with condition number κ ($r_{xy} = -0.86$ and $r_{xy} = -0.88$, respectively). Figure 3 shows that for 2-h predictions, these correlations are slightly weaker

Table 4. Average Validation Results Based on CEG Values (% of Predictions in Zone A) for All Prediction Horizons, Model Types, and Calibration Datasets*

		Calibration Dataset										
Prediction Horizon	Model	1a	2a	3a	4a	5a	1b	2b	3b	4b	5b	Median
5 min	ARX	100	100	100	100	100	100	100	100	100	100	100
	ARMAX	100	99	100	99	100	100	100	100	98	100	100
	BJ	100	97	100	99	100	100	99	100	97	99	99
1 h	ARX	89	88	89	89	89	**90**	**90**	89	**90**	89	89
	ARMAX	98	99	100	97	100	100	100	100	99	98	99
	BJ	89	93	**99**	92	**99**	98	92	**99**	95	98	97
2 h	ARX	70	69	69	69	69	71	**72**	70	71	69	70
	ARMAX	87	93	97	92	**98**	93	96	96	96	92	94
	BJ	73	83	93	81	**95**	94	81	94	91	94	92
Infinite	ARX	**44**	35	34	38	38	**44**	40	36	41	36	38
	ARMAX	74	75	81	77	**89**	81	80	81	84	84	81
	BJ	54	74	73	72	72	74	72	79	**85**	75	73
Median		88	90	95	90	**97**	94	91	95	93	93	

*Shaded values represent the worst validation results for a given model type and prediction horizon; boldfaced values represent the best.

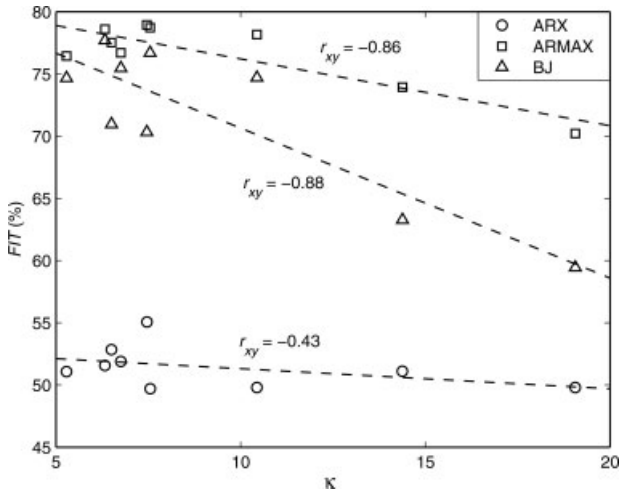

Figure 2. Overall median 1-h predictions for validation data plotted as a function of the condition number κ of the calibration dataset.

Least-squares trend lines and the associated correlation coefficients r_{xy} are also shown.

($r_{xy} = -0.64$ and $r_{xy} = -0.82$ for ARMAX and BJ models, respectively). The performance of the ARX models is correlated with κ to a significantly lesser degree, with $r_{xy} = -0.43$ and $r_{xy} = -0.39$ for 1-h and 2-h predictions, respectively.

Comparisons among prediction metrics

Inspection of Tables 2–4 indicates that the three metrics are, in general, qualitatively similar. In fact, they are highly correlated. Figure 4 shows the plots of each metric vs. each other metric for all model fits in Tables 2–4. The least-squares trend lines and associated Pearson correlation coefficients are also shown. For all the three relationships, $|r_{xy}| \geq 0.96$. The range of values for the three metrics, however, varies greatly. FIT ranges from −4 to 87%, median RAD from 2.5 to 25.3%, and the CEG Zone A from 34 to 100%. FIT is the metric with greatest range and thus is more sensitive than the other two metrics.

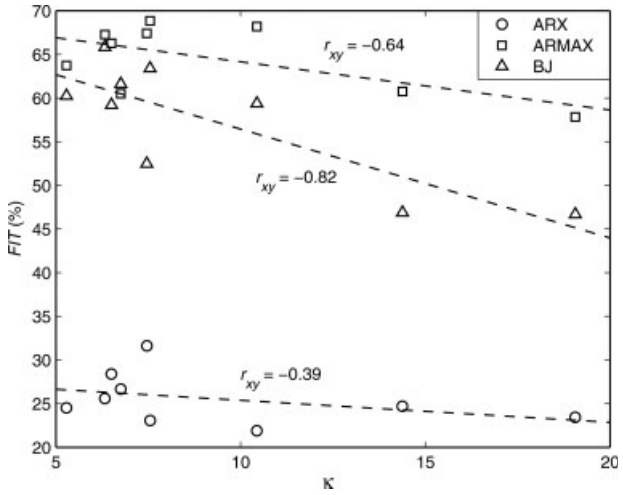

Figure 3. Overall median 2-h predictions for validation data plotted as a function of the condition number κ of the calibration dataset.

