HAL
open science

Coupling the SAEM algorithm and the extended Kalman filter for maximum likelihood estimation in mixed-effects diffusion models

Maud Delattre, Marc Lavielle

To cite this version:

Maud Delattre, Marc Lavielle. Coupling the SAEM algorithm and the extended Kalman filter for maximum likelihood estimation in mixed-effects diffusion models. Statistics and Its Interface, 2013, 6 (4), pp.519–532. hal-00916803

HAL Id: hal-00916803

https://hal.science/hal-00916803

Submitted on 12 Dec 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Statistics and Its Interface Volume 6 (2013) 519–532

Coupling the SAEM algorithm and the extended Kalman filter for maximum likelihood estimation in mixed-effects diffusion models*

Maud Delattre and Marc Lavielle†

We consider some general mixed-effects diffusion models, in which the observations are made at discrete time points and include measurement errors. In these models, the observed likelihood is generally not explicit, making maximum likelihood estimation of the parameters particularly complex. We propose a specific inference methodology for these models. In particular, we combine the SAEM algorithm with the extended Kalman filter to estimate the population parameters. We also provide some tools for estimating the individual parameters, for recovering the individual underlying diffusion trajectories and for evaluating the model. The methods are evaluated on simulations and applied to a pharmacokinetics example.

Keywords and phrases: Stochastic differential equations, Mixed-effects models, SAEM, Extended Kalman filter.

1. INTRODUCTION

Mixed-effects models are standard tools for simultaneously describing repeated measurements taken on a series of individuals. Their structure allows a suitable consideration of the whole variability characterizing such data, which is usually split into intra-individual variability – i.e. the variability occurring within the dynamics of each individual — and between-subjects variability. In a mixed-effects model, the same structural model is used for describing each individual sequence of observations, but the parameters of this model vary randomly among the individuals, which allows to better account for differences between subjects. In a mixed-effects diffusion model, the description of each individual series of observations is based on stochastic differential equations (SDE). Diffusion is known to be a relevant tool for modeling stochastic dynamic phenomena, and is widely used in various fields including finance, physics, biology, physiology, control engineering, etc. When an experiment includes several individuals, a straightforward extension of the individual diffusion model is the mixed-effects diffusion model, which also takes into account variability between individuals. So far, the main applications of such population approaches with SDE-based models are in neuroscience [13, 22] and in pharmacokinetics and pharmacodynamics (PK & PD) [9, 21]. Although many methods are available for the inference in classical fixed-effects diffusion models, there is still a need for a general, fast and easy-to-implement method for inference in mixed-effects diffusion models. Indeed, except in some very specific classes of mixed-effects diffusion models [6, 9], the likelihood of the observations does not have any closed-form expression, making maximum likelihood estimation of the model parameters rather intricate. The difficulty is twofold for computing the observed likelihood since it involves the transition densities of the underlying individual diffusion processes and integrals over the unobserved individual parameters that can rarely be computed in closed form.

Several authors have recently tackled maximum likelihood estimation of parameters in mixed-effects diffusion models, either in the case of noise-free or noisy observations of the diffusion. One of the main approaches developed for these purposes consists in approximating the likelihood of the observations and then maximizing the approximated likelihood with respect to the parameters. A specific combination of the First-Order Conditional Estimation (FOCE) method with the extended Kalman filter has for example been suggested [15, 21, 25]. Other approximations of the likelihood in mixed-effects diffusion models can be found in [23] and [22]. An alternative to such methods is given by EM-type algorithms which iteratively perform maximum likelihood estimation based on the complete log-likelihood rather than the marginal log-likelihood. In particular, specific versions of the SAEM algorithm have been proposed for estimating the population parameters in mixed-effects diffusion models. In [10] the SAEM algorithm is combined with an Euler-Maruyama approximation of the individual processes and in [11], it is coupled with some particle Markov Chain Monte-Carlo methods. In these two versions of SAEM however, simulation of both the random individual parameters and the individual latent processes is required at the

*The research leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement n. 115156, resources of which are composed of financial contributions from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies in kind contribution. The DDMoRe project is also financially supported by contributions from Academic and SME partners.
†Corresponding author.

simulation step, which is computationally cumbersome. In the present work, we propose a new inference methodology for mixed-effects diffusion models which consists in coupling the SAEM algorithm with the extended Kalman filter for estimating the population parameters. In this new version of the SAEM algorithm, we only need to simulate the individual parameters at each iteration. We also provide tools for estimating the individual parameters and the individual diffusion trajectories. We also show that Visual Predictive Checks can be used as a diagnostic tool for detecting misspecifications of the model.

The organization of the present paper is as follows. In Section 2, we introduce and motivate mixed-effects diffusion models. In Section 3, we detail our new estimation methodology based on the extended Kalman filter and the SAEM algorithm. Some numerical illustrations of the proposed methods are provided in Section 4 through a brief Monte-Carlo simulation study and an example on a PK application. Section 5 summarizes the results and discusses the properties of the proposed methodology.

