A review on estimation of stochastic differential equations for pharmacokinetic/pharmacodynamic models

Sophie Donnet, Adeline Samson

A review on estimation of stochastic differential equations for pharmacokinetic/pharmacodynamic models

Sophie Donnet[1] and Adeline Samson[2]

[1] Laboratoire Ceremade, Université Paris Dauphine, France,
[2] UMR CNRS8145, Laboratoire MAP5, Université Paris Descartes, PRES Sorbonne Paris Cité, France

Abstract

This paper is a survey of existing estimation methods for pharmacokinetic/pharmacodynamic (PK/PD) models based on stochastic differential equations (SDEs). Most parametric estimation methods proposed for SDEs require high frequency data and are often poorly suited for PK/PD data which are usually sparse. Moreover, PK/PD experiments generally include not a single individual but a group of subjects, leading to a population estimation approach. This review concentrates on estimation methods which have been applied to PK/PD data, for SDEs observed with and without measurement noise, with a standard or a population approach. Besides, the adopted methodologies highly differ depending on the existence or not of an explicit transition density of the SDE.

Key Words: Stochastic differential equations; Pharmacokinetic; Pharmacodynamic; population approach; maximum likelihood estimation; Kalman Filter; EM algorithm; Hermite expansion; Gauss quadrature; Bayesian estimation

1 Introduction

Pharmacokinetics (PK) aims at describing the relationship between the dose administered and the exposure to the drug, i.e. the total concentration of drug in the body. Pharmacodynamics (PD) quantifies the relationship between the drug exposure and the response to this exposure. PK/PD models are often described by differential systems derived from physiology. In general, the proposed models are deterministic, that is, the observed kinetic/dynamic is driven exclusively by internal deterministic mechanisms. However, real pharmacological processes are always exposed to influences that are not completely understood or not feasible to model explicitly. Ignoring these phenomena in the modeling may affect the estimation of PK/PD parameters and the derived conclusions. Therefore there is an increasing need to extend the deterministic models to models including a stochastic component. A natural extension of deterministic differential equations model is a system of stochastic differential equations (SDEs), where relevant parameters have been modeled as suitable stochastic processes, or stochastic processes have been added to the driving system equations (Ditlevsen and Samson, 2012).

The first papers encouraging the introduction of random fluctuations in PK/PD were published by D'Argenio and Park (1997) and Ramanathan (1999a,b). The authors underline that PK/PD have contributions from both deterministic and stochastic components: drug concentrations follow

determinable trends but the exact concentration at any given time is not completely determined. For example, Ramanathan (1999b) proposes a stochastic one-compartment PK model with a variable elimination rate. More sophisticated PK/PD models have then been proposed with multiple compartments, non-linear or time-inhomogeneous absorption or elimination (see for example Ferrante *et al.*, 2003; Tornøe *et al.*, 2004a; Ditlevsen and De Gaetano, 2005b; Ditlevsen *et al.*, 2005; Picchini *et al.*, 2006).

Parameter estimation for SDE has been highly tackled in the statistical literature, often motivated by financial applications (see Sørensen, 2004, for a review). However, many suggested solutions require high frequency data and are not suited for PK/PD data where designs are usually sparse. Especially, estimation methods based on approximations of the continuous-time observation likelihood (namely the Girsanov formula), which require a high number of data and a small time step between two successive observations, are not adapted. Moreover, PK/PD data are more and more analyzed through a population approach when data from several subjects are considered simultaneously. This yields to PK/PD models with random parameters. Combining SDE with a population approach is quite appealing but raises inference challenges.

In this paper, we concentrate our review on estimation methods adapted to the particular characteristics of PK/PD data. After a short presentation of some examples of PK/PD SDEs in Section 2, we introduce some preliminary comments on the likelihood functions depending on the considered observation model (with and without measurement noise) in Section 3. Section 4 is about estimation methods for standard PK/PD SDE: when SDEs are directly observed, the reader is introduced to techniques based on (i) exact maximum likelihood estimator when explicit solution is available or (ii) Hermite expansion of the transition density, (iii) approximation of the spectral density if the SDE has no explicit solution. When the SDE is observed with an additive measurement error, methods are based on (iv) Kalman filter and its extended version or (v) Monte Carlo approximation of the likelihood. Section 5 discusses estimation methods within a population approach: we detail methods based on (i) exact maximum likelihood estimator when linear SDE with random effect and no measurement noise are considered, (ii) Gauss-Hermite quadrature to approximate the likelihood, possibly coupled with (iii) Hermite expansion of the transition density, (iv) Bayesian approach, (v) Kalman filter and linearization of the likelihood, (vi) Expectation-Maximisation algorithm. The paper finishes with some discussion (Section 6).

2 Stochastic PK/PD models

In this section, we present some stochastic compartmental PK/PD models that have been proposed in the literature. This list is far from being exhaustive but aims at presenting typical situations, each of them involving a different level of statistical inference difficulty.

2.1 From deterministic to stochastic model in PK

Let us first consider a very simple PK model proposed by Ramanathan (1999a), namely a one compartment PK model with first-order elimination k_e and an injected intravenous bolus dose D of drug. The kinetic of the drug concentration C_t in the body at time $t > 0$ is described by the following deterministic differential equation:

$$\frac{dC_t}{dt} = -k_e C_t, \quad C_0 = \frac{D}{V},$$

where V is the volume of the compartment. This equation has an explicit solution: $C_t = \frac{D}{V}e^{-k_e t}$. Now, assume that k_e is not constant in time but randomly fluctuates around a mean value as $k_e + \xi_t$, where ξ_t is a Gaussian white noise process. Then $\xi_t dt$ can be written as γdB_t, where B_t is a Brownian motion and γ is a constant parameter. Incorporating this noise into the deterministic model, C_t becomes a stochastic process, solution of the following SDE:

$$dC_t = -k_e C_t dt + \gamma C_t dB_t, \quad C_0 = \frac{D}{V}. \tag{1}$$

That process –known as geometric Brownian motion– has an explicit expression

$$C_t = \frac{D}{V}e^{-k_e t}\exp\left(\left(-\frac{\gamma^2}{2}\right)t + \gamma B_t\right).$$

This stochastic process, which is log-normal, only takes positive values, which is noticeable when modeling concentration. Parameters to be estimated are $\theta = (k_e, V, \gamma)$.

A stochastic one compartment PK model with first-order absorption has also been considered by Ferrante *et al.* (2003):

$$dC_t = (k_a/V - k_e C_t)dt + \gamma dB_t, \quad C_0 = \frac{D}{V}, \tag{2}$$

where k_a is the absorption rate. This process –known as an Ornstein-Uhlenbeck process– has an explicit expression

$$C_t = \frac{D}{V}e^{-k_e t} + \frac{k_a}{V k_e}(1 - e^{-k_e t}) + \gamma \int_0^t e^{-k_e(t-s)} dB_s,$$

and is Gaussian with explicit mean and variance. Parameters to be estimated are $\theta = (k_e, k_a, V, \gamma)$.

Multi compartments PK models have also been extended to stochastic versions. For example, a two compartments model is described by the deterministic system

$$\begin{array}{rcl} \frac{dC_{1t}}{dt} & = & -k_1 C_{1t} + k_2 C_{2t} - k_e C_{1t} \\ \frac{dC_{2t}}{dt} & = & k_1 C_{1t} - k_2 C_{2t} \end{array} \tag{3}$$

where C_{1t}, C_{2t} are both compartments concentrations, and k_1, k_2 are transfer constants. Cuenod *et al.* (2011) propose a stochastic version of (3) with independent Brownian motions B_{1t}, B_{2t} on each equation and diffusion coefficients γ_1, γ_2:

$$\begin{array}{rcl} dC_{1t} & = & (-k_1 C_{1t} + k_2 C_{2t} - k_e C_{1t})\, dt + \gamma_1 dB_{1t} \\ dC_{2t} & = & (k_1 C_{1t} - k_2 C_{2t})\, dt + \gamma_2 dB_{2t}. \end{array} \tag{4}$$

The yielding stochastic process has an explicit Gaussian solution. Parameters to be estimated are $\theta = (k_1, k_2, k_e, \gamma_1, \gamma_2)$. Ditlevsen and De Gaetano (2005b) add a Brownian motion only on the second compartment C_{2t} of (3), with a diffusion coefficient γ_2:

$$\begin{array}{rcl} dC_{1t} & = & (-k_1 C_{1t} + k_2 C_{2t} - k_e C_{1t})\, dt \\ dC_{2t} & = & (k_1 C_{1t} - k_2 C_{2t})\, dt + \gamma_2 dB_{2t}. \end{array} \tag{5}$$

This leads to a hypoelliptic SDE, that is, a stochastic differential system in which only a few equations include a volatility term, the other equations being of ODE type. No explicit solution

exists for this SDE. Parameters to be estimated are then $\theta = (k_1, k_2, k_e, \gamma_2)$.

SDEs have also been proposed for glucose kinetics using data from intravenous glucose tolerance test. For example, Tornøe *et al.* (2004b) develop a stochastic non-linear multi-compartments glucose minimal model. No explicit solution exists for this complex SDE.

2.2 Stochastic coupled PK/PD models

Several papers propose a stochastic model for simultaneous analysis of PK/PD data. In that case, either only the PK model, or only the PD model or both models are considered stochastic.

For example, Ferrante *et al.* (2005) consider a stochastic PD model (Gompertz) of the bacterial count N_t under antibiotic effect coupled to a two dimensional deterministic PK model for the antibiotic concentration C_t:

$$\begin{array}{rcl} dN_t & = & (r - bN_t \log N_t - kC_t)N_t dt + \gamma dB_t \\ C_t & = & \frac{Dk_a}{V(k_a - k_e)}(e^{-k_e t} - e^{-k_a t}), \end{array} \tag{6}$$

with r the intrinsic growth rate, b the growth deceleration rate, k the bacterial effect of the drug, D the dose of antibiotic, V the volume of distribution, k_a and k_e the absorption and elimination constants, respectively. Log-transformation of this non-linear time-inhomogenous SDE has an explicit solution (see Ferrante *et al.*, 2005). Parameters to be estimated are $\theta = (k_e, k_a, V, r, b, k, \gamma)$.

