Pakistan Journal of Applied Sciences 3 (4): 274-279, 2003
ISSN 1607-8926

Identification of Two Time Series Models of Type 1 Diabetes using Patient's clinical Database

A.Karim El-Jabali and S. L. Alousi
Department of Electrical Engineering, Al-Isra University, Amman, Jordan

ABSTRACT

The use of patient's clinical data to construct models of glucose insulin dynamics is presented. The models were tested and validated. Two Multivariable time series models: Auto-Regressive with Exogenous Input (ARX) and Auto-Regressive Moving Average with Exogenous Input (ARMAX) were developed and validated. It is shown that the ARX is better than the ARMAX in terms of predicting the near future values of glucose concentration. These research models can be used to develop an optimal controller of blood glucose concentration that computes and delivers the required insulin dose.

Key words: Diabetes, modeling, identification, time series

Diabetes: the size of the problem

Diabetes mellitus takes an ever-increasing proportion interest in national health care policies and budgets because the number of people with diabetes grows worldwide, the. The latest World Health Organization estimate for the number of people with diabetes, world-wide, in 2003 is 370 million and about 9% of the global total number of deaths (World Health Organization, 2002). The total health care costs of treating diabetes in the USA in 1997 were US$ 98 billion in addition to intangible costs (pain, anxiety, inconvenience and generally lower quality of life etc). These annual figures are rising as diabetes prevalence increases.

Problem definition

The amount of glucose in the blood is controlled mainly by two hormones insulin and glucagon. Inadequate amount of these hormones can cause blood sugar levels to fall too low (hypoglycemia) or rise too high (hyperglycemia). The pancreas contains alpha and beta cells that produce glucagon and insulin respectively. Other hormones that influence blood sugar levels are cortisol, growth hormone and catecholamines. When blood sugar rises after a meal, the beta cells release insulin. The insulin helps glucose enter body cells, lowering blood levels of glucose to the normal range. When blood sugar drops too low, the alpha cells secrete glucagon. This signals the liver to release stored glycogen and change it back to glucose, raising blood sugar levels to the normal range. The normal range for blood sugar is about 60 to 120 mg dL^{-1}, depending on when a person last ate (Cotran, 1999).

Diabetes occurs when the body cannot use glucose for fuel because either the pancreas is not able to make enough insulin or the insulin that is available is not effective. As a result, glucose builds up in the blood instead of getting into body cells.

The aim of treatment in diabetes is keeping blood sugar in a close-to-normal range and to minimize the frequency and severity of glycemic excursions. To do this, people with diabetes may use multiple injections of insulin each day or use insulin pump, frequent testing of blood glucose, a diet and exercise plan and guidance from health care professionals (Jean_Venable *et al.*, 2001). In this paper, two algorithms that describe the interaction of insulin/glucose are presented and can be used in the treatment.

Mathematical modeling

Mathematical models and computer simulations are finding increased utility and application as the biochemical processes become better established and as the available computing power increases. Biologically realistic mathematical models serve as the basis for the majority of the methods used in quantitative physiologic analyses of medical data (Proceedings of National Institutes of Health Conference, 1989). These models help to test hypotheses about the mechanisms that govern these complex systems, reveal contradictions or incompleteness of data and hypotheses and allow prediction of system performance under untested or presently un-testable conditions. They may also predict and supply the values of experimentally inaccessible variables. Mathematical models need

to be as possible representative, comprehensive and not complicated. The first two requirements give good insight into whatever is being investigated by providing a concise, quantified summary of the observed behavior. The complexity of the model may be restricted by the need for its later implementation particularly for the more complex models implementation is a time-consuming task and requires considerable effort.

The general potential of mathematical models is good when there is sufficient knowledge of the system to allow the formulation of strong hypotheses (David Foster *et al.*, 2002). As the ability to acquire data expands and the sophistication of computing increases, more effective and broader applications may be expected. This includes but not restricted to estimation, prediction, calibration and optimization (NIST/SEMATECH, 2003). Here the estimation is used to obtain the value of the model's parameters (coefficients) for a particular combination of values of the predictor variables, while prediction is used as a test of the model's validity. This is accomplished by dividing the data set into two subsets one for estimation and the other for prediction.

