Montaser Roushdi Ali, E.; Diez, J.; Bondía Company, J. (2017). Stochastic Seasonal Models for Glucose Prediction in the Artificial Pancreas. Journal of Diabetes Science and Technology. 11(6):1124-1131. doi:10.1177/1932296817736074

Stochastic Seasonal Models for Glucose Prediction in the Artificial Pancreas

Eslam Montaser, MSc[1], Jose-Luis Diez, Ph.D.[1], Jorge Bondia, Ph.D.[1]

Author Affiliations: [1]Instituto Universitario de Automática e Informática Industrial, Universitat Politècnica de València, València, Spain

Eslam Montaser	Jose-Luis Diez	Jorge Bondia
Universitat Politècnica de València, C/ Camino de Vera, s/n, 46022 València, Spain	Universitat Politècnica de València, C/ Camino de Vera, s/n, 46022 València, Spain	Universitat Politècnica de València, C/ Camino de Vera, s/n, 46022 València, Spain
+34 963877007	+34 963877007	+34 963877007
emontase@gmail.com	jldiez@isa.upv.es	jbondia@isa.upv.es

Abbreviations: (AP) artificial pancreas, (T1D) type 1 diabetes, (PH) prediction horizon, (CGM) continuous glucose monitoring, (AR) autoregressive, (ARMA) autoregressive-moving-average, (ARX) autoregressive with exogenous inputs, (SARIMA) seasonal-autoregressive-integrated-moving-average, (ARIMA) autoregressive-integrated-moving-average, (SAR) seasonal autoregressive, (SMA) seasonal moving-avarage, (SARIMAX) seasonal-autoregressive-integrated-moving-average with exogenous variables, (ARIMAX) autoregressive-integrated-moving-average with exogenous variables, (ACF) autocorrelation function, (PACF) partial autocorrelation function, (AIC) Akaike information criterion, (MAE) mean absolute error, (RMSE) root mean squared error, (MAPE) mean absolute percentage error, (UI) Theil inequality coefficient, (RMSE) root-mean-square error.

Keywords: artificial pancreas, glucose prediction, seasonal models, stochastic

models, type 1 diabetes

Corresponding Author: Jorge Bondia, Ph.D., Departamento de Ingeniería de Sistemas y Automática, Universitat Politècnica de València, C/ Camí de Vera, s/n, 46022 Valencia, Spain; e-mail address: jbondia@isa.upv.es

Funding Source: This work was funded by the Spanish Ministry of Economy and Competitiveness, grants DPI2013-46982-C2-1-R and DPI2016-78831-C2-1-R, and the European Union through FEDER funds.

Conflict-of-Interest Disclosure: none

Acknowledgements: The authors acknowledge the collaboration of Paolo Rossetti from Hospital Francesc de Borja de Gandia, F.J. Ampudia-Blasco from Hospital Clínico Universitario de Valencia, I. Conget, M. Giménez and C. Quirós from Hospital Clinic de Barcelona and J. Vehí from Universitat de Girona who participated in the implementation and/or design of the study from which data were obtained, as well as the altruist participation of all the patients involved in the study.

Figures and table count: 5 figures and 4 tables.

Abstract

Background:

Linear empirical dynamic models have been widely used for glucose prediction. The extension of the concept of seasonality, characteristic of other domains, is explored here for the improvement of prediction accuracy.

Methods:

Twenty time series of 8-hour postprandial periods (PP) for a same 60g-carbohydrate meal were collected from a closed-loop controller validation study. A single concatenated time series was produced representing a collection of data from similar scenarios, resulting in seasonality. Variability in the resulting time series was representative of worst-case intra-subject variability. Following a leave-one-out cross-validation, seasonal and non-seasonal autoregressive-integrated-moving-average models (SARIMA and ARIMA) were built to analyze the effect of seasonality in the model prediction accuracy. Further improvement achieved from the inclusion of insulin infusion rate as exogenous variable was also analyzed. Prediction horizons (PHs) from 30 to 300 min were considered.

Results:

SARIMA outperformed ARIMA revealing a significant role of seasonality. For a 5-h PH, average MAPE was reduced in 26.62%. Considering individual runs, the improvement ranged from 6.3% to 54.52%. In the best-performing case this reduction amounted to 29.45%. The benefit of seasonality was consistent among different PHs, although lower PHs benefited more, with MAPE reduction over 50% for PHs of 60 and 120 minutes, and over 40% for 180 min. Consideration of insulin infusion rate into the seasonal model further improved performance, with a 61.89% reduction in MAPE for 30-min PH and reductions over 20% for PHs over 180 min.

Conclusions:

Seasonality improved model accuracy allowing for the extension of the PH significantly.

Introduction

An important feature of any artificial pancreas[1,2] is its ability to predict glucose along a given prediction horizon (PH), either as part of the control algorithm itself, such as in systems based on Model Predictive Control (MPC) techniques[3-5], or as part of a monitoring subsystem to predict, for instance, hypoglycemic episodes[6-8]. Model requirements and input information will depend on the specific purpose. For instance, future values of insulin infusion are available during the MPC optimization process where predictions are needed. However, this is not the case in the context of risk prediction in patient monitoring during closed-loop operation.

Linear empirical dynamic models rely on time series as an observation on a dynamic system[9]. These include autoregressive (AR), autoregressive-moving-average (ARMA), and autoregressive models with exogenous inputs (ARX), among others. These models have been widely used in the context of glucose prediction. Gani *et al.*[10] identified 30th-order AR models from continuous glucose monitoring (CGM) data with 1-min sampling time from nine T1D subjects. An average root-mean-square error (RMSE) of 12.6 mg/dL was reported for a 60-min PH, after data smoothing and parameter regularization. Sparacino and colleagues[11] identified a 1st-order time-varying AR model based on data from 28 T1D subjects wearing a microdialysis system with a 3-min sampling time. They demonstrated the feasibility of predicting hypoglycemic events 20-25 min ahead in time, considering a 30-min PH. A median RMSE ranging from 18.33 to 20.32 mg/dL, depending on the selection of a forgetting factor, was reported for that PH. Low-order AR and ARMA models were considered by Eren-Oruklu and associates[12] considering PHs up to 30 minutes in healthy and type 2 diabetes subjects. A sum of squares of glucose prediction error ranging between 10.32 and 12.55 mg/dL was reported, depending on the study, for a 30-min PH. Finan and colleagues[13-15] evaluated ARX models from simulated and clinical ambulatory data with 5-min sampling time. The authors concluded that 60 minutes was a maximum achievable PH in terms of model prediction accuracy. An average RMSE of 26, 34 and 40 mg/dL was reported for 30-, 45- and 60-