2. MODEL

2.1 Fixed-effects diffusion model

Models involving diffusion processes are relevant statistical tools for describing random variability in dynamical systems. They have numerous applications, including the description of the price movement of financial assets, neuronal or population growth, among others.

In the fixed-effects version, these models assume that one trajectory of a diffusion is discretely observed with noise at discrete time points denoted $t_1 < \cdots < t_j < \cdots < t_n$. Let us note $(X(t), t > 0) \in \mathbb{R}^d$ the underlying dynamical process and $y_j \in \mathbb{R}$ the noisy observation of $X(t_j)$, $j = 1, \ldots, n$. The general form of the diffusion model is given by:

$$(1) \qquad \begin{cases} dX(t) = b(X(t), \theta)dt + \gamma(X(t), \theta)dW(t), \\ y_j = g(X(t_j), \theta) + \xi_j, \\ \xi_j \underset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2(\theta)), \quad j = 1, \ldots, n \end{cases}$$

with the initial condition $X(t_1) = x \in \mathbb{R}^d$. Here, $(W(t), t > 0)$ is a standard Wiener process in $\mathbb{R}^d$ and $\xi_j \in \mathbb{R}$ represents the measurement error occurring at the jth observation such that ξ_j and $(W(t), t > 0)$ are independent. The measurement function $g(x, \theta) : \mathbb{R}^d \times \mathbb{R}^p \to \mathbb{R}$, the drift function $b(x, \theta) : \mathbb{R}^d \times \mathbb{R}^p \to \mathbb{R}^d$ and the diffusion function $\gamma(x, \theta) : \mathbb{R}^d \times \mathbb{R}^p \to \mathcal{M}_d(\mathbb{R})$, where $\mathcal{M}_d(\mathbb{R})$ is the set of $d \times d$ matrices with real elements, are known functions depending on an unknown parameter $\theta \in \mathbb{R}^p$.

Estimation of the unknown parameter θ from the observations is challenging. Indeed, an exact evaluation of the likelihood and its maximization with respect to θ are generally intractable. $y_1, \ldots, y_n$ are not independent since they depend on a Markovian latent process. The general expression for the observed likelihood is given by:

$$(2) \qquad p(y, \theta) = p(y_1; \theta) \prod_{j=2}^{n} p(y_j | y_1, \ldots, y_{j-1}; \theta),$$

where $y = (y_1, \ldots, y_n)$ denotes the sequence of observations. Each conditional density in (2) is expressed as an integral over the latent process which rarely has a closed form expression. As a consequence, most diffusion-specific estimation procedures are based on an approximation of (2). One classical and widely-used method is the extended Kalman filter (EKF) which consists in approximating the observations' distribution with a Gaussian distribution by linearization of the state and measurement equations along the diffusion trajectories [14]. More recently, Markov chain Monte Carlo methods such as sequential Monte Carlo methods have been proposed [2, 5, 12]. We refer the reader to [20] for a precise comparison between Monte Carlo methods and Kalman filtering when inferring diffusion models.

2.2 Mixed-effects diffusion model

Let us now consider model (1) with observations coming from several subjects. An adequate adaptation of model (1) in such a context consists of considering as many dynamical systems as individuals, and defining the parameters of the individual dynamical systems as independent random variables, in such a way to correctly reflect the variability between the different trajectories. To standardize notation, we consider N different subjects randomly chosen from a population and we note n_i the number of observations for individual i, so that $t_{i1} < \cdots < t_{i,n_i}$ are subject i's observation time points, $i = 1, \ldots, N$. $(X_i(t), t > 0) \in \mathbb{R}^d$ and $y_{ij} \in \mathbb{R}$ will respectively denote individual i's diffusion and the observation of $X_i(t_{ij})$. The y_{ij}, $i = 1, \ldots, N$, $j = 1, \ldots, n_i$ are governed by a mixed-effects model based on a d-dimensional real-valued system of stochastic differential equations with the general form:

$$(3) \qquad \begin{cases} dX_i(t) = b(X_i(t), \phi_i)dt + \gamma(X_i(t), \phi_i)dW_i(t), \\ y_{ij} = g(X_i(t_{ij}), \phi_i) + \xi_{ij}, \\ \xi_{ij} \underset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2(\phi_i)), \quad j = 1, \ldots, n_i, \ i = 1, \ldots, N, \end{cases}$$

with initial condition $X_i(t_1) = x_{i1} \in \mathbb{R}^d$, $i = 1, \ldots, N$. The ϕ_i are unobserved independent p-dimensional random subject-specific parameters, drawn from a distribution π which depends on a set of parameters $\theta \in \mathbb{R}^q$, called population parameters:

$$\phi_i \sim \pi(\cdot, \theta).$$

$(W_1(t), t > 0), \ldots, (W_N(t), t > 0)$ are standard independent Wiener processes such that $\phi_1, \ldots, \phi_N$ and $(W_1(t), t > 0), \ldots, (W_N(t), t > 0)$ are independent. The ξ_{ij} are independent Gaussian random variables representing residual

errors such that ξ_{ij} and $\phi_1, \ldots, \phi_N$ are independent and ξ_{ij} and $(W_1(t), t > 0), \ldots, (W_N(t), t > 0)$ are too. The measurement function $g(x, \phi) : \mathbb{R}^d \times \mathbb{R}^p \to \mathbb{R}$, the drift function $b(x, \phi) : \mathbb{R}^d \times \mathbb{R}^p \to \mathbb{R}^d$ and the diffusion function $\gamma(x, \phi) : \mathbb{R}^d \times \mathbb{R}^p \to \mathcal{M}_d(\mathbb{R})$, are known functions that are common to the N subjects and depend on the unknown parameters ϕ_i.