Mathematical models of diabetes are of wide diversity. These models represent a range of approaches, including linear (Lehmann *et al.*, 1994) and nonlinear (Candas and Radziuk, 1994; Thomas Briegel and Volker Tresp, 2002); probabilistic (Andreassen *et al.*, 1994) and compartment (Parker *et al.*, 1999) and non-compartment (Cobelli et al., 1996). A spectrum of models is reported in (Naylor *et al.*, 1997). The models of glucose metabolism and insulin kinetics described here are expected to aid in reaching a generalized model.

Data and methods

The data set (Kahn, 1994). consists of a protocol of type I diabetic patient over a period of 75 days. During that time period, the present blood glucose level PGL, dosages of insulin injections (short term acting STI, mid-term MTI and long term LTI acting insulin) and qualitative indication about the amounts of food intake (MEAL) and physical effort exerted (EXERCISE) were recorded. The total number of readings is 560 for each variable. Fig. 1 shows twenty readings of glucose from which it is clear that blood glucose fluctuates in wide range. The average time between readings is 175 min. The goal of this study is to establish effective equations i.e. mathematical model of diabetes dynamics that are consistent with the data. The efficiency of the developed model can be tested by its ability to determine the next glucose level for specified parameters values as well as to perform parameter estimation in which model output are fit to the clinical data. The input to the system in all developed models consists from PGL, STI, MTI, LTI, MEAL and EXERCISE, while the output is the Next Glucose Level (NGL). In this study the data is divided into two sets: one called estimation data used to estimate the formula coefficients and the other called validation data used to find how fit is the given formula in predicting the output. Two multivariable models of glucose/insulin dynamics are developed and compared: Auto-Regressive with Exogenous Input (ARX) and its Moving Average version (ARMAX). The performance of models is computed for 5, 10 and 60 steps ahead using Mean Square Error (MSE), Mean Absolute Error (MAE), Average Error (AE) and Percentage Relative Error (PRE).

Time series models

Time series methods try to discover the properties of the system through successive in time measurements of the input and output variables. When several variables are considered together, the series becomes multivariate. The model in this case seeks to capture the trends and various inter-relationships between the different series. A time series is usually not a very compact representation of a time evolving phenomenon. It is necessary to condense the information and find a parameterization that contains the features that are most relevant for the underlying system. How much information can be recovered from time delayed copies of finite sets of noisy measurements is quite a complicated question and a general answer is not available [16], (Thomas Schreiber, 1999; Silipo *et al.*, 1998).

One formal requirement for almost all time series methods is stationarity. The most common definition of a stationary process found in textbooks (often called strong stationarity) is that all conditional probabilities are constant in time. If nonstationarity is detected, often the time series is discarded as unsuitable for a detailed analysis, or it was split into segments that were short enough to be regarded as stationary (Thomas Schreiber, 1999 and Silipo *et al.*, 1998). There are many methods to investigate the stationarity of the data [Thomas Schreiber, 1999 and Andrew and Harvey, 1992). The application of those methods suggested in (Andrew and Harvey, 1992) to data under consideration were approved the further developing of time series models for diabetics.

The application of time series methods to explore the underlying dynamics of biomedical systems is popular (Thomas Schreiber, 1999 and Prank *et al.*, 1998). There are many forms of time series models among them the Multivariable Auto-Regressive with Exogenous Input (ARX) Model and the Multivariable Auto-Regressive Moving Average with Exogenous Input (ARMAX) Model. Both forms are to be used now in this paper to derive models of diabetes mellitus.