min PH, respectively[15]. This corresponds to an improvement of 9% compared to a zero-order-hold predictor. A variety of linear and nonlinear time-series models were evaluated by Ståhl and Johansson[16] from clinical data from one subject, with non-uniform and sparse sampling (fingerstick measurements) with spline interpolation, in order to produce a short-term blood glucose predictors for up to two-hour-ahead blood glucose prediction. However, many difficulties were met not achieving the required accuracy.

Empirical dynamic models are also widely used in other domains such as business and economic time series. A particular characteristic in these domains is *seasonality*, that is, the existence of regular patterns of changes and fluctuations that repeat periodically[17]. This paper explores the extension of the concept of seasonality for glucose prediction with a proof-of-concept study. The main rationale is that pre-processing of CGM time series (and available additional information) may translate daily events into seasonal phenomena. For instance, glucose concentration tends to peak and then decline in a characteristic way after a meal intake in a particular scenario. In this case, a new pre-processed family of time series can be built from the original CGM data by concatenating postprandial periods (PPs) of fixed length where similarity of behaviors is expected, according to some metrics, which would theoretically produce seasonal time series. This allows for the application of seasonal models that exploit this similarity for more accurate predictions and longer PHs. Seasonal-autoregressive-integrated-moving-average (SARIMA) models are considered in this work and compared to its non-seasonal counterpart in order to investigate the benefit of seasonality into glucose prediction. The use of insulin infusion rate as an exogenous variable is also explored.

Methods

Data overview

CGM time series covering 8-hour PPs for a same meal were collected from the Clinic University Hospital of Valencia, Spain. Data belonged to a closed-loop controller

validation study where 10 T1D subjects underwent an in-hospital 8-hour standardized mixed meal test (60g carbohydrate) on two occasions with a hybrid artificial pancreas with 15-min sampling period. Patients wore two pumps with CGM devices (Paradigm Veo® insulin pump with Enlite-2 sensors®, Medtronic MiniMed, Northridge, CA), which were calibrated 15 minutes before the meal test was administered (lunch at noon). CGM glucose data was available for eight hours after the meal, from 12:00 p.m. until 20:00 p.m. Glucose concentration was also measured every 15 minutes with a reference method (YSI 2300 Stat Plus Glucose Analyzer, YSI Incorporated Life Sciences, Yellow Springs, Ohio, USA).

Despite meal size was controlled in this in-patient study, this didn't prevent the presence of high intra- and inter-individual variability. These were measured by the Coefficient of Variance of the Area-Under-the-Curve for the 8-hour duration of the study (CV-AUC_{8h}), which was computed with the trapezoidal rule. Euclidean distance among paired PPs was also computed to determine time series shape similarity. A sampling period of 15 minutes was considered to match glucose reference measurements and our controller configuration.

SARIMA model

A SARIMA model is an expanded form of its non-seasonal counterpart ARIMA model that includes as new model components seasonal autoregressive (SAR) and seasonal moving-average (SMA) terms. In an empirical dynamic model, an observation at time t is expressed as a linear combination of observations at times $t-1, t-2, \dots, t-p$ (previous p measurements) by the AR component, and as a linear combination of stochastic errors, also called shocks, at times $t, t-1, t-2, \dots, t-q$ by the MA component. In a SARIMA model, SAR and SMA terms are added so that an observation at time t depends on previous observations and stochastic errors at times with lags that are multiples of the seasonality period s. In the context of postprandial glucose prediction, this means that the glucose prediction will depend not only on previous measurements for that PP, but also on previous PPs in the time series.

Given a CGM time series $\{G(t) \mid t = 1, 2, \dots, k\}$, a SARIMA model is expressed as:

$$\nabla_s^D \nabla^d G(t) = c + w(t), \quad (1)$$

$$\emptyset_p(z^{-1})\Phi_P(z^{-s})w(t) = \theta_q(z^{-1})\Theta_Q(z^{-s})\varepsilon(t), \quad (2)$$

where $G(t)$ is the glucose concentration at time t, c is a constant term (intercept), ∇ is the backward difference operator, *i.e.* $\nabla G(t) := G(t) - G(t-1)$, d is the non-seasonal integration order, ∇_s is the seasonal backward difference operator, *i.e.* $\nabla_s G(t) := G(t) - G(t-s)$, D is the seasonal integration order, the input $\varepsilon(t)$ is the stochastic error following a white noise process $\varepsilon(t) \sim WN(0, \sigma^2)$ and $\emptyset_p(z^{-1})$, $\Phi_P(z^{-s})$, $\theta_q(z^{-1})$ and $\Theta_Q(z^{-s})$ are polynomials in the lag (back-shift) operator z^{-1} of degree p, P, q and Q, respectively, defined as

(AR) $\emptyset_p(z^{-1}) := 1 - \emptyset_1 z^{-1} - \emptyset_2 z^{-2} - \cdots - \emptyset_p z^{-p}$,

(SAR) $\Phi_P(z^{-s}) := 1 - \Phi_s z^{-s} - \Phi_{2s} z^{-2s} - \cdots - \Phi_{Ps} z^{-Ps}$,

(MA) $\theta_q(z^{-1}) := 1 + \theta_1 z^{-1} + \theta_2 z^{-2} + \cdots + \theta_q z^{-q}$,

(SMA) $\Theta_Q(z^{-s}) := 1 + \Theta_s z^{-s} + \Theta_{2s} z^{-2s} + \cdots + \Theta_{Qs} z^{-Qs}$.

Model (1)-(2) can be expressed in short form as SARIMA$(p, d, q)(P, D, Q)_s$.