Another example is given by Tornøe *et al.* (2004b) with a stochastic first-order elimination one-compartment PK model and an indirect response PD model (Emax):

$$\begin{array}{rcl} dC_t & = & -k_e C_t + \gamma_C dB_{1t} \\ dR_t & = & \left(k_{in} - k_{out}\left(1 + \frac{C_t}{EC_{50} + C_t} R_t\right)\right) dt + \gamma_R dB_{2t}, \end{array} \tag{7}$$

where R_t is the state variable for the PD response, EC_{50} is the drug concentration causing 50% of maximal stimulation, γ_C and γ_R are two diffusion coefficients. This system has no explicit solution. Parameters to be estimated are $\theta = (k_e, k_{in}, k_{out}, EC_{50}, \gamma_C, \gamma_R)$.

More complex PK/PD systems have also been introduced in the literature. Picchini *et al.* (2006) propose for instance a two dimensional SDE for the oscillations of glycemia occurring in response to hyperinsulinisation and to constant glucose infusion. Tornøe *et al.* (2004a) include in the same model an Emax PD equation. In Ditlevsen *et al.* (2005) a 6-dimensional SDE is proposed to model the auto regulation of renal blood flow and glomerular filtration rate.

In conclusion, a wide variety of SDEs are proposed in PK/PD modeling. Except linear, SDEs have generally no explicit solution and this complicates the parameter estimation. Another difficulty raises from the sparsity of PK/PD data (low frequency and/or partial observations of the multidimensional models). The methods proposed in the literature include these characteristics.

3 Preliminary comments on estimation

We now introduce a general model and the notations used through out the whole paper. Two observations models are considered, depending if the PK/PD SDE is directly observed or observed with an additive measurement noise. The likelihood functions are presented in these two contexts. Corresponding estimation methods are detailed in Section 4. Observation models with a PK/PD

population approach and corresponding estimation methods are gathered in Section 5.

3.1 SDE model

The studied biological process is assumed to be modeled by $(X_t)_{t\geq 0}$ a stochastic process in $\mathcal{X} \in \mathbb{R}^p$ defined as the solution of the multidimensional SDE:

$$\begin{array}{rcl} dX_t & = & \mu(X_t, t, \theta)dt + \Gamma(X_t, \theta)dB_t \\ X_0 & = & x_0(\theta) \end{array} \tag{8}$$

where $\theta \in \Theta \subset \mathbb{R}^d$ is the parameter vector and Θ is the set of parameters, $\mu(X_t, t, \theta) : \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^p$ is the *drift function*, $\Gamma(X_t, \theta) : \mathbb{R}^p \times \mathbb{R}^d \to \mathcal{M}(\mathbb{R}^p)$, a $p \times p$ matrix, is the *volatility function*, $(B_t)_{t\geq 0}$ denotes a p-dimensional Brownian motion and $x_0(\theta)$ is the initial condition. If the drift and the volatility functions are independent of time t, then the system is said to be *homogenous*, otherwise it is *inhomogenous*. The initial condition may be random but for sake of simplicity, we assume it deterministic in this paper. The aim is to estimate the parameter θ from (maybe noisy and partial) observations of X_t.

We assume the solution of SDE (8) to exist, and the law of X_t given θ and $X_s = x_s$ for any $s < t$, to have a strictly positive density, the transition density, w.r.t. the Lebesgue measure on $\mathcal{X}$:

$$x \to p_X(x, t - s|x_s, \theta) > 0, \quad x \in \mathcal{X}.$$

This existence is ensured by regularity properties on the drift and volatility functions (see Lipster and Shiryaev, 2001, for instance). Note that this assumption does not imply the transition density to be explicit, it happens only in few cases.

3.2 Observation models and likelihood functions

In the first papers published on the estimation of PK/PD SDEs, processes were supposed to be exactly observed, no observation noise was considered. This assumption, rather unrealistic for PK/PD data, has the great advantage to simplify the likelihood, as presented in section 3.2.1. In section 3.2.2, the more realistic assumption of discrete and noisy observations of PK/PD SDE is introduced, as well as the corresponding likelihood.

3.2.1 Discretely and directly observed diffusions

Assume that $(X_t)_{t\geq 0}$ is directly observed at discrete times $t_1, \ldots, t_n$. We denote indistinctly by X_j or X_{t_j} the observation at time t_j and $X_{1:n} = (X_{t_1}, \cdots, X_{t_n}) = (X_1, \cdots, X_n)$. Set $\Delta_j = t_j - t_{j-1}$. Let $p_{\underline{X}}$ denote the density of $X_{1:n}$ given θ. By Markov property, $p_{\underline{X}}$ is the product of the $n-1$ transition densities. Thus the likelihood function $\mathcal{L}(\theta; X_{1:n})$ is

$$\mathcal{L}(\theta; X_{1:n}) = p_{\underline{X}}(X_{1:n}|\theta) = \prod_{j=1}^{n} p_X(X_{t_j}, \Delta_j | X_{t_{j-1}}, \theta). \tag{9}$$

When p_X is explicit, the likelihood (9) is explicit and exact maximum likelihood estimators (MLE) can be computed (see Section 4.1.1). Difficulties in the parameter estimation arise when the transition density p_X is not explicit. In that case, approximate estimators have been proposed in the literature (see Section 4.1.2).

3.2.2 Discretely observed diffusions with observation noise

Let $y_{1:n} = (y_1, \ldots, y_n)$ be the noisy observations of the diffusion process $(X_t)_{t\geq 0}$ at time $t_1, \ldots, t_n$. We assume that $y_j \in \mathbb{R}^q$, with $q \leq p$, is derived from the following statistical model

$$\begin{array}{rcl} y_j & = & h(X_{t_j}, \theta) + g(X_{t_j}, \theta)\, \varepsilon_j, \quad \varepsilon_j \sim_{i.i.d.} \mathcal{N}(0, I_q) \\ dX_t & = & \mu(X_t, t, \theta)dt + \Gamma(X_t, \theta)dB_t \\ X_0 & = & x_0(\theta) \end{array} \tag{10}$$

where $h(x, \theta)$ and $g(x, \theta)$ are two functions from $\mathbb{R}^p \times \mathbb{R}^d$ in $\mathbb{R}^q$ and ε_j is a measurement error random q-vector. For the sake of simplicity, we consider the usual assumption of a Gaussian distribution on ε_j with identity variance matrix I_q. Standard choices for g are $g(x, \theta) = \sigma$ resulting in the *homoscedastic model* and $g(x, \theta) = (a + \sigma h(x, \theta))$ resulting in the *heteroscedastic model*. Function h is generally the identity function when X_t in one-dimensional and a linear application modeling that only a restricted number or a linear combination of components of $(X_t)_{t\geq 0}$ is observed, otherwise.

The likelihood function $\mathcal{L}(\theta; y_{1:n})$ computed from data $y_{1:n}$ is the marginal distribution $p_{\underline{Y}}$:

$$\mathcal{L}(\theta; y_{1:n}) = p_{\underline{Y}}(y_{1:n}|\theta) = \int_{\mathcal{X}} p_{\underline{Y}|\underline{X}}(y_{1:n}|X_{1:n}, \theta) p_{\underline{X}}(X_{1:n}|\theta) dX_{1:n}, \tag{11}$$

where $p_{\underline{Y}|\underline{X}}$ is the conditional density of $y_{1:n}$ given $X_{1:n}$. If the transition density is Gaussian and $h(X, \theta)$ is linear, $\mathcal{L}(\theta; y_{1:n})$ has an explicit form (see Section 4.2.1). Otherwise, either the transition density, or the integration over X are not explicit. Then the intractable likelihood $\mathcal{L}(\theta; y_{1:n})$ has to be approximated. These methods are presented in Section 4.2.2.

4 Estimation for PK/PD SDE

We first present estimation methods adapted to directly observed PK/PD SDE (section 4.1). Then estimation for SDE observed with noise is detailed in section 4.2.

4.1 Discretely but directly observed diffusions

Likelihood (9) is explicit for few SDEs and analytic estimators have then been derived, as presented in Section 4.1.1. Approximate estimators suited for general SDEs are reviewed in Section 4.1.2.

4.1.1 Explicit or exact estimators

Ferrante *et al.* (2003) consider the stochastic one compartment PK model (2). They introduce the ideal maximum likelihood estimator (MLE) based on continuous-time observed trajectory $(C_t)_{t\in[0,T]}$. As data belong to an infinite dimensional space, the continuous-time observation likelihood has to be computed with the Girsanov formula (Lipster and Shiryaev, 2001):

$$\mathcal{L}(\theta; (C_t)_{t\in[0,T]}) = \exp\left\{ -\frac{1}{\gamma^2} \left[\int_0^T (k_a/V - k_e C_t) dC_t - \frac{1}{2} \int_0^T (k_a/V - k_e C_t)^2 dt \right] \right\}$$

where $\theta = (k_e, k_a, V, \gamma)$. The MLE based on continuous data $(C_t)_{t\in[0,T]}$ is defined as $\hat{\theta}_{[0,T]} = \arg\max_{\theta\in\Theta}\{\mathcal{L}(\theta; (C_t)_{t\in[0,T]})\}$. It can be proved that $\hat{\theta}_{[0,T]}$ is asymptotically unbiased, normally distributed and efficient (Prakasa Rao, 1999). However, PK/PD data are measured at discrete

times rather than continuously. Discretisation of $\hat{\theta}_{[0,T]}$ can be considered but it is generally difficult to study the properties of such estimators analytically, especially when the number of data is small and fixed. As an alternative, Ferrante *et al.* (2003) consider the likelihood (9) based on discrete-time observations $C_{1:n}$. They can compute it exactly because PK model (2) has an explicit transition density. More precisely, the process (C_t) is Gaussian and the transition density is

$$p_C(y, t-s|z, \theta) = P(C_t = y|C_s = z, \theta) = \frac{1}{\sqrt{2\pi v_\theta(t-s)}} \exp\left[-\frac{(y - m_\theta(z, t-s))^2}{2v_\theta(t-s)}\right],$$

where $m_\theta(z, t-s) = \frac{k_a}{Vk_e} + (z - \frac{k_a}{Vk_e}) \exp(-k_e(t-s))$ and $v_\theta(t-s) = \frac{\gamma^2}{2k_e} 1 - \exp(-2k_e(t-s))$. Then the exact likelihood function is $\mathcal{L}(\theta; C_{1:n}) = \prod_{j=1}^{n} p_C(C_j, \Delta_j|C_{j-1}, \theta)$. The MLE based on discrete-time observations is defined as

$$\hat{\theta}_n = \arg\max_{\theta \in \Theta} \mathcal{L}(\theta; C_{1:n}).$$

He is not explicit but can be obtained from a numerical optimization of the likelihood. The MLE $\hat{\theta}_n$ is asymptotic normal and efficient (Lipster and Shiryaev, 2001).