Least-squares trend lines and the associated correlation coefficients r_{xy} are also shown.

Comparisons of the dynamic characteristics of the identified models

Tables 5 and 6 summarize dynamic characteristics of the identified models that are relevant to the diabetes application, namely, the impulse-response characteristics. Table 5 lists two insulin bolus response characteristics: the maximum change in glucose concentration ($\Delta G_{\min}$), and the time-to-maximum change in glucose concentration ($t_{\min}$). Table 6 lists two meal-related characteristics: the peak change in glucose concentration ($\Delta G_{\max}$), and the time-to-peak change in glucose concentration ($t_{\max}$). To facilitate analysis of Tables 5 and 6, the corresponding ranges of the dynamic characteristics for the physiological model are included. These ranges were determined by simulating boluses and meals over the range of glucose concentrations of the calibration datasets, that is, 65 to 200 mg/dL.

Tables 5 and 6 indicate that datasets with higher levels of excitation tended to produce models with more accurately identified dynamic characteristics (cf. Table 1). It is evident from Tables 5 and 6 that the ARX models consistently had poorly identified dynamic characteristics; the magnitudes of $\Delta G_{\min}$ and $\Delta G_{\max}$ were grossly underestimated and the times-to-maximum change, $t_{\min}$ and $t_{\max}$, were very small. The ARMAX and BJ models frequently had reasonable dynamic characteristics for the datasets with high levels of excitation. Regardless of the type of model, the characteristics of the models identified from dataset 1a were inaccurate.

Transient characteristics of model predictions

Figure 5 displays representative 1-h and 2-h-ahead validation results and FIT values for each type of model. The predictions are representative in the sense that their FIT values

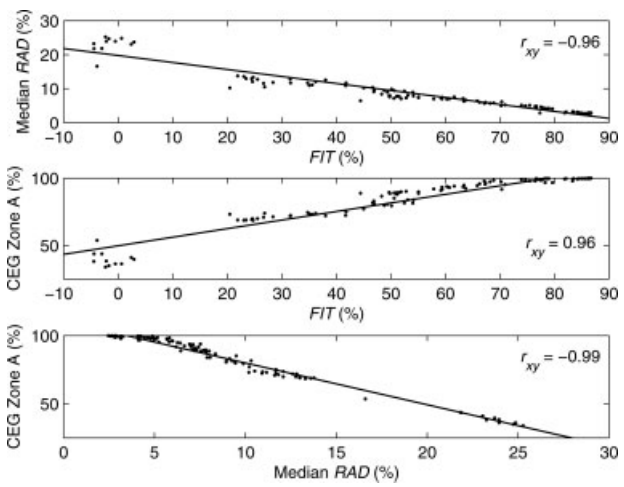

Figure 4. Correlations among prediction metrics.

Each prediction metric is plotted against the other metrics: median RAD vs. FIT (top), CEG Zone A vs. FIT (middle), and CEG Zone A vs. median RAD (bottom). Least-squares trend lines for each relationship and the associated correlation coefficients r_{xy} are also shown.

Table 5. Impulse-Response Characteristics of Identified Models: Insulin Boluses*

Dynamic Characteristic	Model	Calibration Dataset									
		1a	2a	3a	4a	5a	1b	2b	3b	4b	5b
ΔG_{min}† (mg/dL)	ARX	0	−12	−5	−2	−2	−3	−2	−3	−3	−3
	ARMAX	0	−13	**−28**	**−32**	**−28**	**−22**	**−21**	**−23**	**−31**	**−35**
	BJ	−7	−10	−7	−16	**−25**	**−23**	**−25**	**−25**	**−27**	**−21**
		Hovorka model range: −70 to −21									
t_{min}‡ (min)	ARX	5	20	20	20	15	15	15	15	15	25
	ARMAX	450	115	**145**	110	**145**	125	**135**	**135**	**140**	**135**
	BJ	115	**150**	**190**	**185**	**195**	**155**	**155**	**160**	**165**	130
		Hovorka model range: 135 to 290									

*Boldfaced values represent quantities within the corresponding range of the physiological model.
†Maximum glucose change after 1-U insulin impulse (i.e., bolus).
‡Time-to-maximum glucose change after 1-U insulin impulse (i.e., bolus).

are very close to the median FIT values for that prediction horizon and model type (see Table 2). In this case, the models were identified from dataset 5b. The ARX models predict the glucose dynamics immediately following a meal and bolus reasonably accurately but show a strong "shadow" effect, predicting the peak excursion magnitudes reasonably well but significantly delayed.

The ARMAX models show the best predictive capabilities of the three types of models. They predict the glucose excursions fairly accurately, capturing the dynamics immediately following a meal as well as the timing and magnitudes of the peaks.