Exogenous variables

Exogenous variables in model (1)-(2) were also considered. There exist different approaches for incorporating exogenous variables into a model. Denoting as $X(t)$ the exogenous variable, a term $\eta_r(z^{-1})X(t)$, where $\eta_r(z^{-1}) := \eta_0 + \eta_1 z^{-1} + \cdots + \eta_r z^{-r}$, is commonly added to the right-hand-side of equation (2), yielding the so-called ARX, ARMAX, ARIMAX or SARIMAX models depending on the considered structure. In many statistical packages such as R and Eviews, exogenous variables are considered as explanatory variables into a linear regression model with a stochastic error process of certain structure. In this case, a SARIMAX model is expressed as

$$\nabla_s^D \nabla^d G(t) = c + \eta_r(z^{-1})X(t) + w(t) \quad (3)$$

$$\emptyset_P(z^{-1})\Phi_P(z^{-s})w(t) = \theta_q(z^{-1})\Theta_Q(z^{-s})\varepsilon(t) \quad (4)$$

when current and past values of the exogenous variable $X(t)$ are used. In this case, the polynomial $\eta_r(z^{-1})$ represents a finite-impulse-response filter. The rest of components

are defined as in (1)-(2). Model (3)-(4) can be expressed in short form as SARIMAX$(p, d, q, r)(P, D, Q)_s$. Granger causality test[18] can be used to determine the usefulness of including an exogenous input for improving forecasting. In this study, Eviews software, version 9.5, was used and exogenous variables were treated as in (3)-(4).

Identification Procedure

Box-Jenkins methodology[9,19] was used for model building and evaluation (see Figure 1). A leave-one-out cross-validation procedure was considered dividing data into training and validation sets. In order to avoid data from a same patient to appear both in training and validation, data from the validation patient was excluded from the training set. This resulted in 18 PPs for training and 1 for validation, since two CL studies per patient were available. PPs in the training set were randomly ordered at each run according to a random sequence generator (www.random.org). A stationarity analysis was first carried out with the unit-root test (Augmented Dickey-Fuller test)[20]. The backward-difference operator ∇ was applied to the time series as many times as necessary (integration order d) to remove non-stationarity, if present. Sample autocorrelation function (ACF) and partial autocorrelation function (PACF) were used to identify the orders of the autoregressive and moving average terms (p and q, respectively), as well as identifying seasonality (seasonally differencing the time series with the operator ∇_s if necessary). Maximum likelihood was used for parameter estimation. Akaike information criterion (AIC) was used for model selection, which is defined in Eviews software as $AIC := \frac{1}{n}(-2\,L + 2K)$, where L is the value of the log-likelihood, K is the number of free parameters in the model and n is the number of observations. Remark the scaling by $1/n$. For diagnostic checking, ACF and PACF plots for the residuals were analyzed to test the existence of any significant spikes in the confidence interval, Ljung-Box Q test[21] was used for testing randomness at each distinct lag and Jarque-Bera test[22] was used to test the normality of the residuals. Finally, accuracy of the model forecasting was measured with the following metrics:

Mean absolute error: $MAE := \frac{1}{n}\sum_{i=1}^{n}|e_i|$,

Root mean squared error: $RMSE := \sqrt{\frac{1}{n}\sum_{i=1}^{n} {e_i}^2}$,

Mean absolute percentage error: $MAPE := 100\frac{1}{n}\sum_{i=1}^{n}\left|\frac{e_i}{G_i}\right|$

where n is again the number of observations, G_i is the i-th observation, $\hat{G}_i$ is a forecast for G_i and $e_i := G_i - \hat{G}_i$ is the forecasting error.

Results

Figure 2 shows the dataset resulting from the concatenation of the twenty 8-h PPs. The CGM time series had a mean of 136.1 mg/dL, with a standard deviation of 48.48 mg/dL. Despite the same meal was provided, data exhibited high variability with postprandial peaks ranging from 304 mg/dL (P91) to 125 mg/dL (P42) and the incidence of hypoglycemia in some patients (P11, P51, P52, P71, P101), two of them severe (P11, P101), according to CGM values. They were non-normally distributed. Inter-individual variability measured by CV-AUC_{8h} was 21.52%, whereas intra-individual variability was 9.17%. However, the latter spanned from 3.22% (patient 6) to 18.67% (patient 9). Since only two studies per patient were available, intra-patient variability might be underestimated. It is thus considered that worst-case intra-patient variability is represented by the generated time series. Euclidean distance between each pair of PPs was also computed to analyze similarity of time series (see Figure 3), providing similar conclusions. Patient 9 is the most dissimilar among studies (green box in P91-P92, only exceeded by comparatively few yellow-red boxes outside the diagonal (between-patient comparisons). P81, P82 and P91 were the most dissimilar with the rest of periods (higher incidence of yellow-red boxes). Total basal insulin infusion in the 8-h period ranged from 5.21U (P31) to 16.40U (P71). An extended bolus computed from the patient's insulin-to-carbohydrate ratio and open-loop basal infusion rate was additionally administered at meal time.

Both SARIMA and ARIMA models were identified for each run in the cross-validation. Figure 4(a) shows the forecasting accuracy metrics for a 5-h PH for both cases. A high PH

was initially chosen to challenge the model. SARIMA outperform ARIMA in all metrics (mean(SD): MAE(mg/dL) 34.56(19.35) vs. 47.72(24.43); RMSE(mg/dL) 40.02(21.62) vs. 55.02(26.93); MAPE(%) 22.02(9.41) vs. 30.01(13.05); $p<0.05$ in all cases). In the following, the analysis will be restricted to MAPE since the three measures provided the same information. Figure 4(b) shows the obtained MAPE as the PH increases from 30 min to 5 hours, consistently outperforming SARIMA. The identified model structure differed slightly between runs, with AR and MA orders up to 4. No time series differentiation was needed for both ARIMA and SARIMA models. Seasonality with lag 33 (the size of the PP) was obtained in all cases, as expected. SAR and SMA orders were up to 2.

