

Original citation:

Khovanova, N. A., Zhang, Yan and Holt, Tim A. (2014) Generalised stochastic model for characterisation of subcutaneous glucose time series. In: 2014 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), Valencia, Spain, 1-4 Jun 2014. Published in: 2014 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI) pp. 484-487.

Permanent WRAP url:

<http://wrap.warwick.ac.uk/64745>

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions. Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher's statement:

“© 2014 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting /republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.”

A note on versions:

The version presented here may differ from the published version or, version of record, if you wish to cite this item you are advised to consult the publisher's version. Please see the 'permanent WRAP url' above for details on accessing the published version and note that access may require a subscription.

For more information, please contact the WRAP Team at: publications@warwick.ac.uk



<http://wrap.warwick.ac.uk>

Generalised stochastic model for characterisation of subcutaneous glucose time series*

Natasha Khovanova¹, Yan Zhang¹, and Tim A. Holt²

Abstract—A generalised stochastic model with second order differential equations is proposed to describe the response of blood glucose concentration to meals in groups of nondiabetic people and two types of diabetic patients. A variational Bayesian approach is applied in order to infer parameters of the models, and the best model was selected based on the computed log-evidence for each prandial event. The model with a linear structure represents most of the events, while the nonlinear terms need to be included more frequently for Type II diabetic patients. This indicates different physiological mechanisms of glucose absorption for different groups. The deterministic parameters and intensities of stochastic components are compared by groups using the ANOVA test, and the results show significant differences between the groups. This model can potentially be used for long term prediction of the glucose concentration response to external stimuli.

I. INTRODUCTION

According to World Health Organisation, 347 million people worldwide have diabetes and diabetes deaths will double between 2005 and 2030. Diabetes is a group of metabolic diseases that occur when the pancreas does not produce enough insulin (type I diabetes), or when the body cannot effectively use the insulin it produces (type II diabetes). Untreated diabetes is characterised by chronic hyperglycaemia (high levels of blood glucose). Control of hyperglycaemia is important to prevent long term complications, including effects on the eyes, kidneys, peripheral nerves, and cardiovascular system. However, low levels of blood glucose level can also be dangerous, and even life threatening. Therefore, controlling blood glucose levels and reducing dramatic glucose fluctuation through both lifestyle and pharmacological interventions are the main therapeutic goals in the management of diabetes.

For more than half a century researchers have been working towards establishing an effective dynamic model describing and predicting blood glucose concentration levels. Glucose concentration is tightly regulated by a complex neuro-hormonal control system and can be influenced by various factors. Exogenous factors such as food intake, exercise, rate of digestion etc. deviate the glucose level from the basal level, while endogenous factors, including epinephrine, cortisol, growth hormone [1], and pancreatic endocrine hormones (insulin and glucagon) [2], serve as internal regulators to drive the glucose concentration level back to normal and maintain it. Insulin is the main regulator of glucose homeostasis. The glucose-insulin interaction comprises a feedback control system: when the blood glucose level rises, insulin is released from pancreatic β -cells to inhibit hepatic glucose production and promote

glucose utilization so as to bring the glucose rapidly to the pre-perturbation level. If the glucose-insulin system is impaired, high levels of blood glucose (hyperglycaemia) or low levels (hypoglycaemia) may occur. Postprandial (after a meal) glucose test are often performed to monitor the control of diabetes patients. However, according to Cohen et al. [3], measures of 2-h glucose concentrations alone or combined with fasting concentration, body composition etc. are poor predictors of glycaemic exposure. Modern continuous glucose monitoring (CGM) has the advantage of obtaining hundreds of data points over a week, and therefore provides a better opportunity to understand the mechanisms of postprandial glucose change.

II. MODELS AND METHODS

Various models have been developed and simulated in continuous time to describe and predict changes in blood glucose level, and numerous experiments have been designed to define the parameters of these equations. The simplest model pioneered by Bolie in 1961 describing the responses of glucose levels to food intake has the following format

$$\begin{cases} \dot{G}(t) &= -a_1G(t) - a_2I(t) + J(t) \\ \dot{I}(t) &= a_3G(t) - a_4I(t) \end{cases} \quad (1)$$

where G , I denote the plasma glucose and insulin respectively, J is the glucose input, and a_1, a_2, a_3, a_4 are parameters [4]. This model was criticized as too simple to represent the glucose-insulin system mostly due to its assumption of linear relationship between insulin secretion and glucose. Many models have been developed to overcome this problem. The two most widely accepted are Bergman's minimal model [5] and Cobelli's two compartment model [6]. These models include multiple variables in order to reflect and adequately describe glucose-insulin dynamics as a multidimensional system. The limitation is that some parameters of these models are not readily measurable, which poses difficulties in the validation of these models and their applicability for prediction of glucose dynamics. This is because accessible experimental data for individual patients diagnosed with diabetes still remain one-dimensional: only one transient variable (the glucose level) as a function of time is available, representing the response of a complex system to an external influence. In order to overcome this limitation, an entirely different *minimal inference approach* has been employed by this research: development of a multidimensional model represented by inferring parameters of unrecorded variables.

