LUND UNIVERSITY

Maximum likelihood estimation of a time-inhomogeneous stochastic differential model of glucose dynamics

Picchini, Umberto; Ditlevsen, Susanne; De Gaetano, Andrea

Link to publication

Maximum likelihood estimation of a time-inhomogeneous stochastic differential model of glucose dynamics

Umberto Picchini[1,2]   Susanne Ditlevsen[1]   Andrea De Gaetano[2]

[1] Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen Ø, Denmark. email: umberto@math.ku.dk, susanne@math.ku.dk
[2] Biomathematics Laboratory IASI-CNR, Largo A. Gemelli 8, 00168 Rome, Italy; email: andrea.degaetano@gmx.net

Abstract

Stochastic differential equations (SDEs) are assuming an important role in the definition of dynamical models allowing for explanation of internal variability (stochastic noise). SDE models are well-established in many fields, such as investment finance, population dynamics, polymer dynamics, hydrology and neuronal models. The metabolism of glucose and insulin has not yet received much attention from SDE modellers, except from a few recent contributions, because of methodological and implementation difficulties in estimating SDE parameters. **Objectives:** here we propose a new SDE model for the dynamics of glycemia during a euglycemic hyperinsulinemic clamp experiment, introducing system noise in tissue glucose uptake, and apply for its estimation a closed-form Hermite expansion of the transition densities of the solution process. **Results:** the present work estimates the new model parameters using a computationally efficient approximate maximum likelihood approach. By comparison with other currently used methods, the estimation process is very fast, obviating the need to use clusters or expensive mainframes to obtain the quick answers needed for everyday iterative modeling. Furthermore, it can introduce the demonstrably essential concept of system noise in this branch of physiological modeling. **Conclusions:** SDE modeling for metabolic processes is physiologically pertinent and computationally feasible using commonly available resources.

Keywords: stochastic differential equations, dynamical models, non-autonomous differential equations, system noise, parameter estimation, closed-form transition density expansion, Hermite expansion, insulin, euglycemic hyperinsulinemic clamp.

1   Introduction

In this paper we consider a new stochastic differential model for the dynamics of the glycemia observed during a Euglycemic Hyperinsulinemic Clamp procedure (EHC, DeFronzo et al. (1979)) performed on human subjects, derived from a more complex model (Picchini et al. (2006)).

We choose to consider stochastic differential equation (SDE) models because deterministic models (ODE, DDE, PDE) do not accommodate random variations of metabolism. In fact, a deterministic model assumes (i) that the mathematical process X generating the observations (glycemias in our situation) is smooth (continuous and continuously differentiable) in the considered time-frame; and (ii) that the variability of the actual measurements is due only to observation errors, which do not influence the course of the underlying process. An alternative approach would result from the hypothesis that the underlying mathematical process itself is not smooth, at least when considered at a feasible time resolution. Physiologically this would be equivalent to postulating e.g. that the rate of glucose uptake by tissues varies randomly over time around some average level. This assumption leads to an SDE model as a natural extension of classic deterministic models.

Appropriate parameter values in these SDE models are crucial for the characterization of the dynamic phenomena being considered. It is often the case that these parameters are not known accurately. Researchers are naturally interested in obtaining better estimates of the parameters from experimental data: in practice the available data are discrete time series, sampled over some time interval, whereas SDEs are driven by almost surely continuous processes; this may complicate the estimation. Parameter estimation for discretely observed diffusion processes is non-trivial, because of theoretical and implementation difficulties, and during the past decades there has been a great deal of research effort in this area (e.g. Aït-Sahalia (2001, 2002b), Beskos et al. (2006), Bibby et al. (2004), Bibby and Sørensen (1995), Brandt and Santa-Clara (2002), Dacunha-Castelle and Florens-Zmirnou (1986), Ditlevsen

and De Gaetano (2005), Durham and Gallant (2002), Elerian et al. (2001), Gallant and Long (1997), Gouriéroux et al. (1993), Hurn and Lindsay (1999), Kutoyants (2004), Nicolau (2002), Pedersen (1995), Prakasa Rao (1999), Shoji and Ozaki (1998), Sørensen (2000)).

In Picchini et al. (2006) a two-dimensional SDE model considering simultaneously the dynamics of glucose and insulin was analyzed. Parameter estimation proved difficult, so that a two-step procedure was employed, where all parameters in the drift were estimated first on the corresponding deterministic model, and the diffusion parameter was estimated subsequently by Monte Carlo approximations of the likelihood (Pedersen (2001)). This approach was very time consuming, and moreover it would be more appropriate to estimate all parameters in a single optimization pass. This inspired us to simplify the model to a one dimensional formulation, restricting attention to glucose dynamics after the steady state of insulin concentration has been reached. The present work is concerned with computationally efficient estimation of the parameters of this reduced SDE model of the EHC experiment, paying particular attention to those parameters most important for the evaluation of the patient's insulin sensitivity (specifically the insulin-dependent glucose disposal rate K_{xgI}, see Section 2.3). The parameter estimation problem is approached as suggested in Egorov et al. (2003), building on ideas presented in Aït-Sahalia (2002b), which allow to estimate all model parameters in a single pass, in fast computer time.

Define the one-dimensional time-homogeneous (Itô) SDE

$$dX_t = \mu(X_t, \theta)dt + \sigma(X_t, \theta)dW_t, \qquad X_{t_0} = X_0$$

where W is a standard Wiener process (Brownian motion), $\mu(\cdot, \cdot) : \mathbb{R} \times \Theta \to \mathbb{R}$ is the drift term, $\sigma(\cdot, \cdot) : \mathbb{R} \times \Theta \to \mathbb{R}$ is the diffusion term and for simplicity we assume $X_0 = x_0$ fixed (i.e. non-random). The drift and the diffusion are assumed to be known functions depending on an unknown p-dimensional parameter vector $\theta \in \Theta \subset \mathbb{R}^p$. Since X is a Markov process the likelihood function of θ is simply the product of transition densities; the transition densities of X are rarely known but can often be approximated, see the Discussion.

The data consist of observations $x_0, x_1, ..., x_n$ at time-points $0 \leq t_0 < t_1 < ... < t_n$ of X, and are often assumed to be generated without measurement error. This assumption preserves the Markov property of the observations: otherwise, if the observations are contaminated with

measurement error they would no longer be Markov. When the measurement error is small compared to the noise term $\sigma(\cdot, \cdot)$, it can in many situations conveniently be ignored. The diffusion term can be interpreted as the action of many factors, each with a small individual effect, which are not explicitly represented in the deterministic part of the model (the drift term), and which instantaneously affect the X values. Therefore, in the stochastic differential model, the total effect of many small effects (which are not individually modelled) is represented by the diffusion term, while in the drift term the most relevant and generally well-recognized factors affecting the mean structure of the process are explicitly included.

