

# The Importance of Different Frequency Bands in Predicting Subcutaneous Glucose Concentration in Type 1 Diabetic Patients

Yinghui Lu, Andrei V. Gribok, W. Kenneth Ward, and Jaques Reifman\*

**Abstract**—We investigated the relative importance and predictive power of different frequency bands of subcutaneous glucose signals for the short-term (0–50 min) forecasting of glucose concentrations in type 1 diabetic patients with data-driven autoregressive (AR) models. The study data consisted of minute-by-minute glucose signals collected from nine deidentified patients over a five-day period using continuous glucose monitoring devices. AR models were developed using single and pairwise combinations of frequency bands of the glucose signal and compared with a reference model including all bands. The results suggest that: for open-loop applications, there is no need to explicitly represent exogenous inputs, such as meals and insulin intake, in AR models; models based on a single-frequency band, with periods between 60–120 min and 150–500 min, yield good predictive power (error < 3 mg/dL) for prediction horizons of up to 25 min; models based on pairs of bands produce predictions that are indistinguishable from those of the reference model as long as the 60–120 min period band is included; and AR models can be developed on signals of short length (~300 min), i.e., ignoring long circadian rhythms, without any detriment in prediction accuracy. Together, these findings provide insights into efficient development of more effective and parsimonious data-driven models for short-term prediction of glucose concentrations in diabetic patients.

**Index Terms**—Autoregressive prediction models, continuous glucose monitoring (CGM), diabetes, glucose dynamics, glucose signal frequency analysis.

## I. INTRODUCTION

RECENT developments in continuous glucose monitoring (CGM) offer new opportunities and challenges in data

Manuscript received November 3, 2009; revised February 12, 2010; accepted March 25, 2010. Date of publication April 15, 2010; date of current version July 14, 2010. This work was supported in part by the U.S. Army Medical Department, Advanced Medical Technology Initiative, Telemedicine and Advanced Technology Research Center of the U.S. Army Medical Research and Materiel Command, Fort Detrick, MD, and in part by the U.S. Air Force Diabetes Research Program. *Asterisk indicates corresponding author.*

Y. Lu is with the Bioinformatics Cell, Telemedicine and Advanced Technology Research Center, U.S. Army Medical Research and Materiel Command, Fort Detrick, MD 21702 USA (e-mail: ylu@bioanalysis.org).

A. V. Gribok was with the Bioinformatics Cell, Telemedicine and Advanced Technology Research Center, U.S. Army Medical Research and Materiel Command, Fort Detrick, MD 21702 USA. He is currently with the Department of Nuclear Engineering, The University of Tennessee, Knoxville, TN 37996 USA (e-mail: agribok@gmail.com).

W. K. Ward is with the iSense Corporation and Oregon Health and Sciences University, Wilsonville, OR 97070 USA (e-mail: kenward503@msn.com).

\*J. Reifman is with the Bioinformatics Cell, Telemedicine and Advanced Technology Research Center, U.S. Army Medical Research and Materiel Command, Fort Detrick, MD 21702 USA (e-mail: jaques.reifman@us.army.mil).

Digital Object Identifier 10.1109/TBME.2010.2047504

collection and analysis [1], as these CGM devices can sample subcutaneous glucose concentrations as frequently as every minute and store the sampled time series for up to several days for retrospective analysis. On one hand, this abundance of information opens new opportunities in data analysis and the understanding of the mechanisms of glucose regulation, while on the other hand, it poses additional challenges in terms of information processing and interpretation. Long and frequently sampled time-series data collected from CGM devices have naturally invited the use of techniques that require continual availability of glucose data, such as Kalman filtering [2] and techniques that are purely data driven, such as autoregressive (AR) models [3], for short-term prediction of glucose concentrations in diabetic patients [4]–[8]. The obvious advantages of AR models are their analytical tractability and their ability to linearly extrapolate future time-series values, producing a reliable and accurate forecast. However, to yield accurate predictions, AR models need to be fitted to a “training” signal and the fitting procedure should lead to a sequence of AR model coefficients that capture the major frequency components (or bands) in the glucose signal.

The frequency bands in blood glucose signals reflect the physiological mechanisms of glucose regulation, with different mechanisms driving different frequency bands. For example, in healthy individuals, the pulsatile insulin secreted by the pancreas is reflected in patterns of blood glucose signal oscillations with periods between 4 and 15 min [9]. Similarly, the patterns associated with postprandial glucose regulation of healthy individuals have predominant periods ranging from 51 to 112 min [10], with larger amplitude oscillations observed after evening meals and smaller ones after morning meals, reflecting the circadian rhythmicity of glucose regulation [11].

Recently, Rahaghi and Gough suggested that the spectrum of oscillations in blood glucose signals of healthy individuals could be characterized by four major frequency bands [12]. They suggest that the lowest frequency band, corresponding to periods of at least 700 min (Band IV), reflects patterns of glucose regulation associated with ultradian rhythms of long periodicity and circadian rhythms. The first midfrequency band, spanning oscillations with periods from 150 to 500 min (Band III), is believed to be primarily associated with the schedules of exogenous inputs, such as the time of meals and insulin boluses intake, while the second midfrequency band, spanning periods from 60 to 120 min (Band II), is deemed to reflect insulin secretion in response to continuous enteral nutrition, constant intravenous glucose infusion, or ingestion of a meal or insulin

boluses [10], [13]. These oscillations are considered to be intrinsic responses [12] because they can be interpreted as a step, or an impulse, response of the glucose regulatory system. The highest frequency band, with periods between 5 and 15 min (Band I), is assumed to be generated by rapid pulsatile insulin secretion by pancreatic  $\beta$  cells and is best observed in fasting, nondiabetic subjects [9]. However, because  $\beta$  cells are destroyed in type 1 diabetic patients and insulin is not produced, these high frequencies are absent in the blood glucose signals of this patient population [14].