As before, the observed likelihood has a nontrivial form. By independence of the N individuals, the joint probability distribution function (pdf) of the observations is given by:

$$p(y_1, \ldots, y_N; \theta) = \prod_{i=1}^{N} p(y_i; \theta). \tag{4}$$

In any mixed-effects model, the marginal density of the ith data vector $y_i = (y_{i1}, \ldots, y_{i,n_i})$ is obtained by integrating the conditional density of the data given the non-observable random effects ϕ_i with respect to the density of the individual parameters:

$$p(y_i; \theta) = \int p(y_i|\phi_i; \theta)\pi(\phi_i; \theta)d\phi_i, \tag{5}$$

with

$$p(y_i|\phi_i; \theta) = p(y_{i1}|\phi_i; \theta) \prod_{j=2}^{n_i} p(y_{ij}|y_{i1}, \ldots, y_{i,j-1}, \phi_i; \theta).$$

It is worth noting that equation (5) gives a generic expression of the marginal distribution of the observations, which is valid for any conditional distribution of the observations given the individual parameters and for any distribution of the random parameters. In most cases, the ϕ_i's are assumed to be Gaussian random variables, with mean μ and covariance matrix Ω (see the illustrative Section of the present paper for example). In this specific example, the set of population parameters would be $\theta = (\mu, \Omega)$. The extension to models including covariates, such as

$$\phi_i = \beta C_i + \eta_i, \quad \eta_i \underset{i.i.d.}{\sim} \mathcal{N}(0, \Omega),$$

is easy, where β is a $p \times K$ matrix of fixed-effects, C_i a K-vector of covariates and η_i a p-dimensional random-effect, $i = 1, \ldots, N$.

3. SOME INFERENCE METHODOLOGY FOR MIXED-EFFECTS DIFFUSION MODELS

3.1 Maximum likelihood estimation of the population parameters

The important issue in mixed-effects models is to evaluate both the typical trend in the population and the variability between subjects. Thus, the main inference problem does not reduce as in (1) to inferring the individual parameters ϕ_i from the observations of subject i, $i = 1, \ldots, N$, but rather to estimating the pdf of the ϕ_i from the observations of all subjects. More precisely, we aim to compute the maximum likelihood estimate (MLE) of θ. Recall that in mixed-effects diffusion models the likelihood function is not explicit. Thus maximizing (4) with respect to θ is not straightforward. In a general manner, linear and nonlinear mixed-effects models can be seen as incomplete data models in which the random parameters $\phi = (\phi_1, \ldots, \phi_N)$ are the non-observed data, and the population parameters θ are the parameters of the model that need to be estimated from the observations. In general, estimating the parameters of parametric incomplete data models is tricky and some specific algorithms are required. Among these, the EM algorithm [8] is used extensively. The EM algorithm is iterative and each iteration consists of computing and maximizing with respect to θ the conditional expectation $E\left(\log p(y, \phi; \theta)|y, \theta^{(k-1)}\right)$, where $\theta^{(k-1)}$ represents the current estimation of θ. In many situations however, especially when dealing with nonlinear mixed-effects models, a closed form expression of $E\left(\log p(y, \phi; \theta)|y, \theta^{(k-1)}\right)$ is not available. Some variants of EM have been developed to get around this difficulty. The SAEM algorithm [7] is one of them.

3.1.1 General description of the SAEM algorithm

At each iteration of SAEM, the E-step is performed using a stochastic approximation procedure. Let $\theta^{(k-1)}$ denote the current estimate for the population parameters. Iteration k of the SAEM algorithm involves three steps [7]:

- In the simulation step, $\theta^{(k-1)}$ is used to simulate the missing data $\phi_i^{(k)}$ under the conditional distribution $p(\phi_i|y_i, \theta^{(k-1)})$, $i = 1, \ldots, N$.
- In the stochastic approximation step, the simulated data $\phi^{(k)}$ and the observations y are used together to update the stochastic approximation $Q_k(\theta)$ of the conditional expectation $\mathbb{E}(\log p(y, \phi; \theta)|y, \theta^{(k-1)})$ according to:
  (6)
  $$Q_k(\theta) = Q_{k-1}(\theta) + \gamma_k \left[\log p(y, \phi^{(k)}; \theta) - Q_{k-1}(\theta)\right],$$
  where $(\gamma_k)_{k>0}$ is a sequence of positive step sizes decreasing to 0 and starting with $\gamma_1 = 1$.
- In the maximization step, an updated value of the estimate $\theta^{(k)}$ is obtained by maximization of $Q_k(\theta)$ with respect to θ:
  $$\theta^{(k)} = \underset{\theta \in \mathbb{R}^q}{\operatorname{argmax}}\, Q_k(\theta).$$

This procedure is iterated until numerical convergence of the sequence $(\theta^{(k)})_{k>0}$ to some estimate $\hat{\theta}$ is achieved. Convergence results can be found in [7].

