

Prediction of Dynamic Glycemic Trends Using Optimal State Estimation

Matthew W. Percival *,** Wendy C. Bevier ** Howard Zisser ** Lois Jovanovič *,** Dale E. Seborg * Francis J. Doyle III *,**

 * University of California, Santa Barbara, CA 93106-5080 USA (Tel: +1(805)893-8133; e-mail: frank.doyle@icb.ucsb.edu).
 ** Sansum Diabetes Research Institute, 2219 Bath St., Santa Barbara, CA 93105-4321 USA

Abstract:

Efficacious therapeutic regimens to treat type 1 diabetes mellitus require devices capable of continuous feedback control; recent advances in medical technology mean that such devices are now available. Any closed-loop controller would require a predictive aspect to avoid sluggish control related to delays in insulin action, or hypoglycemia from an overdose of insulin. Using clinical data and an adaptive version of the simple Bergman minimal model (Bergman *et al.*, 1979), glycemic prediction was performed. Model parameters were estimated using clinical data. An augmented state Kalman filter was then used to estimate parameters dynamically. Predictive accuracy varied from subject to subject, with median R^2 values for the best validation days of 80% for 30 minute predictions. Such techniques would be useful in a closed-loop control framework for adapting a glycemic controller to subject-based variations in insulin sensitivity.

Keywords:

Biomedical systems; Kalman filters; parameter estimation; physiological models.

1. INTRODUCTION

Since the discovery of insulin in 1921, therapeutic methods for treating those with type 1 diabetes mellitus (T1DM) have undergone much refinement. Until recently, the only method of administering insulin for those in an ambulatory state was the hypodermic syringe; an insulin bolus would be delivered subcutaneously with meals, or at any time glucose levels were above the recommended range. With the development of continuous insulin pumps, a continuous infusion of insulin into the subcutaneous tissue may be effected, which represents a step closer to mimicking the normal physiologic plasma insulin profile.

The amount of insulin to be delivered throughout the day is determined by the subject from discrete measurements of plasma glucose. Intensive insulin therapy, which may require glucose determination as many as 12 times a day, can result in better glycemic control. However, even with this degree of frequency, this form of treatment is not without dangers, since those using intensive insulin therapy are up to three times more likely to have severe hypoglycemic episodes (Diabetes Control & Complications Trials Research Group, 1993). A chronically elevated blood glucose also leads to long-term complications, *e.g.*, macrovascular disease, due to damage to blood vessels (Gerich, 2005). Given the increase in life-threatening events for those adhering to this regimen, the open-loop feedback control methodology is inadequate.

Recent advances in biomedical devices mean that the available technology should be sufficient for a closed-loop device (Hovorka *et al.*, 2006). Reliable real-time subcutaneous glucose sensors have become available, and could be combined with a continuous subcutaneous insulin pump and a control algorithm to produce an artificial β -cell, capable of normalizing glucose concentrations.

Replicating normal physiologic control of glucose concentrations has several significant challenges. The β -cell has the advantage of intravenous delivery and measurement; the artificial equivalent will be implemented through subcutaneous glucose measurements and insulin delivery. This change in route of administration not only introduces large lag times (Hovorka, 2005), but also can be unreliable due to variable insulin kinetics at the injection site. Current treatment does not include the delivery of glucagon, a counter-regulatory hormone present in the normal state, to counteract the effects of an insulin overdose. Accurate prediction of glycemia could help overcome these problems.

The issue of predicting glucose concentrations can be addressed with deterministic models representing the effects of meals and insulin on blood glucose concentration. Such modeling of glucose-insulin interactions can be divided into two broad categories: empirical and physiological. Empirical models are developed from data, and do not require and understanding of the processes that govern

^{*} This study is a joint investigation between University of California, Santa Barbara and Sansum Diabetes Research Institute, supported by the Juvenile Diabetes Research Foundation grants 22-2006-1115, and 22-2006-1108, and the National Institutes of Health grants R21-DK068706, R01-DK068683, and R21-DK069833.