In contrast to blood glucose, there is a limited body of work concerning the frequency analysis of subcutaneous CGM signals, which, due to time delays and signal attenuation from blood-to-interstitial transport, may differ from that of the blood glucose, requiring additional independent studies. By and large, the existing studies of the spectrum of CGM signals are limited to Fourier series analysis [15], [16] and are not guided to educating the development of predictive glucose models. For example, Breton *et al.* performed Fourier analysis for determining the minimal sampling periods of blood and subcutaneous CGM signals of type 1 diabetic patients and concluded that CGM signals of periods shorter than 36 min are nonexistent [15]. This loss of information is attributed to the lowpass filtering effect caused by the blood-to-interstitial transport of glucose concentration. In a separate Fourier analysis study, Miller and Strange [16] suggest that, in type 1 diabetic patients, the amplitude of the second and third harmonics of CGM signals are correlated with mean HbA1c values and that, in type 2 diabetic patients, the frequency content with periods shorter than 72 min is characterized by white noise.

The efficient development of glucose concentration predictive models for type 1 diabetic patients, in particular data-driven AR models, where the model coefficients capture the frequency content of the signal [4], [5], require a more detailed analysis of the frequency components of the CGM glucose signal so that the relative importance and predictive power of the different frequency bands are properly characterized. In particular, the answers to the following key, yet unknown, questions need to be addressed: 1) How to optimize the glucose signal filtering process so as to eliminate uninformative signal components while keeping the important ones? 2) Which frequency components must be present in an AR model to yield accurate, short-term predictions? 3) Whether there is a need to explicitly represent exogenous inputs, such as meals and insulin intake, into the model? and 4) How much data are needed to develop an accurate AR model?

In this paper, we attempt to address the aforementioned questions by first associating the frequency content of subcutaneous glucose signals with those of blood glucose signals and then investigating the relative importance of the different frequency bands of the subcutaneous glucose signal in AR modeling. Based on the four major frequency bands suggested by Rahaghi and Gough [12], we applied subband AR modeling [17] to CGM signals of type 1 diabetic patients and determined the predictive power of the different frequency bands and their dependencies on prediction horizon. We found that, provided enough training data were available, the AR models captured all the frequency

information present in the subcutaneous glucose concentration signal, obviating the need to explicitly represent exogenous inputs into the model, such as meals and insulin, for open-loop applications. We also found that the frequency band associated with the intrinsic response of glucose regulation was indispensable for obtaining accurate predictions up to 50 min ahead, although the energy content of this frequency band in the glucose signal is low ( $\sim 1.5\%$  of the total signal's energy). Finally, we concluded that a training signal as short as 300 min, i.e., one that excludes low circadian rhythm frequencies, was capable of producing accurate predictions, potentially shortening data collection time and expediting model development. Together, these findings provide insights into the development of more effective and parsimonious AR models for short-term predictions of subcutaneous glucose concentrations in diabetic patients.

## II. METHODS

### A. Study Population

In this paper, we analyzed the temporal dynamics and frequency content of subcutaneous glucose time-series data of nine deidentified type 1 diabetic patients. Subcutaneous glucose measurements were collected on a minute-by-minute basis for each of the nine subjects for approximately 5 days with the iSense CGM system [4], [6]. Subjects were confined to the investigational site for the whole duration of the study and limited to mild physical activity. Subjects were included if they were between 18 and 70 year of age, had been diagnosed with type 1 diabetes and treated with insulin for at least 12 months, had a body mass index of  $< 35.0 \text{ kg/m}^2$ , and had glycated hemoglobin (HbA1c) of  $> 6.1\%$ . Subjects were excluded if they had acute and severe illness apart from diabetes, had a clinically significant abnormal electrocardiogram, hematology, or biochemistry-screening test, or had any disease requiring the use of anticoagulants. In addition, subjects were excluded if they were pregnant or lactating.

The subjects were provided three meals per day at 9 A.M., 1 P.M., and 7 P.M. (plus a midafternoon snack at 4 P.M.) and continued their normal insulin therapy, which was provided either by an external continuous insulin pump or by multiple daily subcutaneous injections. In addition, each subject received a bolus of regular or ultra rapid insulin immediately before each meal (excluding the snack) either by subcutaneous injection or via the subcutaneous catheter of the insulin pump. Fig. 1(a) shows the raw CGM signal for one patient (subject #6) in our study collected over 4000 min (66.7 h), where the time points of food and insulin intake are illustrated by vertical lines. The figure illustrates the typical daily variations in glucose concentration, including a drastic increase in concentration between 6 and 9 A.M. although no food or other nutrients were taken during the night. This is due to a circadian rhythm known as the "dawn phenomenon," which is explained by an increase in insulin resistance caused by certain hormones, and occurs in both diabetic and nondiabetic individuals alike [18], [19].