The best performing run was Run 4, with validation data P22. In this case, inspection of the ACF revealed data were stationary (the trend had a non-significant p-value 0.0877) and seasonal at lag 33 with significant p-value 0.0000. Seasonally differenced data was stationary with significant p-value 0.0001, so it was not necessary to take any difference. Model SARIMA$(4,0,4)(1,0,1)_{33}$ was the most appropriate model, with AIC 7.9566. Table 1 shows the estimated model parameters using maximum likelihood estimation. All spikes in the residuals ACF were within the significance limits (white noise). Table 2 shows the Ljung-Box Q test for testing randomness at each distinct lag, also demonstrating that the residuals have no remaining autocorrelations. The tests for residual normality showed that the residuals were approximately normal. A MAPE of 6.73% was obtained for training data. A similar fitting was obtained for ARIMA, with MAPE 7.05%. Figure 5 shows the prediction performance using validation data for a 5-h prediction horizon. A MAPE of 6.62% and a RMSE of 10.28 mg/dL were obtained for SARIMA. For ARIMA, prediction metrics were worse with MAPE 9.39% and RMSE 14.39 mg/dL, as it becomes apparent in Fig. 5.

The effect of considering insulin infusion as exogenous variable for performance improvement was investigated. Besides, insulin infusion information is needed in applications such as MPC. This analysis was carried out only for Run 4 as the best

performing case, challenging further improvement. Insulin infusion signal contained bolus and basal infusion and was expressed in U per sampling period. Granger causality test was applied to test the null hypothesis that CGM does not "Granger cause" insulin infusion and vice versa. The null hypothesis was rejected with a significant p-value of 0.0146. Therefore, inclusion of insulin infusion into the model might improve performance. The order of the exogenous polynomial was computed from the cross-correlation plot and AIC, resulting in the model SARIMAX$(4, 0, 4, 2)(1, 0, 1)_{33}$ with AIC 7.9544. Table 3 shows the estimated parameters for this model. The same procedure was used to derive its non-seasonal counterpart resulting in the model ARIMAX$(4, 0, 4, 2)$ with AIC 7.9952. In the forecasting period, a MAPE of 5.12% and a RMSE of 8.47 mg/dL were obtained for the SARIMAX model for a 5-h PH, compared to 6.62% and 10.28 mg/dL for SARIMA, and 10.51% and 16.17 mg/dL for ARIMAX. Differences among the behavior of the different models can be observed in Figure 5.

Finally, forecasting performance as measured by MAPE and RMSE at different PHs is presented in Table 4. Prediction horizons of 30, 60, 120, 180, 240 and 300 minutes were considered.

Discussion

Training data consisted in a collection of PPs from different patients covering both early and late postprandial phases (8 hours). Time between meals during the day is generally shorter. Nocturnal period was not represented by our data. However, PP has shown to be much more challenging than nocturnal period for an artificial pancreas[23]. Both CV-AUC_{8h} and Euclidean distance (Fig. 3) showed large inter-individual variability and a large range in intra-individual variability, with its worst-case represented by inter-individual variability. Thus, the concatenated time series defines a challenging scenario with a worst-case highly variable patient. Data variability might be attenuated with the use of classification techniques, collecting similar enough postprandial responses into different datasets, with their corresponding prediction model.

A first-order seasonal AR and MA component was identified with seasonality lag 33 in all SARIMA runs due to the concatenated nature of the time series. In all runs, SARIMA outperformed ARIMA revealing a significant role of seasonality. 5-h PH average MAPE was reduced in 26.62%. Considering individual runs, the improvement ranged from 6.3% (Run 7; validation data P41) to 54.52% (Run 3; validation data P21). In the best performing case, according to MAPE (Run 4), this reduction amounted to 29.45%. Prediction improvement by introducing seasonality also becomes apparent from Figure 5. The benefit of seasonality was consistent among different prediction horizons, as illustrated in Figure 4(b) and Table 4 for Run 4. Lower prediction horizons benefited more, with a MAPE reduction over 50% for PHs of 60 and 120 minutes, and over 40% for 180 min. In these case, MAPE was close to 6% and RMSE below 10 mg/dL, doubling these values when seasonality was not considered. In greater PHs benefit of seasonality is still observed, although decreasing due to variability in the time series.

Consideration of insulin infusion rate into the seasonal model further improved performance for Run 4. Although analysis was limited to this case to reduce computational burden, remark it corresponds to the most challenging situation for model improvement since SARIMA model for Run 4 has the best prediction accuracy in the cross-validation study. SARIMAX improved performance as compared to SARIMA with a 61.89% reduction in MAPE (2.90% vs. 7.61%) for 30-min PH to a 7.33% reduction at 2-h PH (5.86% vs. 5.46%) and reductions over 20% for PHs over 180 min, as shown in Table 4. A RMSE below 10 mg/dL was obtained for all PHs. This means that SARIMAX models might allow the increment of prediction horizons in MPC-based artificial pancreas systems. Table 4 also shows that SARIMAX outperformed in all cases its non-seasonal counterpart ARIMAX.

This is a proof-of-concept study and as such it has limitations. It is assumed that mealtime is known, allowing for the construction of concatenated time series with fixed-length PPs. However, to date, meal announcement is a common component of artificial

pancreas systems and, otherwise, meal detection algorithms are incorporated[24-26]. Remark that although focus was put on PPs, this approach can be applied to other fixed-length time series data subsets representing characteristic scenarios where similarity is expected or learned from classification techniques. Another limitation is the data used, which did not correspond to a single patient, although inter-patient variability in the data was representative of worst-case intra-patient variability defining a challenging scenario. A collection of 18 PPs were used for model training at each cross-validation run. Seasonal components of the identified models were first or second order, which means that current meal depends, at most, on the two previous similar meals. Thus, the length of the data used is considered appropriate for this proof-of-concept study. However, further investigation is needed with longer single-patient CGM data and the combination of seasonal modelling with classifiers.

Conclusion

Despite the limitations of this study, seasonality has shown to be an important factor to improve model predictive power allowing for the significant extension of prediction horizons. Further work is now needed for the classification of periods under scenarios yielding "similar enough" glycemic responses to fully exploit the expected benefit of seasonal models.