Fig. 1. Inputs, outputs, and interactions between the compartmental models representing glucose absorption, insulin absorption, and glucose-insulin kinetics.

the system. Finan *et al.* (2007) have shown how a range of empirical models can be developed from simulated data. Physiological models require an understanding of the system kinetics and dynamics. There is currently no consensus on an accurate physiological model covering glycemic variations in those with T1DM, as can be seen from the range of published physiological models (Makroglou *et al.*, 2006). Recently published models such as those of Canonico *et al.* (2006) and Dalla Man *et. al* (2007) are examples of models varying in structure and dynamics for modeling the same system. Some physiological models have been developed using experimental data from normal subjects, *e.g.* Hovorka *et al.* (2002); this casts uncertainty on the validity of model parameters for subjects with T1DM, who have a different physiology.

Even if the physiological model accurately describes the interactions of glucose and insulin, the dynamics are known to vary both across a population, and within an individual. The within-subject variations occur on timescales ranging from minutes (during exercise), to years (as the aging process elicits metabolic changes). Since these changes can potentially change the insulin requirement of an individual by over 100%, dynamic adaptation of model parameters must occur. (Hann *et al.*, 2005) have addressed the issue of real-time parameter estimation using a nonlinear model, with success at predicting up to one hour into the future with T1DM subjects in an intensive care unit.

The goal of this study was to establish a benchmark for glycemic prediction on subjects with T1DM in ambulatory conditions using linear, time-series models derived from the Bergman minimal model (MM) (Bergman *et al.*, 1979). The issue of dynamic parameter estimation is addressed with an augmented state Kalman filter. Performance indices specific to T1DM are applied to glycemic predictions. Also, statistical methods are invoked to establish error bounds for predictions.

2. PHYSIOLOGICAL MODELING

A compartmental approach to physiological modeling was adopted. The interactions between the three major compartments representing glucose absorption (from meals), insulin absorption (from SC infusion) and glucose-insulin kinetics are shown in Fig. 1.

For this proof-of-concept study, the MM was used to represent the effects of glucose-insulin kinetics; this model

requires inputs of plasma insulin concentration and IV glucose absorption rate. The insulin absorption model used was model 10 from Wilinska *et al.* (2005); the model considers degradation of insulin at the injection site, followed by absorption through both fast and slow channels. The meal absorption model used was that of Hovorka *et al.* (2004); the model consists of two identical compartments in series, and considers bioavailability of ingested carbohydrate.

The MM was originally designed for use in an IV glucose tolerance test, not as the kinetic core of a full physiological model. However, it has survived a significant number of independent investigations into its accuracy (Bergman, 2007); since it encompasses the major aspects of glucoseinsulin kinetics, it therefore warrants further use. The model is based around steady-state conditions, whereby hepatic glucose production is balanced by uptake from the brain and the periphery; this is achieved with a basal plasma insulin concentration and results in a basal blood glucose concentration. When glycemia deviates from this basal condition, glucose can effect its own disposal via a reduction in hepatic production. An increase in plasma insulin concentration has a delayed response, through the remote compartment. This represents plasma insulin movement across the capillary endothelium.

The model has two state variables: deviation plasma glucose concentration (G'), defined as the deviation of plasma glucose from basal $(G - G_b)$, and remote (interstitial) insulin concentration (X). The two inputs are the deviation plasma insulin concentration (I'), defined as the deviation of plasma insulin from basal $(I-I_b)$, and the rate of glucose absorption (U). The system is described in equation (1). Parameter p_1 represents glucose sensitivity, and the ratio p_3/p_2 represents insulin sensitivity.

$$\frac{dG'}{dt} = -p_1G' + G'X + U$$

$$\frac{dX}{dt} = -p_2X + p_3I'.$$
(1)

From this bilinear model, three linear models are developed. The principal difference between the bilinear MM and the linearized MM is that there is neither an increase in insulin effectiveness nor a decrease in insulin effectiveness at blood glucose concentrations above or below basal, respectively. Each model uses a different approximation for insulin dependence. Each model is then discretized; this was appropriate since only discrete measurements are available. Additionally, the characteristic sampling period (typically 3-10 minutes) is significantly less than the dominant timescale in the model (approximately 100 minutes). It is therefore reasonable to discretize the linear models, assuming piecewise constant inputs over the period of discretization, Δt , without considerable loss of accuracy. These models are equivalent to first order autoregressive models with two exogenous, first order inputs.