In Ferrante *et al.* (2000, 2005), the same estimation method is used for the PK/PD SDE (6) to model bacterial growth and bacteria-drug interactions. The log transformation $X_t = \log N_t$ of SDE (6) has a Gaussian explicit solution. The transition density $p_X(y, t-s|z)$ is thus explicit and the exact MLE is derived from the likelihood function (9) of the discrete data $(X_{1:n}, C_{1:n})$.

4.1.2 Estimators based on approximation

Transition density of more complex PK/PD SDEs is unknown and approximation has to be used to estimate parameters. Several approaches have been proposed. The simplest one is the Gaussian Euler-Maruyama approximation (Oksendal, 2007) but this scheme is efficient only if $\Delta_j = t_j - t_{j-1}$ is small, which is not usual for PK/PD data. We present here two other strategies suited for PK/PD data, one based on an Hermite expansion of the transition density and the second based on the spectral density. This second approach has been introduced for an hypoelliptic SDE. To our knowledge, this is the only estimation method adapted to hypoelliptic SDE.

Hermite expansion of the transition density. This estimation method has been first used in a PK/PD SDE context by Picchini *et al.* (2008). This approach has been originally proposed by Aït-Sahalia (2002) for unidimensional and time-homogeneous equations and then extended to inhomoegeneous equations by Egorov *et al.* (2003) and to multi-dimensional equations by Aït-Sahalia (2008). Picchini *et al.* (2008) consider a one-dimensional ($q = 1$) inhomogeneous SDE of glucose dynamics after the steady state of insulin concentration has been reached. The transition density is not explicit, neither the likelihood. Their estimation method works as follows.

They assume the volatility function $\Gamma(x, \theta)$ to be bounded below by a strictly positive function and introduce the Lamberti transform of X_t: $\ell(X_t) = \int^{X_t} \frac{du}{\Gamma(u,\theta)}$. The Hermite expansion p_X^S of order S of the unknown transition density p_X is defined by

$$p_X^S(X_j, \Delta_j|X_{j-1}, \theta) = \frac{1}{\sqrt{\Delta_j}\Gamma(X_j, \theta)} f_0(Z_j) \sum_{s=0}^{S} m_s\left(\ell(X_{j-1}, \Delta_j), \theta\right) H_s(Z_j), \tag{12}$$

where H_s is the Hermite polynom of order s for $s = 1 \dots S$, f_0 is the standard normal density

function, $Z_j = \frac{\ell(X_j)-\ell(X_{j-1})}{\sqrt{\Delta_j}}$ and $m_s\left(\ell(X_{j-1}), \Delta_j, \theta\right) = \frac{1}{s!} E\left[H_s\left(Z_j\right) | \ell(X_{j-1})\right]$ is the conditional moment of the Hermite polynom H_s. Then the approximated likelihood of order S is

$$\mathcal{L}^S(\theta, X_{1:n}) = \prod_{j=1}^{n} p_X^S(X_j, \Delta_j | X_{j-1}, \theta) \tag{13}$$

and the estimator of θ is defined as

$$\hat{\theta}_n^S = \arg\max_{\theta \in \Theta} \mathcal{L}^S(\theta, X_{1:n}).$$

Aït-Sahalia (2002) and Egorov *et al.* (2003) prove that under regularity conditions on the drift and volatility functions, $\hat{\theta}_n^S$ is consistent. This method is computationally efficient.

Monte Carlo approximation of the spectral density. Ditlevsen *et al.* (2005, 2007) propose this method for a hypoelliptic PD SDE modeling tubuloglomerular feedback mechanism in a rat nephron. The SDE system consists in 6 equations with Brownian motion on only one coordinate. There is no analytic expression for the transition density and so for the likelihood function.

Instead of estimating the parameters by maximizing an approximation of the likelihood function, Ditlevsen *et al.* (2005) propose an estimator based on the spectral density $\tilde{p}(u, \theta)$, where u is a frequency in the Fourier domain, the spectral density characterizing uniquely the stochastic process. An estimation of θ is obtained by minimizing the distance between the theoretical spectral density $\tilde{p}(u, \theta)$ of the SDE and the estimated spectral density $\hat{p}(u)$ obtained from the data by the periodogram at the Fourier frequency u. The exact spectral density $\tilde{p}(u, \theta)$ of the PD model having no explicit form, the authors propose to approximate it by simulation. If K independent trajectories are simulated with a Euler-Maruyama approximation of the SDE for a fixed parameter value θ, a periodogram $\tilde{p}_k(u, \theta)$ can be estimated for each trajectory k and the spectral density $\tilde{p}(u, \theta)$ is then approximated by Monte Carlo:

$$\tilde{p}^K(u, \theta) = \frac{1}{K} \sum_{k=1}^{K} \tilde{p}_k(u, \theta).$$

The strong law of large numbers ensures convergence of $\tilde{p}^K(u, \theta)$ to $\tilde{p}(u, \theta)$. Then the parameters are estimated by nonlinear least squares as

$$\tilde{\theta}_n^K = \arg\min_{\theta \in \Theta} \sum_{u \in [0, n/(2T)]} \left[\log \hat{p}(u) - \log \tilde{p}^K(u, \theta)\right]^2.$$

Approximate confidence intervals on $\tilde{\theta}_n^K$ can also be obtained by Monte Carlo simulations. No theoretical properties have been proved for this estimator.

4.2 Discretely observed diffusions with observation noise

Noisy observations of PK/PD SDEs as defined by model (10) have been considered in the literature the restriction of a constant diffusion coefficient $\Gamma(X_t, \theta) = \Gamma_\theta$. We first present an exact estimation method proposed for linear SDEs (Section 4.2.1), then two approximate estimation methods for more general SDEs (Section 4.2.2).

4.2.1 Explicit or exact estimators

Linear PK/PD SDEs with homoscedastic measurement noise have been studied by D'Argenio and Park (1997); Tornøe *et al.* (2004a); Favetto and Samson (2010); Cuenod *et al.* (2011). Model (10) reduces to:

$$\begin{aligned} y_j &= H_\theta X_j + \sigma\varepsilon_j, \quad \varepsilon_j \sim_{iid} \mathcal{N}(0, I_q), \quad j = 1, \ldots, n \\ dX_t &= (A_\theta X_t + C_\theta(t))dt + \Gamma_\theta dB_t \end{aligned} \tag{14}$$

setting $h(X_t, \theta) = H_\theta X_t$, with H_θ a $q \times p$ matrix, $g(X_t, \theta) = \sigma$, $\mu(X_t, t, \theta) = A_\theta X_t + C_\theta(t)$. The p-vector $C_\theta(t)$ may depend on time t.

By recursive conditioning, likelihood (11) can be decomposed into

$$\mathcal{L}(\theta; y_{1:n}) = p(y_1; \theta) \prod_{j=2}^{n} p(y_j | y_{1:j-1}; \theta), \tag{15}$$

where $p(y_j|y_{1:j-1};\theta)$ is the conditional law of y_j given $y_{1:j-1}$. In the particular case of model (14), $y_j|y_{1:j-1}$ is Gaussian. Indeed, the solution of the linear SDE (14) is Gaussian, as well as y_j by linearity of $h(x, \theta)$. Let us denote $m_{j|1:j-1}$ and $R_{j|1:j-1}$ the conditional expectation and variance of y_j given $y_{1:j-1}$ respectively. We have:

$$p(y_j|y_{1:j-1};\theta) = \frac{\exp\{-\frac{1}{2}(y_j - {}^t m_{j|1:j-1})(R_{j|1:j-1})^{-1}(y_j - m_{j|1:j-1})\}}{\sqrt{|2\pi R_{j|1:j-1}|}}.$$

Then $m_{j|1:j-1}$ and $R_{j|1:j-1}$ can be computed using the Kalman filtering procedure. Indeed, by linearity of $h(x, \theta)$, we have:

$$\begin{aligned} m_{j|1:j-1} &= E(y_j|y_{1:j-1}) &&= H_\theta E(X_j|y_{1:j-1}) \\ R_{j|1:j-1} &= Var(y_j|y_{1:j-1}) &&= H_\theta\, Var(X_j|y_{1:j-1})\, {}^t H_\theta + \sigma^2 I_q \end{aligned} \tag{16}$$

where ${}^t H_\theta$ is the transposed matrix of H_θ. Consequently, $m_{j|1:j-1}$ and $R_{j|1:j-1}$ rely on the predictive expectations and variances:

$$\hat{X}_{j|1:j-1} = E(X_j|y_{1:j-1}) \quad \text{and} \quad P_{j|1:j-1} = E((X_j - \hat{X}_{j|1:j-1})^t (X_j - \hat{X}_{j|1:j-1}))$$

whose computations are details below.

Predictive moments. Noting that $(X_t)_{t \geq t_j}$ verifies $X_t = \int_{t_{j-1}}^t \mu(X_s, s, \theta)ds + \Gamma_\theta \int_{t_j}^t dBs$, for any $t \geq t_{j-1}$ yields to the following integral equation for the predictive conditional expectation:

$$E(X_t|y_{1:j-1}) = \int_{t_{j-1}}^{t} E(\mu(X_s, s, \theta)|y_{1:j-1})ds. \tag{17}$$

Consequently, drift μ being linear for model (14), $x_t := E(X_t|y_{1:j-1})$ verifies $x_t = \int_{t_{j-1}}^t \mu(x_s, s, \theta)ds$, which can be written in a differential form as

$$\begin{aligned} dx_t/dt &= \mu(x_t, t, \theta) \\ x_{t_{j-1}} &= E(X_{t_{j-1}}|y_{1:j-1}) = \hat{X}_{j-1|1:j-1}. \end{aligned} \tag{18}$$

Using similar arguments, we also obtain an ODE verified by $P_t := Var(X_t|y_{1:j-1})$ for $t \geq t_{j-1}$.

