REVSTAT – Statistical Journal
Volume 3, Number 2, November 2005, 137–153

MIXED EFFECTS IN STOCHASTIC DIFFERENTIAL EQUATION MODELS

Authors: Susanne Ditlevsen
– Department of Biostatistics, University of Copenhagen, Denmark
(sudi@pubhealth.ku.dk)

Andrea De Gaetano
– CNR IASI, Laboratorio di Biomatematica, Universita Cattolica del Sacro Cuore, Italy
(andrea.degaetano@biomatematica.it)

Received: January 2005 Revised: May 2005 Accepted: August 2005

Abstract:

- A class of statistical models is proposed where random effects are incorporated into a stochastic differential equations model, and an expression for the likelihood function is derived. In general, though, it is not possible to find an explicit expression for the likelihood function, but in a very simple example it is derived and explicit maximum likelihood estimators are found. The estimators are evaluated in a simulation study, and illustrated on dissolution data of metoprolol tartrate tablets.

Key-Words:

- *maximum likelihood; pharmacokinetics; population estimates; random effects; repeated measurements; stochastic processes.*

AMS Subject Classification:

- 60H10, 62M99.

1. INTRODUCTION

In biomedical research, studies in which repeated measurements are taken on a series of individuals or experimental animals play an important role. Models including random effects to model this kind of data enjoy an increasing popularity. In these models it is assumed that all responses follow a similar functional form, but with parameters that vary among individuals. The increasing popularity of mixed-effects models lies in the flexible modeling of correlation structures, where the total variation is specifically split in within-group and between-group variation. This will often lead to more precise estimation of population parameters. Especially in pharmacokinetic/pharmacodynamic (PK/PD) modeling most studies include random effects in the models, thereby improving population parameter estimation.

Continuous biological processes are often described by systems of ordinary differential equations (ODE), which unfortunately cannot account for noisy components often present in biological systems, representing the parts of the dynamics that we cannot predict or understand, or that we choose not to include in the explicit modeling. A natural extension is given by systems of stochastic differential equations (SDE), where system noise is modeled by including a diffusion term of some suitable form in the driving equations. In PK/PD modeling the focus is most often on the infinitesimal changes of substances, which naturally leads to a ODE-system. The inter-individual variability is modeled with the random effect, and the intra-individual variability with an additive noise term (possibly after some convenient transformation). However, noise in the differential equations describing the behavior of the system requires an extension of the model class to SDE models.

The theory for mixed-effects models is well developed for deterministic models (without system error), both linear and non-linear ([2, 3, 14, 25]), and standard software for model fitting is available, see e.g. ([18]) and references therein. Early and important references in the pharmacokinetic field are ([21, 22]). Estimating parameters in SDE models is not straightforward, except for simple cases. A natural approach would be likelihood inference, but the transition densities are rarely known, and thus it is usually not possible to write the likelihood function explicitly. A variety of methods for statistical inference in discretely observed diffusion processes has been developed during the past decades, see e.g. ([1, 4, 5, 6, 7, 9, 10, 13, 16, 17, 20, 23, 24]). However, to our knowledge there is practically no theory at present for SDE models with random effects. In ([15]) it is suggested to apply the Kalman filter to approximate the likelihood function for a SDE model with random effects, with a non-linear drift term and a constant diffusion term. Eventually, as SDE models will be more commonly applied to biomedical data, there will be an increasing need for developing a theory including mixed effects, and for results on the estimation of model parameters. In ([8]) methods for PK/PD population modeling are reviewed, but the authors regret

that system noise is not considered since it is difficult to estimate, and that there exists no software at present in the pharmacokinetic field.

In the present paper a class of statistical models is proposed where random effects are incorporated into a diffusion model, and an expression for the likelihood function is derived. In general, though, it is not possible to find an explicit expression for the likelihood function, but in a very simple example it is derived and explicit maximum likelihood estimators are found. The estimators are evaluated in a simulation study and illustrated on experimental data.

2. THE MODEL

Consider the one-dimensional SDE model for some continuous process evolving in M different subjects randomly chosen from a population:

$$\begin{aligned} dX_t^i &= g(X_t^i, \boldsymbol{\theta}, \mathbf{b}^i)\, dt + \sigma(X_t^i, \boldsymbol{\theta}, \mathbf{b}^i)\, dW_t^i \ ; \qquad i = 1, ..., M \\ \mathbf{b}^i &\sim N(0, \boldsymbol{\Sigma}) \\ X_0^i &= x_0^i \end{aligned} \tag{2.1}$$

where $\boldsymbol{\theta}$ is a p-dimensional *fixed effects* parameter (the same for the entire population) and $\mathbf{b}^i$ is a q-dimensional *random effects* parameter (subject specific), which is assumed to follow a normal distribution in the population, with covariance matrix $\boldsymbol{\Sigma}$ that is assumed known up to the parameter vector $\boldsymbol{\Psi}$. The W_t^i are standard Brownian motions. The W_t^i and $\mathbf{b}^j$ are assumed mutually independent for all $1 \le i, j \le M$, and independent of X_0^i. The drift and the diffusion coefficient functions $g(\cdot)$ and $\sigma(\cdot)$ are assumed known up to the parameters, and are assumed sufficiently regular to ensure a unique solution. Let $E \subseteq \mathbb{R}$ denote the state space of X_t^i. Assume that the distribution of X_t^i given $\mathbf{b}^i$ and $X_s^i = x, t > s$, has a strictly positive density w.r.t. the Lebesgue measure on E, which we denote by

$$y \mapsto p\big(y, x, t-s | \mathbf{b}^i, \boldsymbol{\theta}\big) > 0 \ , \qquad y \in E \ . \tag{2.2}$$

Assume the M subjects each are observed at the $(n_i + 1)$ discrete time points $(t_0^i, t_1^i, ..., t_{n_i}^i)$. Let $\mathbf{y}^i$ be the $(n_i + 1)$-dimensional response vector for the i'th subject: $\mathbf{y}^i = (y_0^i, ..., y_{n_i}^i)$, $y(t_j^i) = y_{t_j^i}^i = y_j^i$, and let $\mathbf{y}$ be the N-dimensional total response vector, $N = \sum_{i=1}^M (n_i+1)$. Write $t_j^i - t_{j-1}^i = \Delta_j^i$ for the distance between observation $j-1$ and j for subject i.