The log-likelihood function $l_n(\theta)$ of θ is given by (disregarding the asymptotically irrelevant density of the initial observation X_0, in case it is not fixed):

$$l_n(\theta) = \sum_{i=1}^{n} \ln p(x_i, t_i | x_{i-1}, t_{i-1}; \theta) \tag{1}$$

where $p(X_t, t | X_s, s; \theta)$, $s < t$ is the transition density of X. Under mild regularity conditions the corresponding maximum likelihood estimator (MLE) $\hat{\theta}$ is consistent, asymptotically normally distributed and asymptotically efficient as n tends to infinity (Dacunha-Castelle and Florens-Zmirnou (1986)).

However, often the transition density function $p(\cdot)$ is unknown. One approach to this problem is to compute an approximation to $p(\cdot)$. There are several ways to do that (see Sørensen (2004) and Aït-Sahalia (2007) for a review): a fast and accurate method was suggested by Aït-Sahalia (2001, 2002b) for time-homogeneous SDEs, and was extended by Egorov et al. (2003) to one-dimensional time-inhomogeneous SDEs (i.e. the SDE depends directly on time t, not only through the process values). In this work we consider a new one-dimensional time-inhomogeneous stochastic model of glycemia dynamics. For this model we have been able to estimate parameters rapidly using the approach suggested in Egorov et al. (2003).

2 Material and methods

2.1 Subjects

Data from a previous EHC study were analyzed. Sixteen subjects were enrolled at the Department of Internal Medicine at the Catholic University School of Medicine in Rome. For

one subject the recorded glycemia values were accidentally lost and this subject was therefore discarded from the analysis. All subjects were clinically euthyroid, had no evidence of diabetes mellitus, hyperlipidemia, or renal, cardiac or hepatic dysfunction and were undergoing no drug treatments that could have affected carbohydrate or insulin metabolism. The subjects consumed a weight-maintaining diet consisting of at least 250 g of carbohydrate per day for 1 week before the study. Table 1 reports metabolical characteristics of the subjects. The study protocol followed the guidelines of the Medical Ethics Committee of the Catholic University of Rome Medical School; written informed consent was obtained from all subjects.

2.2 Experimental protocol

Each subject was studied in the postabsorptive state after a 12-14 h overnight fast. Subjects were admitted to the Department of Metabolic Diseases at the Catholic University School of Medicine in Rome the evening before the study. At 07.00 hours on the following morning, the infusion catheter was inserted into an antecubital vein; the sampling catheter was introduced in the contralateral dorsal hand vein and this hand was kept in a heated box (60 $^\circ C$) in order to obtain arterialized blood. A basal blood sample was obtained in which insulin and glucose levels were measured. At 08.00 hours, after a 12-14 h overnight fast, the EHC was performed according to DeFronzo et al. (1979). A priming dose of short-acting human insulin was given during the initial 10 min in a logarithmically decreasing manner so that the plasma insulin was raised acutely to the desired level. During the five-hour clamp procedure, the glucose and insulin levels were monitored every 5 min and every 20 min respectively, and the rate of infusion of a 20% glucose solution was adjusted during the procedure following the published algorithm DeFronzo et al. (1979). Because serum potassium levels tend to fall during this procedure, KCl was given at a rate of 15-20 mEq/h to maintain the serum potassium between 3.5 and 4.5 mEq/l. Serum glucose was measured by the glucose oxidase method using a Beckman Glucose Analyzer II (Beckman Instruments, Fullerton, Calif., USA). Plasma insulin was measured by microparticle enzyme immunoassay (Abbott Imx, Pasadena, Calif., USA).

2.3 SDE model

Consider the one-dimensional time-inhomogeneous (Itô) SDE:

$$dX_t = \mu(X_t, t)dt + \sigma(X_t)dW_t, \qquad X_{t_0} = X_0, \quad t \geq t_0 \qquad (2)$$

where $X_t \equiv X(t)$ represents the glycemia at time t for a given subject, and

$$\mu(X_t, t) = \frac{T_{gx}(t - \tau_g) + T_{ghnet}}{V_g} - K_{xgI}I^*X_t, \qquad (3)$$

$$\sigma(X_t) = \sigma I^* X_t. \qquad (4)$$

Here W is a Wiener process with $W_{t_0} = 0$, X_0 is the recorded glycemia at time $t_0 = 40$ min for the given subject; $T_{gx}(\cdot)$ is an (input or forcing) function representing the variable glucose infusion rate, whose values $\lambda_1, ..., \lambda_m$ change at times $0 = \nu_1 < \nu_2 < \cdots < \nu_m$, and are obtained during the EHC procedure according to the algorithm in DeFronzo et al. (1979). The $T_{gx}(\cdot)$ function is defined as:

$$T_{gx}(t) = \sum_{\nu_j \leq t} \frac{(\lambda_j - \lambda_{j-1}) \cdot (t - \nu_j)^5}{\nu_j + (t - \nu_j)^5}, \qquad t > 0, \quad \lambda_0 = 0, \quad j = 1, ..., m, \qquad (5)$$

where $T_{gx}(t) = 0$ for every $t \in [-\tau_g, 0]$, $t = 0$ being the instant in which the insulin infusion start (to be distinguished from the t_0 instant, which equals 40 min). The exponent of $(t - \nu_j)$ in (5) has been chosen to be the minimum integer such that the average of $\{|T_{gx}(\nu_j) - \lambda_j|\}_{j=1,...,m}$ is less than 3×10^{-3}. Thus $T_{gx}(\cdot)$ depends explicitly on t and so the SDE (2)-(4) is time-inhomogeneous. The other variables and the parameters are defined in Table 2 and Table 3, respectively.

The insulin-dependent glucose disposal rate (K_{xgI}) – which is constrained to be positive, since insulin's only action is to accelerate glucose removal from plasma – is assumed to exhibit substantial irregular oscillations over time. The diffusion coefficient $\sigma(\cdot)$ is obtained by allowing the parameter K_{xgI} to vary randomly as $(K_{xgI} - \xi_t)$, where $\xi(\cdot)$ is a Gaussian white-noise process. The noise term $\xi_t dt$ can be written as σdW_t (see e.g. Øksendal (2000)), where $\sigma > 0$ scales the Wiener process. Notice that the system noise may subsume the effects of terms not explicitly included in the model.