Let J_μ denote the Jacobian of μ w.r.t. x, we get

$$\begin{aligned} dP_t/dt &= J_\mu(x_t,t,\theta)P_t + P_t\,{}^t J_\mu(x_t,t,\theta) + \Gamma_\theta\,{}^t\Gamma_\theta \\ P_{t_{j-1}} &= Var(X_{t_{j-1}}|y_{1:j-1}) = P_{j-1|1:j-1}. \end{aligned} \tag{19}$$

Because of the linearity of μ, equations (18-19) have explicit solutions. Moreover, noting that $\hat{X}_{j|1:j-1} = x_{t_j}$ and $P_{j|1:j-1} = P_{t_j}$, we obtain

$$\begin{aligned} \hat{X}_{j|1:j-1} &= e^{A_\theta \Delta_j}\hat{X}_{j-1|1:j-1} + \tilde{C}_\theta(t_{j-1},\Delta_j) \\ P_{j|1:j-1} &= e^{A_\theta \Delta_j}P_{j-1|1:j-1}e^{\,{}^t A_\theta \Delta_j} + \tilde{\Gamma}_\theta(t_{j-1},\Delta_j) \end{aligned} \tag{20}$$

where $\hat{X}_{j-1|1:j-1} = E(X_{j-1}|y_{1:j-1})$ and $P_{j-1|1:j-1} = Var(X_{j-1}|y_{1:j-1})$ are the filtering moments and $\tilde{C}_\theta(t_{j-1},\Delta_j)$, $\tilde{\Gamma}_\theta(t_{j-1},\Delta_j)$ are given in Tornøe *et al.* (2004a) and Favetto and Samson (2010).

Filtering moments. A similar reasoning yields the update (filtering) step:

$$\begin{aligned} \hat{X}_{j|1:j} &= \hat{X}_{j|1:j-1} + K_j(y_j - H_\theta \hat{X}_{j|1:j-1}), \\ P_{j|1:j} &= (I - K_j H_\theta)P_{j|1:j-1}, \end{aligned}$$

where $K_j = P_{j|1:j-1}\,{}^t H_\theta (H_\theta P_{j|1:j-1}\,{}^t H_\theta + \sigma^2)^{-1}$ is the Kalman gain.

Finally, iterating the computations of predictive and filtering moments (the so-called Kalman filtering) leads to the exact likelihood (15) and the MLE is then defined as

$$\hat{\theta}_n = \arg\max_{\theta\in\Theta}\ \mathcal{L}(\theta; y_{1:n}).$$

Favetto and Samson (2010) prove that $\hat{\theta}_n$, which is the exact MLE, is consistent and asymptotically Gaussian. This estimation method is implemented in the Danish Technical University project CTSM (Kristensen *et al.*, 2001).

This estimator has been used in a PK context by Tornøe *et al.* (2004a) with a stochastic PK/PD model describing the dynamic of insulin and glucose from an euglycaemic clamp study. The PK model of insulin kinetic is a single compartment with linear rates. The insulin effect is modeled by adding an effect compartment. The PK part contains three equations describing the amount of insulin remaining to be absorbed from the tissue, the amount of insulin in the central compartment and the amount of insulin in the effect compartment. The PD of the amount of infused glucose is linked to the insulin concentration with a sigmoidal E_{max} model. The estimation results show that the diffusion coefficients are all estimated to zero, which indicates that the measured data can eventually be modeled by a deterministic PK/PD model.

Another example is given by Cuenod *et al.* (2011). They consider the stochastic two-compartment PK model (4) to describe tissue micro-vascularization and angiogenesis via Dynamic Contrast Enhanced Imaging (DCE-imaging) techniques. Results illustrate that the stochastic PK model is more stable than its deterministic version.

***Remark* 1** *Model (10) with explicit transition density for X_t is a state space model or a hidden Markov model (Cappé* et al.*, 2005). Corresponding estimation methods have not yet been all applied to PK/PD data.*

4.2.2 Approximate estimators

When either the transition density or likelihood $\mathcal{L}(\theta; y_{1:n})$ is not explicit, an approximation has to be proposed. A first solution relies on an extension of the Kalman filter and has been introduced by Tornøe *et al.* (2004b). Another approach was proposed by Ditlevsen and De Gaetano (2005b) with an Monte Carlo approximation of the likelihood to estimate the volatility parameters, the drift parameters being estimated using the deterministic version of the model.

Extended Kalman Filter. This approach, based on a generalization of the Kalman filter, has been proposed by Tornøe *et al.* (2004b) for the general model (10) with the restriction of time-homogeneous drift $\mu(X_t, t, \theta) = \mu(X_t, \theta)$:

$$\begin{array}{rcl} y_j & = & h(X_{t_j}, \theta) + \sigma\, \varepsilon_j, \quad \varepsilon_j \sim_{i.i.d.} \mathcal{N}(0, I_q) \\ dX_t & = & \mu(X_t, \theta)dt + \Gamma_\theta dB_t \\ X_0 & = & x_0(\theta) \end{array} \tag{21}$$

Once again, the likelihood can be decomposed into $\mathcal{L}(\theta; y_{1:n}) = p(y_1; \theta) \prod_{j=2}^{n} p(y_j | y_{1:j-1}; \theta)$. However, whereas each conditional distribution $p(y_j | y_{1:j-1}; \theta)$ is Gaussian for linear model (14), it is not true for model (21). The idea of the extended Kalman Filter (EKF) is thus to approximate this conditional distribution by a Gaussian distribution:

$$p(y_j | y_{1:j-1}, \theta) \simeq \frac{\exp\{-\frac{1}{2}(y_j - {}^t m_{j|1:j-1})(R_{j|1:j-1})^{-1}(y_j - m_{j|1:j-1})\}}{\sqrt{|2\pi R_{j|1:j-1}|}}$$

where, as for linear model, $m_{j|1:j-1} = E(y_j | y_{1:j-1}) = E(h(X_j, \theta) | y_{1:j-1})$ and $R_{j|1:j-1} = Var(y_j | y_{1:j-1}) = Var(h(X_j, \theta) | y_{1:j-1}) + \sigma^2$. But in that case, h being not linear, $m_{j|1:j-1}$ and $R_{j|1:j-1}$ are not explicit. Thus, EKF considers a local linearisation of h around $E(X_j | y_{1:j-1})$. More precisely, let J_h be the Jacobian function of h with respect to x, we can write $h(X_j, \theta) \simeq h(E(X_j | y_{1:j-1}), \theta) + J_h(E(X_j | y_{1:j-1}), \theta)(X_j - E(X_j | y_{1:j-1}))$ and approximate the conditional moments by:

$$\begin{array}{rcl} m_{j|1:j-1} & \simeq & h(\hat{X}_{j|1:j-1}, \theta) \\ R_{j|1:j-1} & \simeq & J_h(\hat{X}_{j|1:j-1}, \theta) P_{j|1:j-1}\, {}^t J_h(\hat{X}_{j|1:j-1}, \theta) + \sigma^2 \end{array} \tag{22}$$

where, as before, $\hat{X}_{j|1:j-1}$ and $P_{j|1:j-1}$ are the conditional predictive expectation and variance.

The difficulty for the nonlinear model (21) arises in the computation of these predictive quantities. Using integral equation (17), one would like to deduce a solution for $\hat{X}_{j|1:j-1}$, as in the linear case. However, μ being not linear, another approximation is used, writting:

$$E(\mu(X_s, \theta) | y_{1:j-1}) \simeq \mu(E(X_s | y_{1:j-1}), \theta).$$

Then, one can say that an approximation of $x_t := E(X_t | y_{1:j-1})$ is a solution of differential equation (18). Similarly, using an approximation of the variance, $P_t := Var[X_t | y_{1:j-1}]$ can be approximate by a solution of differential equation (19). However, except for linear drift μ, equations (18-19) have no close solution. A new approximation is performed in EKF, consisting in a linearization of the ODEs using Taylor expansion of μ. In Tornøe *et al.* (2004b), a first order Taylor expansion is

considered, leading to the following approximate predictive quantities

$$\begin{aligned}\hat{X}_{j|1:j-1} &= \hat{X}_{j-1|1:j-1} + \mu(\hat{X}_{j-1|1:j-1}, \theta)\Delta_j \\ P_{j|1:j-1} &= P_{j-1|1:j-1} + \Delta_j \left(J_\mu(\hat{X}_{j-1|1:j-1}, \theta) P_{j-1|1:j-1} + P_{j-1|1:j-1} J_\mu(\hat{X}_{j-1|1:j-1}, \theta) + {\Gamma_\theta}^t \Gamma_\theta \right)\end{aligned}$$

***Remark* 2** *Delattre and Lavielle (submitted) propose a higher order Taylor approximation of equations equations (18-19) (see Section 5.3.3).*

The update step is performed using similar Gaussian approximation and linearization:

$$\begin{aligned}\hat{X}_{j|1:j} &= \hat{X}_{j|1:j-1} + K_j(y_j - h(\hat{X}_{j|1:j-1}, \theta)) \\ P_{j|1:j} &= (I - K_j J_h(\hat{X}_{j|1:j-1}, \theta)) P_{j|1:j-1}\end{aligned}$$

where K_j is the approximated Kalman gain

$$K_j = {P_{j|1:j-1}}^t J_h(\hat{X}_{j|1:j-1}, \theta)(J_h(\hat{X}_{j|1:j-1}, \theta) {P_{j|1:j-1}}^t J_h(\hat{X}_{j|1:j-1}, \theta) + \sigma^2)^{-1}.$$

Finally, the EKF supplies an iterative procedure to approximate the likelihood, and maximize it yields an estimator of θ. This method is implemented in the CTSM project. However, to our knowledge, no theoretical results have been proved rigorously for this estimator, which can behave badly when both SDE and observation model are highly nonlinear.

This estimation method has been used for PK/PD models. For example, Tornøe *et al.* (2004b) use PK/PD SDE (7) on simulated data to illustrate the properties of the estimator. Then, for the modeling of intravenous glucose tolerance test with a glucose minimal model (GMM), they use a nonlinear two-dimensional SDE with two compartments (glucose and insulin). The observation is partial and consists in discrete noisy measure of the glucose rate. Data are analyzed both with the SDE and the deterministic GMM. The diffusion term is estimated significantly different from zero. This indicates that the deterministic GMM does not fully capture the glucose/insulin dynamics. Another example is given in Kristensen *et al.* (2005) for PK/PD model development.

Monte Carlo approximation of the likelihood. Ditlevsen and De Gaetano (2005b) consider this method with the hypoelliptic two-dimensional PK SDE (5) for the modeling of the uptake of dodecanedoic acid. A diffusion coefficient appears only on one of the two equations. Only the concentration C_1 in the first compartment is observed, with a measurement noise. Ditlevsen and De Gaetano (2005b) consider the heteroscedastic error model

$$y_j = C_{1,t_j} \varepsilon_j$$

where ε_j are iid log-normal variables with variance $\mathrm{var}(\log \varepsilon_j) = \sigma^2$.