Another accepted challenge accompanying modelling is the presence of a strong stochastic component. One of the main drawbacks of the models above is categorising every misfitting as a measurement noise without considering the imperfection of the model itself. The presence of the strong *stochastic* component can be an internal characteristic of the system alone and/or the result of measurement error. In nonlinear systems, noise acts as a driving force and can radically modify deterministic dynamics. While measurement noise is normally considered as a blurring effect to trajectories of the deterministic system, the coupling of noise to nonlinear deterministic equations can lead to non-trivial effects. Therefore, differentiation between system noise and measurement error provides more realistic description of observed variations in system's dynamics, accounts for fluctuations in physiological parameters, and improves estimations of the parameters and states of the system.

We applied the minimal inference technique in order to define a model based on stochastic differential equations which accounts for the nonlinear character of the blood glucose response to food intake, has minimal order and minimal number of unknown parameters, and is capable of predicting the glucose dynamics in response to the external stimuli.

A general form of stochastic differential equations for characterisation and prediction of postprandial (reaction to food intake) blood glucose levels can be written as follows:

$$\frac{d^2}{dt^2}x_t + f_1(x_t) \frac{d}{dt}x_t + f_2(x_t) = \eta_t \quad (2)$$

$$y_t = x_t + \varepsilon_t \quad (3)$$

where

$$f_1(x_t) = \sum_{i=0}^n \theta_{ki} x_t^i, \quad f_2(x_t) = \sum_{i=0}^n \theta_i x_t^i \quad (4)$$

Formula (2) is an evolution equation and formula (3) represents a measurement equation. y_t is the measured time series from Continuous Glucose Monitoring System, and ε_t is the measurement error. x_t is the blood glucose time series, also referred as hidden states, η_t is the dynamical noise. θ_i and θ_{ki} in formula (4) are evolution parameters. The initial conditions are the glucose level measured at time $t = 0$ and its rate of change. $f_1(x_t)$ and $f_2(x_t)$ are polynomial functions of x_t . Increasing the polynomial terms of $f_1(x_t)$ and $f_2(x_t)$ can introduce nonlinearity into the structure. The best structure can be selected by the following variational Bayesian method.

A variational Bayesian minimal inference approach selects the best model by comparing the optimised evidence term for each model. The evidence $p(y|m)$ represents the plausibility of the model by calculating the possibility of getting data y given model m . However, intergrating over the distributions of all the parameters θ_i and θ_{ki} is usually computational expensive, so they are approximated by tractable distributions using mean-field approximation and Laplace approximation. Therefore, the value of the model evidence can only be

approached. The variational Bayesian method constructs a lower bound on the model evidence and optimises this bound using an iterative scheme. The approximation can be obtained by alternating between estimating the distribution over hidden variables for a particular setting of the parameters and then re-estimating the parameters given the distribution over the hidden variables. The optimised model evidence is a key quantity used to choose between different structures of the model, and the model with the largest value of the optimised evidence is chosen [7]. By introducing nonlinear terms into the equations, this model improves Bolie's oversimplified model; at the same time, it does not need the extra data source such as labelled meals required by Cobelli's model. Driven by the measured glucose data only, the inferred parameters describe the dynamics and characteristics of the underlying complex physiological processes influencing the change of blood glucose levels with time. Since the mechanisms of the glucose absorption is different for Type I, Type II diabetes and non-diabetic people, different distributions of parameters for these groups are expected.

III. RESULTS AND DISCUSSION

Fifteen glucose profiles for three categories of people (5 profiles in control group of people without diabetes, 4 profiles in the group of patients with Type I diabetes, 5 profiles in the group of patients with Type II diabetes) were available for the study. An example of a glucose time series for one patient is shown in Figure 1. Subcutaneous glucose values were taken every 5 minutes over 72 hours using CGMS (Continuous Glucose Monitoring System) and represent a number of postprandial peaks corresponding to food intake. We chose all the distinguishable peaks in the profile and analysed them using Eqn.(2)-(4).