Thus model (2)-(5) expresses the variation of plasma glucose concentration when the tissue

6

glucose uptake rate K_{xgI} is perturbed by stochastic noise σdW. At $t_0 = 40$ minutes from the start of the EHC experiment, insulin concentrations appear nearly constant, and thus model (2)-(5) considers no insulinemia dynamics but only the average insulinemia I^* (see Picchini et al. (2005, 2006) for a two-dimensional deterministic and a stochastic approach, respectively). The dynamics of the glucose concentration in its distribution space (V_g) is attributed to the external glucose infusion rate (T_{gx}), to liver glucose output (T_{ghnet}) and to insulin-dependent glucose tissue uptake rate (K_{xgI}), in dependence of the glycemia X_t and the average insulinemia I^*. Infused glucose raises glycemia after a delay τ_g due to the time required to equilibrate the intravenously infused quantity throughout the distribution space: given that recirculation time is of the order of 30 seconds, we explored three reasonable values for τ_g, namely $\tau_g = 1$, 2 and 3 min, in order to perform a sensibility analysis on the other model parameters. For these different τ_g values, we observed no appreciable differences in the parameter estimates: the reported results refer thus to the case $\tau_g = 1$.

The external forcing function $T_{gx}(\cdot)$, which in this model formulation only depends on t, is bounded between 0 and 5 mmol/min/KgBW, approximately corresponding to the pump infusion of 50% glucose in water at a maximal rate of 100 ml in a subject of small size (50 Kg). Thus $\mu(\cdot, \cdot)$ and $\sigma(\cdot)$ fulfill the usual Lipschitz condition and linear growth bound, so that model (2)-(5) has a unique t-continuous solution (see e.g. section 5.2 in Øksendal (2000)). Since $T_{gx}(\cdot)$ changes over time as an external forcing function, the distribution of X_t will depend on t and thus not be stationary.

When $T_{gx}(\cdot)$ is constant, the stationary distribution of $X_\infty := \lim_{t\to\infty} X_t$ is an Inverse Gamma distribution with shape parameter $1 + 2K_{xgI}/\sigma^2 I^*$ and scale parameter $2(T_{gx} + T_{ghnet})/V_g(\sigma I^*)^2$ (Bibby et al. (2005)) (if X is Inverse Gamma, then $1/X$ is Gamma). The asymptotic mean glycemia is $G^* = (T_{ghnet} + T_{gx}^*)/(K_{xgI}I^*V_g)$, where T_{gx}^* is the mean glucose infusion rate over the last hour of the experiment, see Table 1.

The EHC procedure attempts to reach steady-state with a constant blood glucose concentration and infusion rate. Thus, it is reasonable to assume that the process is approaching stationarity towards the end of the experiment, which can be used to determine one parameter from the others. Thus, from the G^* expression above, an estimate of T_{ghnet} is given by

$$T_{ghnet} = K_{xgI}I^*G^*V_g - T_{gx}^*. \tag{6}$$

7

However, this procedure is only valid if stationarity is approximately reached during the experiment, otherwise all parameters should be left unconstrained. We therefore performed two different estimations: one considers T_{ghnet} a free parameter, while the other determines T_{ghnet} from the stationarity conditions given in equation (6).

Measurement error is assumed negligible, i.e. this error is small compared to the magnitude of the system noise.

2.4 Parameter estimation

Denote the vector of unknown parameters with $\theta = (K_{xgI}, T_{ghnet}, V_g, \sigma)$, which we want to estimate. We assume $\tau_g = 1$ as motivated in Section 2.3. For ease of notation we drop the reference to θ when not necessary, that is, we write $f(x)$ instead of $f(x, \theta)$ for a given function f. Suppose that $n + 1$ glycemia observations $x_0, x_1, ..., x_n$ generated from model (2)-(5) are available for a single subject at non-stochastic time-points $40 = t_0 < t_1 < \cdots < t_n$. To compute the log-likelihood (1), the transition density p_X of X can be approximated in closed-form by a Hermite expansion up to a order K (Egorov et al. (2003)). We choose $K = 2$, which has been shown often to be sufficient (Aït-Sahalia (2002b), Jensen and Poulsen (2002), Egorov et al. (2003)). The order $K = 2$ transition density approximation is given by

$$p_X(x, t | x_s, s) \approx p_X^{(2)}(x, t | x_s, s) = \frac{1}{\sigma(x)\Delta^{1/2}} p_Z^{(2)}\left(\frac{\gamma(x, s + \Delta) - y_s}{\Delta^{1/2}}, s + \Delta | y_s, s\right), \quad s < t \quad (7)$$

where $\Delta = t - s$ (and in our case Δ is constantly equal to 5 min), y_s is the value of the transformed process Y at time s, and Y is defined by $Y_t = \gamma(X, t)$, where

$$\gamma(x, t) = \int^x \frac{du}{\sigma(u, t)}$$

and the lower bound of integration is an arbitrary point in the interior of the state-space of X (i.e. the constant of integration is irrelevant). By Itô's lemma, Y is the solution to the following SDE

$$dY_t = \mu_Y(Y_t, t)dt + dW_t,$$

where

$$\mu_Y(y, t) = \frac{\mu(\gamma^{-1}(y, t), t)}{\sigma(\gamma^{-1}(y, t))} + \frac{\partial \gamma}{\partial t}(\gamma^{-1}(y, t), t) - \frac{1}{2}\frac{\partial \sigma}{\partial x}(\gamma^{-1}(y, t)).$$

Define now a further transformation $Z_t = (Y_t - y_s)/\Delta^{1/2}$ whose transition density p_Z can be approximated in closed form as (Egorov et al. (2003))

$$p_Z^{(2)}(z, t|y_s, s) \;\;=\;\; \phi(z) \sum_{k=0}^{4} \beta_k^{(2)}(t, y_s, s) H_k(z) \tag{8}$$

where $\phi(\cdot)$ is the standard normal pdf and the H_k's are the Hermite polynomials given by
$$H_0(z) = 1, \quad H_1(z) = -z, \quad H_2(z) = z^2 - 1, \quad H_3(z) = -z^3 + 3z, \quad H_4(z) = z^4 - 6z^2 + 3.$$
The coefficients $\beta_k^{(2)}$ for model (2)-(5) are given in appendix.

For model (2)-(5) we have

$$Y_t \equiv \gamma(X_t, t) = \frac{1}{\sigma I^*} \ln(X_t) \quad \Rightarrow \quad X_t \equiv \gamma^{-1}(Y_t, t) = e^{\sigma I^* Y_t}$$

thus
$$\mu_Y(Y_t, t) = \frac{T_{gx}(t - \tau_g) + T_{ghnet}}{\sigma V_g I^* \exp(\sigma I^* Y_t)} - \frac{K_{xgI}}{\sigma} - \frac{\sigma I^*}{2}.$$

Once the $\beta_k^{(2)}$'s are obtained, $p_Z^{(2)}$ and $p_X^{(2)}$ can be computed using equations (7)-(8). Thus, we approximate the log-likelihood $l_n(\theta)$ with its order $K = 2$ expansion

$$l_n^{(2)}(\theta) = \sum_{i=1}^{n} \ln p_X^{(2)}(x_i, t_i | x_{i-1}, t_{i-1}), \tag{9}$$

and then $l_n^{(2)}$ can be maximized w.r.t. θ in order to obtain an (approximated) maximum likelihood estimate $\hat{\theta}^{(2)}$ of θ, which is consistent under mild regularity conditions (Egorov et al. (2003)).