The hypoellipticity of PK SDE (5) makes estimation more complex. Indeed, contrary to the two-dimensional PK SDE considered by Favetto and Samson (2010), this hypoelliptic linear SDE has no explicit solution or transition densities. Therefore, the likelihood is not available. Ditlevsen and De Gaetano (2005b) thus propose a two-step estimation method. As the deterministic corresponding PK model has an explicit solution, they estimate PK parameters k_1, k_2 from the deterministic PK model by usual nonlinear least squares estimators. Then, with these parameters fixed, they proceed as a second step to the estimation of γ and σ. However, even when k_1, k_2 are fixed, the

likelihood

$$\mathcal{L}(\gamma, \sigma; y_{1:n}) = \int \prod_{j=1}^{n} p(y_j|C_{1,t_j}, \sigma) p(C_{1,t_1}, \ldots, C_{1,t_n}; \gamma) dC_{1,t_1} \ldots dC_{1,t_n}$$

has no close form due to the hypoellipticity of (5). Therefore, they approximate the likelihood by Monte-Carlo simulations, as first proposed by Pedersen (1995). A large number K of trajectories $(C^{(k)}_{1,1:n})_{k=1,\ldots,K}$ are simulated given γ with the Milstein scheme from SDE (5). Then the likelihood is approximated by

$$\mathcal{L}^K(\gamma, \sigma; y_{1:n}) = \frac{1}{K} \sum_{k=1}^{K} \prod_{j=1}^{n} p(y_j|C^{(k)}_{1,t_j}, \sigma)$$

and numerically maximized. Estimators of γ, σ are asymptotically equivalent to the MLE. The same approach is used in Picchini *et al.* (2006) for modeling the euglycemic hyperinsulinemic clamp. This approach is computationally intensive: $K = 500.000$ was required to obtain a good approximation of the likelihood. It is also more appropriate to estimate all parameters in a single optimization step.

5 SDE for population PK/PD models

PK-PD studies often include not a single individual but a group of subjects. In the first papers considering SDE models on a population of subjects, individuals were treated independently without proposing any global model (see for instance Picchini *et al.*, 2006; Ditlevsen and De Gaetano, 2005b). However, the success of mixed effects models – which allow to discriminate between the inter and intra subjects variability– has encouraged their use for PK/PD SDEs.

We first introduce the population PK/PD SDE in Section 5.1. The estimation methods for population PK/PD SDEs observed without measurement noise are presented in Section 5.2 and for population PK/PD SDEs observed with measurement noise in Section 5.3.

5.1 Population PK/PD SDE

Assume that PK/PD data are available from I subjects. Let $(X^{(i)})_{t\geq 0} \in \mathbb{R}^p$ be the process of interest for subject i, for $i = 1, \ldots I$. We assume that the processes $\{(X^{(i)}_t)_{t\geq 0}, i = 1 \ldots I\}$ verify the same SDE but with individual parameters ϕ_i:

$$\begin{array}{rcll} dX^{(i)}_t & = & \mu(X^{(i)}_t, t, \phi_i, \eta)dt + \Gamma(X^{(i)}_t, \eta)dB^{(i)}_t & X^{(i)}_0 = X_0(\phi_i) \\ \phi_i & \sim_{i.i.d.} & p_\Phi(\cdot; \theta_{pop}), & \end{array}$$

where $\eta \in \mathbb{R}^{d_0}$ is a parameter common to all the individuals, $\phi_i \in \mathbb{R}^d$ are the individual random parameters, also called random effects, distributed with a density p_Φ depending on a population parameter $\theta_{pop} \in \Theta$ and $(B^{(i)}_t)$ are I independent Brownian motions. A standard choice for p_Φ is a Gaussian distribution

$$\phi_i \sim_{i.i.d.} \mathcal{N}(\mu, \Omega)$$

with mean $\mu \in \mathbb{R}^d$ and variance Ω a definite positive matrix of size d. In that case, $\theta_{pop} = (\mu, \Omega)$. However, other distributions can be considered, such as Gamma distributions to ensure the positivity of the parameter, or discrete distributions. The parameters to be estimated are

$\theta = (\theta_{pop}, \eta)$ from direct discrete observations of $X_t^{(i)}$ or noisy discrete observations.

Statistical inference for such models is a challenging issue. Some of the proposed methodologies derive from the non-linear mixed effects models literature whereas Picchini *et al.* (2010) extend their method developed for the observation of one diffusion (Section 4.1.2) to the observation of several individual diffusions. Obviously, the difficulty is not the same if the transition density is explicit or not and if the process is observed directly or with observation noise.

5.2 Population PK/PD SDE without observation noise

We first assume that the processes $(X_t^{(i)})_{t\geq 0}$, for $i = 1, \ldots, I$ are *directly* observed on the I individuals at discrete times t_{ij}, for $j = 1, \ldots n_i$. Set $\Delta_{ij} = t_{ij} - t_{ij-1}$. Let $X_{1:n_i}^{(i)} = (X_{t_{i1}}^{(i)}, \ldots X_{t_{in_i}}^{(i)})$ denote the observations of individual i and $\mathbf{X} = (X_{1:n_i}^{(i)})_{i=1\ldots I}$, the vector of whole data. By independence of the I individuals, the likelihood is:

$$\mathcal{L}(\theta; \mathbf{X}) = \prod_{i=1}^{I} \mathcal{L}_i(\theta; X_{1:n_i}^{(i)}) = \prod_{i=1}^{I} \int_{\Phi} p_{\underline{X}}(X_{1:n_i}^{(i)} | \phi_i, \eta) p_{\Phi}(\phi_i | \theta_{pop}) d\phi_i. \tag{23}$$

where $p_{\underline{X}}(X_{1:n_i}^{(i)} | \phi_i, \eta) = \prod_{j=1}^{n_i} p_X(X_j^{(i)}, \Delta_{ij} | X_{j-1}^{(i)}, \phi_i, \eta)$ similarly as equation (9).

In a few cases, the likelihood (23) can be expressed in a close form, leading to the exact MLE of θ (section 5.2.1). Otherwise, (i) if the transition density $p_X(x, t-s|x_s, \phi, \gamma)$ has an explicit solution but integrating out the ϕ_i's in (23) is unfeasible; or (ii) if the transition density can not be expressed explicitly, likelihood (23) has no close form and approximate solutions have to be proposed. In case (i), the integral has to be approximated (see Section 5.2.2) whereas in case (ii), the transition density must be approximated and maybe also the integral (see Section 5.2.3).

5.2.1 Exact estimator

Ditlevsen and De Gaetano (2005a) consider the one-compartment linear elimination PK SDE (1) with one random effect:

$$\begin{aligned} dC_t^{(i)} &= -k_{e,i} C_t^{(i)} dt + \gamma C_t^{(i)} dB_t^{(i)} \\ k_{e,i} &\sim \mathcal{N}(k_e, \omega^2). \end{aligned} \tag{24}$$

In that case, $\phi_i = k_{e,i}$, $\theta_{pop} = (k_e, \omega^2)$, $\eta = \gamma^2$ and $\theta = (k_e, \omega^2, \gamma^2)$. Data are discrete observations $C_{t_{ij}}^{(i)}$ for $i = 1, \ldots, I$ and $j = 1, \ldots, n_i$. Using a log-transformation of the data, the likelihood (23) can be computed exactly. An estimation method adapted from linear mixed model (Pinheiro and Bates, 2000) is applied. Namely, the likelihood is splitted into two parts, that are independent of and dependent of the random effect ϕ_i, respectively. The integral of the dependent part is the integral of a Gaussian density because SDE (24) is linear and has a log-normal solution. The explicit expression of the likelihood and of the exact MLE of the parameter θ follows. The MLE is Gaussian, and Ditlevsen and De Gaetano (2005a) give explicit variance of the MLE, leading to explicit confidence intervals.

5.2.2 Estimation for explicit transition density SDE

Excepted for particular SDEs such as (24), even if the transition density $p_X(x, t-s|x_s, \phi, \gamma)$ has an explicit expression, the difficulty to estimate θ comes from the integration of the individual random

effects density. 1 strategy is to approximate the likelihood, that is the integral with respect to the random effects density, and then use a standard maximization procedure. This has been proposed by Picchini *et al.* (2010) with a Gaussian quadrature of the likelihood and by Picchini and Ditlevsen (2011) with a Laplace approximation of the likelihood. The first method has been applied to the one-compartment linear elimination PK SDE with one random effect (24) and the second one to the stochastic two-compartment PK model (4) with random effects on k_1, k_2, k_e. We now detail the two techniques.

Gaussian quadrature. We restrict to the case of a one-dimensional Gaussian individual parameter ϕ ($\phi_i \sim_{i.i.d.} \mathcal{N}(\mu, \omega^2)$), we refer the reader to Picchini *et al.* (2010) for the case of any continuous distribution. Assume that $\phi \mapsto p_{\underline{X}}(X^{(i)}_{1:n_i}|\phi, \eta)$ is $\mathcal{C}^{2R}(\mathbb{R})$. Each individual likelihood $\mathcal{L}_i(\theta; X^{(i)}_{1:n_i}) = \int_\Phi p_{\underline{X}}(X^{(i)}_{1:n_i}|\phi_i, \eta) p_\Phi(\phi_i|\theta_{pop}) d\phi_i$ is approximated by the Gauss-Hermite quadrature of order R:

$$\mathcal{L}_i^R(\theta; X^{(i)}_{1:n_i}) = \sum_{r=1}^{R} \pi_r \, p_{\underline{X}}(X^{(i)}_{1:n_i}|\omega\sqrt{2}z_r + \mu, \eta)$$

where $z_r, r = 1, \ldots, R$ are the zeros of the Hermite polynomial $H_R(\cdot)$ of degree R and $\pi_r = \frac{2^{R-1}R!}{R^2(H_{R-1}(z_r))^2}$ are adequate weights. Neither the zeros z_r nor the weights π_r depend on the individual. The approximate likelihood is then defined as

$$\mathcal{L}^R(\theta; \mathbf{X}) = \prod_{i=1}^{I} \mathcal{L}_i^R(\theta; X^{(i)}_{1:n_i}) = \prod_{i=1}^{I} \sum_{r=1}^{R} \pi_r \, p_{\underline{X}}(X^{(i)}_{1:n_i}|\omega\sqrt{2}z_r + \mu, \eta) \tag{25}$$

The convergence of $\mathcal{L}^R(\theta; \mathbf{X})$ towards $\mathcal{L}(\theta; \mathbf{X})$ is ensured when R tends to infinity and the domain of integration is compact. The parameter θ can be estimated by maximizing this approximate likelihood

$$\widehat{\theta}_n^R \quad = \quad (\widehat{\eta}_n^R, \widehat{\omega}_n^R) = \arg\min_{\theta\in\Theta} \left\{ -\sum_{i=1}^{I} \log \sum_{r=1}^{R} \pi_r \, p_{\underline{X}}(X^{(i)}_{1:n_i}|\omega\sqrt{2}z_r + \mu, \eta) \right\}$$

The estimator $\widehat{\theta}_n^R$ is obtained by a standard optimization algorithm. No theoretical convergence has been proved for this estimator.