A variational Bayesian toolbox [8] was applied to extract the dynamic patterns hidden in noisy data, i.e. to fit available glucose time series by Eqn.(2). The nonlinear functions f_1 and f_2 were optimised by minimising the number of nonlinear parameters θ_i and θ_{ki} necessary to describe a prandial event. Various degrees of polynomial nonlinearities in Eqn.(2) were considered starting with the linear case (linear f_1 and f_2). Nonlinear terms were added to the Eqn.(4) until a satisfactory fitting was achieved.

103 out of 146 peaks across all fifteen profiles were modelled by the linear model (Model 1) with f_1 and f_2 being linear functions with one parameter each, i.e. $\theta_i = 0$ and $\theta_{ki} = 0$ for $i > 0$ in Eqn.(4). The rest of the peaks were fit by a nonlinear model (Model 2), where $f_1(x_t) = \theta_{k2}x^2 + \theta_{k1}x + \theta_{k0}$ is a quadratic function with three parameters and $f_2(x_t) = \theta_0$ is a linear function with one parameter. Note that some of the peaks could be fitted well using both models. Introducing nonlinearities to the model can improve the performance in some cases, however the model might be unstable due to overfitting. The trade-off between a better fitting result and a more complex structure needs to be carefully balanced. As discussed in the previous section, model evidence is considered to be the best indicator of selecting the best model [7]. The toolbox [8] defines the

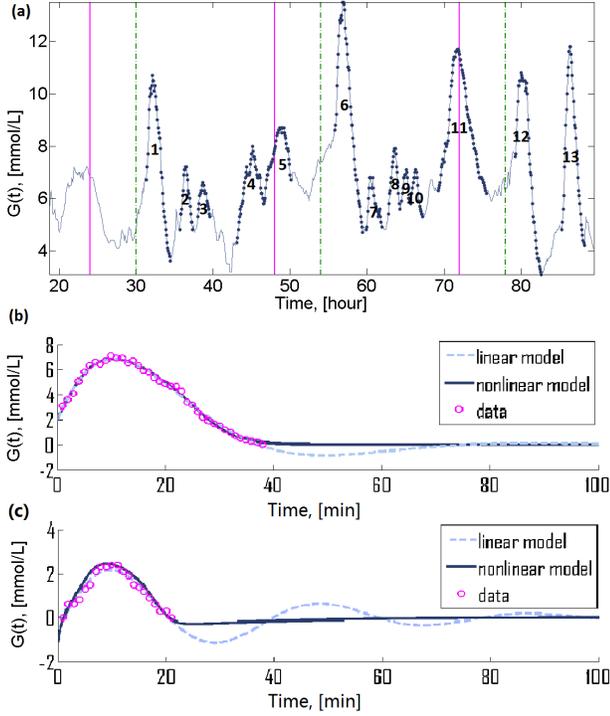


Fig. 1. (a) Concentration of subcutaneous glucose level $G(t)$ measured continuously over 72 hours. Solid line represents the measured glucose values and the dots are the values used for inferring parameters of single prandial events (peaks). The dashed and solid vertical lines corresponds to 6 am and midnight respectively. There were 13 peaks observed over 72 hours as labelled; (b) Fitting for peak 1; (c) Fitting for peak 2.

TABLE I

SUMMARY OF PEAK FITTING USING MODEL 1 AND MODEL 2

	control	type I	type II
Total Peaks	50	47	49
No. of peaks fitted by Model 1	35	34	34
No. of peaks fitted by Model 2	23	24	33

lower bound of the log-evidence as free energy F . There is also another important criterion to choose the correct model based on the physiological background of the model. Since the data describe the blood glucose level response to an external stimulus, the system needs to be stabilised at a reasonable level without too many oscillations for several hours after the food intake. Please refer to Table I for the details of the fitting. From the table, around 70% of the peaks can be fitted using the linear model (Model 1) for all the groups. Less than 50% of the peaks can be fitted using the nonlinear model (Model 2) for the control and Type I group, but around 67% of the peaks can be fitted by Model 2 for the Type II diabetic group, which indicates a strong nonlinear character of the response in Type II diabetic patients.