The maximization of (9) is fast since, unlike most of the other available estimation procedures, simulation of many sample paths from the X process is not required to approximate p_X (see also the Discussion).

3   Results

The parameter T_{ghnet} was either determined from (6), or it was estimated as a free parameter. In the latter case the estimates attained physiologically unacceptable values: this may be due to overparametrization phenomenon, and the corresponding results are therefore not reported.

The results for T_{ghnet} determined by (6) are reported in Table 4. For each subject and for each parameter Table 4 reports individual estimates and the corresponding 95% confidence intervals; see the appendix for details. All parameters were well identified. For one subject the parameter estimates resulted physiologically unacceptable and were thus marked with an 'NA'. Notice that the diffusion coefficient estimates ($\hat{\sigma}$) are all much larger than zero (a $\hat{\sigma} \simeq 0$ would indicate that a deterministic model is sufficient to describe the time-course of glycemia in the given subject). Therefore the dynamical process which best represents the observations is a stochastic process with non-negligible system noise. Pictorial evidence of the diffusion magnitude is given in Figure 1, as described below.

Figure 1 reports, for five subjects, the observed glycemias, the empirical mean of 2000 simulated trajectories of the SDE (2) generated with the Milstein scheme with a stepsize of 0.1 min (Kloeden and Platen (1992)) - which converges to $\mathbb{E}(X_t|X_0)$ when the number of trajectories goes to infinity - as well as one simulated trajectory and the empirical 95% confidence bands of trajectory values, for every simulation time between 40 min and the last measured glycemia value. Moreover, for each simulated glycemia trajectory, the corresponding glucose infusion rates curve was simulated according to the algorithm in DeFronzo et al. (1979) from time $t_0 = 40$ onward: that is, the clamp procedure was virtually performed for each glycemia trajectory. Figure 2 reports the plot of the estimates of K_{xgI} vs σ: a positive correlation is assessed ($r = 0.89$); the slope of the linear regression fit equals 0.112 ($p-value < 0.001$). This significant correlation between tissue glucose-uptake rate (K_{xgI}) and system noise coefficient (σ) may indicate the incorporation of the effects of several physiological mechanisms increasing tissue glucose uptake as glycemia peaks become more frequent in the K_{xgI} coefficient value (in the context of a model with fixed insulinemia).

In order to assess the appropriateness of an order $K = 2$ for the loglikelihood approximation, given a constant Δ of 5 min and the available number of glycemia points, a Monte Carlo study was performed. For a given subject 1000 datasets of glycemia observations and glucose infusion rates were simulated using the parameter estimates given in Table 4, as explained below: each set of glycemias was created by simulating one trajectory from model (2)-(5) using the Milstein scheme, and extracting the simulated glycemias by linear interpolation at the time points $\{t_i\}$ corresponding to the actual measurements. The generated trajectory was "controlled" by a simulated clamp procedure, in the same way described above to produce

Figure 1. The parameter estimation procedure was performed for each simulated dataset, thus obtaining 1000 new estimates for the parameters of the considered subject. The Monte Carlo procedure was performed on a total of five subjects (those corresponding to the most physiologically reasonable estimates) and, for each parameter, Table 5 reports the sample mean of the obtained estimates and the approximate 95% confidence interval (from the 2.5^{th} to the 97.5^{th} percentile). The original estimates (Table 4) fall well within the 95% confidence intervals of the corresponding estimates in Table 5, even though the original estimates result somewhat different from the corresponding Monte Carlo mean values, and the relative 95% confidence intervals are rather wide. In fact, the empirical distributions of the 1000 estimates often appears skewed and some outliers bias the results. In order to protect against those outliers, Table 5 also reports the medians and the 1st-3rd quartiles of the estimates, and it can be appreciated how the medians appear closer to the original estimates, and how skewed the distributions of estimates are. In particular the K_{xgI} parameter, which is very important for our study since it represents an index of "insulin sensitivity", results to be well identified. Thus, a $K = 2$ order of approximation should provide sufficiently accurate results, even though they may also be improved using a smaller Δ, see Stramer and Yan (2007).

4   Discussion

Stochastic differential models allow for a general representation of the system's variability structure, by considering the action of dynamical random terms (stochastic noise) perturbing the system state at each instant. SDE models are also attractive because they represent a generalization of their deterministic counterparts (ODE, PDE, DDE), since when setting the stochastic noise equal to zero (i.e. $\sigma \equiv 0$ in our case) the SDE reduces to a deterministic differential model, for which estimation methods are well-known.

SDE models have been extensively used in the last decades in e.g. investment finance, turbulent diffusion, population dynamics, polymer dynamics, biological waste treatment, neuronal models and hydrology. The metabolism of glucose and insulin has not yet received attention from SDE modellers, except from few recent contributions (Tornøe, Jacobsen, Pedersen, Hansen and Madsen (2004), Tornøe, Jacobsen and Madsen (2004), Andersen and Højbjerre (2005), Picchini et al. (2006)).

However, in general SDE models are not as often implemented as their deterministic coun-

11

terparts, essentially because fast estimation methods are lacking. In fact, there exist estimation tools for general SDE models, but they are computationally expensive. The difficulty when using maximum likelihood-based estimation approaches for SDE models is that the transition density function of the underlying stochastic process is often unknown. One approach is to compute an approximation to the transition density function, e.g.: (i) solving numerically the Kolmogorov partial differential equations satisfied by the transition density (Lo (1988)); (ii) deriving a closed-form Hermite expansion to the transition density (Aït-Sahalia (2001, 2002b), Egorov et al. (2003)); (iii) simulating many times the process to Monte-Carlo integrate the transition density (e.g. Pedersen (1995), Hurn and Lindsay (1999), Brandt and Santa-Clara (2002), Durham and Gallant (2002), Nicolau (2002)): this methodology is known as simulated maximum likelihood (SML). Recently a novel method using exact simulation was proposed (Beskos et al. (2006)).