Parameters of the model are $\boldsymbol{\theta}$ and $\boldsymbol{\Psi}$, which we wish to estimate.

3. MAXIMUM LIKELIHOOD ESTIMATION IN SDE MIXED EFFECTS MODELS

To obtain the marginal density, we integrate the conditional density of the data given the non-observable random effects $\mathbf{b}^i$ with respect to the marginal density of the random effects, using the fact that W_t^i and $\mathbf{b}^i$ are independent. This yields the likelihood

$$L(\boldsymbol{\theta}, \boldsymbol{\Psi}|\mathbf{y}) \;=\; \prod_{i=1}^{M} p(\mathbf{y}^i|\boldsymbol{\theta}, \boldsymbol{\Psi}) \;=\; \prod_{i=1}^{M} \int p(\mathbf{y}^i|\mathbf{b}^i, \boldsymbol{\theta})\, p(\mathbf{b}^i|\boldsymbol{\Psi})\, d\mathbf{b}^i \tag{3.1}$$

where $L(\cdot)$ is the likelihood and $p(\cdot)$ are densities. Now

$$p(\mathbf{y}^i|\mathbf{b}^i, \boldsymbol{\theta}) \;=\; \prod_{j=1}^{n_i} p(y_j^i, y_{j-1}^i, \Delta_j^i|\mathbf{b}^i, \boldsymbol{\theta}) \tag{3.2}$$

since X_t^i given $\mathbf{b}^i$ is Markov, where the transition densities are as in (2.2), and, by hypothesis,

$$p(\mathbf{b}^i|\boldsymbol{\Psi}) \;=\; \frac{\exp\left\{-(\mathbf{b}^i)^T \boldsymbol{\Psi}^{-1} \mathbf{b}^i/2\right\}}{\sqrt{|\boldsymbol{\Psi}|}(2\pi)^{q/2}}\;, \tag{3.3}$$

where ^{T} denotes transposition. Substituting (3.2) and (3.3) into (3.1) we obtain

$$L(\boldsymbol{\theta}, \boldsymbol{\Psi}|\mathbf{y}) \;=\; \prod_{i=1}^{M} \int \prod_{j=1}^{n_i} p(y_j^i, y_{j-1}^i, \Delta_j^i|\mathbf{b}^i, \boldsymbol{\theta}) \frac{\exp\left\{-(\mathbf{b}^i)^T \boldsymbol{\Psi}^{-1} \mathbf{b}^i/2\right\}}{\sqrt{|\boldsymbol{\Psi}|}(2\pi)^{q/2}}\, d\mathbf{b}^i\;. \tag{3.4}$$

Solving the integral yields the marginal likelihood of the parameters, independent of the random effects $\mathbf{b}^i$. Note how it is straightforward to generalize to other distributions for the random effects by letting $p(\mathbf{b}^i|\boldsymbol{\Psi})$ be any distribution depending on the parameter $\boldsymbol{\Psi}$. In general it will not be possible to find an explicit solution, but in simple cases we can find an explicit expression for the likelihood, and even find explicit estimating equations for the maximum likelihood estimators.

3.1. A random effect in Brownian motion with drift

In the simplest pharmacokinetic situation, the metabolism of a compound is modeled as a mono-exponential decay in the following way (first-order kinetics):

$$\frac{dC(t)}{dt} \;=\; -kC(t)\;; \qquad C(0) = D/V \tag{3.5}$$

with solution

$$C(t) \;=\; C(0)\, e^{-kt}$$

where $C(t)$ is the concentration of the compound in plasma at time t after a bolus injection, k is the (positive) rate elimination constant, D is the injected dose at time $t = 0$, and V the apparent volume of distribution of the compound. Now assume that we want to model the erratic behavior of the metabolic processes responsible for the removal of the compound from plasma, by allowing k to vary randomly as $k+\xi(t)$, where $\xi(t)$ is a white noise process. Then $\xi(t)\,dt = \sigma\,dW(t)$ where $W(t)$ is Brownian motion and σ a scaling parameter. Incorporating this into (3.5), writing $X_t = C(t)$ and $\beta = -k$, we obtain the equation

$$dX_t \;=\; \beta X_t\,dt + \sigma X_t\,dW_t\;,$$

which is the equation of geometric Brownian motion. The state space E is given by the positive real line. By applying Itô's formula to the transformation: $Y_t = \log X_t$, we obtain a Brownian motion with linear drift:

$$dY_t \;=\; \Big(\beta - \frac{1}{2}\sigma^2\Big)\,dt + \sigma\,dW_t$$

with solution

$$Y_t \;=\; Y_0 + \Big(\beta - \frac{1}{2}\sigma^2\Big)t + \sigma W_t\;.$$

Assume an experiment is conducted on different subjects where the concentration of a compound in plasma is measured at different time points after a bolus injection. We are interested in estimating the parameters in the population, but expect individual differences in the metabolic processes, and would therefore consider a random effect in β, which leads to the model:

$$\begin{aligned} Y_t^i \;&=\; Y_0^i + \Big(\beta + \beta^i - \frac{1}{2}\sigma^2\Big)t + \sigma W_t^i \\ \beta^i \;&\sim\; N(0,\sigma_\beta^2)\;. \end{aligned} \tag{3.6}$$

Another example where this model naturally arises is provided by the initial growth of bacterial or tumor cell populations, where we expect $\beta > 0$.