3.1.2 SAEM for mixed-effects diffusion models

Although some specific versions of SAEM, presented in [10] and [11], do already exist, the aim of the present work

is to propose a new adaptation of the SAEM algorithm to mixed-effects diffusion models. In [10], the contribution consists in coupling the SAEM algorithm with some Euler-Maruyama approximation of the individual diffusion processes. In [11], the SAEM algorithm is combined with some particle Markov Chain Monte-Carlo methods. The parameter estimations provided by these SAEM versions are reasonable. However, it is required in both algorithms that the individual parameters and the trajectories of the individual diffusion processes be simulated at each iteration, which is computationally intensive. In mixed-effects diffusion models, the latent individual processes $(X_i(t), t > 0)$, $i = 1, \ldots, N$, and the individual parameters ϕ_i, $i = 1, \ldots, N$, are unobserved data. However, simulation of the diffusion trajectories is not necessary since there exists some efficient procedures to approximate the pdf $p(y_i|\phi_i)$ for any given ϕ_i, such as the extended Kalman filter. An appropriate combination of such approximations with SAEM can limit the simulation step of SAEM to the simulation of the individual parameters under $p(\phi_i|y_i; \theta^{(k-1)})$, thus reducing the computational cost. The difficulty comes from having to sample from $p(\phi_i|y_i; \theta^{(k-1)})$ at each iteration of the algorithm, since this is not known in a closed form, especially in mixed-effects diffusion models. Following the idea of [16], a Metropolis-Hastings algorithm is implemented to perform simulation of $\phi_i^{(k)}$, $i = 1, \ldots, N$. Theoretical convergence results about the MCMC-SAEM algorithm obtained by the combination of SAEM with Markov chain Monte Carlo methods in its simulation steps, can be found in [1, 16]. Computation of the acceptance probabilities of the Metropolis-Hastings algorithm requires knowledge of the expression of $p(y_i, \phi_i; \theta)$, and the ability to explicitly calculate it. The expression of $p(y_i, \phi_i; \theta)$ is also necessary to update the function $Q_k(\theta)$ at each iteration of the algorithm. The key to an efficient implementation of the SAEM algorithm is therefore rapid computation of $p(y_i, \phi_i; \theta)$ for any ϕ_i and any θ for all $i = 1, \ldots, N$. Recall that

$$p(y_i, \phi_i; \theta) = p(y_i|\phi_i)p(\phi_i; \theta).$$

Computing $p(\phi_i; \theta)$ is straightforward since the ϕ_i are Gaussian variables, but computing $p(y_i|\phi_i)$ in closed form is generally impossible. Some numerical approximation $\tilde{p}(y_i|\phi_i)$ of $p(y_i|\phi_i)$ with Gaussian densities is performed with the extended Kalman filter.

a) The continuous-discrete extended Kalman filter The continuous-discrete extended Kalman filter performs state estimation from continuous-time diffusion models with discrete-time measurements, such as model (1), by linearization of the drift function b, the volatility function γ and the measurement function g, leading to a Gaussian approximation of the processes and observations' densities. In the present paragraph, we focus on a single individual and a given parameter set ϕ. To simplify the notation, we omit the index i and the dependence of the results on ϕ. In the following, $X_{k|k'} \in \mathbb{R}^d$ and $P_{k|k'} \in \mathcal{S}_d^+(\mathbb{R})$ will denote the mean and the covariance matrix of the conditional probability distribution of $X(t_k)|y_1, \ldots, y_{k'}$, where $\mathcal{S}_d^+(\mathbb{R})$ stands for the set of $d \times d$ symmetric and positive-definite real-valued matrices. The EKF consists of a recursive method, divided into two steps, *Prediction* and *Filtering*, repeated $n-1$ times ($j = 2, \ldots, n$).

- In the *predictive cycle*, the distribution of $X(t_j)|y_1, \ldots, y_{j-1}$ is approximated by a Gaussian distribution. This requires computing an estimate $\hat{X}_{j|j-1}$ of the distribution's mean $X_{j|j-1}$ and an estimate $\hat{P}_{j|j-1}$ of its covariance matrix $P_{j|j-1}$. For this, the following system of equations, ("moment equations"), is solved in time interval $[t_{j-1}, t_j]$:

$$(7) \quad \begin{cases} \dot{x} = b(x(t)), \\ \dot{P} = B(x(t))P(t) + P(t)B(x(t))^T + \Omega(x(t)), \end{cases}$$

with initial conditions

$$\begin{cases} x(t_{j-1}) = \hat{X}_{j-1|j-1}, \\ P(t_{j-1}) = \hat{P}_{j-1|j-1}. \end{cases}$$