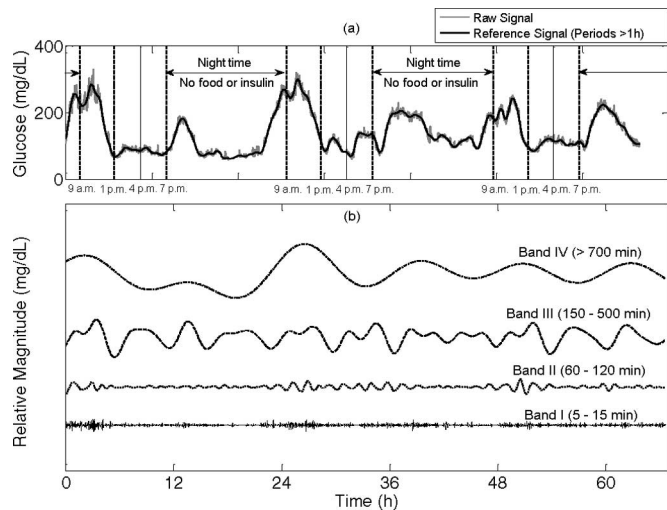


Fig. 1. Subcutaneous glucose concentration of a typical type 1 diabetic patient. (a) Raw and reference (i.e., filtered) glucose signals. (b) Relative magnitude of the corresponding bandpass-filtered time series of the original raw signal representing the four major frequency bands.

### B. Frequency Analysis

To study the relative importance and predictive power of the four major frequency bands suggested by Rahaghi and Gough [12] for subcutaneous CGM signals, we developed four bandpass filters, where each filter only passed glucose signals in one of the four period bands (Band I: 5–15 min, Band II: 60–120 min, Band III: 150–500 min, and Band IV:  $\geq 700$  min) of a raw CGM time-series signal. Fig. 1(b) shows the corresponding four bandpass-filtered signals extracted from the raw time-series signal in Fig. 1(a). For example, through bandpass filtering of Band IV, only patterns associated with circadian rhythms, such as the 24-h dawn phenomenon, and ultradian rhythms with periodicities  $\geq 700$  min were extracted from the raw signal. However, because the Band I frequencies in Fig. 1(b) are expected to consist mainly of measurement noise devoid of any significant physiological information [15], [16], we performed limited analysis for this high-frequency band.

We also developed multibandpass filters that only passed the frequencies associated with each of the pairwise combinations of three of the four bands, II, III, and IV. For example, for filters that passed the pairwise combination of bands II+III, we used a multibandpass filter that only passed the frequencies in each of these two bands, eliminating all other frequencies, including those in the 120–150 min gap between these bands. To obtain a reference glucose concentration signal and corresponding model against which all other subband signals and models could be compared with, we used a lowpass filter with a cut-off frequency of  $1/3600$  Hz (equivalent to a period of 60 min), which removed the high-frequency content of the CGM signal up to the lower bound of Band II. This cut-off frequency was selected for two reasons. First, as discussed above, Fourier series analysis from different studies of diabetic patients provides substantial evidence that CGM signals with periods shorter than 36 min are nonexistent [15] and that periods shorter than 72 min are characterized by white noise [16]. Second, our group has shown

that, to obtain consistent AR coefficients and robust models from CGM signals, it is necessary to remove frequencies with periods shorter than  $\sim 90$  min [4]. The resulting reference signal is illustrated in Fig. 1(a).

Finally, to analyze the overall frequency content of glucose concentration signals of type 1 diabetic patients, we used the Welch's method with a Hamming window of 50% overlap to compute the power spectral density (PSD) of the raw CGM signals [20].

### C. Autoregressive Modeling

An AR model is a type of linear model where a future signal  $y_{n+1}$ , at discrete time  $n+1$ , is represented by a linear combination of previous signal observations  $y_{n-i}$ ,  $i = 0, 1, 2, \dots, m-1$ , plus white noise  $\varepsilon_{n+1}$

$$y_{n+1} = \sum_{i=0}^{m-1} b_i y_{n-i} + \varepsilon_{n+1} \quad (1)$$

where  $m$  denotes the order of the model, i.e., the number of previous observations used to represent  $y_{n+1}$ , and  $b_i$  represents fixed model coefficients. The coefficients  $b_i$  describe the temporal correlations between each of the previous signals  $y_{n-i}$ ,  $i = 0, 1, 2, \dots, m-1$ , and the next one  $y_{n+1}$ , and capture the frequency content of the underlying signal [3]. Therefore, we may use the set of coefficients  $b_i$  and previous signals  $y_{n-i}$  to make one-step-ahead predictions for  $y_{n+1}$ , i.e.,

$$\hat{y}_{n+1} = \sum_{i=0}^{m-1} b_i y_{n-i} \quad (2)$$

where  $\hat{y}_{n+1}$  denotes the predicted value for  $y_{n+1}$ . Equation (2) can also be used to make  $k$ -step-ahead predictions, with  $k = 2, 3, \dots$ , by iteratively substituting the  $k-1$  predicted values for the corresponding  $k-1$  yet unobserved signals. For example, we may make two-step-ahead predictions  $\hat{y}_{n+2}$  by substituting the predicted value  $\hat{y}_{n+1}$  for its unobserved signal  $y_{n+1}$  in the right-hand side of (2).

An AR model can also be represented in the frequency domain. By performing a  $Z$ -transform of (1), we can convert its discrete-time representation into a frequency-domain representation

$$H(z) = \frac{1}{1 + \sum_{i=0}^{m-1} -b_i z^{-(i+1)}} \quad (3)$$

where  $H(z)$  denotes the corresponding transfer function of the AR model. Because the coefficients  $b_i$  in an AR model capture the frequency content of the glucose signal, the PSD  $P(\omega)$  of this all-pole transfer function  $H(z)$

$$P(\omega) = \frac{1}{\left| 1 + \sum_{i=0}^{m-1} -b_i e^{-j\omega(i+1)} \right|^2} \quad (4)$$

where  $\omega$  denotes the radian frequency, can be used to approximate the spectrum of the underlying glucose signal [21].

