2008 American Control Conference
Westin Seattle Hotel, Seattle, Washington, USA
June 11-13, 2008

Identification of Empirical Dynamic Models From Type 1 Diabetes Subject Data

Daniel A. Finan,† Cesar C. Palerm,† Francis J. Doyle III,†
Howard Zisser,‡ Lois Jovanovič,‡ Wendy C. Bevier,‡ and Dale E. Seborg†

Abstract— **Empirical linear dynamic models have been identified from ambulatory data from two type 1 diabetes subjects in order to determine approximately how far into the future the models could be expected to make reasonably accurate predictions. For a prediction horizon of 30 minutes, FIT values (related to R^2 values) of the model predictions for validation data were 46% for one subject and 60% for the other subject. These FIT values correspond to root mean square errors of 14 and 24 mg/dL, respectively. Longer prediction horizons resulted in substantially worse predictions for these ambulatory subject data.**

I. INTRODUCTION

Type 1 diabetes mellitus (T1DM) is an autoimmune disease characterized by a negligible production of endogenous insulin from the pancreatic β-cells. This insufficient insulin production results in inadequate regulation of glucose concentration. Both high and low blood glucose levels (*hyperglycemia* and *hypoglycemia*, respectively) are deleterious to one's health. Sustained hyperglycemia is associated with a significant increase in the chances of developing long-term complications such as microvascular disease [1] (e.g., eye, kidney, and nerve disease) and cardiovascular disease. Acute hypoglycemia, on the other hand, can cause immediate health threats such as seizures and coma.

T1DM subjects can significantly reduce their risk of complications due to both hyperglycemia and hypoglycemia by regulating their blood glucose more effectively [2]. Conceptually, an artificial β-cell would achieve improved glycemic control automatically. This biomedical device would consist of a continuous subcutaneous glucose monitor, a controller, and a continuous subcutaneous insulin infusion (CSII) pump [3]. The controller would automatically adjust the insulin dose in order to regulate glucose levels based on current and past glucose and insulin infusion information, and deliver this dose via the CSII pump. Although some types of controllers would not require a model of the subject's glucose-insulin dynamics (e.g., a proportional-integral-derivative, or PID controller), model-based controllers are well suited for this application given their ability to explicitly handle input constraints (e.g., the insulin delivery rate is bounded by zero and a maximum value, and its absolute rate of change is also bounded).

This work was supported by the National Institutes of Health, grant R21–DK069833–02, and by the Juvenile Diabetes Research Foundation, grant 22–2006–1115.

†Dept. of Chemical Engineering, University of California, Santa Barbara, CA 93106–5080

‡Sansum Diabetes Research Institute, Santa Barbara, CA 93105

The performance of a model-based controller is related to the predictive capability of the model [4]. In general, predictive models involve an inherent trade-off between the accuracy of the predictions and the length of the prediction horizon. The longer the prediction horizon, the more useful the prediction is to a model-based controller, but the less accurate it tends to be. This paper investigates the relationship between model prediction accuracy and the length of the prediction horizon.

II. PREDICTIVE MODELS FOR T1DM SUBJECT DATA

Although there have been myriad simulation studies exploring the prediction of future glucose trends, only a few have utilized actual T1DM subject data. Furthermore, direct comparisons among these studies are difficult to make because they often used different sampling intervals, prediction horizons, and/or metrics to quantify prediction accuracy. Bellazzi et al. [5] used non-uniformly and sparsely sampled T1DM subject data collected in ambulatory conditions, linearly interpolated at 2-h intervals, to identify low-order autoregressive exogenous input (ARX) models whose inputs included meals (a binary variable indicating the presence or absence of a meal) and a filtered insulin input. They investigated 1-step, 2-step, and 3-step (i.e., 2-h, 4-h, and 6-h) prediction horizons. The authors summarize their modeling results by reporting 1-step ahead prediction metrics for the best-case subject, the worst-case subject, and the mean case of their 60-subject data bank. The mean prediction errors were 2 mg/dL for the best-case subject, 35 mg/dL for the worst-case subject, and 19 ± 8 mg/dL for the mean subject. The marked difference between the results for the best- and worst-case subjects illustrates a fundamental and significant inter-subject variability. The authors concede that these results are at least somewhat positively biased due to the linear interpolation.