Laplace approximation. When the individual parameters are of dimension d greater than 1, Picchini and Ditlevsen (2011) use a Laplace approximation of the likelihood. More precisely, the log-likelihood $\log \mathcal{L}_i(\theta; X^{(i)}_{1:n_i})$ is approximated by

$$\log \mathcal{L}_i^L(\theta; X^{(i)}_{1:n_i}) = \log p_{\underline{X}}(X^{(i)}_{1:n_i}|\hat{\phi}_i, \eta) + \log p_\Phi(\hat{\phi}_i|\theta_{pop}) + \frac{d}{2}\log(2\pi) - \frac{1}{2}\log|-H(\hat{\phi}_i|\theta)|$$

where $\hat{\phi}_i = \arg\max_{\phi_i} \left(\log p_{\underline{X}}(X^{(i)}_{1:n_i}|\phi_i, \eta) + \log p_\Phi(\phi_i|\theta_{pop}) \right)$, $|\cdot|$ denotes the determinant of a matrix and $H(\phi_i|\theta)$ is the Hessian matrix:

$$H(\phi_i|\theta) = \partial^2 \left(\log p_{\underline{X}}(X^{(i)}_{1:n_i}|\phi_i, \eta) + \log p_\Phi(\phi_i|\theta_{pop}) \right) / \partial\phi_i \partial^t \phi_i.$$

Then the approximate likelihood is numerically optimized. This methods requires the computation of the I Hessian matrices $H(\phi_i|\theta)$ at each step of the optimization, which can be burdensome. The authors give in their paper various numerical and practical solutions to tackle that point.

5.2.3 Estimation for unknown transition density SDE

When the transition density has no explicit expression, Picchini *et al.* (2010) propose to approximate it with a Hermite expansion, as detailed in Section 4.1.2. Then this approximate transition density is injected in the approximate marginal distribution (25).

More precisely, the transition likelihood is approximated by $p_X^S(X_j^{(i)}, \Delta_{ij}|X_{j-1}^{(i)}, \phi_i, \eta)$ defined by equation (12) where S is the order of the Hermite expansion. The conditional likelihood $p_{\underline{X}}(X_{1:n_i}^{(i)}|\phi_i, \eta)$ of subject i can be approximated by:

$$p_{\underline{X}}^S(X_{1:n_i}^{(i)}|\phi_i, \eta) = \prod_{j=2}^{n_i} p_X^S(X_j^{(i)}, \Delta_{ij}|X_{j-1}^{(i)}, \phi_i, \eta) \tag{26}$$

exactly as equation (13). Then, the integration over the individual parameters ϕ_i is obtained using the Gauss-Hermite quadrature (25). Finally, the approximate likelihood is:

$$\mathcal{L}^{R,S}(\theta; \mathbf{X}) = \prod_{i=1}^{I} \sum_{r=1}^{R} \pi_r \, p_{\underline{X}}^S(X_{1:n_i}^{(i)}|\omega\sqrt{2}z_r + \mu, \eta).$$

As before, the MLE estimator of θ is obtained applying an optimization routine on the approximate likelihood:

$$\widehat{\theta}_n^{R,S} = \arg\min_{\theta\in\Theta} \left\{ -\log \mathcal{L}^{R,S}(\theta; \mathbf{X}) \right\}$$

No theoretical result has been proved but Picchini and Ditlevsen (2011) illustrate the efficiency of $\widehat{\theta}_n^{R,S}$ on several uni- and multi-dimensional examples (note that the multidimensionality comes from both the dimension of ϕ and the dimension of the diffusion X_t), especially for the stochastic two-compartment PK model (4) with random effects on k_1, k_2, k_e.

5.3 Population SDE models with measurement error

We now consider that the processes $X_t^{(i)}$, for $i = 1, \ldots, I$, are observed with a measurement error. Let $y_j^{(i)}$ denote the j-th observation of subject i at time t_{ij}. The model is the following:

$$\begin{array}{rcll} y_j^{(i)} & = & h(X_{t_{ij}}^{(i)}) + g(X_{t_{ij}}^{(i)}, \sigma)\varepsilon_{ij}, & \varepsilon_{ij} \sim \mathcal{N}(0, I_q) \\ dX_t^{(i)} & = & \mu(X_t^{(i)}, t, \phi_i, \eta)dt + \Gamma(X_t^{(i)}, \eta)dB_t^{(i)}, & X_0^{(i)} = X_0(\phi_i) \\ \phi_i & \sim & p_\Phi(\cdot; \theta_{pop}) & \end{array} \tag{27}$$

The parameters to be estimated are $\theta = (\theta_{pop}, \eta, \sigma)$. Let us denote $\mathbf{y} = (y_{1:n_i}^{(i)})_{i=1\ldots I}$, the whole vector of data. The likelihood is:

$$\mathcal{L}(\theta; \mathbf{y}) = \prod_{i=1}^{I} \int_\Phi \left(\int_{\mathcal{X}} p_{\underline{Y}}(y_{1:n_i}^{(i)}|X_{1:n_i}^{(i)}, \sigma) p_{\underline{X}}(X_{1:n_i}^{(i)}|\phi_i, \eta) dX_{1:n_i}^{(i)} \right) p_\Phi(\phi_i|\theta_{pop}) d\phi_i \tag{28}$$

where $p_{\underline{Y}}(y_{1:n_i}^{(i)}|X_{1:n_i}^{(i)}, \sigma) = \prod_{j=1}^{n_i} p(y_j^{(i)}|X_j^{(i)}, \sigma)$ is the conditional density of the observations y given the diffusion X. We also denote $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_I)$.

Difficulties pointed out with direct observations of $X_t^{(i)}$ still hold but are exacerbated by the necessity to integrate over the hidden trajectories $X_{1:n_i}^{(i)}$. Suggested solutions are derived from the ones presented in Section 4.2 and combined with estimation methods developed for nonlinear mixed effects models. We first present, in Section 5.3.1, a Bayesian approach, proposed by Donnet

et al. (2010). Then we detail how the extended Kalman Filter (EKF) can be coupled with the First Order Conditional Estimate (FOCE) algorithm to estimate θ, as first proposed by Overgaard *et al.* (2005); Tornøe *et al.* (2005) (Section 5.3.2). Finally, estimation methods based on the Expectation-Mazimization (EM) algorithm, proposed by Donnet and Samson (2008, submitted); Delattre and Lavielle (submitted), are gathered in Section 5.3.3.

5.3.1 Bayesian inference

In a Bayesian inference, a prior distribution is set on the parameters of interest $\theta \sim \pi(\cdot)$. The aim is to estimate the posterior distribution of θ given the observation $\mathbf{y}$, which is given by the Bayes formula $p(\theta|\mathbf{y}) \propto f(\mathbf{y};\theta)\pi(\theta)$. Estimators of θ are then obtained with the posterior mean $E(\theta|\mathbf{y})$ or the median of the posterior distribution, for instance. Except in very few cases, the posterior distribution $p(\theta|\mathbf{y})$ has no explicit expression. A standard strategy is to generate a sample from that distribution of interest, using a Monte Carlo Markov Chain (MCMC) algorithm (Robert and Casella, 2004). This approach has been studied for modeling growth curves with a population Gompertzian SDE by Donnet *et al.* (2010).

The principle of the MCMC algorithm is to generate a Markov Chain whose marginal stationary distribution is the distribution of interest $p(\theta|\mathbf{y})$. In model (27), the simplest MCMC algorithm is to alternatively generate, at iteration k:

$$[1].\ \boldsymbol{\phi}^{(k)}|\mathbf{X}^{(k-1)},\mathbf{y},\theta^{(k-1)} \qquad [2].\ \mathbf{X}^{(k)}|\boldsymbol{\phi}^{(k)},\mathbf{y},\theta^{(k-1)} \qquad [3].\ \theta^{(k-1)}|\mathbf{X}^{(k)},\boldsymbol{\phi}^{(k)},\mathbf{y},$$

with stationary distribution $p(\boldsymbol{\phi},\mathbf{X},\theta|\mathbf{y})$. One can prove that the stationary distribution of the Markov Chain produced by this algorithm is $p(\boldsymbol{\phi},\mathbf{X},\theta|\mathbf{y})$ (Robert and Casella, 2004). As a consequence, after a high number of iterations, the sampled parameters $\theta^{(k)}$ are assumed to be distributed under the marginal posterior distribution $p(\theta|\mathbf{y})$. However, in population PK-PD SDE, the conditional distributions $p(\boldsymbol{\phi}|\mathbf{X},\theta,\mathbf{y})$, $p(\mathbf{X}|\boldsymbol{\phi},\mathbf{y},\theta)$ and $p(\theta|\mathbf{X},\boldsymbol{\phi},\mathbf{y})$ are not explicit. Thus, the direct simulation is replaced by a Metropolis-Hastings algorithm, that is, a candidate is simulated with a *proposal distribution* q and accepted with a probability whose expression ensures the stationarity of the distribution of interest $p(\boldsymbol{\phi},\mathbf{X},\theta|\mathbf{y})$. For instance, the Metropolis-Hastings algorithm for step [1] which generates $\boldsymbol{\phi}^{(k)}|\mathbf{X}^{(k-1)},\mathbf{y},\theta^{(k-1)}$ at iteration k of the MCMC algorithm is

- Propose a candidate $\boldsymbol{\phi}^c$ with a proposal distribution $\boldsymbol{\phi}^c \sim q(\cdot|\boldsymbol{\phi}^{(k-1)})$
- Compute the probability of acceptation:

$$\rho(\boldsymbol{\phi}^c|\boldsymbol{\phi}^{(k-1)}) \quad = \quad \min\left\{1, \frac{p(\boldsymbol{\phi}^c|\theta^{(k)},\mathbf{X}^{(k-1)},\mathbf{y})}{p(\boldsymbol{\phi}^{(k-1)}|\theta^{(k)},\mathbf{X}^{(k-1)},\mathbf{y})}\frac{q(\boldsymbol{\phi}^{(k-1)}|\boldsymbol{\phi}^c)}{q(\boldsymbol{\phi}^c|\boldsymbol{\phi}^{(k-1)})}\right\}$$

- Set

$$\boldsymbol{\phi}^{(k)} = \begin{cases} \boldsymbol{\phi}^c & \text{with probability } \rho(\boldsymbol{\phi}^c|\boldsymbol{\phi}^{(k-1)}) \\ \boldsymbol{\phi}^{(k-1)} & \text{with probability } 1-\rho(\boldsymbol{\phi}^c|\boldsymbol{\phi}^{(k-1)}) \end{cases}$$

Similar algorithms are proposed to simulate steps [2] and [3]. Metropolis-Hastings does not require the exact expression of the targeted distribution. Thanks to the Bayes formula, only the expression of $p_{\underline{Y}}(y^{(i)}_{1:n_i}|X^{(i)}_{1:n_i},\sigma)$, $p_{\underline{X}}(X^{(i)}_{1:n_i}|\phi_i,\eta)$ and $p_\Phi(\phi_i|\theta_{pop})$ are used. We refer the reader to Donnet *et al.* (2010) for more details.