Funding Source

This work was funded by the Spanish Ministry of Economy and Competitiveness, grants DPI2013-46982-C2-1-R and DPI2016-78831-C2-1-R, and the European Union through FEDER funds.

Acknowledgements

The authors acknowledge the collaboration of Paolo Rossetti from Hospital Francesc de Borja de Gandia, F.J. Ampudia-Blasco from Hospital Clínico Universitario de Valencia, I. Conget, M. Giménez and C. Quirós from Hospital Clinic de Barcelona and J. Vehí from

Universitat de Girona who participated in the implementation and/or design of the study from which data were obtained, as well as the altruist participation of all the patients involved in the study.

Conflict-of-Interest Disclosure

None

References

1. Haidar A. The Artificial Pancreas: How Closed-Loop Control Is Revolutionizing Diabetes. IEEE Control Systems Magazine 2016;36(5):28–47.

2. Kovatchev B, Tamborlane WV, Cefalu WT, Cobelli C. The Artificial Pancreas in 2016: A Digital Treatment Ecosystem for Diabetes. Diabetes Care 2016;39(7):1123–6.

3. Hovorka R, Canonico V, Chassin LJ, et al. Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes. Physiological Measurement 2004;25:905–20.

4. Magni L, Raimondo D, Bossi L, et al. Model Predictive Control of Type 1 Diabetes: An in Silico Trial. Journal of Diabetes Science and Technology 2007;1(6):804–12.

5. Grosman B, Dassau E, Zisser HC, Jovanovic L, Doyle IIII FJ. Zone Model Predictive Control: A Strategy to Minimize Hyper- and Hypoglycemic Events. Journal of Diabetes Science and Technology 2010;4(4):961–75.

6. Palerm C, Willis J, Desemone J, Bequette BW. Hypoglycemia prediction and detection using optimal estimation. Diabetes Technology & Therapeutics 2001;7(1):3–14.

7. Cameron F, Niemeyer G, Gundy-Burlet K, Buckingham B. Statistical Hypoglycemia Prediction. Journal of Diabetes Science and Technology 2008;2(4):612–21.

8. Harvey RA, Dassau E, Zisser H, Seborg DE, Jovanovic L, Doyle FJ. Design of the Health Monitoring System for the Artificial Pancreas: Low Glucose Prediction Module. Journal of Diabetes Science and Technology 2012;6(6):1345–54.

9. Box GEP, Jenkins GM, Reinsel GC, Ljung GM. Time series analysis. Fifth. John Wiley & Sons, Inc., Hoboken, NJ; 2015.

10. Gani A, Gribok A, Rajaraman S, Ward W, Reifman J. Predicting Subcutaneous Glucose Concentration in Humans: Data-Driven Glucose Modeling. IEEE Transactions on Biomedical Engineering 2009;56(2):246–54.

11. Sparacino G, Zanderigo F, Corazza S, Maran A, Facchinetti A, Cobelli C. Glucose Concentration can be Predicted Ahead in Time From Continuous Glucose Monitoring Sensor Time-Series. IEEE Transactions on Biomedical Engineering 2007;54(5):931–7.

12. Eren-Oruklu M, Cinar A, Quinn L, Smith D. Estimation of future glucose concentrations with subject-specific recursive linear models. Diabetes Technology & Therapeutics 2009;11(4):243–53.

13. Finan DA, Zisser H, Jovanovic L, Bevier WC, Seborg DE. Practical Issues in the Identification of Empirical Models from Simulated Type 1 Diabetes Data. Diabetes Technology & Therapeutics 2007;9(5):438–50.

14. Finan DA, Palerm CC, III FJD, et al. Identification of Empirical Dynamic Models From Type 1 Diabetes Subject Data. 2008. p. 8.

15. Finan DA, Doyle FJ, Palerm CC, et al. Experimental evaluation of a recursive model identification technique for type 1 diabetes. Journal of Diabetes Science and Technology 2009;3(5):1192–202.

16. Ståhl F, Johansson R. Diabetes mellitus modeling and short-term prediction based on blood glucose measurements. Mathematical Biosciences 2009;217:101–17.

17. Hylleberg S. Modelling seasonality. New York : Oxford University Press; 1992.

18. Engle RF, Granger C. Co-integration and error correction: representation, estimation, and testing. Econometrica: journal of the Econometric Society 1987;55(2):251–76.

19. Jenkins GM, Reinsel GC. Time series analysis: forecasting and control. Holden-Day, San Francisco; 1976.

20. Ghysels E, Lee HS, Noh J. Testing for unit roots in seasonal time series: Some theoretical extensions and a Monte Carlo investigation. Journal of Econometrics 1994;62(2):415–42.

21. Ljung GM, Box GEP. On a measure of lack of fit in time series models. Biometrika 1978;65(2):297–303.

22. Jarque CM, Bera AK. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. Econ Lett 1980;6(3):255–9.

23. Gingras V, Taleb N, Roy-Fleming A, Legault L, Rabasa-Lhoret R. The challenges of Achieving Postprandial Glucose Control using Closed-Loop Systems in Patients with Type 1 Diabetes. Diabetes Obesity and Metabolism 2017; Accepted Author Manuscript. doi:10.1111/dom.13052

24. Dassau E, Bequette BW, Buckingham BA, Doyle FJ. Detection of a Meal Using Continuous Glucose Monitoring: Implications for an artificial beta-cell. Diabetes Care 2007;31(2):295–300.

25. Weimer J, Chen S, Peleckis A, Rickels MR, Lee I. Physiology-Invariant Meal Detection for Type 1 Diabetes. Diabetes Technology & Therapeutics 2016;18(10):616–24.

26. Samadi S, Turksoy K, Hajizadeh I, Feng J, Sevil M, Cinar A. Meal Detection and Carbohydrate Estimation Using Continuous Glucose Sensor Data. IEEE J Biomed Health Inform 2017;21(3):619–27.

Table 1.

Model parameters for model SARIMA (4, 0, 4)(1, 0, 1)$_{33}$ in best-performing Run 4, following notation in (1)-(2). σ^2 is the estimate of the error variance from the maximum likelihood estimation.