Model 1. A linearization about basal conditions eliminates the bilinear term, since $X_b = 0$, and hence the resulting linear model is independent of plasma insulin concentration. The model equation is given in equation 2. So that p_1 is uniquely identified, a glucose absorption correction factor K_G is added; the definitions of the parameters used in equation 2 are given in equation 3.



Fig. 2. Parameter values and standard errors for each dataset and subject using model 2. The size of the standard error for parameters β_2 and β_3 show that the sign of the parameter can not be consistently identified.

$$G'_{k} = \beta_1 G'_{k-1} + \beta_2 U_{k-1} \tag{2}$$

$$\beta_1 = \exp(-p_1 \Delta t)$$

$$\beta_2 = \frac{K_G}{p_1} (1 - \exp(-p_1 \Delta t))$$
(3)

Model 2. An additional term is added to include dependence upon remote insulin concentration as in equation 4. A remote insulin concentration correction factor, K_{X1} , is added in order to uniquely identify p_1 . Equations 3 and 5 define the parameters used in the model.

$$G'_{k} = \beta_1 G'_{k-1} + \beta_2 U_{k-1} + \beta_3 X_{k-1} \tag{4}$$

$$\beta_3 = -\frac{K_{X1}}{p_1} (1 - \exp(-p_1 \Delta t)) \tag{5}$$

Model 3. Since the time constant $1/p_2$ of the remote insulin dynamics is significantly smaller than the time constant $1/p_1$ of the glucose dynamics, the remote insulin is assumed rapidly equilibrating; therefore the remote insulin concentration is $X = \frac{p_3}{p_2}I'$. The model described by equation 6, and the parameters are defined by equations 3 and 7.

$$G'_{k} = \beta_1 G'_{k-1} + \beta_2 U_{k-1} + \beta_3 I_{k-1} \tag{6}$$

$$\beta_3 = -\frac{K_{X2}p_3}{p_1p_2}(1 - \exp(-p_1\Delta t)) \tag{7}$$

3. PARAMETER ESTIMATION

The linear models were fitted to clinical data using the parameter estimation techniques described below. The clinical data were obtained from subjects with T1DM using an Institutional Review Board approved protocol. Informed, witnessed consent was obtained from subjects. Data were acquired using the Continuous Glucose Monitoring System (CGMS®, Medtronic Minimed, Northridge CA) and comprised CGMS readings, insulin pump records, and subject-reported estimates of time and carbohydrate content of meals. Twenty-seven datasets were analyzed; each dataset was from midnight to midnight, and was collected from three subjects. For each of the linear models, a least squares estimate of parameters was performed on each set of data. An estimate of the confidence limits of the parameter was then obtained to quantify the validity of the model fit to the data. The confidence limits used were 95%, drawn from a t-distribution.

Fig. 2 shows the estimated parameters and their associated standard error for model 2. Parameter β_2 should be positive, and was correctly identified in 15 of the 27 datasets. Parameter β_3 should be negative, and was correctly identified in 14 of the 27 datasets. Models 1 and 3 were marginally more successful in correctly identifying parameter sign, but the size of the standard error in the parameter estimates indicates that the 95% limits around β_2 and β_3 make the sign of the parameter statistically uncertain.

The dynamic estimation of parameters was performed using an augmented state Kalman filter, an extension of the ideas originally introduced by Kalman (1960). The estimator phase of the algorithm was achieved using either of the linear models, and the solution of the discrete algebraic Riccati equation provided the Kalman gain and correction (Åström and Wittenmark, 1997). Using a linear timeinvariant (LTI) state-space model structure (equation 8), parameters were included one at a time (equation 9); the subscript i represents the single parameter that is being used to augment the state vector.

$$\begin{aligned}
x_k &= Ax_{k-1} + Bu_{k-1} + w_{k-1} \\
y_k &= Cx_k + v_k
\end{aligned} \tag{8}$$

$$x_k = \left[G_k \ \beta_{i,k} \right]^T \tag{9}$$

Matrices A, B, and C are given in equation (10). Process noise w_{k-1} and measurement noise v_k are assumed Gaussian, zero mean, with covariance matrices Q, and R, respectively. Application of an observability test (Chen, 1999) for the system given in equation (10) shows that the state variables, *i.e.*, G_k and $\beta_{i,k}$, are not observable. Therefore, it is impossible to determine which state is the source of the noise. Logically, if some uncertainty is assumed to be on the measured state variable, the balance can be applied to the augmented state variable.