This algorithm can be used as soon as the transition density $p(\mathbf{X}^{(k-1)}|\boldsymbol{\phi}^{(k-1)}, \theta^c)$ of the SDE is explicit. If not, Donnet *et al.* (2010) propose to perform the Bayesian inference on an approximate model issued from the Euler-Maruyama approximation of the solution of the SDE. They study the distance between the approximate posterior distribution $p_h(\theta|\mathbf{y})$, where h is the step size of the Euler-Maruyama approximation, and the exact one $p(\theta|\mathbf{y})$. This distance is bounded by h.

5.3.2 Linearisation and extended Kalman filter

Overgaard *et al.* (2005); Tornøe *et al.* (2005) propose to adapt the well-known First Order Conditional Estimate (FOCE) algorithm to population SDEs. FOCE has been introduced by Lindstrom and Bates (1990) for nonlinear mixed models with deterministic regression function. The algorithm is based on a local linearization of the regression function, leading to a linear mixed model, for which estimation of parameters is more easy. Overgaard *et al.* (2005); Tornøe *et al.* (2005) use this idea, coupled with the extended Kalman filter (EKF) already mentioned in Section 4.2.2. Their methodology has been used in a collection of papers, see for instance Mortensen *et al.* (2007); Overgaard *et al.* (2007); Klim *et al.* (2009); Berglund *et al.* (2011). We detail their approach below.

Following the idea of recursive conditioning used for standard SDEs, the likelihood $\mathcal{L}(\theta; \mathbf{y})$ (28) can be re-written as:

$$\mathcal{L}(\theta; \mathbf{y}) = \prod_{i=1}^{I} \mathcal{L}_i(\theta; y_{1:n_i}^{(i)}) = \prod_{i=1}^{I} \int_{\Phi} \left[\prod_{j=1}^{n_i} p(y_j^{(i)}|y_{1:j-1}^{(i)}, \phi_i, \sigma, \eta) \right] p_{\Phi}(\phi_i|\theta_{pop}) d\phi_i$$

where $y_{1:j}^{(i)} = (y_1^{(i)}, \ldots, y_j^{(i)})$ and $y_0^{(i)} = x_0(\phi_i)$ by convention. As in Section 4.2.2, the EKF approximates $p(y_j^{(i)}|y_{1:j-1}^{(i)}, \phi_i, \eta)$ by a Gaussian distribution, whose mean $m_{j|1:j-1}^{(i)}$ and variance $R_{j|1:j-1}^{(i)}$ are computed recursively (see equation (22)) and depend on ϕ_i, σ and η. The approximate likelihood for subject i is:

$$\begin{aligned} \mathcal{L}_i^{EKF}(\theta; y_{1:n_i}^{(i)}) &= \int_{\Phi} \prod_{j=1}^{n_i} \frac{\exp\{-\frac{1}{2}(y_j^{(i)} - m_{j|1:j-1}^{(i)})^T (R_{j|1:j-1}^{(i)})^{-1} (y_j^{(i)} - m_{j|1:j-1}^{(i)})\}}{\sqrt{|2\pi R_{j|1:j-1}^{(i)}|}} p_{\Phi}(\phi_i|\theta_{pop}) d\phi_i \\ &= \int_{\Phi} e^{\ell_i(\phi_i)} d\phi_i \end{aligned}$$

where $\ell_i(\phi_i)$ is the approximate conditional log-density of the random effects given the observation (recall that $m_{j|1:j-1}^{(i)}$ and $R_{j|1:j-1}^{(i)}$ depend on ϕ_i). Then ℓ_i is approximated by a second-order Taylor expansion, where the expansion is made around the value $\hat{\phi}_i$ that maximizes ℓ_i (FOCE algorithm). At the value $\hat{\phi}_i$, the first derivative of ℓ_i is zero. So finally, the likelihood approximated by EKF and by the Taylor expansion reduces to

$$\mathcal{L}^{EKF,FOCE}(\theta, \mathbf{y}) = \prod_{i=1}^{I} \frac{1}{\sqrt{2\pi}} e^{\ell_i(\hat{\phi}_i)} |-H(\ell_i)|^{-1/2}$$

where $H(\ell_i)$ is the Hessian of ℓ_i. The Hessian is not computed exactly, but with an approximation similar to FOCE (see Pinheiro and Bates, 2000, for more details). Once that approximation has been done, the maximization of $\mathcal{L}^{EKF,FOCE}(\theta, \mathbf{y})$ is performed by FOCE. This methodology suffers from the drawbacks inherent to EKF and FOCE, and no theoretical convergence has been proved. This algorithm is implemented in NONMEM (Tornøe *et al.*, 2005), in a Matlab package (Mortensen

et al., 2007) and in the R package PSM (Klim *et al.*, 2009).

5.3.3 EM algorithm for mixed effects models defined by SDE

Another approach to estimate parameters of mixed effects models is based on the Expectation-Maximization (EM) algorithm, originally proposed by Dempster *et al.* (1977). Stochastic versions of EM have been adapted to nonlinear mixed models (Walker, 1996; Kuhn and Lavielle, 2005). Donnet and Samson (2008, submitted) and Delattre and Lavielle (submitted) propose to extend it to population SDEs. This approach has been applied to a one-compartment PK model (2) with three random effects and measurement noise. We now detail their method.

The EM algorithm is a way to circumvent the problem of integrating the likelihood (28) over the non-observed data $\mathbf{X}$ and $\boldsymbol{\phi}$, in situations where the maximization of the conditional expectation of the likelihood of the complete data $(\mathbf{y}, \mathbf{X}, \boldsymbol{\phi})$ is possible:

$$Q(\theta|\theta') = E\left[\log p(\mathbf{y}, \mathbf{X}, \boldsymbol{\phi}; \theta)|\mathbf{y}; \theta'\right].$$

EM is an iterative procedure: at the ℓ-th iteration, given the current value $\widehat{\theta}_{\ell-1}$ of the parameters, the E-step evaluates $Q_\ell(\theta) = Q(\theta \,|\, \widehat{\theta}_{\ell-1})$ while the M-step updates $\widehat{\theta}_{\ell-1}$ by maximizing $Q_\ell(\theta)$. This algorithm is not explicit for population SDEs: the E-step is not explicit and the conditional distribution $p(\mathbf{X}, \boldsymbol{\phi}|\mathbf{y}; \theta)$ is not explicit. A stochastic version of the EM algorithm has been proposed by Kuhn and Lavielle (2005), the Stochastic Approximation EM algorithm (SAEM), for nonlinear mixed model with deterministic regression function by Kuhn and Lavielle (2005) and for mixed model based on SDEs by Donnet and Samson (2008, submitted) and Delattre and Lavielle (submitted).

The SAEM algorithm for population SDE (27) proceeds as follows. The E-step is divided into a simulation step (S-step) of the non-observed data $(\mathbf{X}^{(\ell)}, \boldsymbol{\phi}^{(\ell)})$ with the conditional distribution $p(\mathbf{X}, \boldsymbol{\phi} \,|\mathbf{y}; \widehat{\theta}_{\ell-1})$ and a stochastic approximation step (SA-step):

$$Q_\ell(\theta) = Q_{\ell-1}(\theta) + \alpha_\ell \left[\log p(\mathbf{y}, \mathbf{X}^{(\ell)}, \boldsymbol{\phi}^{(\ell)}; \theta) - Q_{\ell-1}(\theta)\right],$$

where $(\alpha_\ell)_{\ell \in \mathbb{N}}$ is a sequence of positive numbers decreasing to zero. The simulation step can not be performed exactly, for the reasons exposed in Section 5.3.1. When the transition density $p_X(x, t-s|x_s, \phi, \eta)$ is explicit, Donnet and Samson (2008) propose to use a MCMC algorithm similar to the one used in the Bayesian framework. An improvement has also been proposed more recently by Donnet and Samson (submitted) based on the Particle MCMC (PMCMC) algorithm proposed by Andrieu *et al.* (2010), which combined the MCMC algorithm with particle filter techniques. The theoretical convergence of both algorithms SAEM-MCMC and SAEM-PMCMC have been established if the complete log-likelihood belongs to the exponential family (with respect to the parameter θ).

When the transition density is not explicit, Donnet and Samson (2008) propose to approximate the solution of the SDE with the Euler-Maruyama scheme. More precisely, if the time interval between two observation points is small, the transition density $p_X(x, t-s|x_s, \phi, \eta)$ can be approximated by a Gaussian distribution. If the time intervals between two observation instants are too large to obtain a good approximation of the transition density, a natural approach is to introduce a set of auxiliary latent data points between every pair of observations, as first proposed by Pedersen (1995). Under general assumptions, Donnet and Samson (2008) prove that the SAEM-

MCMC algorithm theoretically converges to a maximum of the likelihood of the Euler-Maruyama approximate model. The error between the estimator issued from this approximate model and the "true" model (27) can be bound by a function of the step size of the Euler-Maruyama approximation. However, in practice the use of a Euler-Maruyama scheme implies the introduction of a large amount non-observed data, which can slow the convergence rate.

An alternative has been proposed by Delattre and Lavielle (submitted) with an approximation of the E-step of EM by the EKF algorithm. Their SAEM algorithm is based on the simulation of only the individual parameters ϕ_i under the conditional distribution $p(\phi_i|y^{(i)};\theta)$, the diffusion trajectories are not simulated and are directly integrated out in the conditional distribution $p(\phi_i|y^{(i)};\theta)$. Since direct simulation of this conditional distribution is generally impossible, a MCMC algorithm is used with a Metropolis-Hastings approach. However, the acceptance probability computed in Metropolis-Hastings requires the knowledge of the expression of $p(y^{(i)}|\phi_i,\sigma)$ which has no close form. Delattre and Lavielle (submitted) propose a Gaussian approximation of this conditional likelihood, based on the EKF. The implementation of the EKF is similar to (22) but with a higher order Taylor expansion for the prediction equations. This algorithm is much more less computationally intensive than those proposed by Donnet and Samson (2008, submitted) but the theoretical convergence of their estimator is not proved.