Table 1: ARX performance evaluation for 5, 10 and 60 steps prediction

Model	ARX model			
Performance	MSE(x10^3) mg/dL	MAE mg/dL	AE mg/dL	PRE %
5 step ahead	0.00375	4.3295	4.3295	0.2405
10 step ahead	1.4140	21.668	-15.262	1.5025
60 step ahead	1434.8	29.739	-1.563	10.26

Table 2: ARMAX performance evaluation for 5, 10 and 60 steps prediction

Model	ARMAX			
Performance	MSE(x10^3) mg/dL	MAE mg/dL	A E mg/dL	PRE %
5 step ahead	645.9164	19.033	14.6018	0.9131
10 step ahead	1.2788	25.828	-8.4937	2.0166
60 step ahead	1387.5	30.461	-1.52	12.34

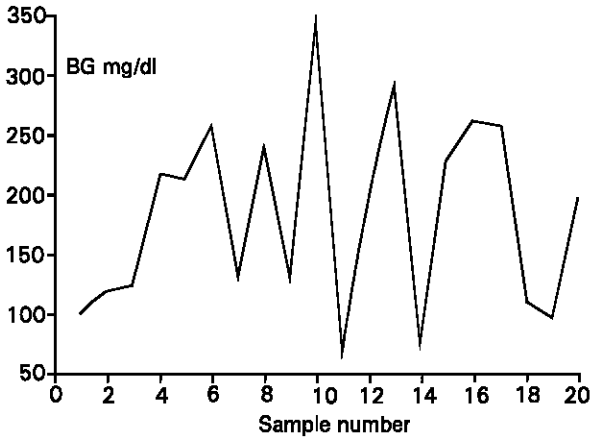

Fig. 1: The first twenty readings of blood glucose

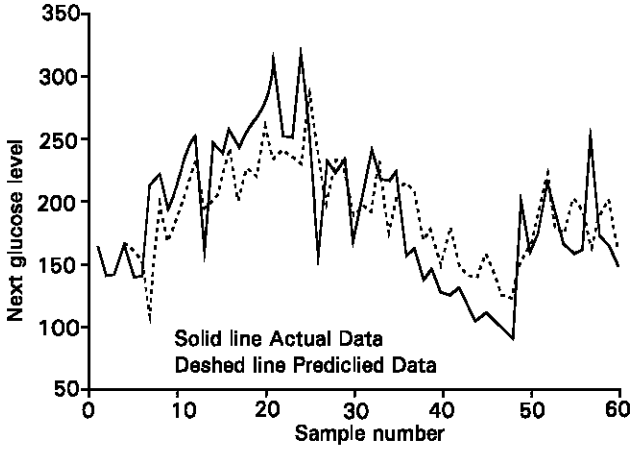

Fig. 2: Actual and Predicted Data for ARX Model

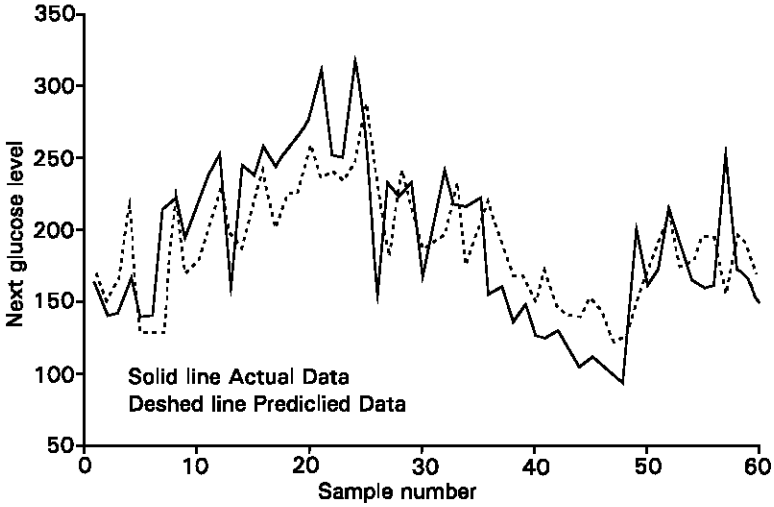

Fig. 3: Error between Actual and Predicted Data for ARX Model

ARX model

A common question that justifies the use of ARX models is whether present and future glucose values can be predicted from recent blood glucose history. The indication that there is significant statistical dependence between the individual successive glycemic measurements gives a motivation to use the record of patient to find a model that forecasts at least near future values of glucose. The last may be a helpful tool for the patient and physician in