$$A = \begin{bmatrix} \beta_1 & 0\\ 0 & \beta_{i,k} \end{bmatrix} B = \begin{bmatrix} \beta_2 & \beta_3\\ 0 & 0 \end{bmatrix} C = \begin{bmatrix} 1 & 0 \end{bmatrix}$$
(10)

In practice, for any change in the augmented state to be observed, there must be correlation between the noise on the state G and the augmented state β_i ; hence the matrix Q has off-diagonal elements. In tuning the Kalman filter, Qmust still be symmetric and positive semi-definite for the solution of the algebraic Riccati equation. The conditions on the elements of Q are given in equation (11).

$$Q = \begin{bmatrix} a & b \\ b & c \end{bmatrix} ac - b^2 \ge 0 \tag{11}$$

4. PERFORMANCE METRICS

Two common examples of performance metrics used for quantifying correlation between a model and data are the coefficient of determination (R^2) , and the Median Relative Absolute Deviation (MRAD). R^2 values range from 100%, indicating a perfect fit, to $-\infty$; a value of 0% is equivalent



Fig. 3. Strong correlation was observed between R^2 and MRAD performance metrics, with this correlation deteriorating as the prediction horizon increased.



Fig. 4. No correlation was observed between R^2 and the percentage of points in zone A of the rEGA.

to predicting the mean of the data across the set. MRAD ranges from 0%, representing a perfect fit, to 100%.

Although these metrics are mathematically sound, they do not necessarily quantify performance in a manner relevant to diabetes. To address this issue, Clarke *et al.* (1987) developed point error grid analysis (pEGA). With this metric, the error between prediction and measurement is assigned a zone, depending upon its clinical significance. This analysis is useful since a qualitative index is assigned to model predictive performance, in terms that are useful to a clinician.

Kovatchev *et al.* (2004) developed rate error grid analysis, which analyzed how well the rate-of-change of glucose concentrations is being predicted. The principle is identical to pEGA, but the analysis yields additional information since correlation of predicted and measured glycemic rateof-change is being quantified.

The large number of simulations produced in validating and calibrating parameter sets and tuning the variance matrices provided opportunity to compare these metrics. This is necessary since there is little consensus among published models on a standard performance metric when fit-



Fig. 5. The R^2 score for the 60 minute prediction horizon for subject 2 using model 2. Columns represent validation of all the obtained parameter sets for that subject.

ting models to clinical data. Over 12,000 simulations were run, and the performance metrics were plotted against each other. Examples of these plots are shown in Figs. 3 and 4. Correlation was observed between all point metrics, with the correlation strongest when the metric performance was best. This implies that caution must be exercised when rating mediocre fits based on only one metric. The rEGA was uncorrelated with all point metrics, indicating that this analysis characterizes a desirable aspect of model fit not otherwise quantified.

5. GLYCEMIC PREDICTION

After the calibration procedure of obtaining optimal model parameters for each data set, each set of parameters was then validated against all other data sets for that subject. Model predictions of up to 180 minutes were made. A bias term, \hat{d}_k , was also added to the predictions, to correct for persistent error. The bias for prediction horizon P was defined as the error between the measurement at time kand the prediction at time k made at time k - P. This bias was then added to the prediction made at time k for time k + P, as shown in equation 12.

$$\hat{d}_k = G_k - \hat{G}_{k|k-P}$$

$$\hat{G}_{k+P|k,bias} = \hat{G}_{k+P|k} + \hat{d}_k$$
(12)

Using the results of the predictions with constant parameter values as a benchmark, the Kalman filter was then tuned in an attempt to improve upon the model predictions. The parameters β_2 and β_3 were subject to most investigation since they represent an inaccurate glucose absorption profile (indicative of inaccurate subjectreported carbohydrate intake) and a change in insulin sensitivity, respectively.

An example of the R^2 values obtained for a 60 minute prediction horizon for each calibration set of parameters validated against each data set is shown in Fig. 5. Horizontal trends show that some data will be better predicted regardless of which parameter set is used. A vertical trend would imply that a universal set of parameters was found by that calibration day; vertical trends are very weak, indicating that the data, rather than the parameters de-



Fig. 6. The prediction error is plotted for subject 3, day 7, using model 2, and parameters from calibration day 4. Performance degrades significantly as the prediction horizon increases in the presence of input stimuli.