Bremer and Gough [6] used 10-min data from ambulatory T1DM patients to identify autoregressive (AR) models. They explored 1-step, 2-step, and 3-step (i.e., 10-min, 20-min, and 30-min) prediction horizons, and report that the 1-step predictions are accurate and that "for certain data, 20-min or 30-min predictions may also be acceptable." They provided no quantification of their predictions of T1DM subject data.

Hovorka et al. [7] performed experiments in 10 T1DM patients under clinical conditions, using their own physiological model to make predictions of 15-min glucose data up to 4 steps (i.e., 60 min) into the future. The glucose

978-1-4244-2079-7/08/$25.00 ©2008 AACC.

was measured intravenously, but delayed by 30 min to mimic subcutaneous measurement. The model parameters were recursively estimated using a sophisticated Bayesian method. The predictions of the resulting models had root mean square error (RMSE) values of 8.6, 13.0, and 17.3 mg/dL for 2-step, 3-step, and 4-step (i.e., 30-min, 45-min, and 60-min) predictions, respectively.

Sparacino et al. [8] collected 48 h of continuous (3-min) glucose data from 28 ambulatory T1DM subjects. In their retrospective analysis, they recursively identified simple polynomial and AR models from these time-series data. They investigated prediction horizons of 10 and 15 steps (i.e., 30 and 45 min). Their best models had RMSE values of 18 and 32 mg/dL for the 30-min and 45-min prediction horizons, respectively.

Reifman et al. [9] developed AR models from continuous glucose data from nine T1DM subjects. The dataset for each subject spanned approximately five days, during which the subjects were "confined to the investigational site" and "limited to mild physical activity." For models identified from individual patients, RMSE values of 30-min and 60-min predictions of validation data ranged from 18–30 mg/dL and 25–41 mg/dL, respectively. For models identified from other subjects, mean RMSE values of 30-min and 60-min predictions of validation data ranged from 21–31 mg/dL and 28–42 mg/dL, respectively.

III. METHODS

A. Subject and Simulated Data

For each of two T1DM subjects (denoted by A and B) in ambulatory conditions, two datasets were collected that spanned multiple days and consisted of continuous (5-min) glucose measurements, insulin pump records, and subject-reported estimates of the times and carbohydrate (CHO) content of meals. An inherent characteristic of these ambulatory data was the absence of any reliable or quantitative measure of other factors that greatly affect glucose concentration, such as exercise and stress. Thus, as a "reality check," a nonlinear physiological model [7,10] of type 1 diabetes was used to generate more deterministic datasets corresponding to each of the four subject datasets. These simulated datasets were generated by using the actual inputs (i.e., insulin infusion rates and subject-reported meals) as the inputs to the model. Thus, these simulated datasets are "more deterministic" in the sense that neither exercise-related nor stress-related disturbances, measured or unmeasured, influence the glucose concentration.

The model was roughly tuned for each dataset by using the subject's weight as a model parameter and adjusting the insulin sensitivity parameters in the model such that the inputs resulted in (approximately) the corresponding experimental mean glucose concentration. Also, white noise was added to the simulated glucose concentration with zero mean and a standard deviation of 3.3 mg/dL. Table I summarizes the datasets investigated in this paper. Fig. 1 shows a representative subject dataset (B_1) with its corresponding simulated dataset (B_{1S}). For certain sections of the dataset, e.g. samples ∼550–700, the model mimics the data very well. However, this is not usually the case. Fig. 1 illustrates that there is much unexplained variability in the ambulatory subject data.

TABLE I
DESCRIPTION OF DATASETS

Type	Dataset	Number of Samples	Mean Glucose (mg/dL)	Standard Deviation (mg/dL)
SUBJECT DATA	A_1	1195	85.5	26.0
	A_2	801	106.1	32.8
	B_1	1375	146.3	62.0
	B_2	1000	151.1	50.0
SIMULATED DATA	A_{1S}	1195	88.0	46.0
	A_{2S}	801	103.6	46.0
	B_{1S}	1375	149.4	70.6
	B_{2S}	1000	150.0	81.1

Each dataset was then divided in half. The first half (the calibration portion of the data, or "Cal") was used for model identification; the second half (the validation portion of the data, or "Val") for model validation.