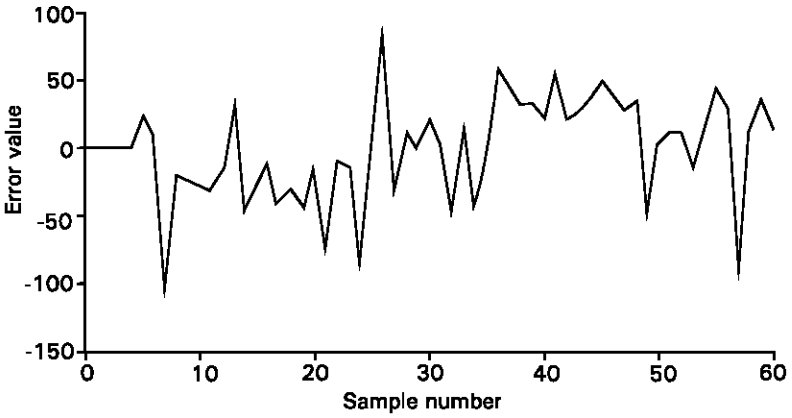

Fig. 4: The first ten actual and predicted data for ARX model

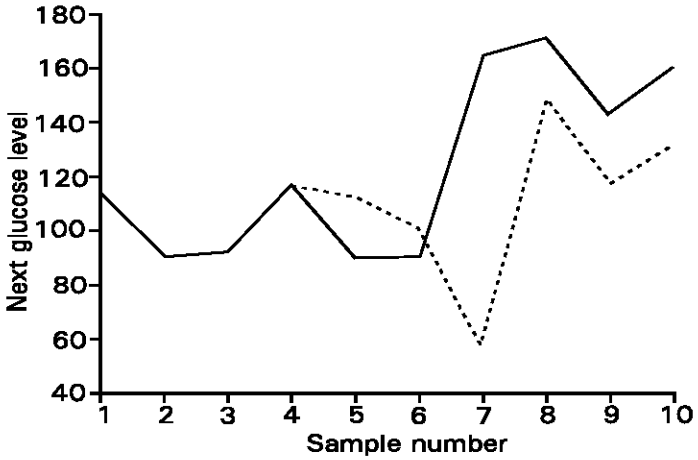

Fig. 5: Actual and Predicted Data for ARMAX Model

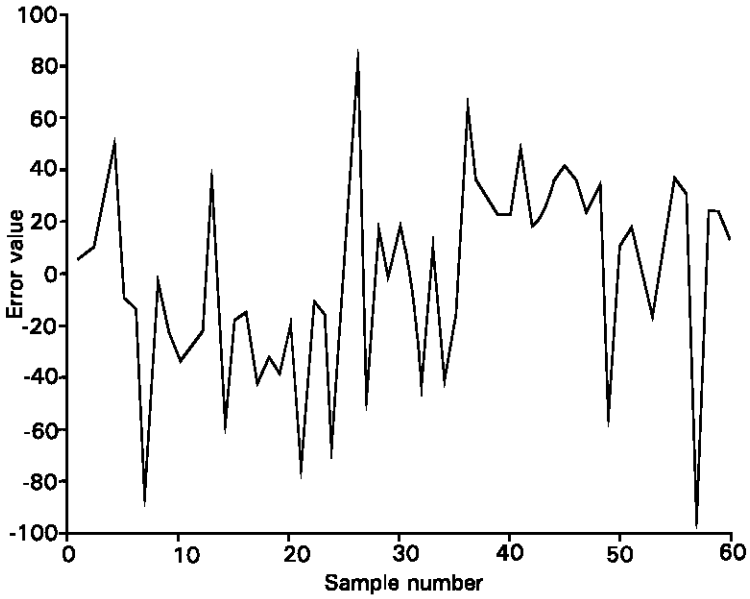

Fig. 6: Error between Actual and Predicted Data ARMAX Model

managing the complications of the disease. In this study, the problem is a multidimensional with five inputs and one output. A multivariable ARX model is given by