Each of these techniques have been successfully implemented by the aforementioned authors, but each has their limitations. Aït-Sahalia (2002a) notes that methods (i) and (iii) above are computationally intense and poorly accurate. Conversely Durham and Gallant (2002) build on their importance sampling ideas in order to improve the performance of Pedersen's (1995) (or equivalently Brandt and Santa-Clara's (2002)) method and point out that method (ii) above, while general, accurate and fast, can be actually applied to a small number of multidimensional models. In fact, the coefficients to the transition density expansion employed in method (ii) are difficult to obtain for multidimensional nonlinear irreducible models (that is for processes having non-commutative noise, see e.g. Kloeden and Platen (1992)). Therefore, instead of considering an already developed deterministic model for glycemia and insulinemia observations (Picchini et al. (2005)) to accomodate stochastic-noise, we reverted to a simpler model, considering only glycemia observations in near steady state conditions (see Picchini et al. (2006) for a different stochastic approach). Thus we were able to apply the extension of the Aït-Sahalia's method for time-inhomogeneous SDEs proposed by Egorov et al. (2003) in order to estimate all parameters simultaneously: in fact, once the coefficients of the transition density expansion have been determined to the desired order, the loglikelihood function can be expressed in closed form as an expression depending on the glycemia observations, the glucose infusion rates, the unknown parameters and time. In this way the simulation of a large number of process trajectories is not required. The likelihood function can then

be optimized rapidly (few seconds are required on a 3.0 GHz Intel Pentium IV with 512 MB of RAM), and it has been proved that this method leads to consistent estimates of the true parameter values (Egorov et al. (2003)). In fact, in a previous work (Picchini et al. (2006)) we were unable to estimate simultaneously all the parameters of a two-dimensional SDE, because we were using a "simulated maximum likelihood" approach, which was computationally demanding.

In this paper we adopted the procedure described above to estimate the parameters of a SDE model of the glycemia dynamics observed during a Euglycemic Hyperinsulinemic Clamp procedure (EHC), performed on human subjects. The EHC is widely considered the tool of choice for the assessment of insulin sensitivity, in spite of its labor-intensive execution, due to the simple interpretation which is usually attributed to the obtained results (DeFronzo et al. (1979), Zierler (1999)). The favor with which the EHC is viewed in this context makes it so that many databases of clamp results have been built in recent years by several diabetological research groups. The present work goes in the direction of trying to enhance the value of this large collection of experimental data, improving, by means of a suitable mathematical model, on the rather elementary determination of insulin sensitivity from clamp data, typically used by clinical researchers (DeFronzo et al. (1979)).

In conclusion, the present work considers a method which, while easily applicable only to models of moderate complexity, is however fast, robust and can introduce the demonstrably essential concept of system noise in this branch of physiological modeling. By comparison with other currently used methods (e.g. Bibby and Sørensen (1995), Pedersen (1995), Durham and Gallant (2002)), using an Aït-Sahalia/Egorov *et al.* expansion requires previous explicit calculation of the expansion coefficients: this can be often done without problems using any one of several symbolic calculus programs. Once this step is completed for a given model, the parameter estimation process takes few seconds even on a normal, single PC, without making recourse to clusters or expensive computers.

Appendix

Density expansion coefficients

Here we report the expressions for the parameters $\beta_k^{(2)}$ appearing in equation (8). For model (2)-(5) we have

$$
\begin{aligned}
\beta_0^{(2)}(t, y_s, s) &= 1 \\
\beta_1^{(2)}(t, y_s, s) &= -\Delta^{1/2}\psi - \frac{\Delta^{3/2}}{4}(2\psi_{01} + 2\psi\psi_{10} + \psi_{20}) \\
\beta_2^{(2)}(t, y_s, s) &= \frac{\Delta}{2}(\psi^2 + \psi_{10}) + \frac{\Delta^2}{12}(6\psi\psi_{01} + 6\psi^2\psi_{10} + 4\psi_{10}^2 + 4\psi_{11} + 7\psi\psi_{20} + 2\psi_{30}) \\
\beta_3^{(2)}(t, y_s, s) &= -\frac{\Delta^{3/2}}{6}(\psi^3 + 3\psi\psi_{10} + \psi_{20}) \\
\beta_4^{(2)}(t, y_s, s) &= \frac{\Delta^2}{24}(\psi^4 + 6\psi^2\psi_{10} + 3\psi_{10}^2 + 4\psi\psi_{20} + \psi_{30})
\end{aligned}
$$

where

$$
\begin{aligned}
\psi &= \mu_Y(y_s, s) = \frac{T_{gx}(s - \tau_g) + T_{ghnet}}{\sigma V_g I^* \exp(\sigma I^* y_s)} - \frac{K_{xgI}}{\sigma} - \frac{\sigma I^*}{2} \\
\psi_{01} &= \frac{\partial \mu_Y(y, s)}{\partial s}\Big|_{y=y_s} = \frac{1}{V_g \sigma I^* \exp(\sigma I^* y_s)} \sum_{j;\nu_j \leq (s - \tau_g)} \frac{5(\lambda_j - \lambda_{j-1})(s - \tau_g - \nu_j)^4 \nu_j}{(\nu_j + (s - \tau_g - \nu_j)^5)^2} \\
\psi_{10} &= \frac{\partial \mu_Y(y, s)}{\partial y}\Big|_{y=y_s} = -\frac{T_{gx}(s - \tau_g) + T_{ghnet}}{V_g \exp(\sigma I^* y_s)} \\
\psi_{11} &= \frac{\partial^2 \mu_Y(y, s)}{\partial y \partial s}\Big|_{y=y_s} = -\frac{1}{V_g \exp(\sigma I^* y_s)} \sum_{j;\nu_j \leq (s - \tau_g)} \frac{5(\lambda_j - \lambda_{j-1})(s - \tau_g - \nu_j)^4 \nu_j}{(\nu_j + (s - \tau_g - \nu_j)^5)^2} \\
\psi_{20} &= \frac{\partial^2 \mu_Y(y, s)}{\partial y^2}\Big|_{y=y_s} = \frac{(T_{gx}(s - \tau_g) + T_{ghnet})\sigma I^*}{V_g \exp(\sigma I^* y_s)} \\
\psi_{30} &= \frac{\partial^3 \mu_Y(y, s)}{\partial y^3}\Big|_{y=y_s} = -\frac{(T_{gx}(s - \tau_g) + T_{ghnet})(\sigma I^*)^2}{V_g \exp(\sigma I^* y_s)}
\end{aligned}
$$

and the index j in the summations spans the glucose infusion rates recorded up to time $s - \tau_g$.