In this simple example we have $\boldsymbol{\theta} = (\beta, \sigma^2)$ and $\boldsymbol{\Psi} = \sigma_\beta^2$. We wish to estimate $\boldsymbol{\zeta} = (\beta, \sigma^2, \sigma_\beta^2)$. The conditional distribution $(Y_t^i | Y_0^i = y_0^i; \beta, \sigma^2, \beta^i)$ is Gaussian with

$$\begin{aligned} E\big[Y_t^i | Y_0^i = y_0^i; \beta, \sigma^2, \beta^i\big] \;&=\; y_0^i + \Big(\beta + \beta^i - \frac{1}{2}\sigma^2\Big)t \\ \mathrm{Var}\big[Y_t^i | Y_0^i = y_0^i; \beta, \sigma^2, \beta^i\big] \;&=\; \sigma^2 t \end{aligned}$$

so the conditional transition density is given by

$$p(y_j^i, y_{j-1}^i, \Delta_j^i; \beta, \sigma^2, \beta^i) = \frac{1}{\sqrt{2\pi\sigma^2\Delta_j^i}} \exp\left\{ -\frac{\Big(y_j^i - y_{j-1}^i - \Big(\beta + \beta^i - \frac{1}{2}\sigma^2\Big)\Delta_j^i\Big)^2}{2\sigma^2\Delta_j^i} \right\}.$$

We will find the likelihood (3.4):

$$L(\boldsymbol{\zeta}|\mathbf{y}) \;=\; \prod_{i=1}^{M} \int p(\mathbf{y}^i | \beta^i, \beta, \sigma^2)\, p(\beta^i | \sigma_\beta^2)\; d\beta^i\;.$$

The computation would be much simplified if we assumed equidistant observations, that is $\Delta_j^i = \Delta$ for all times and subjects, but unfortunately this is rarely the case in real data. Not only will measurements often be taken with varying time gaps, but different subjects might be measured at different time points. In general $n_l \neq n_k$, and $\Delta_j^l \neq \Delta_j^k \neq \Delta_i^k$ (unbalanced data).

Due to the simplicity of the model, techniques adapted from linear regression with a random regression coefficient can be applied, (see e.g. [18]). Define the precision factor: $\eta^2 = \sigma^2/\sigma_\beta^2$. The conditional densities can be written as follows:

$$\begin{aligned} p(\mathbf{y}^i|\beta^i,\beta,\sigma^2) &= \prod_{j=1}^{n_i} p(y_j^i, y_{j-1}^i, \Delta_j^i|\beta^i,\beta,\sigma^2) = \\ &= \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi\sigma^2\Delta_j^i}} \exp\left\{-\frac{\left(y_j^i - y_{j-1}^i - \left(\beta+\beta^i-\tfrac{1}{2}\sigma^2\right)\Delta_j^i\right)^2}{2\sigma^2\Delta_j^i}\right\} \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{n_i}{2}}} \exp\left\{-\frac{\sum_j \frac{1}{\Delta_j^i}\left(y_j^i - y_{j-1}^i - \left(\beta+\beta^i-\tfrac{1}{2}\sigma^2\right)\Delta_j^i\right)^2}{2\sigma^2}\right\} \prod_{j=1}^{n_i}\frac{1}{\sqrt{\Delta_j^i}} \end{aligned}$$

and

$$p(\beta^i|\sigma_\beta^2) = \frac{1}{\sqrt{2\pi\sigma_\beta^2}} \exp\left\{-\frac{(\beta^i)^2}{2\sigma_\beta^2}\right\} = \frac{(\eta^2)^{\frac{1}{2}}}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(\eta\beta^i)^2}{2\sigma^2}\right\}.$$

For ease of notation we define the parameter function $\alpha = \beta - \frac{1}{2}\sigma^2$ and the quantities $\Delta^i = \left(\prod_{j=1}^{n_i}\Delta_j^i\right)^{\frac{1}{n_i}}$ and $T^i = \sum_{j=1}^{n_i}\Delta_j^i$. The last sum is simply the length of the observation interval for the i'th subject; $t_{n_i} - t_0$. We obtain

$$L(\boldsymbol{\zeta}|\mathbf{y}) = \prod_{i=1}^{M} \frac{(\eta^2)^{\frac{1}{2}}}{(2\pi\sigma^2\Delta^i)^{\frac{n_i}{2}}} \int \frac{\exp\left\{-\dfrac{\sum_j \frac{1}{\Delta_j^i}\left(y_j^i - y_{j-1}^i - (\alpha+\beta^i)\Delta_j^i\right)^2 + (\eta\beta^i)^2}{2\sigma^2}\right\}}{\sqrt{2\pi\sigma^2}} \mathrm{d}\beta^i.$$

Solving the last integral yields the marginal likelihood of the parameters, independent of the random effects β^i. Define the vectors

$$\begin{aligned} \tilde{\mathbf{y}}^i &= \left((\Delta_1^i)^{-\frac{1}{2}}(y_1^i - y_0^i), ..., (\Delta_{n_i}^i)^{-\frac{1}{2}}(y_{n_i}^i - y_{n_i-1}^i), 0\right)^T \\ \tilde{\mathbf{x}}^i &= \left((\Delta_1^i)^{\frac{1}{2}}, ..., (\Delta_{n_i}^i)^{\frac{1}{2}}, 0\right)^T \\ \tilde{\mathbf{z}}^i &= \left((\Delta_1^i)^{\frac{1}{2}}, ..., (\Delta_{n_i}^i)^{\frac{1}{2}}, \eta\right)^T \end{aligned}$$