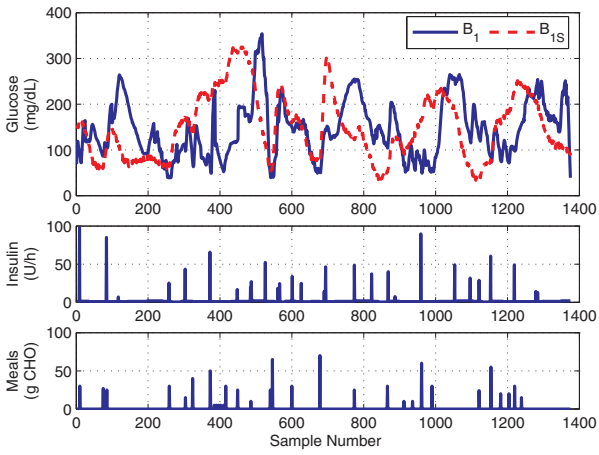

Fig. 1. Datasets B_1 and B_{1S} shown with the insulin and meal inputs.

B. Empirical Models and Identification Techniques

Three types of empirical dynamic models were identified from the datasets: autoregressive (AR), autoregressive exogenous input (ARX), and autoregressive moving average exogenous input (ARMAX). The latter is the most complex of the three types of models, and is described by

$$A(q^{-1})G(t) = B_i(q^{-1})u_i(t) + B_m(q^{-1})u_m(t) + C(q^{-1})e(t), \quad (1)$$

where G is the output (i.e., glucose concentration) and u_i and u_m are the inputs (i.e., insulin infusion rate and meal CHO, respectively). A, B_i, B_m, and C are polynomials in the backward shift operator q^{-1}, i.e., $q^{-1}x(t) \equiv x(t-1)$. White-noise disturbances are represented by e.

The AR and ARX models are special cases of the ARMAX model obtained by specifying $B_i = B_m = 0$ and

$C = 1$ for AR models, and $C = 1$ for ARX models. Note that AR models do not model the effect of the inputs; they simply describe the dependence of future outputs on past outputs. For this paper, third-order models were identified; thus, the AR, ARX, and ARMAX models had 3, 9, and 12 parameters, respectively. Other model orders were investigated and gave similar results.

A typical model identification procedure consists of estimating model parameters such that the 1-step ahead prediction errors are minimized, or, more precisely, such that the sum of the squares of the 1-step ahead prediction errors is minimized [11]. This *least-squares solution* is available in closed form for AR and ARX models, and thus a global minimum is guaranteed. However, because the 1-step ahead predictions of ARMAX models depend on past prediction errors (which depend on the model parameters to be estimated), a closed-form solution is not available [12]. In this case, a more computationally intensive iterative prediction error method is employed to search for a *local* minimum in the parameter space. Thus, the global minimum is not guaranteed.

For model predictive control purposes, the sum of the squares of the 1-step ahead prediction errors is perhaps the wrong criterion to minimize. Compared to the sampling interval of continuous glucose monitors (e.g., 5 min), the glucose-insulin dynamics in type 1 diabetes are significantly slower. This caveat motivates the investigation of estimating model parameters based on optimizing prediction errors with a longer prediction horizon (PH), such as 3 or 6 steps (i.e., 15 or 30 min). In this case the least-squares solution to the AR and ARX problems, as well as the ARMAX problem, requires an iterative prediction error method for the same reason as the 1-step ahead ARMAX problem [11–13]. Thus, the multi-step ahead least-squares prediction error criterion to minimize is

$$J_{LS} = \sum_{t=1}^{N} \left[G(t+p) - \hat{G}(t+p|t) \right]^2, \qquad (2)$$

where $G(t+p)$ is the measured glucose concentration at time $t+p$, $\hat{G}(t+p|t)$ is the predicted glucose concentration at time $t+p$ based on information up to time t, and p is the prediction horizon (in steps). In this paper, values of $p = 3$ and $p = 6$ are investigated in addition to the standard case $p = 1$.