To investigate the relative importance and predictive power of each of the three Bands (II, III, and IV) associated with

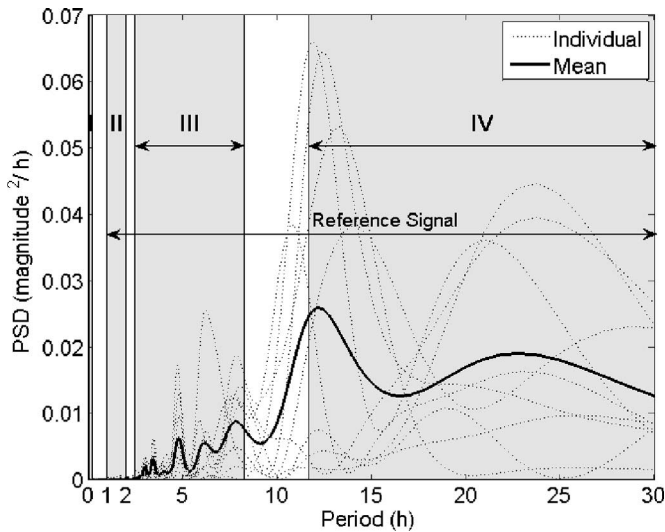


Fig. 2. Power spectral densities (PSD) of the raw glucose time series of nine patients and their averaged spectrum estimated using the Welch's method. The four gray areas correspond to the four modeled frequency bands.

the different dynamics of glucose regulation, we applied the method of regularized least squares to fit the coefficients  $b_i$  and obtain AR models of order 30 ( $m = 30$ ), as proposed in [4]. Regularization yields smoothly varying AR-model coefficients  $b_i$ , a requirement for obtaining stable and accurate models with clinically acceptable time lags [4]. For all calculations, we used consecutive 2000 min (or 2000 data points) of the glucose signal of a subject to fit the AR model. For each subject, we developed separate AR models for each of the three bands as well as for the three pairwise combinations of bands. For comparison purpose, we also developed (reference) AR models for each subject using the reference signal consisting of all three bands. Moreover, to further investigate the glucose dynamics within the different frequency bands and their combinations and compare them to the spectrum estimated from the raw CGM signal using the Welch's method, we calculated the PSD using (4) for each of the AR models.

The predictive power was quantified by root mean squared error (RMSE) deviations, defined as the square root of the mean of the squared differences between the predicted value  $\hat{y}_n$  and the observed value  $y_n$ . The predictive performance of the models was evaluated for each subject using their corresponding testing data between 2000 and 4000 min, and each model was evaluated for different prediction horizons, ranging from 1 to 50 min.

### III. RESULTS

Fig. 2 shows the PSDs estimated using the raw CGM signals for each of the nine patients in the study, where the signal energy was plotted as a function of the signal period instead of frequency to facilitate physiological interpretation. The majority of the signal's energy fell within the two longest period ranges, Bands III and IV. In contrast, Band II contained a relatively small amount of the total energy ( $\sim 1.5\%$ ), while the energy in Band I was only  $\sim 0.6\%$  of the total, which supports the conclusion that, as in blood glucose signals [14], CGM signals of type 1

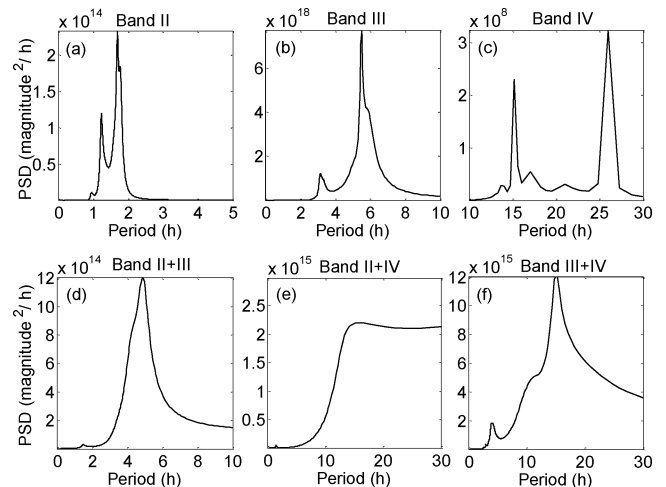


Fig. 3. (a)–(c) Mean power spectral densities (PSD) of single-frequency bands and (d)–(f) pairwise-frequency bands estimated through the AR-model coefficients.

diabetic patients lack high-frequency pulsatile insulin secretion patterns.

Fig. 2 also illustrates the differences in the signal spectrum profile for each of the frequency bands. For example, Band IV was characterized by two distinct peaks at  $\sim 12$  and  $\sim 24$  h associated with well-established ultradian and circadian periods [11]. Similarly, Band III was characterized by five distinct peaks at approximately 3.0, 3.5, 5.0, 6.0, and 8.0 h. The periods at 3.0 and 6.0 h exactly coincided with meal and insulin schedules at 1 P.M., 4 P.M., and 7 P.M., whereas the periods at 3.5 and 5.0 h were likely related to the 4.0 h time interval between meal and insulin schedules at 9 A.M. and 1 P.M. We speculate that the 8.0 h period is not associated with meal and insulin schedules, but rather with ultradian oscillations associated with sleep [13]. Finally, analysis of the spectrum profile for Band II indicated a number of peaks, including two predominant ones with periods at  $\sim 80$  and  $\sim 105$  min (results not shown in the scale used in Fig. 2). This finding is corroborated by the work of Simon *et al.* [10], who identified the predominant period of the intrinsic regulatory response in blood glucose to vary from 51 to 112 min.