Here, $\Omega(x) = \gamma(x)\gamma(x)^T$ and $B(x)$ stands for the Jacobian matrix of $b(x)$ with respect to x. The solution is given by:

$$\begin{cases} \hat{X}_{j|j-1} = x(t_j), \\ \hat{P}_{j|j-1} = P(t_j). \end{cases}$$

- In the *filtering cycle*, the pdf of $X(t_j)|y_1, \ldots, y_j$ is approximated by a Gaussian pdf $\tilde{p}(X(t_j)|y_1, \ldots, y_j)$, which means computing an estimate $\hat{X}_{j|j}$ of the distribution's mean $X_{j|j}$ and an estimate $\hat{P}_{j|j}$ of its covariance matrix $P_{j|j}$. This requires the calculation of the Kalman gain, defined as:

$$K_j = \hat{P}_{j|j-1}G(\hat{X}_{j|j-1})^T(G(\hat{X}_{j|j-1}) \times \hat{P}_{j|j-1}G(\hat{X}_{j|j-1})^T + \sigma^2)^{-1}.$$

Then, $\hat{X}_{j|j}$ and $\hat{P}_{j|j}$ are given by:

$$\begin{cases} \hat{X}_{j|j} = \hat{X}_{j|j-1} + K_j(y_j - g(\hat{X}_{j|j-1})), \\ \hat{P}_{j|j} = (I_d - K_jG(\hat{X}_{j|j-1}))\hat{P}_{j|j-1}, \end{cases}$$

where $G(x)$ is the Jacobian matrix of $g(x)$ with respect to x and I_d denotes the $d \times d$ identity matrix.

Once the EKF is calculated, the conditional distribution of $y_j|y_1, \ldots, y_{j-1}$ can be approximated with a Gaussian distribution with mean $g(\hat{X}_{j|j-1})$ and covariance $G(\hat{X}_{j|j-1})\hat{P}_{j|j-1}G(\hat{X}_{j|j-1})^T + \sigma^2$. Deriving the approximation $\tilde{p}(y)$ of the joint pdf $p(y)$ of the data vector y is straightforward.

b) Resolution of the moment ODEs Implementation of the continuous-discrete EKF is difficult since the moment or-

Table 1. SAEM combined with EKF for the estimation of the population parameters in mixed-effects diffusion models

At a generic iteration k ($k = 1, \ldots, K$) do the following:

1. **for** $i = 1, \ldots, N$ **do**
   - use the EKF (a) and Mazzoni's numerical resolution method of the moment ODEs (b) to compute $\tilde{p}(y_i|\phi_i; \theta^{(k-1)})$
   - sample $\phi_i^{(k)}$ using a single step of the Metropolis-Hastings algorithm targeting
   $$\tilde{p}(\phi_i|y_i; \theta^{(k-1)}) \propto \tilde{p}(y_i|\phi_i; \theta^{(k-1)}) p(\phi_i; \theta^{(k-1)})$$

   end for
2. maximize
   $$Q_k(\theta) = Q_{k-1}(\theta) + \gamma_k \left[\log \tilde{p}(y, \phi^{(k)}; \theta) - Q_{k-1}(\theta)\right]$$
   where
   $$\tilde{p}(y, \phi^{(k)}; \theta) = \prod_{i=1}^{N} \tilde{p}(y_i|\phi_i^{(k)}; \theta) p(\phi_i; \theta)$$

dinary differential equations (7) generally do not have any closed form solution. We suggest the use of a simplified version of the numerical resolution method of these ODEs proposed by Mazzoni in [19], based on higher-order Taylor approximations. As described in [19], we can provide an approximation $\tilde{X}_{k|k'}$ of $\hat{X}_{k|k'}$ and an approximation $\tilde{P}_{k|k'}$ of $\hat{P}_{k|k'}$ by setting at each iteration of the extended Kalman filter:

$$\begin{aligned}
&\tilde{X}_{j|j-1} \\
&\quad = \tilde{X}_{j-1|j-1} + \left(I - B(\tilde{X}_{j-1|j-1})\frac{\Delta t_j}{2}\right)^{-1} b(\tilde{X}_{j-1|j-1})\Delta t_j, \\
&\tilde{P}_{j|j-1} \\
&\quad = \tilde{P}_{j-1|j-1} + M_{\tau_j}\Big(B(X_{\tau_j})\tilde{P}_{j-1|j-1} + \tilde{P}_{j-1|j-1}B(X_{\tau_j})^T \\
&\qquad + \Omega(X_{\tau_j})\Big)M_{\tau_j}^T \Delta t_j,
\end{aligned}$$

where

$$\begin{aligned}
&\Delta t_j = t_j - t_{j-1}, \quad \tau_j = t_{j-1} + \frac{\Delta t_j}{2}, \\
&M_{\tau_j} = \left(I - X_{\tau_j}\frac{\Delta t_j}{2}\right)^{-1}, \\
&X_{\tau_j} = \frac{1}{2}\left(\tilde{X}_{j-1|j-1} + \tilde{X}_{j|j-1} - B(X_{\tau_j})b(X_{\tau_j})\frac{\Delta t_j^2}{4}\right).
\end{aligned}$$

The approximations given by this resolution method has been shown to be of order $\mathcal{O}(\Delta t^2)$. See [19] for more details on these approximations.