The identification problem represented by the minimization of J_{LS} in (2) is a nonlinear least-squares problem. Given initial values for the model parameters, this problem can be solved readily using, for example, a Gauss-Newton method. The initial values of the parameters were taken to be the corresponding values obtained from the standard 1-step ahead optimization. A brief comment on the effect of the initial values follows in the Results section.

C. Prediction Accuracy

The metric used to quantify the accuracy of model predictions in this paper is the FIT value,

$$FIT = \left(1 - \frac{\|G - \hat{G}\|}{\|G - \bar{G}\|} \right) \times 100\%, \qquad (3)$$

where G is the vector of measured glucose values, $\hat{G}$ is the vector of the corresponding glucose predictions, $\bar{G}$ is the vector of mean measured values, and the indicated norms are Euclidean. Thus, the FIT value is a measure of how much variability in the data is explained by the model predictions. When appropriate, the RMSE is also reported to facilitate comparison with previous studies.

IV. RESULTS

Modeling results for the AR and ARX models for the simulated data are reported in Tables II and III, respectively. In general, the predictions do not significantly vary with either the calibration PH p or the type of model. The predictions do vary substantially, however, with the validation PH, the portion of the data used for the validation, and the virtual subject. Predictions deteriorate as the validation PH is extended to 24 steps (120 min), especially for subject for A_S, whose FIT values for the validation portion of either dataset are either very low or negative.

The validation portions of the datasets usually show lower FIT values than the corresponding calibration portions. This trend is true in general for identification problems, but not always. Dataset B_{2S}, for example, shows higher FIT values for the validation portion than for the calibration portion. This result is counterintuitive but possible if, for example, the validation data exhibit a higher degree of linearity than the calibration data.

The predictions vary perhaps most markedly with the virtual subject. Combining the AR and ARX results, the ranges (low−high) of FIT values for the validation sections using 6-step, 12-step, and 24-step (i.e., 30-min, 60-min, and 120-min) predictions are, respectively,

- Subject A_S: $54\% - 68\%$, $20\% - 41\%$, and $-28\% - 8\%$
- Subject B_S: $75\% - 86\%$, $55\% - 77\%$, and $27\% - 59\%$.

Tables II and III can be further summarized in terms of the modeling improvements obtained by using more sophisticated identification techniques. For instance, compared to AR models, ARX models result in improvements of 12-step (60-min) ahead predictions of validation data of $6.3 \pm 4.8\%$ (FIT values, mean $\pm$ standard deviation). Optimizing model parameters using $p = 6$ resulted in improvements of 12-step ahead predictions of validation data of $5.4 \pm 2.9\%$ compared to $p = 1$ (the standard least-squares solution) and $2.5 \pm 1.8\%$ compared to $p = 3$.

Fig. 2 shows how the ARX model predictions of the validation sections vary with the validation PH. The FIT values decline with increasing PH, the curves for subject A_S having a much more severe slope than for B_S. The two curves for each subject group together closely.

TABLE II

AR RESULTS (FIT VALUES, %) FOR THE SIMULATED DATASETS. HIGHLIGHTED VALUES REPRESENT CALIBRATION FITS.

Dataset/ Model[a]		Validation PH (Steps) 1 Cal	1 Val	3 Cal	3 Val	6 Cal	6 Val	12 Cal	12 Val	24 Cal	24 Val
A_{1S}	1	88	89	84	79	76	64	60	38	32	2
	3	88	89	84	81	77	66	61	39	33	4
	6	87	90	84	82	77	67	61	41	34	5
A_{2S}	1	88	85	81	74	68	54	43	20	5	-28
	3	87	84	81	76	70	57	46	23	8	-24
	6	87	84	81	76	71	59	48	27	11	-20
B_{1S}	1	93	92	88	86	78	75	60	55	35	27
	3	93	92	88	86	79	76	63	57	37	30
	6	93	92	88	85	80	76	63	58	38	31
B_{2S}	1	92	94	86	89	76	80	56	64	27	41
	3	92	94	87	90	78	82	59	66	30	43
	6	91	94	87	89	78	82	60	67	32	44

[a] The model is identified from (the calibration portion of) the listed dataset, optimizing the p-step ahead predictions, where $p = 1,\ 3,$ or 6.