Fig. 7. A: 30 minute prediction horizon, $R^2=85\%$. B: 90 minute prediction horizon, $R^2=8\%$. Error bars are also shown. Data was from subject 3, day 7, using model 2, and parameters from calibration day 4.

termine the quality of model fit. This trend was observed for all subjects and models.

Fig. 6 shows how the error develops in time. This plot shows that it is the large deviations from basal conditions that are particularly challenging to predict.

The validation day corresponding to the median R^2 for the best calibration day is shown in Fig. 7, along with prediction horizons of 30 and 90 minutes. The error bars, developed from the 95% confidence limits for the parameter estimates indicate that although the fit seems reasonable to the eye, the true level of uncertainty is large.

Fig. 8 shows the inclusion of output additive disturbance in order to bias the prediction for a two hour prediction horizon. In prolonged mismatch, the prediction with bias captures the magnitude of the peaks and nadirs. However, the R^2 decreases since there is a delay effect in capturing the extreme blood glucose concentrations.



Fig. 8. Prediction horizons of 120 minutes with a bias term $(R^2 = 62\%)$, and without a bias term $(R^2 = 71\%)$. Data was from from subject 3, day 7, using model 2, and parameters from calibration day 4.



Fig. 9. Parameter β_3 was adapted using a Kalman filter. Predictions of 30, 90, and 180 minutes were made, with R^2 values of 86%, 18%, and -300%, respectively. Data was from subject 3, day 7, using model 2, and initialization used day 4 calibration parameters. The plot shows that in response to stimuli, a dynamic parameter estimate is obtained.

Fig. 9 shows dynamic variation of parameter β_3 after tuning of the Kalman filter. The persistent increase in the parameter at a time of great excitation provides strong evidence that the *a priori* parameter estimate is not correct at that time of day. The same trend would be observed with β_2 : the filter is simply correcting the estimate based on the mismatch. In this case, the model does not expect blood glucose to rise so high. The rise could be explained by either a decrease in insulin sensitivity, or an underestimate of meal carbohydrate content. The underestimate in carbohydrate content could also be explained by a higher carbohydrate bioavailability than that described by the meal model. Since parameter variation occurs at the time of the meal, before any change in insulin concentration occurs, it is more likely that the meal size has been underestimated. Comparing the R^2 values from Figs. 7 and 9 for 30 minute and 90 minute prediction horizons, there is a nominal improvement when Kalman filtering is used.

6. CONCLUSIONS

Although simple, the linear versions of the MM provide reasonable glycemic prediction for horizons of 30–60 minutes. This observation is not surprising; since the dynamics of the system are relatively slow, a passive model is not too erroneous over such a timespan. Predictions of the order 90 minutes, which is the lag time between insulin infusion and its effect on glycemia, are inaccurate with these models.

The differences between the linear models developed are not statistically significant: no single model is consistently more accurate in predicting glycemia. Similarly, some data sets exhibit simpler dynamics, which can more easily be captured. Some subjects are harder to model than others, which could be explained by inadequacies in sensor accuracy and/or carbohydrate estimates.

Parameter variation can be inferred, and a well-tuned Kalman filter can show reasonable changes in parameters, albeit *a posteriori*. In order to deconvolute the effects of different parameters, controlled experiments would be required, where meals and insulin boluses were staggered by the system lag time. Nevertheless, learning that a mismatch between measurements and predictions has occurred could be useful in a run-to-run scenario; insulin infusion protocols could be modified to avoid hypo- and hyperglycemic events if the mismatch were systematic.

The calculated error bars give a statistically significant means of quantifying model fit beyond a qualitative judgment. Although some predictions appear statistically sound up to 60 minutes ahead, the error bars show an unacceptably large error, particularly at low blood glucose, at horizons as short as 15 minutes.

Future work would extend the presented parameter estimation techniques to more complex physiological models. Adaptive parameters would be useful in a run-to-run control scenario. Ultimately, a validated, adaptive model would form the basis of a closed-loop controller forming part of the communication algorithm in a closed-loop artificial β -cell.