Approximated parameter confidence intervals

Denote with θ the vector of the *free* unknown parameters. Under mild regularity conditions $\hat{\theta}^{(2)}$ is a consistent estimator of θ (Egorov et al. (2003)), and an approximated confidence region for θ can be obtained from the following:

$$
i_n^{1/2}(\theta)(\hat{\theta}^{(2)} - \theta) \approx G_n^{-1/2}(\theta) \times N(0_k, I_k) \tag{10}
$$

where $N(\cdot,\cdot)$ denotes the normal distribution, k is the dimension of θ, 0_k and I_k are the k-dimensional array of zeros and the $k \times k$ identity matrix, respectively, $G_n(\theta) = i_n^{-1/2}(\theta)H_n(\theta)i_n^{-1/2}(\theta)$ is a $(k \times k)$ matrix, $H_n(\theta) = -\sum_{i=1}^n l_i''(\theta)$ and $i_n(\theta) = \text{diag}\sum_{i=1}^n \mathbb{E}_\theta[l_i'(\theta)l_i'(\theta)^T]$ (the *expected Fisher information*). Here $l_i(\theta) = \ln p_X^{(2)}(x_i, t_i | x_{i-1}, t_{i-1})$, and the superscripts $'$ and T denote differentiation with respect to θ and transposition. We obtained $l_i''(\theta)$ via a symbolic calculus program (the expressions for the second partial derivatives of $l_i(\cdot)$ are not reported since they are considerably lengthy), and $\mathbb{E}_\theta[l_i'(\theta)l_i'(\theta)^T]$ was substituted with $-l_i''(\theta)$, so we considered the *observed Fisher information* in place of the expected Fisher information, since it often makes little difference numerically (e.g. Barndorff-Nielsen and Sørensen (1994), p. 133). Therefore, using these settings we have that $G_n(\theta)$ equals I_k and (10) becomes

$$\left(-\text{diag}\sum_{i=1}^n l_i''(\theta)\right)^{1/2}(\hat{\theta}^{(2)} - \theta) \approx N(0_k, I_k) \tag{11}$$

and substituting $l_i''(\theta)$ with $l_i''(\hat{\theta}^{(2)})$ we have that, approximately,

$$\left(-\sum_{i=1}^n l_i''(\hat{\theta}_j^{(2)})\right)^{1/2}(\hat{\theta}_j^{(2)} - \theta_j) \approx T_{n-p} \tag{12}$$

where T_g is the Student distribution with g degrees of freedom and θ_j is the jth element in the p-dimensional vector of free parameters θ (thus, $p = 3$ when T_{ghnet} is determined from the other parameters and $p = 4$ otherwise). Then we used (12) to derive the approximated 95% confidence intervals in Table 4.

References

Aït-Sahalia, Y. (2001), Closed-form likelihood expansion for multivariate diffusions, Technical report, Princeton University. Forthcoming in *Ann. Stat.*

Aït-Sahalia, Y. (2002*a*), 'Comment on G. Durham and A. Gallant "Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes"', *J. Bus. Econom. Statist.* **20**, 317–321.

Aït-Sahalia, Y. (2002*b*), 'Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach', *Econometrica* **70**(1), 223–262.

Aït-Sahalia, Y. (2007), Estimating continuous-time models with discretely sampled data, *in* R. Blundell, W. Newey and T. Persson, eds, 'Advances In Economics And Econometrics, Theory And Applications, Ninth World Congress', Econometric Society Monographs, Cambridge University Press.

Andersen, K. and Højbjerre, M. (2005), 'A population-based bayesian approach to the minimal model of glucose and insulin homeostasis', *Stat. Med.* **24**(15), 2381–2400.

Barndorff-Nielsen, O. and Sørensen, M. (1994), 'A review of some aspects of asymptotic likelihood theory for stochastic processes', *Int. Stat. Rev.* **62**(1), 133–165.

Beskos, A., Papaspiliopoulos, O., Roberts, G. and Fearnhead, P. (2006), 'Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes', *J. Roy. Statist. Soc. B* **68**, 333–382.

Bibby, B., Jacobsen, M. and Sørensen, M. (2004), Estimating functions for discretely sampled diffusion-type models. `http://home.uchicago.edu/~lhansen/estfct3.pdf`.

Bibby, B., Skovgaard, I. and Sørensen, M. (2005), 'Diffusion-type models with given marginal distribution and autocorrelation function', *Bernoulli* **11**(2), 191–220.

Bibby, B. and Sørensen, M. (1995), 'Martingale estimation functions for discretely observed diffusion processes', *Bernoulli* **1**(1/2), 17–39.

Brandt, M. and Santa-Clara, P. (2002), 'Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets', *J. Financ. Econ.* **63**, 161–210.

Dacunha-Castelle, D. and Florens-Zmirnou, D. (1986), 'Estimation of the coefficients of a diffusion from discrete observations', *Stochastics* **19**, 263–284.

DeFronzo, R., Tobin, J. and Andres, R. (1979), 'Glucose clamp technique: a method for quantifying insulin secretion and resistance', *Am. J. Physiol.* **237**, E214–E223.

Ditlevsen, S. and De Gaetano, A. (2005), 'Mixed effects in stochastic differential equations models', *REVSTAT - Statistical Journal* **3**(2), 137–153.

Durham, G. and Gallant, A. (2002), 'Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes', *J. Bus. Econ. Stat.* **20**(3), 297–316.

Egorov, A., Li, H. and Xu, Y. (2003), 'Maximum likelihood estimation of time-inhomogeneous diffusions', *J. Econometrics* **114**, 107–139.

Elerian, O., Shephard, N. and Chib, S. (2001), 'Likelihood inference for discretely observed nonlinear diffusions', *Econometrica* **69**, 959–993.

Gallant, A. and Long, J. (1997), 'Estimating stochastic differential equations efficiently by minimum chi-square', *Biometrika* **84**, 124–141.

Gouriéroux, C., Monfort, A. and Renault, E. (1993), 'Indirect inference', *J. Appl. Econom.* **8**, 85–118.

Hurn, A. and Lindsay, K. (1999), 'Estimating the parameters of stochastic differential equations', *Math. Comput. Simulat.* **48**, 373–384.

Jensen, B. and Poulsen, R. (2002), 'Transition densities of diffusion processes: numerical comparison of approximation techniques', *Journal of Derivatives* **9**, 1–15.

Kloeden, P. and Platen, E. (1992), *Numerical Solution Of Stochastic Differential Equations*, Springer.

Kutoyants, Y. (2004), *Statistical Inference For Ergodic Diffusion Processes*, Springer.

Lo, A. (1988), 'Maximum likelihood estimation of generalized Ito processes with discretely-sample data', *Economet. Theor.* **4**, 231–247.

Nicolau, J. (2002), 'A new technique for simulating the likelihood of stochastic differential equations', *Economet. J.* **5**, 91–103.