TABLE III

ARX RESULTS (FIT VALUES, %) FOR THE SIMULATED DATASETS. HIGHLIGHTED VALUES REPRESENT CALIBRATION FITS.

Dataset/ Model[a]		Validation PH (Steps) 1 Cal	1 Val	3 Cal	3 Val	6 Cal	6 Val	12 Cal	12 Val	24 Cal	24 Val
A_{1S}	1	88	89	84	80	77	65	61	38	34	5
	3	88	89	84	81	77	66	61	39	34	5
	6	87	90	84	82	78	68	62	41	36	8
A_{2S}	1	88	85	81	75	69	56	45	23	9	-22
	3	87	84	82	77	72	60	50	27	15	-17
	6	86	84	81	78	73	63	55	33	23	-7
B_{1S}	1	93	92	88	88	80	81	64	66	40	41
	3	93	92	89	88	83	82	69	70	47	48
	6	93	92	88	88	83	82	73	72	54	54
B_{2S}	1	92	94	87	90	78	82	60	69	32	49
	3	92	94	88	91	81	85	64	73	37	54
	6	91	94	88	90	81	86	67	77	42	59

[a] The model is identified from (the calibration portion of) the listed dataset, optimizing the p-step ahead predictions, where $p = 1,\ 3,$ or 6.

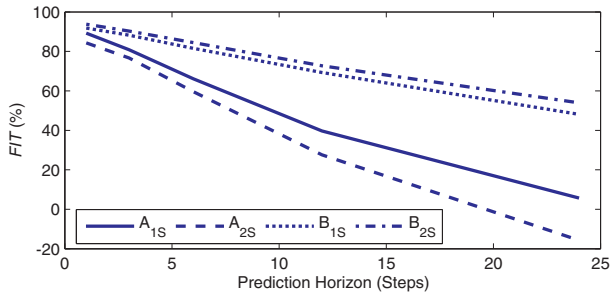

Fig. 2. FIT values for validation portions of the simulated datasets for ARX models.

TABLE IV

AR RESULTS (FIT VALUES, %) FOR THE SUBJECT DATASETS. HIGHLIGHTED VALUES REPRESENT CALIBRATION FITS.

Dataset/ Model[a]		Validation PH (Steps) 1 Cal	1 Val	3 Cal	3 Val	6 Cal	6 Val	12 Cal	12 Val	24 Cal	24 Val
A_1	1	84	85	64	65	45	41	20	8	-3	-15
	3	84	85	64	65	45	42	21	9	-2	-14
	6	84	85	64	66	45	43	21	11	0	-11
A_2	1	81	87	59	69	39	47	13	16	-7	-12
	3	81	87	59	69	39	47	13	17	-6	-11
	6	81	88	58	70	40	49	14	17	-6	-10
B_1	1	86	91	69	77	50	59	26	36	2	17
	3	86	91	69	77	50	59	26	36	3	17
	6	86	91	69	77	50	59	26	36	3	16
B_2	1	91	92	77	75	59	55	30	29	4	2
	3	91	92	77	76	59	55	30	30	4	3
	6	90	91	77	75	60	55	31	30	6	4

[a] The model is identified from (the calibration portion of) the listed dataset, optimizing the p-step ahead predictions, where $p = 1,\ 3,$ or 6.

Modeling results for the AR and ARX models using the subject data are listed in Tables IV and V, respectively. As for the simulated data, the prediction accuracy is largely insensitive to the calibration PH or the type of model (AR versus ARX). The prediction accuracy does vary substantially, however, with the validation PH, the portion of the data used in the validation, and the T1DM subject. Again, as for the simulated data, the model predictions deteriorate as the validation PH is extended to 24 steps, especially for subject A, whose FIT values for the validation portions are either very low or negative.

Again, the prediction accuracy varies significantly with the subject, although to a lesser extent. Combining the AR and ARX results, the ranges (low−high) of FIT values for the validation sections using 6-step, 12-step, and 24-step predictions are, respectively,

- Subject A: $41\% - 51\%$, $7\% - 24\%$, and $-17\% - 4\%$
- Subject B: $55\% - 62\%$, $29\% - 41\%$, and $2\% - 17\%$.