Fig. 3 shows the mean PSDs (averaged over the nine subjects) for Bands II, III, and IV and their pairwise combinations estimated from the corresponding AR-model coefficients using (4). The results show that, as expected, each AR model only captured the frequency content of the corresponding frequency band(s) of the CGM signals. Comparison with Fig. 2 indicated that the PSDs estimated by the subband AR models were, in general, less resolved than the ones obtained from the raw signal. For instance, Fig. 2 shows five salient periods in Band III while Fig. 3(b) depicts only two of them, at  $\sim 3.0$  and  $\sim 6.0$  h. This is mainly attributed to the constraint, or regularization, imposed on the fitting of the AR model coefficients  $b_i$  [4].

Fig. 4 shows the predictive performance of the subband AR models, where RMSEs, averaged over the nine subjects, were plotted as a function of prediction horizon for the reference model as well as for the single-band models [see Fig. 4(a)] and the pairwise-band models [see Fig. 4(b)]. As expected, Fig. 4(a)

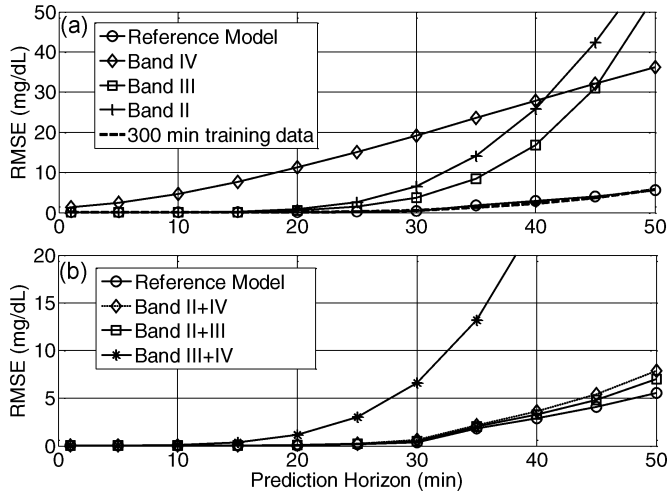


Fig. 4. Average RMSE of the autoregressive model predictions as a function of prediction horizon for the reference model based on all frequency bands, and for models based on: (a) single-frequency bands and (b) pairwise-frequency bands.

shows that the reference model, which captured the full dynamic range of the glucose signal, yielded the smallest RMSEs across the 0–50 min prediction horizons. For prediction horizons of <25 min, the AR models obtained with Band II or Band III frequencies had essentially the same predictive performance as the model obtained with the reference signal. This result is instructive because it suggests that, for short prediction horizons, middle-frequency dynamics, resulting from intrinsic oscillations or schedules of exogenous inputs, are sufficient to produce accurate models and that low-frequency dynamics, associated with circadian rhythms, may not be necessary. For longer prediction horizons (25–50 min), Band III models outperformed Band II models because the former contained more low-frequency content required for predicting longer horizons. This was evident by the performance of Band IV model, which had the worst predictive performance for short horizons (<40 min) while outperforming Band II and III models for prediction horizons of >45 min. We also confirmed that Band I signals were not informative, as simulations showed that models based solely on this frequency band consistently underperformed the other models (results not shown).

Fig. 4(b) compares the performance of the pairwise-band models. In stark contrast with the single-band models, the results indicate that when we considered frequencies from Band II, with either Band III or Band IV, the resulting pairwise models were as predictive over the 0–50-min prediction horizon as those obtained with the reference model. These results suggest that although a frequency band may not have sufficient predictive power by itself, in combination with other bands, the resulting models could be very accurate. Importantly, they also indicate that the energy content of a signal alone was not necessarily an appropriate metric for indicating its predictive power (as discussed earlier, Band II accounts for only ~1.5% of the total energy of the glucose signal). Conversely, models constructed using frequencies from Bands III+IV, which account

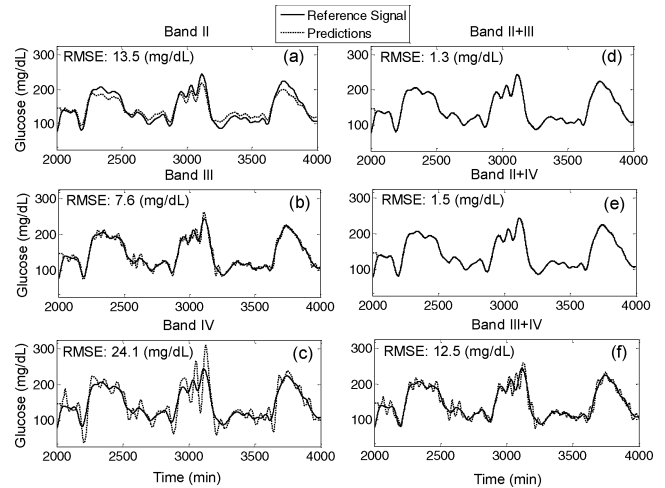


Fig. 5. (a)–(c) Reference glucose signal for subject #6 (solid line), 35-min-ahead autoregressive model predictions (dotted line), and associated RMSE for models based on single-frequency bands and (d)–(f) pairwise-frequency bands.

for ~93.0% of the signal’s total energy, showed inferior performance when compared with the other pairwise models and some of the single-band models.

These results suggest that the prediction accuracy of AR models is highly dependent on which frequencies of the CGM signal are captured by the AR model coefficients  $b_i$  and weakly dependent on the resolution of the captured frequencies. For example, as discussed earlier, the AR models for Band III only captured two [see Fig. 3(b)] of the five salient periods in this frequency band (see Fig. 2). Nevertheless, for a 25-min prediction horizon, the Band III AR models yielded an average RMSE of <2 mg/dL [see Fig. 4(a)].