Parameter	Value	Std. Error	t-Statistic	p-value
c	134.1109	14.70842	9.117968	0.0000
$\emptyset_1$	3.124690	0.008934	349.7433	0.0000
$\emptyset_2$	-3.783549	0.015602	-242.5068	0.0000
$\emptyset_3$	2.044158	0.015166	134.7894	0.0000
$\emptyset_4$	-0.399508	0.008156	-48.98071	0.0000
Φ_{33}	0.912586	0.056899	16.03866	0.0000
θ_1	-1.827147	0.307422	-5.943459	0.0000
θ_2	1.124422	0.347364	3.237018	0.0013
θ_3	-0.177478	0.087565	-2.026818	0.0431
θ_4	0.086584	0.056154	1.541905	0.1236
Θ_{33}	-0.826937	0.078243	-10.56878	0.0000
σ^2	157.4450	53.98418	2.916503	0.0037

Table 2.

Ljung-Box test for the training residuals of Run 4 model.

Lag	12	24	36	48
Q-Stat	2.8661	10.575	21.358	30.513
P-Value	0.239	0.719	0.723	0.801

Table 3.

Model parameters for model SARIMAX (4, 0, 4, 2)(1, 0, 1)$_{33}$ in best-performing Run 4, following notation in (3)-(4). σ^2 is the estimate of the error variance from the maximum likelihood estimation.

Variable	Coefficient	Std.Error	t- Statistic	Prob.
c	131.3957	17.53808	7.492023	**0.0000**
η_0	1.059158	0.307175	3.448056	**0.0006**
η_1	0.933659	0.376779	2.478004	**0.0135**
η_2	0.223623	0.323739	0.678182	**0.4979**
$\emptyset_1$	3.245163	0.006145	528.1297	**0.0000**
$\emptyset_2$	-4.107178	0.007966	-515.5566	**0.0000**
$\emptyset_3$	2.356905	0.011389	206.9404	**0.0000**
$\emptyset_4$	-0.50644	0.008237	-61.48038	**0.0000**
Φ_{33}	0.938280	0.042357	22.15150	**0.0000**
θ_1	-2.004745	0.288557	-6.947493	**0.0000**
θ_2	1.400848	0.376337	3.722324	**0.0002**
θ_3	-0.275892	0.112957	-2.442438	**0.0149**
θ_4	0.043012	0.049410	0.870522	**0.3844**
Θ_{33}	-0.838930	0.066691	-12.57943	**0.0000**
σ^2	154.4792	56.49393	2.734440	**0.0064**

Table 4.
Prediction accuracy measured by MAPE (%) and RMSE (mg/dL), in parenthesis, of seasonal versus non-seasonal counterparts for different prediction horizons.

Model \ PH (min)	30	60	120	180	240	300
No exogenous inputs						
SARIMA(4,0,4)(1,0,1)$_{33}$	7.61 (9.8955)	5.97 (8.5567)	5.46 (8.1011)	6.07 (8.9592)	6.95 (10.7894)	6.62 (10.2870)
ARIMA(4,0,4)	9.00 (9.8259)	13.19 (18.2491)	12.47 (17.3091)	10.51 (15.3702)	9.79 (14.8312)	9.39 (14.3960)
Difference* (%)	15.44 (0.7083)	54.74 (53.1117)	56.21 (53.1975)	42.25 (41.7106)	29.01 (27.2520)	29,50 (28.5427)
Continuous subcutaneous insulin infusion (CSII) as exogenous input						
SARIMAX(4,0,4,2)(1,0,1)$_{33}$	2.90 (3.6264)	3.20 (4.6225)	5.86 (9.1595)	4.49 (7.7283)	4.82 (8.2186)	5.12 (8.4743)
ARIMAX(4,0,4,2)	7.97 (9.1372)	12.99 (18.8784)	12.95 (18.4822)	11.10 (16.5420)	10.58 (16.2076)	10.51 (16.1783)
Difference (%)**	63.61 (60.3117)	75.37 (75.5143)	54.75 (50.4415)	59.55 (53.2807)	54.44 (49,2917)	51.28 (47.6193)

* $100(|SARIMA - ARIMA|)/ARIMA$; ** $100(|SARIMAX - ARIMAX|)/ARIMAX$

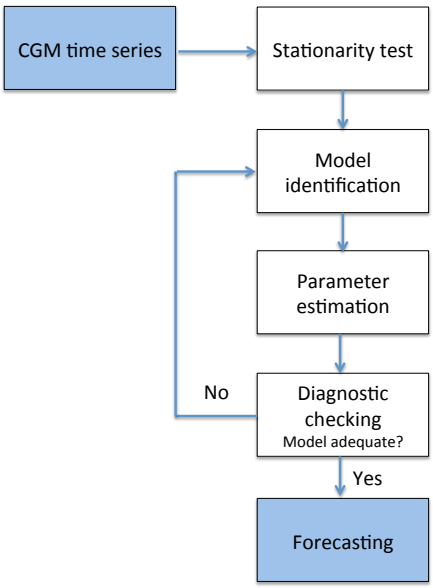


Fig. 1. Steps for building a good model through Box-Jenkins methodology.