where ^{T} indicates transposition. Then

$$\|\tilde{\mathbf{y}}^i - \tilde{\mathbf{x}}^i\alpha - \tilde{\mathbf{z}}^i\beta^i\|^2 = \sum_{j=1}^{n_i}\frac{1}{\Delta_j^i}\left(y_j^i - y_{j-1}^i - (\alpha+\beta^i)\Delta_j^i\right)^2 + (\eta\beta^i)^2$$

such that, splitting the sum of squares into two parts that are independent of and dependent on the random effects, respectively, and noting that the integral of the dependent part is simply the integral of a normal density up to a constant, the likelihood function can be expressed as

$$
\begin{aligned}
L(\boldsymbol{\zeta}|\mathbf{y}) &= \prod_{i=1}^{M} \frac{(\eta^2)^{\frac{1}{2}}}{(2\pi\sigma^2\Delta^i)^{\frac{n_i}{2}}} \int \frac{\exp\left\{-\dfrac{\|\tilde{\mathbf{y}}^i - \tilde{\mathbf{x}}^i\alpha - \tilde{\mathbf{z}}^i\beta^i\|^2}{2\sigma^2}\right\}}{\sqrt{2\pi\sigma^2}}\,\mathrm{d}\beta^i \\
&= \prod_{i=1}^{M} \frac{(\eta^2)^{\frac{1}{2}}}{(2\pi\sigma^2\Delta^i)^{\frac{n_i}{2}}}\ \exp\left\{-\frac{\|\tilde{\mathbf{y}}^i - \tilde{\mathbf{x}}^i\alpha - \tilde{\mathbf{z}}^i\hat{\beta}^i\|^2}{2\sigma^2}\right\} \frac{1}{\sqrt{T^i+\eta^2}} \\
&= \frac{(\eta^2)^{\frac{M}{2}}}{(2\pi\sigma^2)^{\frac{N-M}{2}}} \prod_{i=1}^{M} \frac{1}{(\Delta^i)^{\frac{n_i}{2}}\sqrt{T^i+\eta^2}} \times \\
&\quad \exp\left\{-\frac{\sum_i \left(\sum_j \frac{1}{\Delta_j^i}\big(y_j^i - y_{j-1}^i - (\alpha+\hat{\beta}^i)\Delta_j^i\big)^2 + (\eta\hat{\beta}^i)^2\right)}{2\sigma^2}\right\} \\
&= \frac{(\eta^2)^{\frac{M}{2}}}{(2\pi\sigma^2)^{\frac{N-M}{2}}} \prod_{i=1}^{M} \frac{1}{(\Delta^i)^{\frac{n_i}{2}}\sqrt{T^i+\eta^2}} \times \\
&\quad \exp\left\{-\frac{\sum_{i,j}\frac{1}{\Delta_j^i}(y_j^i - y_{j-1}^i - \alpha\Delta_j^i)^2 - \sum_i (y_{n_i}^i - y_0^i - \alpha T^i)^2 (T^i+\eta^2)^{-1}}{2\sigma^2}\right\}, \qquad (3.7)
\end{aligned}
$$

where $\hat{\beta}^i$ minimizes the sum of squares $\|\tilde{\mathbf{y}}^i - \tilde{\mathbf{x}}^i\alpha - \tilde{\mathbf{z}}^i\beta^i\|^2$ for fixed α, and is obtained from standard regression theory:

$$
\hat{\beta}^i = ((\tilde{\mathbf{z}}^i)^T\tilde{\mathbf{z}}^i)^{-1}(\tilde{\mathbf{z}}^i)^T(\tilde{\mathbf{y}}^i - \tilde{\mathbf{x}}^i\alpha) = \frac{\sum_{j=1}^{n_i}(y_j^i - y_{j-1}^i - \alpha\Delta_j^i)}{\sum_{j=1}^{n_i}\Delta_j^i + \eta^2} = \frac{y_{n_i}^i - y_0^i - \alpha T^i}{T^i+\eta^2}.
$$

These directly provide predictors of the random effects given the parameters. The log-likelihood is

$$
\begin{aligned}
\log L(\boldsymbol{\zeta}|\mathbf{y}) &= \frac{M}{2}\log\eta^2 - \frac{N-M}{2}\log(2\pi\sigma^2) - \frac{1}{2}\sum_{i=1}^{M}\log\left((\Delta^i)^{n_i}(T^i+\eta^2)\right) \\
&\quad - \frac{\sum_{i,j}\frac{1}{\Delta_j^i}(y_j^i - y_{j-1}^i - \alpha\Delta_j^i)^2 - \sum_i (y_{n_i}^i - y_0^i - \alpha T^i)^2 (T^i+\eta^2)^{-1}}{2\sigma^2}. \qquad (3.8)
\end{aligned}
$$

The derivatives of the log-likelihood function with respect to the parameters yield the *score functions* whose zeros will provide the maximum likelihood estimators of the parameters. Straightforward calculations yield the estimating equations,

where the estimator of a parameter is indicated with a hat, e.g. $\hat{\beta}$:

$$
\begin{aligned}
0 &= \sum_{i=1}^{M} \left(\frac{y_{n_i}^i - y_0^i - \hat{\alpha} T^i}{T^i + \hat{\eta}^2} \right) \\
0 &= \sum_{i=1}^{M} \left(\frac{\hat{\sigma_\beta}^2 T^i}{T^i + \hat{\eta}^2} - \frac{(y_{n_i}^i - y_0^i - \hat{\alpha} T^i)^2}{(T^i + \hat{\eta}^2)^2} \right) \\
0 &= \sum_{i=1}^{M} \left(\sum_{j=1}^{n_i} \left(\frac{(y_j^i - y_{j-1}^i - \hat{\alpha} \Delta_j^i)^2}{\Delta_j^i} \right) - \frac{(y_{n_i}^i - y_0^i - \hat{\alpha} T^i)^2}{T^i + \hat{\eta}^2} \right) - \hat{\sigma}^2 (N - M) \; .
\end{aligned}
$$