To provide further insights into the role of the different frequency bands in AR-model predictions, we analyzed 35-min-ahead predictions for a typical patient (subject #6) for models based on the six possible combinations of single and pairwise frequency bands, as illustrated in Fig. 5.

Fig. 5(a) shows that while the Band II model was capable of yielding smooth predictions for the high-frequency oscillations, such as those around 3000 min, it either underpredicted or overpredicted the low-frequency trends, systematically generating a prediction bias. Conversely, Fig. 5(b) shows that the Band III model yielded accurate, albeit nonsmooth, predictions for the low-frequency trends while overpredicting the high-frequency excursions. These were further accentuated by the Band IV model [see Fig. 5(c)], which produced predictions with large oscillations. In contrast, Fig. 5(d) and (e), show that when Band II was combined with, either Band III or Band IV, the resulting models were extremely accurate, yielding negligible RMSEs. Importantly, the exclusion of the Band III frequencies [see Fig. 5(e)], which are deemed to reflect the schedules of exogenous inputs, such as meals and insulin intake, did not negatively impact the predictive power of the Band II+IV model. This supports the finding of Finan *et al.* [22], which suggests that providing the model with timing information about exogenous

inputs may be redundant. Moreover, the exclusion of the circadian rhythms in Band IV from the model did not impact the predictive power of the Band II+III model [see Fig. 5(d)]. This observation is important from a model development point of view, because it indicates that the training data can be obtained from time segments much shorter than 24 h. This conclusion is supported by simulations shown in Fig. 4(a), where the RMSE profile over the entire prediction horizon for a model trained on only 300 min of the reference signal (dashed line) coincided with that of the reference model trained on the entire 2000 min. Fig. 5(f) shows that a model based on the combination of Bands III+IV did not produce accurate predictions (RMSE = 12.5 mg/dL), supporting our finding that the inclusion of Band II frequencies was required to achieve high-fidelity models. Finally, we confirmed that models based on Band I frequencies lacked predictive power, yielding 35-min-ahead prediction RMSEs ranging from 25 mg/dL for pairwise-band models to 247 mg/dL for single-band models (results not shown).

#### IV. CONCLUSION AND DISCUSSION

In this paper, we linked the different frequency bands of the subcutaneous glucose signal with the corresponding ones from blood glucose and their physiological mechanisms of glucose regulation. Using subband analysis of CGM signals from type 1 diabetic patients, we investigated the relative importance and predictive power of these frequency bands for short-term AR-model predictions by addressing four outstanding issues: 1) How to optimize the glucose signal filtering process so as to eliminate uninformative signal components while keeping the important ones? 2) Which frequency components must be present in an AR model to yield accurate, short-term predictions? 3) Whether there is a need to explicitly represent exogenous inputs, such as meals and insulin intake, into the model? and 4) How much data are needed to develop an accurate AR model?

First, we showed that the high-frequency signals in Band I associated with pulsatile insulin secretion in healthy individuals are nonexistent and, hence, lack predictive power in the forecast of type 1 diabetic patients. This is supported by the work of Gough *et al.* [14] who show that high-frequency signals with periods shorter than 18 min should be treated as noise in type 1 diabetic patients. Moreover, the work of Breton *et al.* [15] suggests that signals with periods shorter than 36 min are noninformative in type 1 diabetic patients, supporting our recommendation to filter out signals with frequencies above 1/3600 Hz, i.e., periods below 60 min, before the development of AR models. Such filtering has been found to be required to yield consistent AR coefficients and robust models [4].

We hypothesize that similar filtering would be required to develop AR models for type 2 diabetic patients because while pulsatile secretion of insulin is not completely absent in this patient population it is drastically attenuated [9]. This hypothesis is corroborated by Miller and Strange [16] who suggest that, for type 2 diabetic patients, CGM signals with periods shorter than 72 min are characterized by white noise, and by our recent work, which shows that when signals with periods

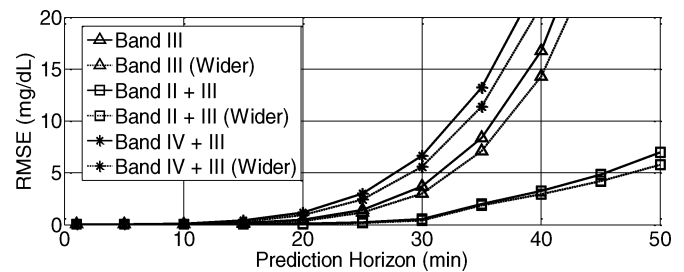


Fig. 6. Comparison of the average RMSE of glucose predictions for models based on the original range of Band III (150–500 min) and the wider range of Band III (120–700 min), where the latter eliminates the gaps between the bands.

below  $\sim 90$  min are filtered out the resulting AR models become portable from individual-to-individual regardless of the type of diabetes [5]. It is not evident, however, whether this filtering should be performed in signals from nondiabetic, fasting individuals with healthy  $\beta$  cells and prominent periods between 5 and 15 min [12].

Second, we showed that, among the single-band models, those based on Band II frequencies, which account for the intrinsic response of glucose regulation, and Band III frequencies, which represent the responses to external schedules, resulted in modest prediction errors ( $< 3$  mg/dL) for prediction horizons of up to 25 min. For models based on pairs of frequency bands, the inclusion of Band II, with either Bands III or IV, was imperative to achieving accurate predictions over the 0–50 min range of prediction horizons, matching those obtained by the reference model including all three bands. Overall, the pairwise-band models containing Band II outperformed each of the single-band models. These results strongly suggest that, for accurate AR-model predictions, Band II frequencies must be captured from the CGM signal. While this finding should have been expected from the well-established glucose dynamics in response to meals and insulin represented by the Band II frequencies, it is somewhat surprising that the energy content of these frequencies only account for  $\sim 1.5\%$  of the signal's total energy.