The BJ models also predict the data reasonably accurately but are susceptible to erratic predictions as seen immediately following lunch in the 1-h-ahead prediction. These erratic predictions might be the results of convergence of the model parameters to local minima.

To contrast the characteristics of the models identified from dataset 1a with those identified from the other nine datasets, representative transient responses for ARMAX models identified from datasets 1a and 3b for 1-h and 2-h prediction horizons are plotted in Figure 6. Although both models capture accurately the timing and magnitude of the breakfast and lunch excursions, the model identified from dataset 3b predicts the dinner excursion much more accurately.

Validation of models for more realistic conditions

The datasets used to identify the models were characterized by a high degree of glucose control (i.e., 65 mg/dL $< G <$ 200 mg/dL), which would likely be an aspect of an ethical, clinically acceptable protocol for model identification. However, this degree of control is representative of only a small fraction of type 1 diabetes subjects in everyday, ambulatory conditions. Studies using continuous glucose monitoring devices have shown that the average member of the control group (i.e., those blinded to the continuous glucose measurements) spends much time outside the 65–200 mg/dL range. Figure 7 summarizes the results for the control group of one such study.[8] Thus, three additional datasets, R_1–R_3, were generated so that the times spent in respective glucose ranges were comparable to realistic values. These datasets are also summarized in Figure 7.

Validation results for the ARMAX models identified from dataset 3b for the more realistic datasets are listed in Table 7. When compared with the fits of the same models on the original datasets (see Table 2), the model fits to the more realistic datasets are comparable and in some cases even better (i.e., larger FIT). This may be explained by the fact that the more realistic datasets have more variability about their means, and thus have more variability to be explained by the models. This nuance is especially pertinent using relatively short prediction horizons. The FIT values of the infinite-step-ahead predictions for the more realistic datasets are significantly lower than those of the original datasets, illustrating the nonlinearity of the model, especially in the hypo- and hyperglycemic regions. Representative 1-h and 2-h-ahead validation results for each of the realistic datasets are shown in Figure 8. The dynamic trends are captured nicely, and the predictions are likely suitable for acceptable model-based controller performance.

Table 6. Impulse-Response Characteristics of Identified Models: Meals*

Dynamic Characteristic	Model	Calibration Dataset									
		1a	2a	3a	4a	5a	1b	2b	3b	4b	5b
ΔG_{max}† (mg/dL)	ARX	4	7	5	5	5	4	5	5	5	5
	ARMAX	11	21	**26**	**27**	**24**	**23**	**24**	**25**	**28**	**28**
	BJ	5	17	17	19	**23**	**23**	**24**	**25**	**26**	19
		Hovorka model range: 23–39									
t_{max}‡ (min)	ARX	40	25	30	25	25	40	50	50	45	45
	ARMAX	100	115	**125**	115	**130**	115	120	**130**	**125**	**125**
	BJ	80	**125**	**125**	**125**	**125**	120	**125**	**125**	**125**	115
		Hovorka model range: 125–180									

*Boldfaced values represent quantities within the corresponding range of the physiological model.
†Peak glucose change after 10-g CHO meal impulse.
‡Time-to-peak glucose change after 10-g CHO meal impulse.

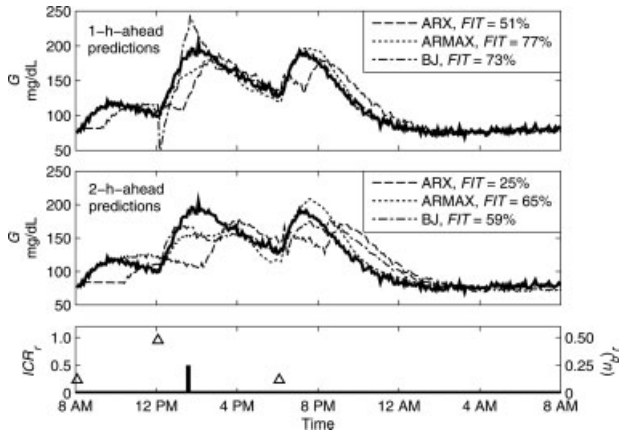

Figure 5. Representative validation results for each type of model for 1-h-ahead (top) and 2-h-ahead (middle) predictions.

The calibration dataset is 5b and the validation dataset is 3b. The inputs are shown (bottom) as the ICR for each meal relative to the nominal ICR for each meal, ICR_r (triangles, left axis), and the correction bolus relative to the nominal lunch bolus $(u_b)_r$ (bar, right axis).

Conclusions

The performance of a model-based controller for an artificial β-cell depends highly on the accuracy and robustness of the model used to predict future glucose trends. To identify an accurate dynamic model, or suite of models, the calibration dataset(s) must have good input excitation and a clear causal relationship between the two most important inputs, insulin boluses and meal CHO, and the output, blood glucose. For typical type 1 diabetes data, these two inputs occur simultaneously and in a specified, constant ratio, thus making their effects on the glucose concentration indistinguishable. Consequently, the degree of input excitation is poor and the identified models are inaccurate. On the other hand, separated inputs (i.e., meals without boluses and vice versa) will likely result in accurate models but can also result in unsafe and clinically unacceptable hypoglycemia or hyperglycemia.