For the linear model (Model 1), two parameters are inferred to capture the dynamics of the glucose level change. $\sqrt{\theta_0}$ is the natural frequency of the system, which determines how fast the system oscillates, and $\theta_{k0}/2\sqrt{\theta_0}$ is the damping length describing how oscillations in a system decay after

TABLE II

ANOVA TEST RESULTS (P VALUES) FOR THE PARAMETERS AND NOISE INTENSITIES, COMPARED BETWEEN THREE GROUPS FOR MODEL 1

	control-type I	type I-type II	control-type II
θ_{k0}	0.0211	0.5798	0.0029
θ_0	5.44E-5	0.6872	2.59E-5
I_{system}	1.85E-4	0.0051	0.0538
$I_{measure}$	4.22E-8	9.64E-4	8.32E-5

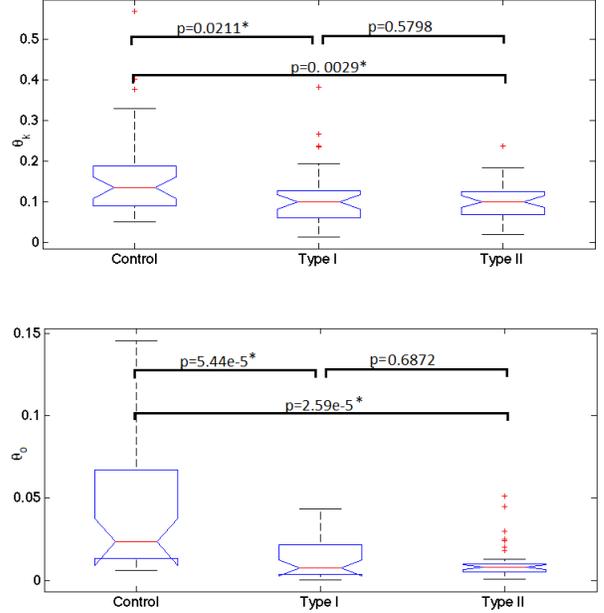


Fig. 2. ANOVA test result for (a) θ_0 and (b) θ_k compared among three groups using Model 1 (linear model)

a disturbance. Comparing the inferred parameters among groups may bring insights of different characteristics of the glucose level variation. The parameters θ_{k0} and θ_0 in Model 1 for 103 peaks were compared by groups using ANOVA which is a parametric test to analyse the differences between groups. If the p-value from the test between two groups is below 0.05, the two groups are considered to be significantly different from each other. The result are shown in Table II. For both parameters, the mean values of the Control group are significantly different from Type I or Type II group, while Type I and Type II group do not show significant differences. Please refer to Figure 2 for details. This conclusion supports the results of [9], which shows larger inter-dependence between neighbouring values in diabetes groups compared with the control group. Another way to determine the plausibility of the models is to compare the intensity of the system noise and measurement error, which are shown in Figure 3. Comparing differences between all three groups estimated by ANOVA, the mean value of the system noise intensity for the control group is the smallest, which suggests that this linear model fits the data from control group better than Type I or Type II groups.

Model 2 fits 80 peaks well, including 43 peaks which

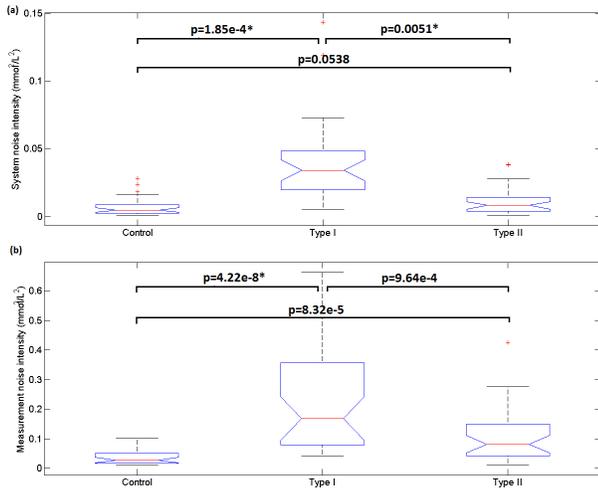


Fig. 3. ANOVA test result for (a) system (b) measurement noise intensities compared among three groups using Model 1 (linear model)

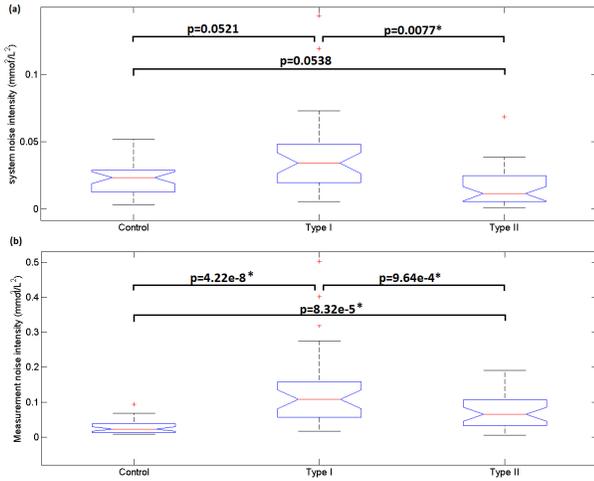


Fig. 4. ANOVA test result for (a) system (b) measurement noise intensities compared among three groups using Model 2 (nonlinear model)