While the stratification of the glucose signal frequency into the four bands used here follows those suggested by Rahaghi and Gough [12], we note that other studies suggest slightly different ranges for the frequency bands and examine their underlying driving mechanisms from different perspectives [9]–[13], [15], [16], [18], [19]. Moreover, the chosen bands are not exhaustive, as there are uncovered frequency ranges, or gaps, between the selected bands. To determine the effect of the uncovered frequencies in our results, we extended each side of Band III to cover the entire range between 120 and 700 min, i.e., we closed the gaps between Band II and Band III and between Band III and Band IV, and recomputed the corresponding model predictions. Fig. 6 shows that, as expected, the wider Band III (120–700 min versus the original 150–500 min) produced smaller prediction errors, however, the improvements were marginal for both the single-band and the pairwise-band models. These results suggest that short-term AR-model predictions are not very

sensitive to slight variations in the selection of the ranges of the frequency bands.

Third, we found that the frequency content of Bands II, III, and IV capture the dynamics of glucose regulation of type 1 diabetic patients, obviating the need to explicitly represent and model exogenous inputs, such as caloric content of meals, amount and type of insulin, and their schedules, for short-term (0–50 min), open-loop, glucose concentration predictions. This is likely attributed to many factors: the blood-to-interstitial transport attenuation of the high-frequencies in the CGM signal below our Band II cutoff; the multiple periodicity of the insulin regimen in our study, where patients received daily bolus insulin and either continuous basal insulin or multiple long-acting insulin injections; and the observation that the most important periods for short-term predictions are those between 60 and 120 min (Band II), which encompass the time constants associated with the responses to meals and insulin intake.

The lack of a requirement to explicitly represent exogenous inputs for short-term predictions with autoregressive models is also supported by our recent work [5], which shows that an AR model based on the frequencies in Bands II–IV from one diabetic patient was able to accurately predict the glucose concentration of 33 other patients from three distinct studies, regardless of—among other factors—the different meals and insulin intake regimens. Moreover, Finan *et al.* have also reported that autoregressive models with exogenous inputs, i.e., ARX models, do not perform better than AR models that exclude such information [22]. Because information about these exogenous inputs is usually not readily available, the ability of AR models to make accurate predictions solely on the basis of prior CGM signals—for example, for open-loop alarms to prevent hypo- and hyper-glycemic episodes—provides a significant advantage over more traditional first-principle models that require such inputs. Conversely, for a model to be useful for a closed-loop artificial pancreas, it must represent and infer the effects of exogenous inputs.

Fourth, we found that AR models could be developed based on time-series signals of short length ( $\sim 300$  min), i.e., excluding circadian rhythm information, without any detriment in prediction accuracy. This observation simplifies and expedites model development by significantly shortening data-collection time.

In conclusion, we have analyzed the relative importance of the different frequency bands of subcutaneous glucose signals in type 1 diabetic patients within the context of developing data-driven AR models. This analysis provides insights, which should be useful for efficient development of more effective and parsimonious data-driven models for short-term prediction of glucose concentrations in diabetic patients.

#### ACKNOWLEDGMENT

The authors would like to thank G. Steil and S. Rajaraman for providing stimulating feedback and S. Ramakrishnan for helping with the figures.

The opinions and assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the U.S. Army or of the U.S. Department

of Defense. This paper has been approved for public release with unlimited distribution.