$$A(q)y(t) = B(q)u(t\text{-}nk) + e(t)$$

where u(t),y(t) and e(t) are input output and random disturbance vectors respectively and A(q), B(q) are the corresponding polynomials (Prank *et al.*, 1998) in the delay operator q^{-1}. After testing polynomials with different dimensions it was found that the following values of the model's parameters give the best results:

A (q) = 1 - 0.2787 q^-1 - 0.1093 q^-2 + 0.1125 q^-3
B1 (q) = 0.1813 q^-2 + 0.3393 q^-3 - 0.1359 q^-4
B2 (q) = -2.709 q^-2 + 1.813 q^-3 - 0.195 q^-4
B3 (q) = -0.8138 q^-2 + 2.086 q^-3 + 2.087 q^-4
B4 (q) = 0.5189 q^-2 + 1.762 q^-3 + 1.026 q^-4
B5 (q) = -4.937 q^-2 + 45.88 q^-3 + 7.849 q^-4.

(i.e. three past values of glucose history and four past values of different types of insulin, meal and exercise can be used to model the dynamics). After that, the performance of the obtained ARX model was evaluated with validation data set that differs from the estimation data set. The new data set was applied to the model to test its ability to predict the actual measured glucose level. The predicted output of the model and the actual data are depicted in (Fig. 2).The error (residuals) between the predicted and actual data is shown in (Fig. 3). The performance of the model is computed for 5, 10 and 60 steps ahead using MSE, MAE, AE and PRE. The results are summarized in Table 1.

From these figures and table, it is seen that the developed model's output replicates the system's output for the first four prediction samples. This fact is shown more clearly in (Fig. 4). This is equivalent to about ten future hours which is a good indication about the model validity. After that the model becomes less reliable and its performance tends to go far away from the actual data.

ARMAX model

A general multivariable ARMAX model structure is

A(q)y(t) = B(q)u(t-nk) + C(q)e(t)

where, A(q) and B(q) are like those for ARX while C(q) is the polynomial representing the weight of the disturbance e(t). The orders of A, B and C are to be selected and the coefficients of them are to be estimated using least square regression algorithm. After many iterations, the following best result (relatively) of parameters is reached:

A (q) = 1 + 0.05073 q^-1 - 0.04447 q^-2 + 0.2591 q^-3
B1 (q) = 0.005721 q^-1 + 0.3144 q^-2 + 0.4302 q^-3
B2 (q) = -0.5042 q^-1 - 2.41 q^-2 + 0.6396 q^-3
B3 (q) = 0.01116 q^-1 - 0.3916 q^-2 + 2.815 q^-3
B4 (q) = 1.083 q^-1 + 1.123 q^-2 + 1.281 q^-3
B5 (q) = 30.55 q^-1 + 21.67 q^-2 + 37 q^-3
C (q) = 1 + 0.3856 q^-1 + 0.1649 q^-2 + 0.09075 q^-3

Hence, the introducing of C (q) error matrix with three past values reduces the required past values of the input variables to three. As in the case of ARX, the obtained ARMAX model was tested with new data set. The predicted output by the model and the actual data are depicted in (Fig. 5). The residual (error) between the predicted and actual data is shown in (Fig. 6). As in the cases of NLSR and ARX, the performance is evaluated for 5, 10 and 60 steps ahead and the results are summarized in Table 2. It is clear that the performance of developed ARMAX model is not better than that of the ARX model. It gives a slightly increased error in the prediction of the near future values with relative percentage error. However, none of the first four values concise with the original value of the systems output as with the ARX model.

Discussion, conclusions and future work

Better understanding of the complex diabetes dynamics by including the effect of more variables (as meal and exercise) in the model is important for modeling and therefore for treatment. Moreover applying variety of time series

models as showed that these models have important properties from which the use of patient's history of a prior recorded information.