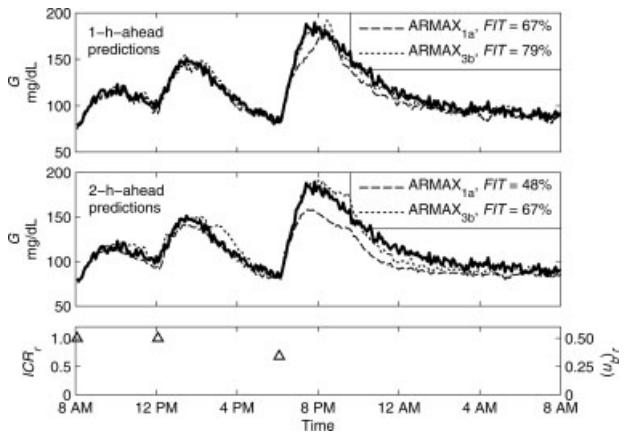

Figure 6. Representative validation results for the ARMAX models identified from datasets 1a and 3b.

The validation dataset is 2a. 1-h-ahead (top) and 2-h-ahead (middle) predictions. The inputs are shown (bottom) as the ICR for each meal relative to the nominal ICR for each meal, ICR_r (triangles, left axis); there was no correction bolus for this particular validation dataset.

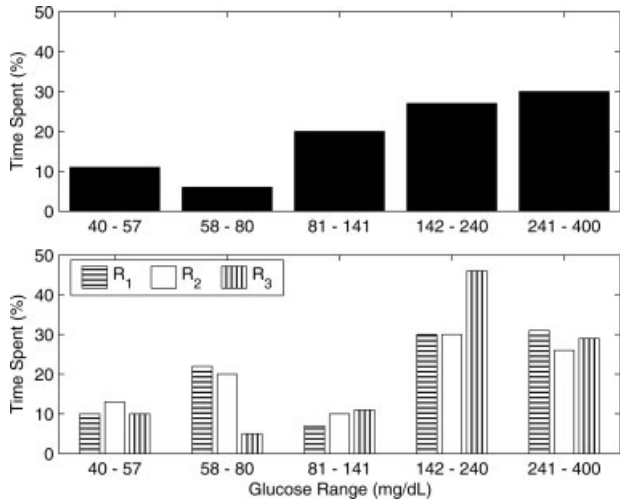

Figure 7. Average time spent by diabetes subjects in various glucose ranges[8] (top) and time spent in the same ranges for the more realistic datasets (bottom).

Table 7. Validation Results (FIT, %) for the ARMAX Models Identified from Dataset 3b for the More Realistic Validation Datasets

Prediction Horizon	R_1	R_2	R_3
5 min	88	89	78
1 h	82	80	68
2 h	70	67	52
Infinite	38	38	27

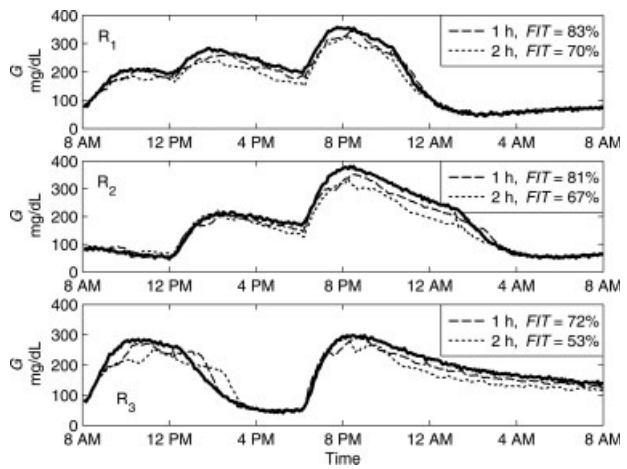

Figure 8. Representative validation results for the ARMAX models identified from dataset 3b for the more realistic datasets.

1-h-ahead (dashed) and 2-h-ahead (dotted) predictions of R_1 (top), R_2 (middle), and R_3 (bottom).

In this extensive simulation study, a compromise between good input excitation (e.g., varying the insulin-to-carbohydrate ratio and/or administering a correction bolus in a state of mild hyperglycemia) and safe levels of glucose (e.g., 65 mg/dL $< G <$ 200 mg/dL) resulted in a clinically acceptable method to obtain accurate predictive models, at least, relative to those obtained from typical type 1 diabetes data. These accurate models were validated on other, similar datasets with good results, and further validated on more realistic simulation data in which the time spent in hypoglycemia and hyperglycemia constituted a significant portion of each day. The validation of the best models on these latter datasets also gave acceptable predictions for a range of horizons including 1 h and 2 h, which may be appropriate for a model-based controller.