#### REFERENCES

- [1] D. C. Klonoff, "Continuous glucose monitoring—Roadmap for 21st century diabetes therapy," *Diabetes Care*, vol. 28, no. 5, pp. 1231–1239, May 2005.
- [2] C. C. Palerm and B. W. Bequette, "Hypoglycemia detection and prediction using continuous glucose monitoring—A study on hypoglycemia clamp data," *J. Diabetes Sci. Technol.*, vol. 1, no. 5, pp. 624–629, Sep. 2007.
- [3] G. Box, G. Jenkins, and G. Reinsel, *Time Series Analysis: Forecasting and Control*. Upper Saddle River, NJ: Prentice-Hall, 1994.
- [4] A. Gani, A. V. Gribok, S. Rajaraman, W. K. Ward, and J. Reifman, "Predicting subcutaneous glucose concentration in humans: Data-driven glucose modeling," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 2, pp. 246–254, Feb. 2009.
- [5] A. Gani, A. Gribok, Y. Lu, W. K. Ward, R. A. Vigerksy, and J. Reifman, "Universal glucose models for predicting subcutaneous glucose concentration in humans," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 1, pp. 157–165, Jan. 2010.
- [6] J. Reifman, S. Rajaraman, A. V. Gribok, and W. K. Ward, "Predictive monitoring for improved management of glucose levels," *J. Diabetes Sci. Technol.*, vol. 1, no. 4, pp. 478–486, Jul. 2007.
- [7] 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, no. 5, pp. 931–937, May 2007.
- [8] T. Bremer and D. A. Gough, "Is blood glucose predictable from previous values? A solicitation for data," *Diabetes*, vol. 48, no. 3, pp. 445–451, Mar. 1999.
- [9] N. Porksen, "The in vivo regulation of pulsatile insulin secretion," *Diabetologia*, vol. 45, no. 1, pp. 3–20, Jan. 2002.
- [10] C. Simon, M. Follenius, and G. Brandenberger, "Postprandial oscillations of plasma glucose, insulin and C-peptide in man," *Diabetologia*, vol. 30, no. 10, pp. 769–773, Oct. 1987.
- [11] E. Van Cauter, E. T. Shapiro, H. Tillil, and K. S. Polonsky, "Circadian modulation of glucose and insulin responses to meals: relationship to cortisol rhythm," *Amer. J. Physiol. Endocrinol. Metab.*, vol. 262, no. 4, pp. E467–475, Apr. 1992.
- [12] F. N. Rahaghi and D. A. Gough, "Blood glucose dynamics," *Diabetes Technol. Ther.*, vol. 10, no. 2, pp. 81–94, Apr. 2008.
- [13] C. Simon, "Ultradian pulsatility of plasma glucose and insulin secretion rate: Circadian and sleep modulation," *Hormone Res.*, vol. 49, no. 3–4, pp. 185–190, Mar. 1998.
- [14] D. A. Gough, K. Kreutz-Delgado, and T. M. Bremer, "Frequency characterization of blood glucose dynamics," *Ann. Biomed. Eng.*, vol. 31, no. 1, pp. 91–97, Jan. 2003.
- [15] M. Breton, D. P. Shields, and B. P. Kovatchev, "Optimum subcutaneous glucose sampling and Fourier analysis of continuous glucose monitors," *J. Diabetes Sci. Technol.*, vol. 2, no. 3, pp. 495–500, May 2008.
- [16] M. Miller and P. Strange, "Use of Fourier models for analysis and interpretation of continuous glucose monitoring glucose profiles," *J. Diabetes Sci. Technol.*, vol. 1, no. 5, pp. 630–638, Sep. 2007.
- [17] S. Rao and W. A. Pearlman, "Analysis of linear prediction, coding, and spectral estimation from subbands," *IEEE Trans. Inf. Theory*, vol. 42, no. 4, pp. 1160–1178, Jul. 1996.
- [18] G. B. Bolli and J. E. Gerich, "The dawn phenomenon—A common occurrence in both non-insulin-dependent and insulin-dependent diabetes mellitus," *New Engl. J. Med.*, vol. 310, no. 12, pp. 746–750, Mar. 1984.
- [19] G. Perriello, P. De Feo, E. Torlone, C. Fanelli, F. Santeusano, P. Brunetti, and G. B. Bolli, "The dawn phenomenon in type 1 (insulin-dependent) diabetes mellitus: Magnitude, frequency, variability, and dependency on glucose counterregulation and insulin sensitivity," *Diabetologia*, vol. 34, no. 1, pp. 21–28, Jan. 1991.
- [20] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*. New York: Wiley, 1996.
- [21] J. Makhoul, "Linear prediction—Tutorial review," *Proc. IEEE*, vol. 63, no. 4, pp. 561–580, Apr. 1975.
- [22] D. A. Finan, C. C. Palerm, F. J. Doyle III, H. Zisser, L. Jovanovic, W. C. Bevier, and D. E. Seborg, "Identification of empirical dynamic models from type 1 diabetes subject data," in *Proc. Amer. Control Conf.*, Seattle, WA, 2008, pp. 2099–2104.



**Yinghui Lu** received the B.S. degree from the Zhejiang University, Hangzhou, China, in 1996, the M.S. degree from The University of Tennessee, Knoxville, in 2002, and the Ph.D. degree from the Georgia Institute of Technology, Atlanta, in 2007, all in electrical engineering.

He is currently a Postdoctoral Fellow in the Bioinformatics Cell, Telemedicine and Advanced Technology Research Center, U.S. Army Medical Research and Materiel Command, Fort Detrick, MD. His research interests include digital signal processing, intelligent control systems, ultrasonics, and medical informatics.



**Andrei V. Gribok** received the B.S. degree in systems science and the M.S. degree in nuclear engineering from Moscow Institute of Physics and Engineering, Moscow, Russia, both in 1987, and the Ph.D. degree in biological physics from Moscow Institute of Biological Physics, Moscow, in 1996.

He is currently a Research Assistant Professor in the Department of Nuclear Engineering, The University of Tennessee, Knoxville and is also on an interpersonnel agreement assignment with the Telemedicine and Advanced Technology Research

Center, U.S. Army Medical Research and Materiel Command, Fort Detrick, MD. His research interests include inverse and ill-posed problems in engineering, statistical learning, and model misspecification in statistics.



**W. Kenneth Ward** received the Medical degree from the University of Colorado, Boulder, in 1978. Thereafter, he was awarded a Research Fellowship in diabetes and endocrinology at the University of Washington, Seattle.

He is currently a Diabetes Specialist and Biosensor Scientist for iSense Corporation, Wilsonville, OR, where he is engaged in collecting data used in the study while collaborating with the staff at the Profil Institute, Neuss, Germany. His current research interests include development of minimally invasive and

fully implantable biosensors, artificial pancreas algorithms, biomaterials, and the foreign body response to implanted devices.



**Jaques Reifman** received the B.S. degree in civil engineering from Rio de Janeiro State University, Rio de Janeiro, Brazil, in 1980, the B.B.A. degree in business administration from Rio de Janeiro Federal University, Rio de Janeiro, in 1985, and the M.S.E. and Ph.D. degrees in nuclear engineering from the University of Michigan, Ann Arbor, in 1985 and 1989, respectively.

He is currently a Senior Research Scientist in the Bioinformatics Cell, Telemedicine and Advanced Technology Research Center, U.S. Army Medical Research and Materiel Command (USAMRMC), Fort Detrick, MD, where he is also the Director of two organizations which he founded: the Department of Defense Biotechnology High Performance Computing Software Applications Institute for Force Health Protection and the USAMRMC Bioinformatics Cell. His research interests include physiological signal processing, statistical pattern recognition, artificial intelligence, data mining, biomathematical modeling, systems biology, bioinformatics, genomics, and proteomics.