Exploring the predictability of blood glucose levels from the time course of the patient's records provides a good tool to judge whether the proposed models have acceptable performance or not. The ARX model showed its power to predict near future values of the glucose concentration with acceptable accuracy, while ARMAX model is less accurate in obtaining the expected value. Both models are of good usefulness for programmed insulin delivery devices. However, these models can not recommend how much insulin to deliver. The problem of development of an optimal controller that will use blood glucose measurements, qualitative information about food intake, physical exercise and past insulin infusion rates to compute the proper required insulin dose can be considered in future work. Another point in the future research is the development of a more generalized model after comparing the models derived from data base of many patients.

REFERENCES

Andreassen, S., J.J. Benn, R. Hovorka, K.G. Olesen and E.R. Carson, 1994. A probabilistic approach to glucose prediction and insulin dose adjustment: description of a metabolic model and pilot evaluation study. Comput Methods Programs Biomed, 41: 153-163.

Andrew, C. and Harvey, 1992. Time Series Models, 2nd edition; Harvester Wheatsheaf.

Brian Hipszer, 2001. A type 1 Diabetic Model; MSc. thesis, Drexel University.

Bremer, Troy, Gough and A. David, 1999. Is blood glucose predictable from previous values? Diabetes.

Cotran, R., 1999. Pathological Basis of Disease, 6th edition; W.B. Saunders Company.

Candas, B. and J. Radziuk, 1994. An adaptive plasma glucose controller based on a nonlinear insulin/glucose model. IEEE Trans Blamed Eng., 41: 116-124.

Cobelli, C., G. Toffolo and E. Ferrannini, 1996. A model of glucose kinetics and their control by insulin, compartmental and noncompartmental approaches. Math Biosci., 71: 291-316.

David Foster, J. Arthur and Jr. Atkinson, 2002. Principles of Pharmacokinetic Data Analysis: Modeling and Simulation, Philadelphia, Pennsylvania.

Jean_Venable, R., Goode and D. Pharm, 2001. New Advances in the Treatment of Diabetes; Medscape Portals.

Kahn, M., 1994. Artificial Intelligence in Medicine AIM-94;Washington University.

Lehmann, E.D., I. Hermanyi and T. Deutsch, 1994. Retrospective validation of a physiological model of glucose-insulin interaction in type 1 diabetes mellitus. Med. Eng. Phys., 16: 193-202.

NIST/SEMATECH, 2003. e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/.

Naylor, J.S., A.S. Hodel, A.M. Albisser, J.H. Evers, J.H. Strickland and D.A. Schumacher, 1997. Comparison of parameterized models for computer-based estimation of diabetic patient. glucose response. Med. Inform., 22: 21-34.

Proceedings of National Institutes of Health Conference, 1989. Modeling in Biomedical Research: An Assessment of Current and Potential Approaches: Applications to Studies in Cardiovascular/Pulmonary Function and Diabetes A, USA.

Parker, R.S., F.J. Doyle and N.A. Peppas, 1999. A Model-based Algorithm for Blood Glucose Control in Type 1 Diabetic Patients; IEEE. TA. BME.

Prank, Klaus, Jurgens, Clemens *et al.*, 1998. Predictive neural networks for learning the time course of blood glucose levels ; Neural Computation.

Raimondas Ciegis,. Some Algorithms in Mathematical Modeling of Diabetes Mellitus; International J. Informatica.

Silipo, R., G. Deco, R. Vergassola and H. Bartsch, 1998. Dynamics extraction in multivariate biomedical time series; Biological Cybernetics, 79: 15 and 13.

Thomas Briegel and Volker Tresp, 2002. A Nonlinear State Space Model for the Blood Glucose Metabolism of Diabetic; Automatisierungstechnik 50: 5; Oldenbourg Verlag.

Thomas Schreiber, 1999. Interdisciplinary application of nonlinear time series methods; Physics Reports.

World Health Organization, 2002. Fact Sheet No. 236.