Øksendal, B. (2000), *Stochastic Differential Equations: An Introduction With Applications*, Springer.

Pedersen, A. (1995), 'A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations', *Scand. J. Statist.* **22**(1), 55–71.

Pedersen, A. (2001), Likelihood inference by Monte Carlo methods for incompletely discretely observed diffusion processes, Technical report, Department of Biostatistics, University of Aarhus, Denmark.

Picchini, U., De Gaetano, A., Panunzi, S., Ditlevsen, S. and Mingrone, G. (2005), 'A mathematical model of the euglycemic hyperinsulinemic clamp', *Theor. Biol. Med. Model.* **2**(44).

Picchini, U., Ditlevsen, S. and De Gaetano, A. (2006), 'Modeling the euglycemic hyperinsulinemic clamp by stochastic differential equations', *J. Math. Biol.* **53**(5), 771–796.

Prakasa Rao, B. (1999), *Statistical Inference For Diffusion Type Processes*, Arnold, London and Oxford University press.

Shoji, I. and Ozaki, T. (1998), 'Estimation for nonlinear stochastic differential equations by a local linearization method', *Stoch. Anal. Appl.* **16**, 733–752.

Sørensen, H. (2004), 'Parametric inference for diffusion processes observed at discrete points in time: a survey', *Int. Stat. Rev.* **72**(3), 337–354.

Sørensen, M. (2000), 'Prediction-based estimating functions', *Economet. J.* **3**, 123–147.

Stramer, O. and Yan, J. (2007), 'On simulated likelihood of discretely observed diffusion processes and comparison to closed-form approximation', *J. Comput. Graph. Stat.* **16**(3), 672–691.

Tornøe, C., Jacobsen, J. and Madsen, H. (2004), 'Grey-box pharmacokinetic/pharmacodynamic modelling of a euglycaemic clamp study', *J. Math. Biol.* **48**, 591–604.

Tornøe, C., Jacobsen, J., Pedersen, O., Hansen, T. and Madsen, H. (2004), 'Grey-box modelling of pharmacokinetic/pharmacodynamic systems', *J. Pharmacokinet. Phar.* **31**(5), 401–417.

Zierler, K. (1999), 'Whole body glucose metabolism', *Am. J. Physiol.* **276**, E409–E426.

Table 1: Metabolic characteristics for the subjects: T_{gx}^* and G^* are the mean glucose infusion rate and the mean glicemia over the last 60 min of the clamp procedure; I^* is the mean insulinemia over the time-interval 40-300 min.

Subject	G^* [mM]	I^* [pM]	T_{gx}^* [mmol/min/kgBW]
1	3.68	473.00	0.062
2	4.76	612.00	0.027
3	5.28	688.48	0.054
4	5.09	630.82	0.028
5	20.03	507.35	0.029
6	19.33	467.39	0.033
7	5.33	596.60	0.026
8	18.51	531.08	0.051
9	5.79	526.29	0.032
10	5.37	489.03	0.055
11	5.35	501.33	0.015
12	22.59	472.70	0.059
13	4.81	609.91	0.027
14	4.53	684.88	0.057
15	3.67	482.83	0.038

Variables	
$t\ [min]$	time from insulin infusion start
$X_t\ [mM]$	plasma glucose concentration at time t
$T_{gx}(t)\ [mmol/min/kgBW]$	glucose infusion rate at time t

Table 2: Definitions of the variables.

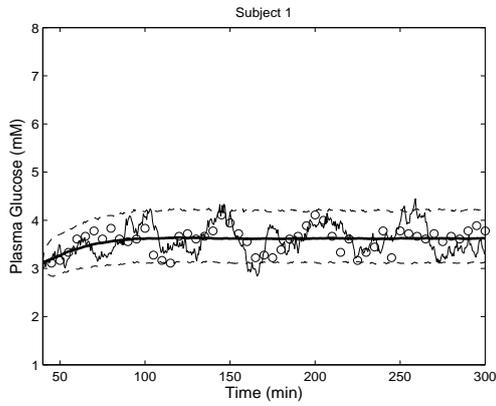| Parameters	
$K_{xgI}\ [min^{-1}/pM]$	insulin-dependent apparent first-order rate constant for glucose tissue uptake at insulinemia I^*
$T_{ghnet}\ [mmol/min/kgBW]$	zero-order net Hepatic Glucose Output
$V_g\ [L/kgBW]$	volume of distribution for glucose
$\tau_g\ [min]$	discrete (distributional) delay of the change in glycemia following glucose infusion
$\sigma\ [pM]^{-1}[min]^{-1/2}$	diffusion coefficient

Table 3: Definitions of the parameters.

Table 4: Individual parameter estimates for the SDE model (2)-(5), when T_{ghnet} is determined, and 95% confidence intervals (only for the free parameters).
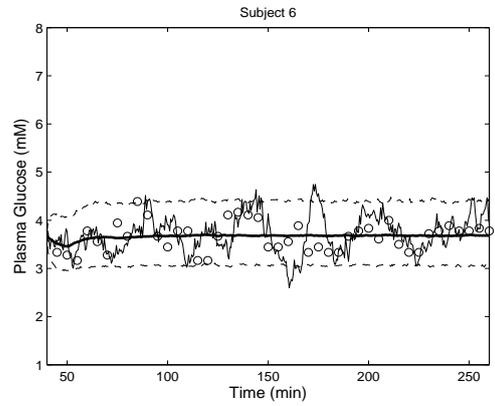
Subject	$\hat{K}_{xg1} \times 10^4$	$\hat{T}_{ghnet}$	$\hat{V}_g$	$\hat{\sigma} \times 10^5$
1	2.951 [2.874, 3.028]	0.175	0.463 [0.451, 0.474]	7.542 [7.395, 7.688]
2	31.259 [31.095, 31.423]	3.599	0.398 [0.396, 0.400]	30.560 [30.463, 30.657]
3	25.775 [25.673, 25.876]	9.218	0.99 [0.986, 0.994]	31.684 [31.603, 31.765]
4	32.198 [31.858, 32.539]	0.269	0.029 [0.028, 0.029]	61.448 [61.098, 61.799]
5	1.373 [1.248, 1.498]	0.329	0.99 [0.922, 1.058]	9.645 [9.262, 10.027]
6	5.782 [5.658, 5.906]	0.085	0.117 [0.115, 0.120]	12.213 [12.023, 12.402]
7	5.553 [5.519, 5.587]	0.346	0.211 [0.210, 0.213]	5.878 [5.856, 5.900]
8	NA	NA	NA	NA
9	3.594 [3.472, 3.714]	1.051	0.99 [0.952, 1.028]	13.891 [13.588, 14.194]
10	2.531 [2.455, 2.607]	0.192	0.371 [0.359, 0.383]	6.717 [6.585, 6.849]
11	11.298 [11.185, 11.410]	0.253	0.089 [0.088, 0.090]	13.679 [13.600, 13.758]
12	1.890 [1.815, 1.965]	0.247	0.99 [0.952, 1.028]	4.978 [4.828, 5.128]
13	12.430 [12.335, 12.525]	3.582	0.99 [0.982, 0.997]	16.821 [16.742, 16.901]
14	23.383 [23.221, 23.545]	1.093	0.158 [0.157, 0.159]	23.878 [23.771, 23.985]
15	1.248 [1.189, 1.307]	0.114	0.688 [0.659, 0.718]	5.088 [4.927, 5.249]