can not be fitted by Model 1. Comparing the four parameters (θ_{k0} , θ_{k1} , θ_{k2} and θ_0) between these three groups by ANOVA, the mean values of all the parameters are significantly different from each other between the Control and Type II groups, and show no differences between Type I and Type II. This result is similar to the linear model, which indicates similar characteristics between diabetic glucose variations. It is shown (Fig. 4) that the mean value of the system noise intensity of Type II is the smallest of the three groups, which is different from the linear model where the control group has the smallest mean value of system noise intensity. This test result confirms the suggestion from [9] that nonlinear models in response to external stimuli for some diabetic patients are better than linear models. As shown in Fig. 3, the Type II diabetes group has smaller system noise intensity in Model 2, which indicates that a nonlinear model fits Type II data better than other groups. Compared with

TABLE III
ANOVA TEST RESULTS (P VALUES) OF THE PARAMETERS AND NOISE INTENSITIES, COMPARED BETWEEN THREE GROUPS FOR MODEL 2

	control vs type I	type I vs type II	control vs type II
θ_{k0}	0.1109	0.0838	1.33E-5
θ_{k1}	9.46E-4	0.4259	1.22E-5
θ_{k2}	0.0907	0.5719	0.0381
θ_0	0.0237	0.5819	1.95E-4
I_{system}	0.0521	0.0077	0.0538
I_{meas}	4.22E-8	9.64E-4	8.32E-5

Fig. 4, the Control group has lower system noise intensity in Model 1.

IV. CONCLUSION

In summary, a generalised stochastic model for glucose time series characterisation has been developed. It can be used to describe the responses of blood glucose levels to food intake in people with or without diabetes. The majority of the prandial events can be fitted well using a linear model, but some events need a nonlinear damping term $f_1(x_t) = \theta_{k2}x^2 + \theta_{k1}x + \theta_{k0}$. This nonlinear damping term is required more often in diabetes patients which indicates a stronger nonlinear character of the response in diabetes patients compared with people from the control group. Values of parameters vary from peak to peak for a single profile, but the distributions of the values show significant differences between diabetes patients and the Control group, while the differences between Type I and Type II patients are not as obvious.

REFERENCES

- [1] Cobelli C., Dalla Man C., Sparacino G., Magni L., De Nicolao G., Kovatchev B. P. (2009). Diabetes: models, signals, and control. *Biomedical Engineering, IEEE Reviews in*, 2, 54-96.
- [2] Makroglou A., Li J., Kuang Y. (2006). Mathematical models and software tools for the glucose-insulin regulatory system and diabetes: an overview. *Applied numerical mathematics*, 56(3), 559-573.
- [3] Cohen O., Basu R., Bock G., Dalla Man C., Campioni M., Basu A., Toffolo G., Cobelli C., Rizza R. A. (2006). Prediction of Postprandial Glycemic Exposure Utility of fasting and 2-h glucose measurements alone and in combination with assessment of body composition, fitness, and strength. *Diabetes care*, 29(12), 2708-2713.
- [4] Bolie VW. Coefficients of normal blood glucose regulation. *J Appl Physiol* 1961;16:783788. [PubMed: 13870789]
- [5] Bergman R. N. (1989). Toward physiological understanding of glucose tolerance: minimal-model approach. *Diabetes*, 38(12), 1512-1527.
- [6] Cobelli C., Caumo A., Omenetto M. (1999). Minimal model SGoverestimation and Slunderestimation: improved accuracy by a Bayesian two-compartment model. *American Journal of Physiology-Endocrinology And Metabolism*, 277(3), 481-488.
- [7] Friston, K., Stephan, K., Li, B., Daunizeau, J. (2010). Generalised filtering. *Mathematical Problems in Engineering*, 2010.
- [8] Toolbox SPM: The toolbox is an extension of SPM package developed at UCL, UK. <http://www.fil.ion.ucl.ac.uk/spm/>.
- [9] Khovanova N. A., Khovanov I. A., Sbrano L., Griffiths F., Holt T. A. (2012). Characterisation of linear predictability and non-stationarity of subcutaneous glucose profiles. *Computer Methods and Programs in Biomedicine*, 110(3), 260-267.