It is worth noting that since this method only requires knowledge of the Jacobian function of the drift function of the dynamical process, it is easily implementable in many diffusion models. Note also that this approach can be easily extended to models involving multidimensional observations.

Table 1 summarises the proposed adaptation of the SAEM algorithm for conducting the estimation of the population parameters in mixed-effects diffusion models of the form of equation (3).

3.2 Some other estimation tasks

3.2.1 Estimation of the Fisher information matrix

When an estimate $\hat{\theta}$ of θ has been obtained, the standard errors of its components can be derived by computing the Fisher information matrix $I(\hat{\theta}) = -\partial^2 \log(p(y;\theta))/\partial\theta\partial\theta'|_{\theta=\hat{\theta}}$. Once again, due to the complex expression of the likelihood, $I(\hat{\theta})$ is not known in closed form. As suggested in [17], we therefore estimate it with a stochastic approximation procedure, based on Louis' formula:

$$\begin{aligned}
\frac{\partial^2 \log(p(y;\theta))}{\partial\theta\partial\theta'} &= \mathbb{E}_\theta\left[\frac{\partial^2 \log(p(y,\phi;\theta))}{\partial\theta\partial\theta'}\bigg| y, \theta\right] \\
&\quad + \mathrm{Cov}_\theta\left[\frac{\partial \log(p(y,\phi;\theta))}{\partial\theta}\bigg| y, \theta\right],
\end{aligned}$$

where

$$\begin{aligned}
&\mathrm{Cov}_\theta\left[\frac{\partial \log(p(y,\phi;\theta))}{\partial\theta}\bigg| y, \theta\right] \\
&\quad = \mathbb{E}_\theta\left[\left(\frac{\partial \log(p(y,\phi;\theta))}{\partial\theta}\right)\left(\frac{\partial \log(p(y,\phi;\theta))}{\partial\theta}\right)'\bigg| y, \theta\right] \\
&\qquad - \mathbb{E}_\theta\left[\frac{\partial \log(p(y,\phi;\theta))}{\partial\theta}\bigg| y, \theta\right]\mathbb{E}_\theta\left[\frac{\partial \log(p(y,\phi;\theta))}{\partial\theta}\bigg| y, \theta\right]'.
\end{aligned}$$

$\partial^2 \log(p(y;\hat{\theta}))/\partial\theta\partial\theta'$ is therefore approximated by the sequence $(H_k)_{k>0}$, defined as

$$\Delta_k = \Delta_{k-1} + \gamma_k \left[\frac{\partial \log(p(y, \phi^{(k)}; \hat{\theta}))}{\partial \theta} - \Delta_{k-1} \right],$$

$$D_k = D_{k-1} + \gamma_k \left[\frac{\partial^2 \log(p(y, \phi^{(k)}; \hat{\theta}))}{\partial \theta \partial \theta'} - D_{k-1} \right],$$

$$G_k = G_{k-1} + \gamma_k \left[\frac{\partial \log(p(y, \phi^{(k)}; \hat{\theta}))}{\partial \theta} \frac{\partial \log(p(y, \phi^{(k)}; \hat{\theta}))}{\partial \theta}' - G_{k-1} \right],$$

$$H_k = D_k + G_k - \Delta_k \Delta_k'.$$

The $\phi^{(k)}$ are simulated under $\tilde{p}(\phi_i|y, \hat{\theta})$ via the Metropolis-Hastings algorithm previously used for the SAEM algorithm. The only difference with SAEM is that the population parameter θ is not updated. The derivatives and second derivatives of $\log(p(y, \phi; \theta))$ are computed in a closed form when $p(y, \phi; \theta)$ can be decomposed as $p(y|\phi)p(\phi; \theta)$. If a component of ϕ has no variability, this decomposition is no longer possible and some of the derivatives are obtained by central differences.

3.2.2 Estimation of the individual parameters

Once the population parameters have been estimated, the distribution of the ϕ_i is fully defined. Estimates of ϕ_i can then be derived from the conditional distribution

$$p(\phi_i|y_i, \hat{\theta}) = \frac{p(y_i|\phi_i, \hat{\theta})p(\phi_i, \hat{\theta})}{p(y_i, \hat{\theta})}.$$

For instance,

1. the Maximum A Posteriori (MAP) is the mode of this conditional distribution:

$$\begin{aligned} \hat{\phi}_i^{\mathrm{MAP}} &= \underset{\phi_i}{\operatorname{argmax}}\, p(\phi_i|y_i, \hat{\theta}) \\ &= \underset{\phi_i}{\operatorname{argmax}}\, p(y_i|\phi_i)p(\phi_i, \hat{\theta}). \end{aligned} \tag{8}$$

   Maximization of the right-hand term in (8) is intractable in mixed-effects models based on SDEs, and requires a numerical optimization procedure.
2. the conditional mean is the expectation of this conditional distribution:

$$\hat{\phi}_i^{\mathrm{E}} = \mathbb{E}(\phi_i|y_i, \hat{\theta}). \tag{9}$$

   Once again, the expression of $\mathbb{E}(\phi_i|y_i, \hat{\theta})$ is not explicitly given in our models, so this conditional mean of ϕ_i is estimated with a Metropolis-Hastings algorithm.