Table 5: Monte Carlo study for the parameters of the SDE model (2)-(5) over 1000 simulations. For each subject the first row reports the sample means and the approximate 95% confidence intervals; the second row reports the medians and the $1^{st} - 3^{rd}$ quartiles and, for ease of comparison, the third row reports the true parameter values (see also Table 4). See main text for details.
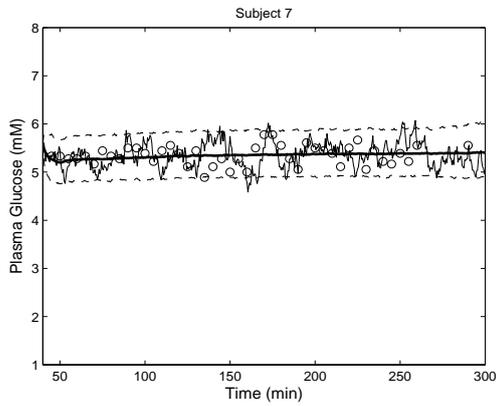
| Subject	$\hat{K}_{xgI} \times 10^4$	$\hat{T}_{ghnet}$	$\hat{V}_g$	$\hat{\sigma} \times 10^5$
1	3.524 [1.392, 10.940]	0.196 [0.059, 0.462]	0.539 [0.118, 1.000]	8.551 [5.801, 14.851]
	2.772 [2.194, 3.559]	0.168 [0.118, 0.249]	0.467 [0.340, 0.726]	7.945 [7.209, 8.857]
	2.951	0.175	0.463	7.542
6	12.171 [2.330, 50.00]	0.643 [0, 7.941]	0.428 [0.019, 1.000]	24.064 [8.266, 100.000]
	5.514 [3.843, 11.649]	0.240 [0.112, 0.557]	0.282 [0.108, 0.871]	14.822 [12.299, 21.200]
	5.782	0.085	0.117	12.213
7	18.235 [1.878, 50.00]	1.739 [0.125, 15.623]	0.373 [0.019, 1.000]	17.695 [4.032, 57.299]
	8.885 [3.689, 30.632]	0.462 [0.264, 0.693]	0.204 [0.076, 0.649]	8.689 [5.972, 22.819]
	5.553	0.346	0.211	5.878
9	7.200 [1.370, 50.00]	1.271 [0.147, 14.406]	0.734 [0.079, 1.000]	22.311 [10.196, 118.88]
	3.195 [2.482, 4.815]	0.689 [0.428, 0.975]	0.99 [0.443, 1.000]	14.752 [12.780, 17.964]
	3.594	1.051	0.99	13.891
10	2.722 [1.258, 5.512]	0.252 [0.056, 0.690]	0.477 [0.147, 1.000]	7.296 [5.298, 10.301]
	2.433 [1.969, 3.022]	0.203 [0.134, 0.317]	0.400 [0.286, 0.600]	7.075 [6.477, 7.822]
	2.531	0.192	0.371	6.717

(a) Subject 1

(b) Subject 6

(c) Subject 7
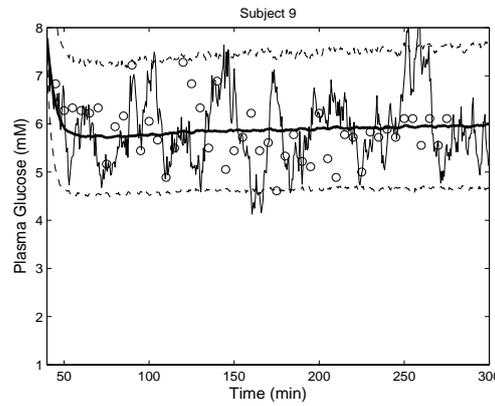
(d) Subject 9
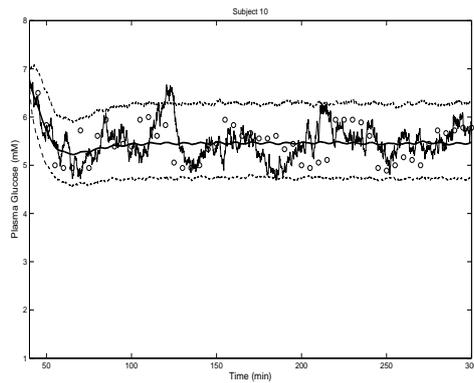
(e) Subject 10

Figure 1: Observed glycemias (∘), empirical mean curves of the stochastic process defined by (2)-(5) (bold solid lines), 95% empirical confidence curves (dashed lines) and one simulated trajectory, when T_{ghnet} is determined.
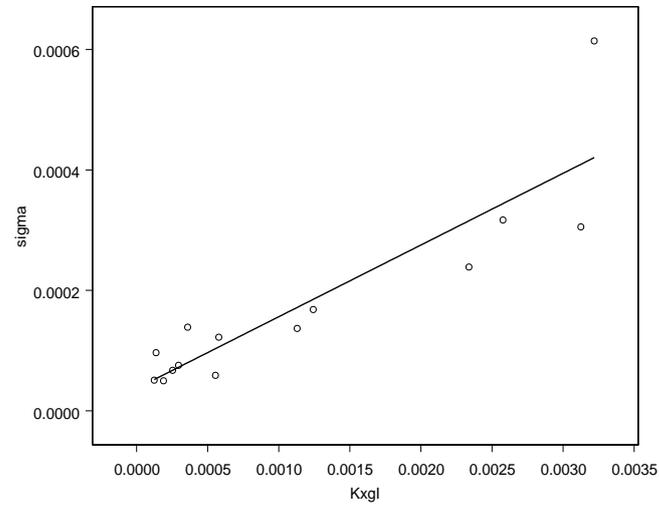
24

Figure 2: Plot of $\hat{\sigma}$ vs $\hat{K}_{xgI}$ and linear regression fit. Correlation coefficient $r = 0.89$; linear regression slope $\beta = 0.112$ ($p - value < 0.001$).