Acknowledgments

The authors gratefully acknowledge financial support from the National Institutes of Health, grant R21-DK069833-02, and from the Juvenile Diabetes Research Foundation, grant 22-2006-1115.

Literature Cited

1. Parker RS, Doyle FJ III, Peppas NA. The intravenous route to blood glucose control. *IEEE Eng Med Biol Mag*. 2001;20:65–73.
2. The Diabetes Control and Complications Trial (DCCT) Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl J Med*. 1993;329:977–986.
3. The Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications (DCCT/EDIC) Study Research Group. Intensive diabetes treatment and cardiovascular disease in patients with type 1 diabetes. *N Engl J Med*. 2005;353: 2643–2653.
4. Centers for Disease Control and Prevention. *National Diabetes Fact Sheet: General Information and National Estimates on Diabetes in the United States, 2005*. Atlanta, GA: U.S. Department of Health and Human Services, 2005.
5. Garg S, Zisser H, Schwartz S, Bailey T, Kaplan R, Ellis S, Jovanovic L. Improvement in glycemic excursions with a transcutaneous, real-time continuous glucose sensor. *Diabetes Care*. 2006;29:44–50.
6. Deiss D, Bolinder J, Riveline J-P, Battelino T, Bosi E, Tubiana-Rufi N, Kerr D, Phillip M. Improved glycemic control in poorly controlled patients with type 1 diabetes using real-time continuous glucose monitoring. *Diabetes Care*. 2006;29:2730–2732.
7. Lagarde WH, Barrows FP, Davenport ML, Kang M, Guess HA, Calikoglu AS. Continuous subcutaneous glucose monitoring in children with type 1 diabetes mellitus: a single-blind, randomized, controlled trial. *Pediatr Diabetes*. 2006;7:159–164.
8. Garg SK, Schwartz S, Edelman SV. Improved glucose excursions using an implantable real-time continuous glucose sensor in adults with type 1 diabetes. *Diabetes Care*. 2004;27:734–738.
9. Einhorn D. Advances in diabetes for the millennium: insulin treatment and glucose monitoring. *MedGenMed*. 2004;6 Suppl 3:8.
10. Kubiak T, Hermanns N, Schreckling HJ, Kulzer B, Haak T. Assessment of hypoglycaemia awareness using continuous glucose monitoring. *Diabet Med*. 2004;21:487–490.
11. Weinzimer SA, Tamborlane WV, Chase HP, Garg SK. Continuous glucose monitoring in type 1 diabetes. *Curr Diab Rep*. 2004;4:95–100.
12. Schaepelynck-Belicar P, Vague P, Simonin G, Lassmann-Vague V. Improved metabolic control in diabetic adolescents using the continuous glucose monitoring system (CGMS). *Diabetes Metab*. 2003;29: 608–612.
13. Ludvigsson J, Hanas R. Continuous subcutaneous glucose monitoring improved metabolic control in pediatric patients with type 1 diabetes: a controlled crossover study. *Pediatrics*. 2003;111:933–938.
14. Kaufman FR, Austin J, Neinstein A, Jeng L, Halvorson M, Devoe DJ, Pitukcheewanont P. Nocturnal hypoglycemia detected with the continuous glucose monitoring system in pediatric patients with type 1 diabetes. *J Pediatr*. 2002;141:625–630.
15. Schiaffini R, Ciampalini P, Fierabracci A, Spera S, Borrelli P, Bottazzo GF, Crino A. The continuous glucose monitoring system (CGMS) in type 1 diabetic children is the way to reduce hypoglycemic risk. *Diabetes Metab Res Rev*. 2002;18:324–329.
16. Kerr D. Continuous blood glucose monitoring: detection and prevention of hypoglycaemia. *Int J Clin Pract Suppl*. 2001;123:43–46.
17. Bode BW, Hirsch IB. Using the continuous glucose monitoring system to improve the management of type 1 diabetes. *Diabetes Technol Ther*. 2000;2Suppl 1:S43–S48.