If we assume that each subject is observed in the same time interval, that is, we assume $T^i = T$ for all $1 \leq i \leq M$, this simplifies to the explicit estimators:

$$\hat{\beta} = \hat{\alpha} + \frac{\hat{\sigma}^2}{2} \tag{3.9}$$

$$\hat{\sigma}^2 = \frac{1}{N - 2M} \left((N - M) \mathrm{SSQ}_\Delta - M \mathrm{SSQ}_\mathrm{T} \right) \tag{3.10}$$

$$\hat{\sigma}_\beta^2 = \frac{N - M}{T(N - 2M)} \left(\mathrm{SSQ}_\mathrm{T} - \mathrm{SSQ}_\Delta \right) \tag{3.11}$$

where

$$\hat{\alpha} = \frac{1}{MT} \sum_{i=1}^{M} (y_{n_i}^i - y_0^i) \tag{3.12}$$

$$\mathrm{SSQ}_\mathrm{T} = \frac{1}{MT} \sum_{i=1}^{M} (y_{n_i}^i - y_0^i - \hat{\alpha} T)^2 \tag{3.13}$$

$$\mathrm{SSQ}_\Delta = \frac{1}{N - M} \sum_{i=1}^{M} \sum_{j=1}^{n_i} \left(\frac{\left(y_j^i - y_{j-1}^i - \hat{\alpha} \Delta_j^i \right)^2}{\Delta_j^i} \right) . \tag{3.14}$$

The asymptotic variances of the estimators estimated from the inverted Fisher information evaluated at the optimum is given by:

$$\hat{\mathrm{Var}}(\hat{\beta}) = \frac{\hat{\sigma}_\beta^2 T + \hat{\sigma}^2}{MT} + \frac{\hat{\sigma}^4}{2(N - 2M)} \tag{3.15}$$

$$\hat{\mathrm{Var}}(\hat{\sigma}) = \frac{\hat{\sigma}^2}{2(N - 2M)} \tag{3.16}$$

$$\hat{\mathrm{Var}}(\hat{\sigma}_\beta) = \frac{(\hat{\sigma}_\beta^2 T + \hat{\sigma}^2)^2}{2MT^2 \hat{\sigma}_\beta^2} + \frac{\hat{\sigma}^4}{2(N - 2M) T^2 \hat{\sigma}_\beta^2} \; . \tag{3.17}$$

There will only be positive solutions for the variance parameters in the data set if

$$\frac{M}{N - M} \mathrm{SSQ}_\mathrm{T} < \mathrm{SSQ}_\Delta < \mathrm{SSQ}_\mathrm{T} \; . \tag{3.18}$$

The last inequality ensures existence of the estimator of the random effect variance parameter σ_β^2, and can be interpreted in the following way: For simplicity assume $\Delta_j^i = \Delta$ for all i, j. Define $a_j^i = (y_j^i - y_{j-1}^i - \hat{\alpha}\Delta)$, the increment for subject i from observation $j-1$ to observation j subtracted the expected increment in the population. Then

$$\mathrm{SSQ}_\mathrm{T} = \frac{1}{MT}\sum_{i=1}^{M}\left(\sum_{j=1}^{n_i} a_j^i\right)^2 \quad \text{and} \quad \mathrm{SSQ}_\Delta = \frac{1}{MT}\sum_{i=1}^{M}\sum_{j=1}^{n_i}(a_j^i)^2 \ .$$

For SSQ_Δ to be smaller than SSQ_T, it is required that at least for one i, $\sum_{j=1}^{n_i}(a_j^i)^2 < (\sum_{j=1}^{n_i} a_j^i)^2$, which e.g. will be the case if all a_j^i are of the same sign. If this is the case it means that all observed increments are either above or under the expected increments for the population, which indicates that the decay rate for this specific subject most probably is different from the general population decay rate β, that is $\beta^i \neq 0$. On the other hand, to estimate the system noise parameter σ^2, we require $\frac{\Delta}{T}\mathrm{SSQ}_\mathrm{T} < \mathrm{SSQ}_\Delta$. The left hand side increases when the number of measured points for each subject decreases. In this case it is natural that we have more information on variation between subjects than variation within subjects.

Considering model (3.6) with $\sigma_\beta^2 = 0$, such that $\beta^i = 0$ for all i (no random effects), leads to the log-likelihood function

$$\log L(\beta, \sigma^2|\mathbf{y}) \ = \ -\frac{N-M}{2}\log(2\pi\sigma^2) - \sum_i \frac{n_i}{2}\log(\Delta^i) - \sum_{i,j}\frac{(y_j^i - y_{j-1}^i - \alpha\Delta_j^i)^2}{2\sigma^2\Delta_j^i}$$

which could also be derived from (3.8) by letting $\eta^2 \to \infty$. This leads to the maximum likelihood estimators

$$\hat{\beta} \ = \ \hat{\alpha} + \hat{\sigma}^2/2 \tag{3.19}$$

$$\hat{\sigma}^2 \ = \ \mathrm{SSQ}_\Delta \ . \tag{3.20}$$

The asymptotic variances of the estimators estimated from the inverted Fisher information evaluated at the optimum is given by:

$$\hat{\mathrm{Var}}(\hat{\beta}) \ = \ \frac{\hat{\sigma}^2}{MT} + \frac{\hat{\sigma}^4}{2(N-M)} \tag{3.21}$$

$$\hat{\mathrm{Var}}(\hat{\sigma}) \ = \ \frac{\hat{\sigma}^4}{2(N-M)} \ . \tag{3.22}$$

3.2. Simulation results

To check the estimators a simulation study was performed. Six sets of parameter values were used to investigate the behavior of the estimators for different relations among the variance components, namely for $\sigma^2 \gg \sigma_\beta^2$, $\sigma^2 \approx \sigma_\beta^2$, and $\sigma^2 \ll \sigma_\beta^2$, respectively, and for two different values of β, consistent with physiologically observed decay rate values. Moreover, two sets of values for the experimental designs were investigated, namely for $M \gg n$ and $M \ll n$, respectively. The values used in the different simulations are reported in Table 1.