ACKNOWLEDGEMENTS

The author would like to thank C.C. Palerm, D.A. Finan, and M.T. Dokucu for their insightful commentary.

REFERENCES

- K.J. Åström and B. Wittenmark. Computer-controlled systems. Prentice Hall, Upper Saddle River, NJ, 3rd edition, 1997.
- R.N. Bergman, Y.Z. Ider, and C.R. Bowden. Quantitative estimation of insulin sensitivity. *American Journal of Physiology: Endocrinology, Metabolism, and Gastrointestinal Physiology*, 5(6):E667–E677, 1979.
- R.N. Bergman. Orchestration of glucose homeostasis: from a small acorn to the California oak. *Diabetes*, 56:1489– 1501, 2007.
- C.T. Chen. *Linear System Theory and Design.* Oxford University Press, New York NY, 1999.
- W.L. Clarke, D. Cox, L.A. Gonder-Frederick and W. Carter. Evaluating clinical accuracy of systems

for self-monitoring of blood glucose. *Diabetes Care*, 10(5):622–628, 1987.

- V. Canonico, P.G. Fabietti, M.M. Benedetti, M.O. Federici and E. Sarti. Virtual type 1 diabetic patient for feedback control systems. *Diabetes Research and Clinical Practice*, 74:S187–S190, 2006.
- C. Dalla Man, R.A. Rizza and C. Cobelli. Meal simulation model of the glucose-insulin system. *IEEE Transactions* in *Biomedical Engineering*, in press, 2007.
- Diabetes Control & Complications Trials Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *New England Journal of Medicine*, 329:977–986, 1993.
- D.A. Finan, H. Zisser, L. Jovanovič, W.C. Bevier, and D.E. Seborg. Practical issues in the identification of empirical models from simulated type 1 diabetes data. *Diabetes Technology and Therapeutics*, 9(5):438–450, 2007.
- J.E. Gerich. The importance of tight glycemic control. American Journal of Medicine, 118(S1):7–1, 2005.
- C.E. Hann, J.G. Chase, J. Lin, T. Lotz, C.V. Doran, and G.M. Shaw. Integral-based parameter identification for long-term dynamic verification of a glucose-insulin system model. *Computer Methods and Programs in Biomedicine*, 77(3):259–270, 2005.
- R. Hovorka, F. Shojaee-Moradie, P.V. Carroll, L.J. Chassin, I.J. Gowrie, N.C. Jackson, R.S. Tudor, A.M. Umpleby, and R.H.Jones. Partitioning glucose distribution/transport, disposal, and endogenous production during IVGTT. *American Journal of Endocrinology and Metabolism*, 282(5):E992–E1007, 2002.
- R. Hovorka, V. Canonico, L.J. Chassin, U. Haueter, M. Massi-Benedetti, F.M. Orsini, T.R. Pieber, H.C. Schaller, L. Schaupp, T. Vering and M.E. Wilinska. Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes. *Physiological Measurement*, 25(4):905–920, 2004.
- R. Hovorka. Management of diabetes using adaptive control. International Journal of Adaptive Control, 19(5):309–325, 2005.
- R. Hovorka, M.E. Wilinska, L.J. Chassin and D.B. Dunger. Roadmap to the artificial pancreas. *Diabetes Research* and *Clinical Practice*, 74:S178–S182, 2006.
- R.E. Kalman. A New approach to Linear Filtering and Prediction Problems *Journal of Basic Engineering*, 82:35–45, 1960.
- B.P. Kovatchev, L.A. Gonder-Frederick, D.J. Cox and W.L. Clarke. Evaluating the accuracy of continuous glucose-monitoring sensors: continuous glucoseerror grid analysis illustrated by TheraSense Freestyle Navigator data. *Diabetes Care*, 27(8):1922–1928, 2004.
- A. Makroglou, J. Li, and Y. Kuang. Mathematical models and software tools for the glucose-insulin regulatory system and diabetes: an overview. *Applied Numerical Mathematics*, 56(3-4):559–573, 2006.
- M.E. Wilinska, L.J. Chassin, H.C. Schaller, L. Schaupp, T.R. Pieber and R. Hovorka. Insulin kinetics in type–1 diabetes: continuous and bolus delivery of rapid acting insulin. *IEEE Transactions in Biomedical Engineering*, 52(1):3–12, 2005.