18. Bode BW, Gross TM, Thornton KR, Mastrototaro JJ. Continuous glucose monitoring used to adjust diabetes therapy improves glycosylated hemoglobin: a pilot study. *Diabetes Res Clin Pract*. 1999; 46:183–190.
19. Heptulla RA, Allen HF, Gross TM, Reiter EO. Continuous glucose monitoring in children with type 1 diabetes: before and after insulin pump therapy. *Pediatr Diabetes*. 2004;5:10–15.
20. Gross TM, Kayne D, King A, Rother C, Juth S. A bolus calculator is an effective means of controlling postprandial glycemia in patients on insulin pump therapy. *Diabetes Technol Ther*. 2003;5:365–369.
21. Pickup J, Keen H. Continuous subcutaneous insulin infusion at 25 years. *Diabetes Care*. 2002;25:593–598.
22. Bode BW, Sabbah HT, Gross TM, Fredrickson LP, Davidson PC. Diabetes management in the new millennium using insulin pump therapy. *Diabetes Metab Res Rev*. 2002;18Suppl 1:S14–S20.
23. Zinman B. Basal insulin replacement and use of rapid-acting insulin analogues in patients with type 1 diabetes. *Endocr Pract*. 2000;6: 88–92.
24. Boland EA, Grey M, Oesterle A, Fredrickson L, Tamborlane WV. Continuous subcutaneous insulin infusion—a new way to lower risk of severe hypoglycemia, improve metabolic control, and enhance coping in adolescents with type 1 diabetes. *Diabetes Care*. 1999;22: 1779–1784.
25. Ljung L. *System Identification: Theory For The User*, 2nd ed. Upper Saddle River, NJ: Prentice Hall PTR, 1999.
26. Finan DA, Zisser H, Jovanovic L, Bevier WC, Seborg DE. Practical issues in the identification of empirical models from simulated type 1 diabetes data. *Diabetes Technol Ther*. 2007;9:438–450.
27. Seborg DE, Edgar TF, Mellichamp DA. *Process Dynamics and Control*, 2nd ed. New York: Wiley, Inc., 2004.
28. Bremer T, Gough DA. Is blood glucose predictable from previous values? *Diabetes*. 1999;48:445–451.
29. 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.
30. Parker RS, Doyle FJ III, Peppas NA. A model-based algorithm for blood glucose control in type I diabetic patients. *IEEE Trans Biomed Eng*. 1999;46:148–157.
31. Parker RS, Gatzke EP, Doyle FJ III. Advanced model predictive control (MPC) for type I diabetic patient blood glucose control. *Proc Am Control Conf*. 2002:3483–3487.
32. Bellazzi R, Siviero C, Stefanelli M, De Nicolao G. Adaptive controllers for intelligent monitoring. *Artif Intell Med*. 1995;7:515–540.
33. Riva A, Bellazzi R. Learning temporal probabilistic causal models from longitudinal data. *Artif Intell Med*. 1996;8:217–234.
34. El-Khatib FH. System identification and adaptive closed-loop glucose control in a type 1 diabetic swine model. Ph.D. Dissertation, *University of Illinois at Urbana-Champaign*, 2005.
35. Trajanoski Z, Regittnig W, Wach P. Neural predictive controller for closed-loop control of glucose using the subcutaneous route: a simulation study. *Control Eng Pract*. 1997;5:1727–1730.
36. Trajanoski Z, Regittnig W, Wach P. Simulation studies on neural predictive control of glucose using the subcutaneous route. *Comput Methods Programs Biomed*. 1998;56:133–139.
37. Trajanoski Z, Wach P. Neural predictive controller for insulin delivery using the subcutaneous route. *IEEE Trans Biomed Eng*. 1998;45:1122–1134.
38. El-Jabali AK. Neural network modeling and control of type 1 diabetes mellitus. *Bioprocess Biosyst Eng*. 2005;27:75–79.