Table 1: Values used in the different simulations.

	Parameter values used in simulations				
	β	σ^2	σ_β^2	M	n
1	-0.02	0.02	0.02	10	50
2	-0.02	0.2	0.02	10	50
3	-0.02	0.02	0.2	10	50
4	-0.02	0.02	0.02	50	10
5	-0.02	0.2	0.02	50	10
6	-0.02	0.02	0.2	50	10
7	-0.2	0.02	0.02	10	50
8	-0.2	0.2	0.02	10	50
9	-0.2	0.02	0.2	10	50
10	-0.2	0.02	0.02	50	10
11	-0.2	0.2	0.02	50	10
12	-0.2	0.02	0.2	50	10

For each of these 12 sets of values, 1.000 data sets were generated from model (3.6), by simulating trajectories according to the Milstein scheme with a step size of 0.01, see Kloeden and Platen (1999), and retaining the observation points at equidistant time points depending on the chosen n. For all simulations the total length of the simulation interval was 100, and the initial value was $\log(100)$. On the simulated data sets, parameters were estimated using Equations (3.9) to (3.14). Parameters were also estimated assuming (wrongly) the model with no random effects by Equations (3.19) and (3.20). Results are reported in Table 2, where the 95% confidence intervals are the 2.5% and 97.5% empirical quantiles of estimates, and are given in brackets.

In all 12.000 simulations the estimators existed ($\hat{\sigma}^2, \hat{\sigma}_\beta^2 > 0$), but for $\beta = -0.02$ a considerable part of the estimates were positive, reflected in the large 97.5% quantiles for $\hat{\beta}$. Not surprisingly, β is more difficult to estimate when σ_β^2 is large. The diffusion parameter σ^2 is well determined with 95% of estimates lying

no more than 11% from the true value, whereas σ_β^2 is more difficult to estimate and depends on the size of M: for small M, the distribution of estimates is right-skewed with wide confidence limits; for larger M, σ_β^2 is better determined.

Table 2: Mean of estimates (95% CI) from simulations of model (3.6). For each set of values, 1.000 data sets were generated and the parameters were estimated using (3.9) to (3.11) (assuming random effects) and (3.19) and (3.20) (assuming no random effects). For all simulations $T = 100$ and $Y_0^i = \log(100)$. See also main text.

	Assuming random effects		
	$\hat{\beta}$	$\hat{\sigma}^2$	$\hat{\sigma}_\beta^2$
1	-0.018 (-0.094;0.060)	0.020 (0.018;0.022)	0.018 (0.006;0.035)
2	-0.019 (-0.097;0.056)	0.200 (0.178;0.222)	0.018 (0.006;0.034)
3	-0.027 (-0.263;0.210)	0.020 (0.018;0.022)	0.178 (0.063;0.317)
4	-0.021 (-0.054;0.012)	0.020 (0.018;0.022)	0.020 (0.013;0.027)
5	-0.021 (-0.057;0.017)	0.200 (0.178;0.221)	0.020 (0.013;0.028)
6	-0.020 (-0.119;0.080)	0.020 (0.018;0.022)	0.197 (0.137;0.267)
7	-0.199 (-0.276;-0.126)	0.020 (0.018;0.022)	0.017 (0.006;0.034)
8	-0.198 (-0.281;-0.123)	0.201 (0.180;0.223)	0.017 (0.005;0.035)
9	-0.198 (-0.440;0.029)	0.020 (0.018;0.022)	0.175 (0.066;0.337)
10	-0.200 (-0.233;-0.167)	0.020 (0.018;0.022)	0.020 (0.013;0.026)
11	-0.201 (-0.236;-0.163)	0.200 (0.178;0.222)	0.020 (0.013;0.027)
12	-0.202 (-0.302;-0.096)	0.020 (0.018;0.022)	0.195 (0.135;0.260)

	Assuming no random effects (wrong model)		
	$\hat{\beta}$	$\hat{\sigma}^2$	-
1	0.000 (-0.076;0.079)	0.057 (0.033;0.091)	
2	-0.001 (-0.082;0.076)	0.236 (0.203;0.275)	
3	0.155 (-0.089;0.404)	0.384 (0.151;0.667)	
4	0.087 (0.039;0.141)	0.237 (0.166;0.315)	
5	0.088 (0.033;0.144)	0.418 (0.339;0.510)	
6	1.075 (0.729;1.458)	2.209 (1.543;2.985)	
7	-0.181 (-0.260;-0.107)	0.056 (0.033;0.089)	
8	-0.181 (-0.264;-0.105)	0.236 (0.202;0.277)	
9	-0.020 (-0.292;0.257)	0.377 (0.156;0.708)	
10	-0.092 (-0.138;-0.042)	0.237 (0.170;0.310)	
11	-0.092 (-0.142;-0.039)	0.418 (0.343;0.499)	
12	0.884 (0.547;1.273)	2.192 (1.519;2.907)	

If a model with no random effects is wrongly assumed, both β and σ^2 are poorly estimated. The estimates are worse for large σ_β^2 and large M, as expected. This illustrates the need to include random effects in the modelling process if they are present in the data.