Neither the increased complexity of the ARX models compared to the AR models, nor the extension of the calibration PH resulted in statistically significant improvements in predictions of validation data (for each comparison, the standard deviation was greater than the mean improvement).

Fig. 3 shows how the ARX model predictions of the validation sections vary with the validation PH. The FIT values are significantly lower than the corresponding values for the simulated data (see Fig. 2). Again, the two curves for each subject group together closely.

Tables VI and VII show the results for the simulated and subject data, respectively, using ARMAX models. Only the 1-step ahead predictions were optimized in identifying the ARMAX models.

For the simulated data, the increased complexity of the ARMAX models results in significantly improved model predictions compared to the AR and ARX models. The 12-

TABLE V

ARX RESULTS (FIT VALUES, %) FOR THE SUBJECT DATASETS. HIGHLIGHTED VALUES REPRESENT CALIBRATION FITS.

Dataset/ Model[a]		Validation PH (Steps) 1 Cal	1 Val	3 Cal	3 Val	6 Cal	6 Val	12 Cal	12 Val	24 Cal	24 Val
	1	84	85	66	64	47	41	24	7	1	-17
A_1	3	84	85	66	65	48	42	25	9	3	-14
	6	84	85	66	65	48	43	26	10	4	-13
	1	81	87	59	71	40	51	16	24	-2	-4
A_2	3	81	83	60	66	41	44	17	18	2	-11
	6	81	83	59	65	41	41	19	13	4	-16
	1	86	91	70	77	51	59	26	36	2	17
B_1	3	86	91	71	77	52	60	28	37	4	17
	6	86	90	70	77	53	60	30	38	5	17
	1	91	92	78	78	61	61	33	40	6	12
B_2	3	90	91	78	78	62	62	35	41	8	11
	6	90	91	78	78	63	62	36	41	7	9

[a] The model is identified from (the calibration portion of) the listed dataset, optimizing the p-step ahead predictions, where $p = 1,\ 3,$ or 6.

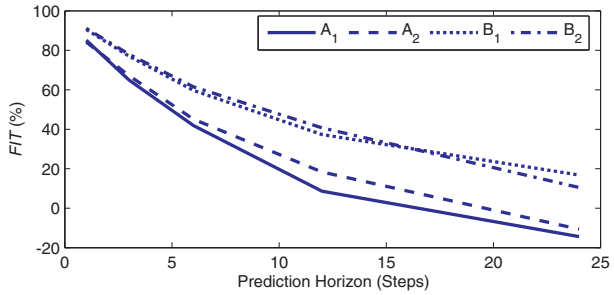

Fig. 3. FIT values for validation portions of the subject datasets for ARX models.

step ahead predictions of validation data for the ARMAX models were $24.5 \pm 8.9\%$ (FIT values, mean $\pm$ standard deviation) higher than those of the AR models, and $18.2 \pm 8.9\%$ higher than those of the ARX models.

For the subject data, the predictions for the ARMAX models were not statistically different than those of the AR and ARX models (again, the standard deviation was greater than the mean improvement for each comparison).

Table VIII illustrates the differences in 6-step and 12-step validation results due to the type of model used for each subject. For the virtual subjects A_S and B_S, improvement is obtained in the predictions as the model complexity increases from AR to ARX to ARMAX. This trend is not observed in the T1DM subjects; in fact, the ARX models explain the most variability in the validation data for both prediction horizons.

The RMSE values listed in Table VIII are comparable and in some cases superior to those obtained in previous studies [8,9] that used ambulatory T1DM subject data.

Figs. 4 and 5 show representative 6-step and 12-step ahead validation results for the ARX models for datasets B_{2S} and B_2, respectively. The process dynamics immediately following a meal are often captured reasonably well, but the model predictions show a strong "shadow" effect which becomes more pronounced with increasing PH.

TABLE VI

ARMAX RESULTS (FIT VALUES, %) FOR THE SIMULATED DATASETS. HIGHLIGHTED VALUES REPRESENT CALIBRATION FITS.