Since the exact expression of $p(\phi_i|y_i, \hat{\theta})$ is not available, both estimations require the use of the EKF, so the calculations are based on the approximation $\tilde{p}(\phi_i|y_i, \hat{\theta})$ of $p(\phi_i|y_i, \hat{\theta})$.

3.2.3 Estimation of the latent process

Using the individual parameter estimate $\hat{\phi}_i$ obtained either as the conditional mode or the conditional mean of $p(\phi_i|y_i, \hat{\theta})$, we can now recover the latent process $(X_i(t), t \geq 0)$ at measurement times $(t_{i1}, \ldots, t_{i,n_i})$. Note that the extended Kalman filter already provides an estimate of $X_i(t_{ij})$ defined as its conditional expectation given the observations of subject i up to time t_{ij}: $\mathbb{E}(X_i(t_{ij})|y_{i1}, \ldots, y_{ij}, \hat{\phi}_i, \theta)$. However, this only takes into account the "past" information relative to $X_i(t_{ij})$. Such an estimation can be improved by using the fixed-interval Kalman smoother [14] which incorporates the "future" observations relative to $X_i(t_{ij})$, and estimates $X_i(t_{ij})$ by

$$\hat{X}_i(t_{ij}) = \underset{X_i(t_{ij})}{\operatorname{argmax}}\, p(X_i(t_{ij})|y_{i1}, \ldots, y_{i,n_i}, \hat{\phi}_i, \theta).$$

3.2.4 Model diagnostics and model selection

Some diagnostic tools are also needed for practical applications. Indeed, it is not enough just to obtain a good fit with a given model in order to keep it. It is mandatory to assess whether the model is capable of producing the observed data.

A formal statistical test requires the use of some relevant statistics whose distribution is known under the null hypothesis, i.e., under the proposed model. The choice of test statistics is essential as it must enable us to uncover various types of misspecifications: a misspecified statistical model, a misspecified structural model, a misspecified diffusion model, etc. Deriving in closed form the probability distribution of these statistics can be extremely difficult for mixed effects diffusion models. On the other hand, a very good approximation of the distribution can easily be obtained using a Monte Carlo approach.

Here, we do not propose a complete methodology for model diagnostics, but suggestions based on visual criteria. The visual predictive check (VPC) is a powerful model diagnostic method that can be used for evaluating the performance of a given model [3, 4]. The VPC is constructed from a large number of simulations from the model. These simulations are used to estimate a prediction interval of the test statistics. We can then visually compare the observed test statistics with this prediction interval. The test statistics should summarize the trend and the variability of the variables of the model, i.e., the observations and the state variables of the dynamical system. Following the approach described in [3, 4], we use certain representative percentiles of the variables (typically 10th, 50th and 90th) computed on successive time intervals. We will see in the next section, dedicated to numerical experiments, how the VPC behaves in various situations.

In most practical cases, several models may be identified as possible candidates for modeling the observed data. Some criterion is then required to compare these different models if only one (or several) of them can be selected. Classical

information criteria such as AIC and BIC can be computed using the approximated probability distribution function

$$\tilde{p}(y;\hat{\theta}) = \prod_{i=1}^{N} \mathbb{E}\left(\tilde{p}(y_i,\phi_i;\hat{\theta})\right), \tag{10}$$

where the expectation is approximated with Monte Carlo integration using an Importance Sampling scheme [18].

3.3 Other existing methods

Several alternative methods for parameter estimation in mixed effects diffusion models have been proposed during the last decade. Since we are interested in these models as tools for practical purposes, we will only comment on methods whose performance has been illustrated with numerical experiments.

The PSM package for R provides functions for estimation of mixed-effects models using stochastic differential equations [15]. The likelihood function is evaluated using Kalman filter (KF) for linear systems or EKF for nonlinear systems. The maximization is based on the First-Order Conditional Estimation (FOCE) method which approximates the nonlinear statistical model with a Gaussian linear model. Contrary to SAEM, FOCE can be very sensitive to the initial values and does not converge to the MLE when the model is nonlinear. Furthermore, though the developers of PSM claim that their method can be used for nonlinear systems, all the numerical examples available with the documentation only consider linear systems.

The method proposed in [23] assumes that there is no residual error in the model. The conditional transition density of the diffusion process is approximated by a Hermite expansion, then the likelihood is numerically optimized using a Laplace approximation in order to handle statistical models with multiple random effects. The approach is interesting but still of limited interest for practical applications, since it assumes that the residual error component is negligible compared to the stochastic component of the model.