3.3. Application to Metoprolol Tartrate dissolution data

The method was applied on Metoprolol Tartrate dissolution data taken from [19], where the percentage of released drug of four types of tablet formulations of 100-mg Metoprolol Tartrate are tabulated at 5-min intervals up to 30 minutes and at 45 minutes after the onset of the experiments, except for the Slow Dissolving Test Formulation, where measurements were taken up to two hours, for details see [19]. Each experiment was repeated six times. The data were also analysed in [12, 13]. In [12], they found that the formulation closest to an exponential behavior was the Slow Dissolving Test Formulation, which is used here to illustrate the methods. Only data up to 45 minutes are used.

The data are illustrated in Figure 1. The percentage of Metoprolol not yet dissolved is modeled as (3.6), where y_j^i are the log-transformed measured percentages for experiment i at time point j. Moreover, the measurement at 30 minutes for experiment four was removed in the analysis since the dissolution process cannot go backwards, see Figure 1. Finally $M = 6$ and $N = 41$.

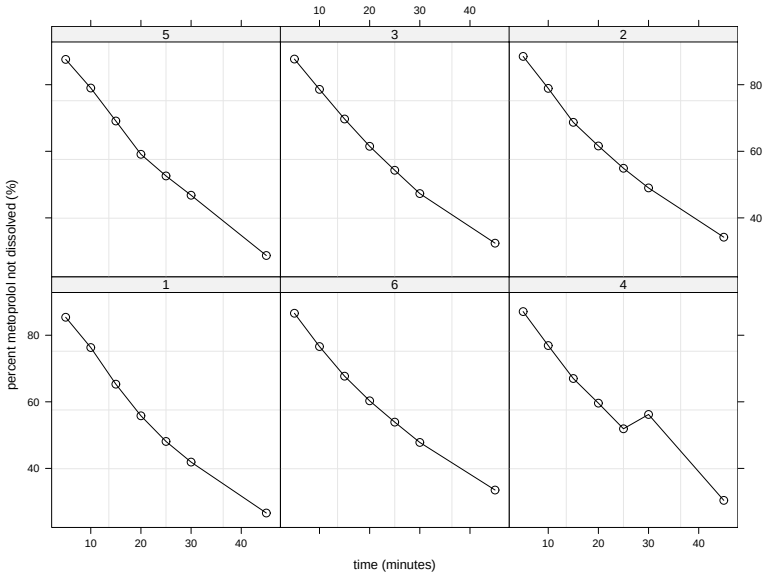

Figure 1: Dissolution profiles of Metoprolol Tartrate tablets. Data is taken from [19].

The data set yields the following quantities: $\hat{\alpha} = -0.026, \mathrm{SSQ}_{\mathrm{T}} = 0.000166$ and $\mathrm{SSQ}_{\Delta} = 0.0000699$ such that condition (3.18) is fulfilled. Estimates and their standard errors are reported in Table 3. The estimates of β are in agreement with comparable values found in [12, 13, 19]. Since $\hat{\sigma}_\beta^2$ is small compared to $\hat{\sigma}^2$, the estimates in the model without random effects only change slightly.

Table 3: Metoprolol data estimates using Equations (3.9) to (3.11) (assuming random effects) and (3.19) and (3.20) (assuming no random effects). The standard errors were estimated using Equations (3.15), (3.16), (3.17), (3.21) and (3.22).

	Assuming random effects		**Assuming no random effects**	
	estimate	std error	estimate	std error
$\hat{\beta}$	-0.02594	0.00083	-0.02593	0.00054
$\hat{\sigma}$	0.00707	0.00093	0.00836	0.00001
$\hat{\sigma}_\beta$	0.00171	0.00071	-	-

Figure 2 shows simulated trajectories from the random effects model with the estimated parameters, and the observed points from two of the six dissolution profiles.

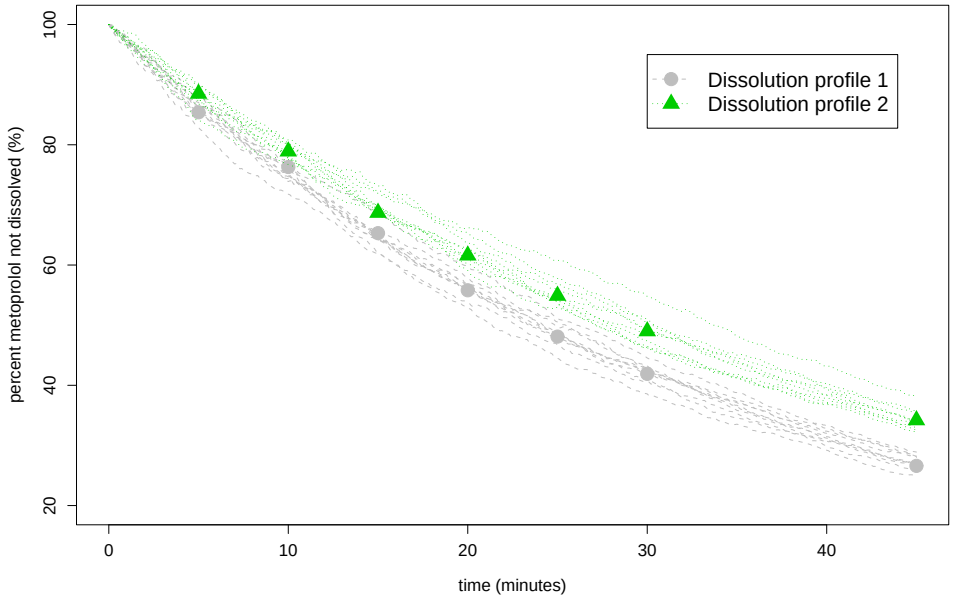

Figure 2: Simulated trajectories from model (3.6), incorporating the estimated parameters and random effect estimates for two of the dissolution profiles. The points are the observed data for the same two dissolution profiles.