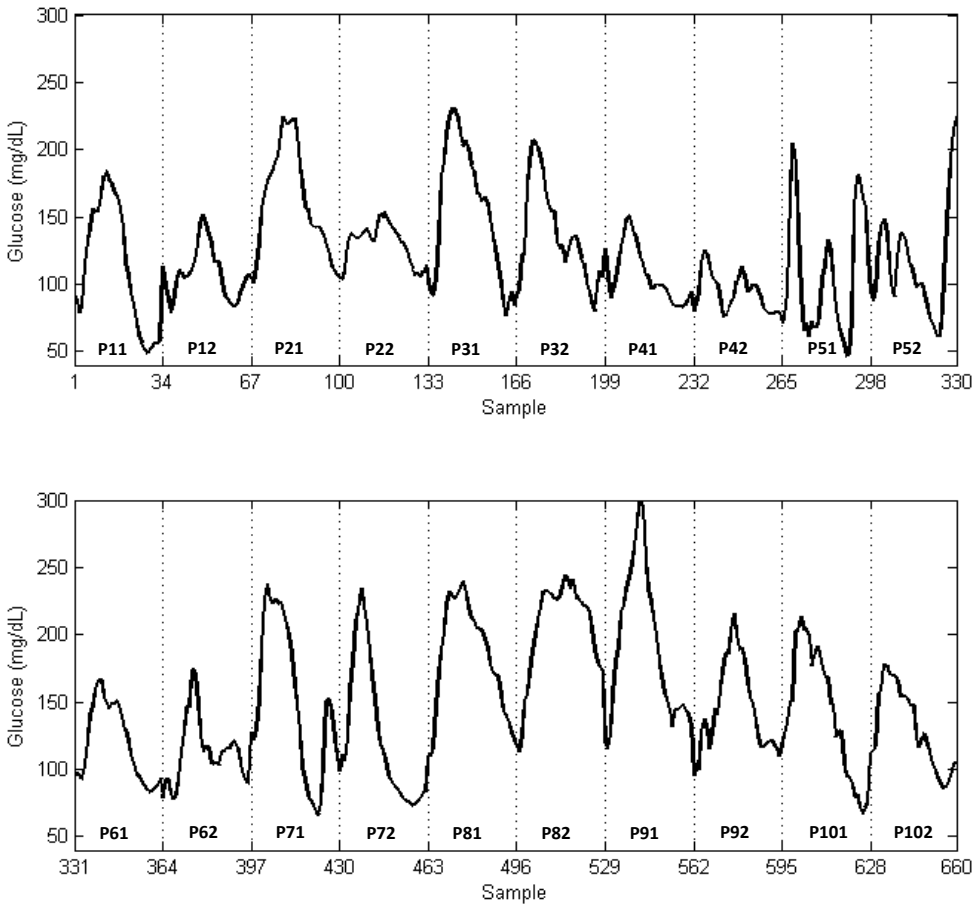


Fig. 2. CGM time series resulting from the concatenation of twenty 8-h postprandial periods for a same 60g carbohydrate meal. The notation P*ij* is used to name the different periods, where i is the number of the patient, $i \in \{1, \ldots, 10\}$, and j is the number of the study per patient, $j \in \{1,2\}$. Sampling period is 15 minutes, yielding 33 samples per postprandial period.

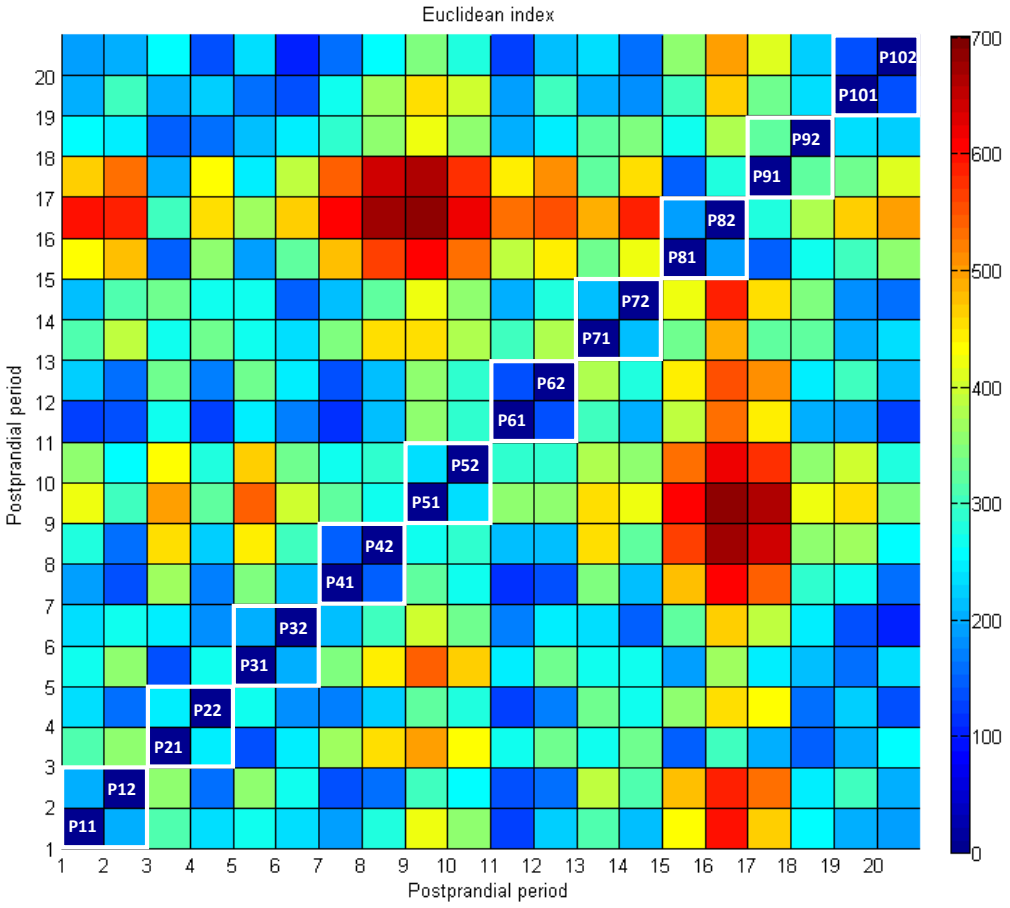
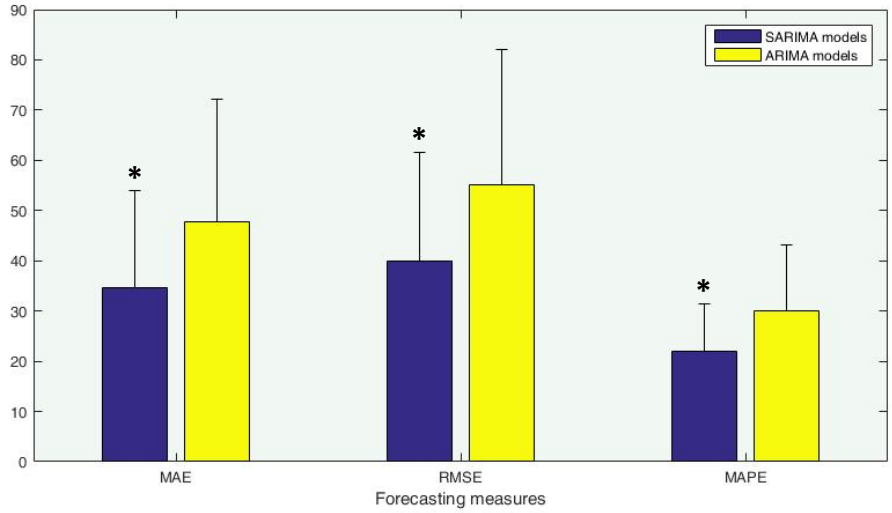