Dataset/ Model[a]	Validation PH (Steps) 1 Cal	1 Val	3 Cal	3 Val	6 Cal	6 Val	12 Cal	12 Val	24 Cal	24 Val
A_{1S}	90	89	89	85	86	76	78	57	59	37
A_{2S}	90	87	88	84	85	76	77	57	58	24
B_{1S}	95	94	94	93	92	92	89	88	79	79
B_{2S}	94	95	92	93	88	90	76	80	45	56

[a] The model is identified from (the calibration portion of) the listed dataset, optimizing the 1-step ahead predictions.

TABLE VII

ARMAX RESULTS (FIT VALUES, %) FOR THE SUBJECT DATASETS. HIGHLIGHTED VALUES REPRESENT CALIBRATION FITS.

Dataset/ Model[a]	Validation PH (Steps) 1 Cal	1 Val	3 Cal	3 Val	6 Cal	6 Val	12 Cal	12 Val	24 Cal	24 Val
A_1	84	86	67	66	49	44	28	10	4	-21
A_2	81	85	59	60	40	29	17	-8	6	-46
B_1	86	91	71	78	53	60	28	34	-6	3
B_2	91	92	79	78	64	61	37	36	3	-8

[a] The model is identified from (the calibration portion of) the listed dataset, optimizing the 1-step ahead predictions.

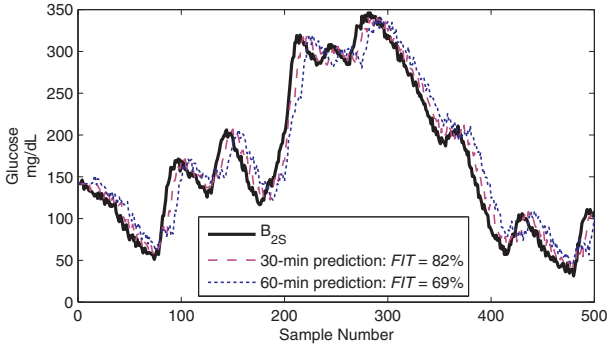

Fig. 4. Representative ARX validation results for Subject B_S. The dataset shown is the validation portion of dataset B_{2S}.

The effects of the initial parameter values in the nonlinear least-squares optimizations were investigated by performing optimizations using sets of initial parameter values significantly different from those obtained from the 1-step optimizations. These perturbations in initial values had little or no effect on the resulting model or its prediction capability.

V. DISCUSSION

For the simulated data, the best modeling results were achieved with ARMAX models. Average FIT values for 30-min predictions of validation simulated data were 76% and 91% for virtual subjects A_S and B_S, respectively. The

TABLE VIII

MEAN VALIDATION RESULTS FOR EACH SUBJECT-MODEL IDENTIFIED OPTIMIZING THE 1-STEP PREDICTIONS.

Subject[a]	Model	FIT (%) PH (Steps) 6	FIT (%) PH (Steps) 12	RMSE (mg/dL) PH (Steps) 6	RMSE (mg/dL) PH (Steps) 12
A_S	AR	59	29	18	31
	ARX	60	31	17	30
	ARMAX	76	57	11	19
B_S	AR	78	59	18	32
	ARX	82	67	16	28
	ARMAX	91	84	7	11
A	AR	44	12	14	22
	ARX	46	16	14	21
	ARMAX	36	1	16	25
B	AR	57	33	25	38
	ARX	60	38	24	36
	ARMAX	60	35	24	37

[a] "Subjects" A_S and B_S are virtual subjects whose data was simulated using the physiological model.

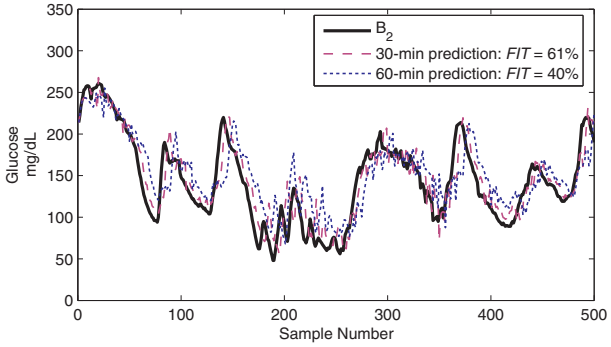

Fig. 5. Representative ARX validation results for Subject B. The dataset shown is the validation portion of dataset B_2.

best modeling results for the subject data were significantly worse than for the simulated data, and were achieved with ARX models. Average FIT values for 30-min predictions of validation subject data were 46% and 60% for subjects A and B, respectively.