4. SUMMARY

In the present paper we propose to extend random effects techniques to the estimation of parameters in SDE models. We believe this extension to be both relevant and needed. It is relevant because as the sophistication of builders and users of mathematical models of biological processes increases, there will be a progressive growth of the use of stochastic differential equations to represent noisy processes. When only few observations can be collected from any given human or animal experimental subject, as is usually the case, recourse to random or mixed effects models will be necessary.

Statistical inference for this class of models is not straightforward. In the present work, a very simple model gave rise to explicit expressions for the likelihood function and for the maximum likelihood estimators. This model is in its deterministic version frequently employed in pharmacokinetics (e.g. to represent drug elimination from plasma or initial tumor cell population growth), and the proposed development is therefore not only of academic interest. However, it is often the case that more complicated models with nonlinearities and/or several compartments are necessary to plausibly represent the system under observation.

Unfortunately, in general it will not be possible to find an explicit expression for the likelihood function (3.4) since the transition densities are rarely known. One possibility could be to approximate the likelihood function numerically, and then optimize the approximated likelihood function directly. It is obviously necessary to find other estimation procedures if the proposed model class is to be of interest to a wider audience.

ACKNOWLEDGMENTS

Work supported in part by the European Community's Human Potential Programme under contract HPRN-CT-2000-00100, DYNSTOCH.

REFERENCES

[1] Bibby, B.M. and Sørensen, M. (1995). Martingale estimation functions for discretely observed diffusion processes. *Bernoulli*, **1**, 1/2, 017–039.

[2] Breslow, N.E. and Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *JASA*, **88**, 421, 9–25.

[3] Diggle, P.J.; Heagerty, P.; Liang, K.-Y. and Zeger, S.L. (2002). *Analysis of Longtitudinal Data*, Oxford University Press, 2nd edition.

[4] Ditlevsen, S. and Sørensen, M. (2004). Inference for observations of integrated diffusion processes. *Scand. J. Statist*, **31**, 3, 417–429.

[5] Ditlevsen, S. and De Gaetano, A. (2005). Stochastic vs. deterministic uptake of dodecanedioic acid by isolated rat livers. *Bulletin of Mathematical Biology*, **67**, 547–561.

[6] Ditlevsen, S.; Yip, K.P. and Holstein-Rathlou, N.H. (2005). Parameter estimation in a stochastic model of the tubuloglomerular feedback mechanism in a rat nephron. *Mathematical Biosciences*, **194**, 49–69.

[7] Elerain, O.; Chib, S. and Shepard, N. (2001). Likelihood inference for discretely observed non-linear diffusions. *Econometrica*, **69**, 959–993.

[8] Jelliffe, R.; Schumitzky, A. and Van Guilder, M. (2000). Population pharmacokinetics/pharmacodynamics modeling: Parametric and nonparametric methods. *Therapeutic Drug Monitoring*, **22**, 354–365.

[9] Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist*, **24**, 211–229.

[10] Kessler, M. and Sørensen, M. (1999). Estimating equations based on eigenfunctions for a discretely observed diffusion process. *Bernoulli*, **5**, 2, 299–314.

[11] Kloeden, P.E. and Platen, E. (1999). *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, Berlin and Heidelberg GmbH & Co. K.

[12] Lansky, P. and Weiss, M. (2003). Classification of dissolution profiles in terms of fractional dissolution rate and a novel measure of heterogeneity. *J. Pharmaceutical Sciences*, **92**, 1632–1647.

[13] Lansky, P.; Lanska, V. and Weiss, M. (2004). A stochastic differential equation model for drug dissolution and its parameters. *J. Controlled Release*, **100**, 267–274.

[14] Lindstrom, M.J. and Bates, D.M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics*, **46**, 673–687.

[15] Overgaard, R.V.; Jonsson, N.; Tornøe, C.W. and Madsen, H. (2005). Non-linear mixed-effects models with stochastic differential equations. Implementation of an estimation algorithm. *Journal of Pharmacokinetics and Pharmacodynamics*, **32**, 85–107.

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

[17] Pedersen, A.R. (2001). *Likelihood inference by Monte Carlo methods for incompletely discretely observed diffusion processes*, Technical Report 1, Department of Biostatistics, Univ. of Aarhus.

[18] Pinheiro, J.C. and Bates, D.M. (2000). *Mixed-Effects Models in S and S-PLUS*, Springer-Verlag, New York.

[19] Polli, J.E.; Singh Rekhi, G.; Augsburger, L.L. and Shah, V.P. (1997). Methods to compare dissolution profiles and a rationales for wide dissolution specifications for metoprolol tartrate tablets. *J. Pharmaceutical Sciences*, **86**, 6, 690–700.

[20] Prakasa Rao, B.L.S. (1999). *Statistical Inference for Diffusion Type Processes*, Arnold Publishers.

[21] Sheiner, L.B. and Beal, S.L. (1980). Evaluation of methods for estimating population pharmacokinetic parameters. I. Michaelis-Menten model: Routine clinical pharmacokinetic data. *J. Pharmacokinet. Biopharm.*, **8**(6), 553–571.

[22] Sheiner, L.B. and Beal, S.L. (1981). Evaluation of methods for estimating population pharmacokinetic parameters. II. Biexponential model and experimental pharmacokinetic data. *J. Pharmacokinet. Biopharm.*, **9**(5), 635–651.

[23] Sørensen, H. (2003). Simulated likelihood approximations for stochastic volatility models. *Scand. J. Statist.*, **30**, 257–276.

[24] Sørensen, M. (2000). Prediction-based estimating functions. *Econometrics J.*, **3**, 123–147.

[25] Vonesh, E.F. and Chinchilli, V.M. (1997). *Linear and Nonlinear Models for the Analysis of Repeated Measurements*, Marcel Dekker, New York.