39. Prank K, Jurgens C, von zur Muhlen A, Brabant G. Predictive neural networks for learning the time course of blood glucose levels from the complex interaction of counterregulatory hormones. *Neural Comput*. 1998;10:941–953.
40. Tresp V, Moody J, Delong W-R. Neural network modeling of physiological processes. In: Hanson SJ, Petsche T, Kearns M, Rivest RL, editors. *Computational Learning Theory and Natural Learning Systems*. Cambridge, MA: The MIT Press, 1994;Vol.II:363–378.
41. Pender JE. *Modelling of blood glucose levels using artificial neural networks*. Ph.D. Dissertation, University of Strathclyde, UK, 1997.
42. Sandham WA Hamilton DJ, Japp A, Patterson K. Neural network and neuro-fuzzy systems for improving diabetes therapy. *Conf Proc IEEE Eng Med Biol Soc*. 1998;20:1438–1441.
43. Liszka-Hackzell JJ. Prediction of blood glucose levels in diabetic patients using a hybrid AI technique. *Comput Biomed Res*. 1999;32:132–144.
44. Mougiakakou SG, Nikita KS. A neural network approach for insulin regime and dose adjustment in type 1 diabetes. *Diabetes Technol Ther*. 2000;2:381–389.
45. Otto E, Semotok C, Andrysek J, Basir O. An intelligent diabetes software prototype: predicting blood glucose levels and recommending regimen changes. *Diabetes Technol Ther*. 2000;2:569–576.
46. Schlotthauer G, Gamero LG, Torres ME, Nicolini GA. Modeling, identification and nonlinear model predictive control of type I diabetic patient. *Med Eng Phys*. 2005;28:240–250.
47. Bellazzi R, Ironi L, Guglielmann R, Stefanelli M. Qualitative models and fuzzy systems: an integrated approach for learning from data. *Artif Intell Med*. 1998;14:5–28.
48. Florian JA Jr, Parker RS. Empirical modeling for glucose control in critical care and diabetes. *Eur J Control*. 2005;11:601–616.
49. Yamaguchi M, Kaseda C, Yamazaki K, Kobayashi M. Prediction of blood glucose level of type 1 diabetics using response surface methodology and data mining. *Med Bio Eng Comput*. 2006;44:451–457.
50. Bergman RN, Ider YZ, Bowden CR, Cobelli C. Quantitative estimation of insulin sensitivity. *Am J Physiol Endocrinol Metab*. 1979;236:E667–E677.
51. Bergman RN, Phillips LS, Cobelli C. Physiologic evaluation of factors controlling glucose tolerance in man. *J Clin Invest*. 1981;68:1456–1467.
52. Sorensen JT. *A physiologic model of glucose metabolism in man and its use to design and assess improved insulin therapies for diabetes*. Ph.D. Dissertation, Massachusetts Institute of Technology, 1985.
53. Makroglou A, Li J, Kuang Y. Mathematical models and software tools for the glucose-insulin regulatory system and diabetes: an overview. *Appl Numer Math*. 2006;56:559–573.
54. Hovorka R, Shojaee-Moradie F, Carroll PV, Chassin LJ, Gowrie IJ, Jackson NC, Tudor RS, Umpleby AM, Jones RH. Partitioning glucose distribution/transport, disposal, and endogenous production during IVGTT. *Am J Physiol Endocrinol Metab*. 2002;282:E992–E1007.
55. Hovorka R, Canonico V, Chassin LJ, Haueter U, Massi-Benedetti M, Federici MO, Pieber TR, Schaller HC, Schaupp L, Vering T, Wilinska ME. Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes. *Physiol Meas*. 2004;25:905–920.
56. Wilinska ME, Chassin LJ, Schaller HC, Schaupp L, Pieber TR, Hovorka R. Insulin kinetics in type-1 diabetes: continuous and bolus delivery of rapid acting insulin. *IEEE Trans Biomed Eng*. 2005;52:3–12.
57. Söderström T, Stoica P. *System Identification*. New York: Prentice Hall, Inc., 1989.
58. The Diabetes Research in Children Network (DirecNet) Study Group. Accuracy of the modified continuous glucose monitoring system (CGMS®) sensor in an outpatient setting: results from a Diabetes Research in Children Network (DirecNet) study. *Diabetes Technol Ther*. 2005;7:109–114.
59. Weinstein RL, Schwartz SL, Brazg RL, Bugler JR, Peyser TA, McGarraugh GV. Accuracy of the 5-day freestyle navigator continuous glucose monitoring system: comparison with frequent laboratory reference measurements. *Diabetes Care*. 2007;30:1125–1130.
60. Clarke WL. The original clarke error grid analysis (EGA). *Diabetes Technol Ther*. 2005;7:776–779.
61. Clarke WL, Cox D, Gonder-Frederick LA, Carter W, Pohl SL. Evaluating clinical accuracy of systems for self-monitoring of blood glucose. *Diabetes Care*. 1987;10:622–628.
62. Clarke WL, Anderson S, Farhy L, Breton M, Gonder-Frederick L, Cox D, Kovatchev B. Evaluating the clinical accuracy of two continuous glucose sensors using continuous glucose-error grid analysis. *Diabetes Care*. 2005;28:2412–2417.
63. Kovatchev BP, Gonder-Frederick LA, Cox DJ, Clarke WL. Evaluating the accuracy of continuous glucose-monitoring sensors. *Diabetes Care*. 2004;27:1922–1928.
64. Sachedina N, Pickup JC. Performance assessment of the medtronic-minimed continuous glucose monitoring system and its use for measurement of glycaemic control in type 1 diabetic subjects. *Diabet Med*. 2003;20:1012–1015.
65. Wentholt IM, Hoekstra JB, DeVries JH. A critical appraisal of the continuous glucose-error grid analysis. *Diabetes Care*. 2006;29:1805–1811.
66. Gross TM, Bode BW, Einhorn D, Kayne DM, Reed JH, White NH, Mastrototaro JJ. Performance evaluation of the MiniMed continuous glucose monitoring system during patient home use. *Diabetes Technol Ther*. 2000;2:49–56.