Fig. 3. Similarity among postprandial periods in the CGM time series as measured by the Euclidean distance between paired periods. Data is shown according to the color scale in the right. White boxes in the diagonal indicate periods corresponding to a same patient.

(a)



(b)

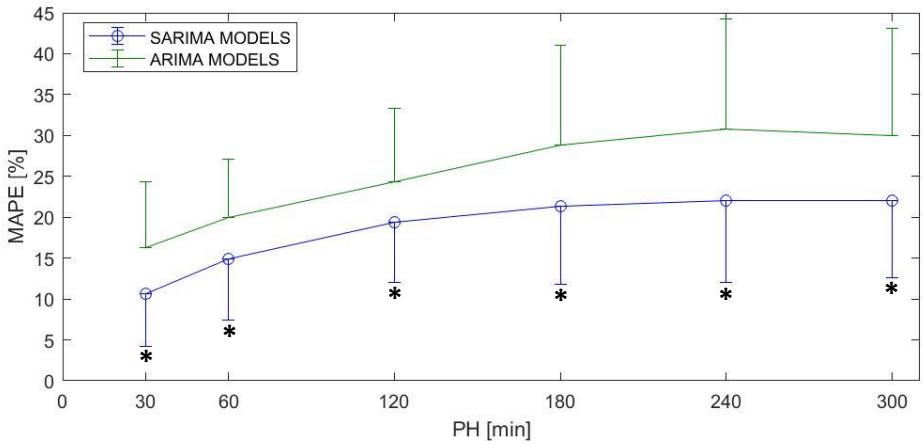


Fig. 4. Forecasting performance: (a) Mean and standard deviation of forecasting measures for the 20-fold cross-validation and a 5-h PH: MAE(mg/dL) is Mean Absolute Error; RMSE(md/dL) is Root Mean Square Error; MAPE(%) is Mean Absolute Percentage Error; (b) Mean and standard deviation of MAPE(%) for increasing values of the prediction horizon. * $p<0.05$.

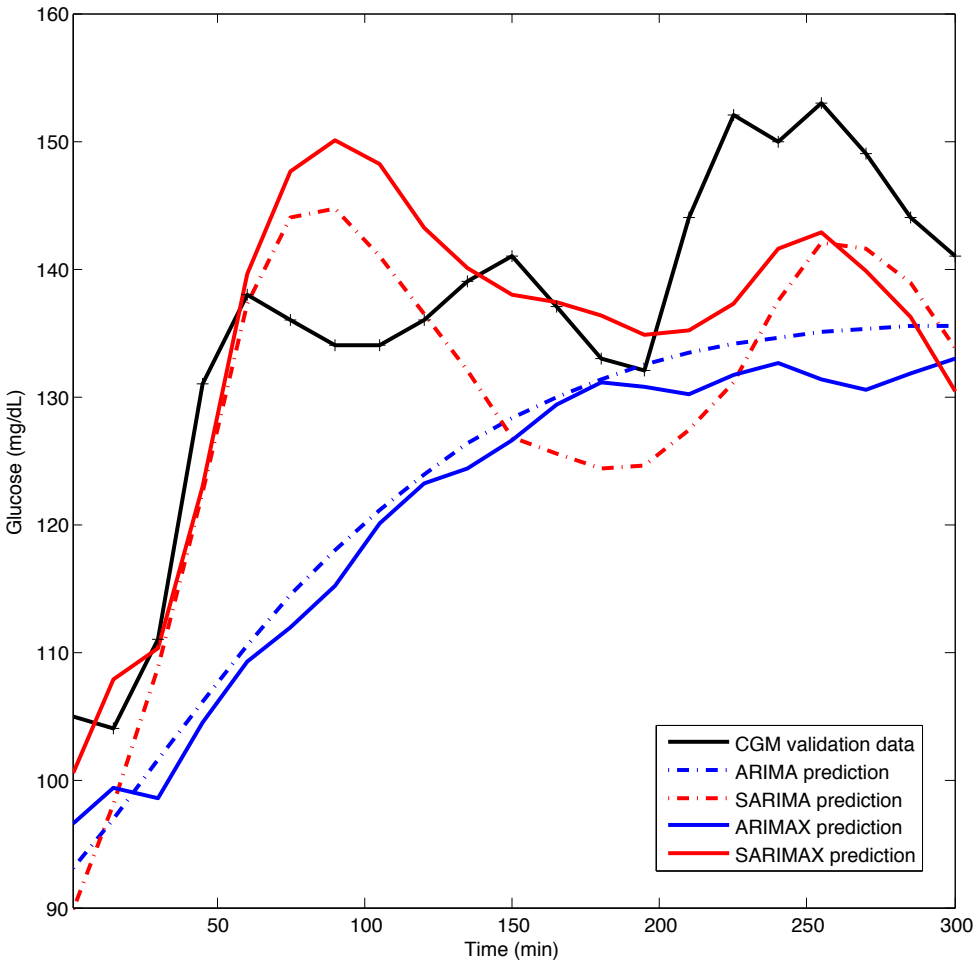


Fig. 5. Forecasting of models ARIMA(4, 0, 4), SARIMA(4, 0, 4)$(1, 0, 1)_{33}$, ARIMAX(4, 0, 4, 2) and SARIMAX(4, 0, 4, 2)$(1, 0, 1)_{33}$ for Run 4 considering a 5-h prediction horizon.