This disparity between the modeling results for the simulated and subject data may be due in part to the nature of the ambulatory subject data, which are often characterized by unexplained excursions. These excursions are evidenced by comparisons with the corresponding simulated data, obtained from the deterministic physiological model.

For both subject and simulated data, prediction accuracy deteriorates quickly as the prediction horizon is extended. For reasonably reliable predictions, the maximum horizon in this study was 60 min. However, predictions over horizons of only 15–30 min may well be suitable for model-based control.

Identifying dynamic models based on optimizing the p-step ahead predictions, $p > 1$, resulted in only a marginal improvement, if any, in the predictive quality of the models. This modest improvement is most likely not worth the increased computational load, compared to a standard 1-step least-squares identification method. This conclusion is especially pertinent to recursive estimation schemes in which model parameters are updated on-line at each sampling instant.

As the complexity of the models increases, i.e., AR to ARX to ARMAX, little improvement is effected in their capabilities to predict the subject data. Thus, AR or ARX models are recommended for use with ambulatory subject data. ARX models are especially attractive given their potential to be incorporated into a model-based controller.

VI. ACKNOWLEDGMENTS

Financial support from the National Institutes of Health, grant R21–DK069833–02, and from the Juvenile Diabetes Research Foundation, grant 22–2006–1115, is gratefully acknowledged.

REFERENCES

[1] 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.*, vol. 329, 1993, pp. 977–986.

[2] Centers for Disease Control and Prevention, National Diabetes Fact Sheet: General Information and National Estimates on Diabetes in the United States, 2005, U.S. Department of Health and Human Services, Atlanta, GA, Tech. Rep., 2005.

[3] B.W. Bequette, A Critical Assessment of Algorithms and Challenges in the Development of a Closed-Loop Artificial Pancreas, *Diabetes Technol. Ther.*, vol. 7, 2005, pp. 28–47.

[4] R.S. Parker and F.J. Doyle III, Control-Relevant Modeling in Drug Delivery, *Adv. Drug Deliv. Rev.*, vol. 48, 2001, pp. 211–228.

[5] R. Bellazzi, C. Siviero, M. Stefanelli, and G. De Nicolao, Adaptive Controllers for Intelligent Monitoring, *Artif. Intell. Med.*, vol. 7, 1995, pp. 515–540.

[6] T. Bremer and D.A. Gough, Is Blood Glucose Predictable from Previous Values?, *Diabetes*, vol. 48, 1999, pp. 445–451.

[7] R. Hovorka, V. Canonico, L.J. Chassin, U. Haueter, M. Massi-Benedetti, M.O. Federici, 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, *Physiol. Meas.*, vol. 25, 2004, pp. 905–920.

[8] G. Sparacino, F. Zanderigo, S. Corazza, A. Maran, A. Facchinetti, and C. Cobelli, Glucose Concentration Can Be Predicted Ahead in Time from Continuous Glucose Monitoring Sensor Time-Series, *IEEE Trans. Biomed. Eng.*, vol. 54, 2007, pp. 931–937.

[9] J. Reifman, S. Rajaraman, A. Gribok, and W.K. Ward, Predictive Monitoring for Improved Management of Glucose Levels, *J. Diabetes Sci. Technol.*, vol. 1, 2007, pp. 478–486.

[10] 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, *Am. J. Physiol. Endocrinol. Metab.*, vol. 282, 2002, pp. E992–E1007.

[11] L. Ljung, *System Identification: Theory for the User*, 2nd ed., Prentice Hall PTR, Upper Saddle River, NJ; 1999.

[12] D.S. Shook, C. Mohtadi, and S.L. Shah, A Control-Relevant Identification Strategy for GPC, *IEEE Trans. Automat. Control*, vol. 37, 1992, pp. 975–980.

[13] D.S. Shook, C. Mohtadi, and S.L. Shah, Identification for Long-Range Predictive Control, *IEE Proc. D*, vol. 138, 1991, pp. 75–84.