6 Conclusion and discussion

Due to the specific nature of PK/PD data, standard estimation methods for SDE models proposed in the financial literature can not be directly applied. In this paper, we restrict our review to PK/PD oriented estimation methods, eliminating those relying on continuous observation of the SDE. Moreover, dynamical systems used in biology are quickly of high complexity as soon as a precise description of the biological processes of interest is considered. As a consequence, ad hoc estimation methods have been proposed in the literature, sometimes suffering from a lack of theoretical justifications but always with practical objectives.

In most papers, estimators are obtained by maximizing the likelihood. If the transition density of the SDE is explicit, exact MLE can be computed, using Kalman filter if SDE is observed with measurement noise. Otherwise, if the transition density is not explicit, the likelihood has to be approximated. We can find in the literature easy to implement methods with theoretical validity (Hermite expansion) when the SDE is directly observed, easy to implement methods but without theoretical validity (extended Kalman filter) and methods with theoretical validity but computationally intensive (Monte Carlo approximation).

In a population context, there exist four estimation methods: Hermite expansion of the marginal likelihood, the combination of FOCE with the extended Kalman filtering and the use of a stochastic version of the EM algorithm and Bayesian approach.

Note that Hermite expansion can not be directly extended when the diffusion is observed with noise whereas Kalman filter, Extended Kalman filter and FOCE methods are strictly designed for models with observation noise. Methods relying on the EM algorithm can be applied to models with or without noise, provided the complete likelihood belongs to the exponential family. Bayesian approach is finally the most adaptable method since it can be applied to any model (one or several trajectories, with or without noise observation) as soon as the transition density is explicit. If not, a Euler approximation has to be performed and the introduction of set of auxiliary latent data

points between every pair of observations can dramatically jeopardize the convergence rate of the algorithm of simulation. We propose in Table 6 a resume of existing methods for population SDE models.

[Table 1 about here.]

References

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

Aït-Sahalia, Y. (2008). Closed-form likelihood expansions for multivariate diffusions. *Ann. Statist.* **36**, 906–937.

Andrieu, C., Doucet, A. and Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *J. R. Statist. Soc. B* pp. 1–33.

Berglund, M., Sunnåker, M., Adiels, M., Jirstrand, M. and Wennberg, B. (2011). Investigations of a compartmental model for leucine kinetics using non-linear mixed effects models with ordinary and stochastic differential equations. *Math Med Biol* .

Cappé, O., Moulines, E. and Ryden, T. (2005). *Inference in Hidden Markov Models (Springer Series in Statistics)*. Springer-Verlag New York, USA.

Cuenod, C., Favetto, B., Genon-Catalot, V., Rozenholc, Y. and Samson, A. (2011). Parameter estimation and change-point detection from dynamic contrast enhanced mri data using stochastic differential equations. *Math. Biosci.* **233**, 1–76.

D'Argenio, D. and Park, K. (1997). Uncertain pharmacokinetic/pharmacodynamic systems: design, estimation and control. *Control. Eng. Practice* **5**, 1707–1716.

Delattre, M. and Lavielle, M. (submitted). Coupling the saem algorithm and the extended kalman filter for maximum likelihood estimation in mixed-effects diffusion models .

Dempster, A., Laird, N. and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Jr. R. Stat. Soc. B* **39**, 1–38.

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

Ditlevsen, S. and De Gaetano, A. (2005b). Stochastic vs. deterministic uptake of dodecanedioic acid by isolated rat livers. *Bull. Math. Biol.* **67**, 547–561.

Ditlevsen, S. and Samson, A. (2012). Introduction to stochastic models in biology. In *Bachar, Batzel and Ditlevsen (Eds.), Stochastic Biomathematical Models with Applications to the Insulin-Glucose System and Neuronal Modeling*. Springer.

Ditlevsen, S., Yip, K. and Holstein-Rathlou, N. (2005). Parameter estimation in a stochastic model of the tubuloglomerular feedback mechanism in a rat nephron. *Math. Biosc.* **194**, 49–69.

Ditlevsen, S., Yip, K.-P., Marsh, D. J. and Holstein-Rathlou, N.-H. (2007). Parameter estimation of feedback gain in a stochastic model of renal hemodynamics: differences between spontaneously hypertensive and sprague-dawley rats. *Am J Physiol Renal Physiol* **292**, 607–616.

Donnet, S., Foulley, J. and Samson, A. (2010). Bayesian analysis of growth curves using mixed models defined by stochastic differential equations. *Biometrics* **66**, 733–741.

Donnet, S. and Samson, A. (2008). Parametric inference for mixed models defined by stochastic differential equations. *ESAIM P&S* **12**, 196–218.

Donnet, S. and Samson, A. (submitted). EM algorithm coupled with particle filter for maximum likelihood parameter estimation of stochastic differential mixed-effects models. *Prepublication MAP5 2010-24* .

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

Favetto, B. and Samson, A. (2010). Parameter estimation for a bidimensional partially observed Ornstein-Uhlenbeck process with biological application. *Scand. J. Statist.* **7**, 200–220.

Ferrante, L., Bompadre, S. and Leone, L. (2003). A stochastic compartmental model with long lasting infusion. *Biometrical journal* **45**, 182–194.

Ferrante, L., Bompadre, S., Leone, L. and Montanari, M. (2005). A stochastic formulation of the gompertzian growth model for in vitro bactericidal kinetics: parameter estimation and extinction probability. *Biometrical journal* **47**, 309–318.

Ferrante, L., Bompadre, S., Possati, L. and Leone, L. (2000). Parameter estimation in a gompertzian stochastic model for tumor growth. *Biometrics* **56**, 1076–1081.

Klim, S., Mortensen, S. B., Kristensen, N., Overgaard, R. and Madsen, H. (2009). Population stochastic modelling (PSM) – An R package for mixed-effects models based on stochastic differential equations. *Computer methods and programs in biomedicine* **94**, 279–289.

Kristensen, N., Madsen, H. and Ingwersen, S. (2005). Using stochastic differential equations for PK/PD model development. *J. Pharmacokinet. Pharmacodyn.* **32**, 109–141.

Kristensen, N., Melgaard, H. and Madsen, H. (2001). *CTSM 2.0, UserÕs Guide.* http://www.imm.dtu.dk/CTSM.

Kuhn, E. and Lavielle, M. (2005). Maximum likelihood estimation in nonlinear mixed effects models. *Comput. Statist. Data Anal.* **49**, 1020–1038.

Lindstrom, M. and Bates, D. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics* **46**, 673–87.

Lipster, R. and Shiryaev, A. (2001). *Statistics of random processes I : general theory.* Springer.

Mortensen, S., Klim, S., Dammann, B., Kristensen, N., Madsen, H. and Overgaard, R. (2007). A matlab framework for estimation of NLME models using stochastic differential equations. applications for estimation of insulin secretion rates. *J. Pharmacokinet. Pharmacodyn.* **34**, 623–642.

Oksendal, B. (2007). *Stochastic differential equations: an introduction with applications.* Springer-Verlag, Berlin-Heidelberg.

Overgaard, R., Jonsson, N., Tornøe, C. and Madsen, H. (2005). Non-linear mixed-effects models with stochastic differential equations: Implementation of an estimation algorithm. *J Pharmacokinet. Pharmacodyn.* **32**, 85–107.

Overgaard, R. V., Holford, N., Rytved, K. A. and Madsen, H. (2007). PKPD model of interleukin-21 effects on thermoregulation in monkeys–application and evaluation of stochastic differential equations. *Pharm Res* **24**, 298–309.

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

Picchini, U., De Gaetano, A. and Ditlevsen, S. (2010). Stochastic differential mixed-effects models. *Scand. J. Statist.* **37**, 67–90.

Picchini, U. and Ditlevsen, S. (2011). Practicle estimation of high dimensional stochastic differential mixed-effects models. *Comput. Statist. Data Anal.* **55**, 1426–1444.

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

Picchini, U., Ditlevsen, S. and De Gaetano, A. (2008). Maximum likelihood estimation of a time-inhomogeneous stochastic differential model of glucose dynamics. *Math. Med. Biol.* **25**, 141–155.

Pinheiro, J. and Bates, D. (2000). *Mixed-effect models in S and Splus.* Springer-Verlag.

Prakasa Rao, B. (1999). *Statistical Inference for Diffusion Type Processes.* Arnold Publisher.

Ramanathan, M. (1999a). An application of ito's lemma in population pharmacokinetics and pharmacodynamics. *Pharm Res* **16**, 584–586.

Ramanathan, M. (1999b). A method for estimating pharmacokinetic risks of concentration-dependent drug interactions from preclinical data. *Drug Metabolism and disposition* **27**, 1479–1487.

Robert, C. P. and Casella, G. (2004). *Monte Carlo statistical methods.* Springer Texts in Statistics. Springer-Verlag, New York, second edn. ISBN 0-387-21239-6.

Sørensen, H. (2004). Parametric inference for diffusion processes observed at discrete points in time: A survey. *International Statistical Review* **72**, 337–354.

Tornøe, C., Overgaard, R., Agersø, H., Nielsen, H., Madsen, H. and Jonsson, E. (2005). Stochastic differential equations in NONMEM: implementation, application, and comparison with ordinary differential equations. *Pharm. Res.* **22**, 1247–58.

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

Tornøe, C. W., Jacobsen, J. L., Pedersen, O., Hansen, T. and Madsen, H. (2004b). Grey-box modelling of pharmacokinetic/pharmacodynamic systems. *J Pharmacokinet Pharmacodyn* **31**, 401–417.

Walker, S. (1996). An EM algorithm for non-linear random effects models. *Biometrics* **52**, 934–944.

Obs. Noise	Without		With	
Explicit Transition	With	Without	With	Without
Hermite approximation Picchini *et al.* (2010)	X	X		
Stochastic EM	X		X	
Stochastic EM with Euler approximation		X		X
FOCE			X	
FOCE with Extended Kalman Filtering				X
Bayesian Inference	X		X	
Bayesian Inference with Euler approximation		X		X

Table 1: SDE for population PK/PD models: a syntetic resume