Appendix: Physiological Model

The physiological model[54–56] consists of three subsystems representing plasma glucose, plasma and subcutaneous insulin, and insulin action. The glucose subsystem is partitioned into two compartments representing the masses of glucose in the plasma and in a "nonaccessible" compartment described by Eqs. A1 and A2, respectively.

$$\frac{dQ_1(t)}{dt} = -\left[\frac{F_{01}^c}{V_G G(t)} + x_1(t)\right] Q_1(t) + k_{12} Q_2(t) - F_R + U_G(t) + EGP_0[1 - x_3(t)] \quad \text{(A1)}$$

$$\frac{dQ_2(t)}{dt} = x_1(t) Q_1(t) - [k_{12} + x_2(t)] Q_2(t) \quad \text{(A2)}$$

$$G(t) = Q_1(t)/V_G \quad \text{(A3)}$$

where: Q_1 and Q_2 are the masses of glucose in the plasma and nonaccessible compartments, respectively,

k_{12} is a rate constant for glucose transport,

V_G is the distribution volume of the plasma,

EGP_0 is the endogenous glucose production extrapolated to zero insulin concentration, and

G is the plasma glucose concentration.

Furthermore, F_{01}^c is the total non-insulin-dependent glucose flux, dependent upon glucose concentration:

$$F_{01}^c = \begin{cases} F_{01} & G \geq 80\,\text{mg/dL} \\ F_{01} G/80 & \text{otherwise} \end{cases} \quad \text{(A4)}$$

F_R is the renal glucose clearance above the glucose threshold of 160 mg/dL:

$$F_R = \begin{cases} 0.003(G - 160) V_G & G \geq 160\,\text{mg/dL} \\ 0 & \text{otherwise} \end{cases} \quad \text{(A5)}$$

and U_G is the glucose gut absorption rate from a meal:

$$U_G(t) = \frac{D_G A_G t e^{-t/t_{\max,G}}}{t_{\max,G}^2} \quad \text{(A6)}$$

where: D_G is the mass of carbohydrates in the meal,

A_G is the carbohydrate bioavailability, and

$t_{max,G}$ is the time-of-maximum appearance rate of glucose in the plasma.

The insulin subsystem[56] describes subcutaneous-to-intravenous insulin absorption using parallel fast and slow channels, and also includes a degree of insulin degradation at the injection site.

$$\frac{dQ_{1a}(t)}{dt} = ku(t) - k_{a1}Q_{1a}(t) - LD_a(t) \tag{A7}$$

$$\frac{dQ_{1b}(t)}{dt} = (1-k)u(t) - k_{a2}Q_{1b}(t) - LD_b(t) \tag{A8}$$

$$\frac{dQ_2(t)}{dt} = k_{a1}Q_{1a}(t) - k_{a1}Q_2(t) \tag{A9}$$

$$\frac{dQ_3(t)}{dt} = k_{a1}Q_2(t) + k_{a2}Q_{1b}(t) - k_eQ_3(t) \tag{A10}$$

$$LD_a(t) = V_{MAX,LD}Q_{1a}(t)/\left[k_{M,LD} + Q_{1a}(t)\right] \tag{A11}$$

$$LD_b(t) = V_{MAX,LD}Q_{1b}(t)/\left[k_{M,LD} + Q_{1b}(t)\right] \tag{A12}$$

$$I = Q_3(t)/V_I \tag{A13}$$

where: Q_{1a} and Q_{1b} represent the masses of insulin in the accessible subcutaneous compartments of the slow and fast channel, respectively,

Q_2 is the mass of insulin in the non-accessible subcutaneous compartment,

Q_3 represents the mass of insulin in the plasma,

u is the rate of subcutaneous insulin infusion,

k_{a1} and k_{a2} are insulin transfer rates,

k_e is the rate constant for the elimination of insulin from the plasma,

LD_a and LD_b are the rates of local degradation at the injection site for the slow and fast channel, respectively, described by Michaelis-Menten dynamics,

$V_{MAX,LD}$ is the maximum rate of insulin degradation,

$k_{M,LD}$ is the Michaelis-Menten constant (i.e., the mass of insulin at which insulin degradation is half of $V_{MAX,LD}$),

k is the fraction of subcutaneously infused insulin that passes through the slow channel,

V_I is the insulin distribution volume, and

I is the concentration of insulin in the plasma.

The insulin action subsystem represents three actions $x_1(t)$, $x_2(t)$, and $x_3(t)$ of insulin on the kinetics of glucose distribution/transport, disposal, and endogenous production, respectively:

$$\frac{dx_1(t)}{dt} = -k_{a1}x_1(t) + k_{b1}I(t) \tag{A14}$$

$$\frac{dx_2(t)}{dt} = -k_{a2}x_2(t) + k_{b2}I(t) \tag{A15}$$

$$\frac{dx_3(t)}{dt} = -k_{a3}x_3(t) + k_{b3}I(t) \tag{A16}$$

where: k_{ai} and k_{bi}, $i = 1,2,3$, represent deactivation and activation rate constants, respectively.

Manuscript received Aug. 6, 2007, and revision received Sept. 5, 2008.