Convergence of the SAEM-PMCMC algorithm towards the maximum likelihood estimator is ensured when the transition density of the SDE is explicit. Furthermore, the algorithm can be implemented and the computational time is reasonable. When the transition density is not explicit, the author defines an approximate diffusion by the recursive Euler-Maruyama scheme which is difficult to use in the practice. Indeed, such an approximation adds intermediate latent variables to the model which are tricky to handle properly within a MCMC or PMCMC approach.

Our method also has its pros and cons. The use of SAEM avoids any approximation of the statistical model and allows to handle a large number of random effects. Furthermore, it is fast and less sensitive to the initial values in most situations. Combining SAEM with the Kalman system is then recommended for linear dynamical systems. Using EKF for nonlinear systems is only recommended when the linear approximation is valid. This is usually the case for physiological systems which are stable and for which the state variables randomly fluctuate around a fix value, or randomly decrease to 0. A Particle MCMC could replace EKF for more general situations, but a usable version of the SAEM-PMCMC algorithm proposed in [9] then needs to be developed. Extension to multidimensional observations proposed in [20] is straightforward with SAEM-EKF since SAEM already handles multidimensional observations.

4. NUMERICAL EXPERIMENTS

4.1 Diffusion models for dynamical systems with linear transfers

Some authors recently underlined that extending the classical ordinary differential equations (ODEs)-based models for pharmacokinetics with stochastic differential equations, could lead to a better description of the fluctuations, usually occurring around the theoretical drug's dynamics [21]. They often suggest to describe the evolution of the drug concentration over time by means of stochastic differential equations rather than ordinary differential equations. We emphasize that diffusion models are more appropriate than ODE-based models for the description of random fluctuations in general dynamical systems. Nevertheless, we suggest to describe the fluctuations within these systems by considering the transfer rates as diffusion processes rather than the components of the system.

Dynamical systems have many applications in various fields. It is, for example a relevant representation to model viral dynamics, population flows, interactions between cells or drugs' pharmacokinetics. Dynamical systems involving linear transfers between different entities are usually modelled by means of a system of ODEs with the following general form:

$$dX(t) = KX(t)dt, \tag{11}$$

where $X(t)$ is a vector whose d^{th} component represents condition of dth entity at time t and $K = (K_{l,l'})_{l,l'}$ is a deterministic matrix defined as:

$$\begin{cases} K_{l,l'} = k_{l,l'} & \text{if} \quad l \neq l', \\ K_{l,l} = -k_{l0} - \sum_{l'} k_{l,l'} & \text{otherwise}, \end{cases} \tag{12}$$

with $k_{l,l'}$ representing the transfer rate from entity l to entity l', and k_{l0} the elimination rate from entity l. An example of such a dynamical system with 3 components is schematized in Figure 1.

In this particular example, matrix K would be defined as

$$K = \begin{pmatrix} -k_{10} - k_{12} - k_{13} & k_{21} & k_{31} \\ k_{12} & -k_{20} - k_{21} - k_{23} & k_{32} \\ k_{13} & k_{23} & -k_{30} - k_{31} - k_{32} \end{pmatrix}.$$

A model defined by equations (11) and (12) is a deterministic model which assumes the transfers to take place at the

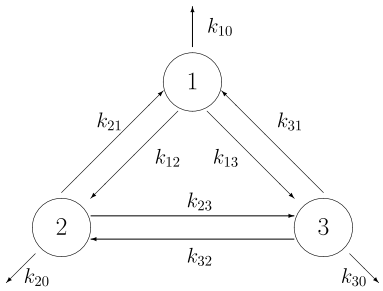

Figure 1. Example of dynamical system with 3 components (circle) and linear transfers between components (arrows).

same rate at all times. This is however often a restrictive assumption since in reality, dynamical systems usually exhibit some random behavior. It is then reasonable to consider that the transfers are not constant but randomly fluctuate over the time. This new assumption leads to the following dynamical system:

$$dX(t) = K(t)X(t)dt, \tag{13}$$

where K has the same structure as in (12) but where some components $k_{l,l'}$ are stochastic processes which take non negative values and randomly fluctuate around a typical value $k^\star_{l,l'}$.

Let us now illustrate the construction of such diffusion models using some specific examples in pharmacokinetics.

Example 1 (Bolus with linear elimination)**.** The following ordinary differential equation

$$dA(t) = -kA(t)dt,$$

is usually used to describe the kinetics of a drug administered by rapid injection (IV bolus) in the plasma. In bolus-specific compartmental models, plasma is assimilated to a single compartment of the human body. $A(t)$ represents the amount of the drug substance in plasma at time t after injection, and k is the elimination rate constant. Now assume that the drug's dynamics is perturbed and define k as a diffusion process which takes non negative values and which fluctuates around a typical value $k^\star$. In this example, non negativity of $k(t)$ is ensured by defining the logarithm of the transfer rate as an Ornstein-Uhlenbeck diffusion process

$$d\log k(t) = -\alpha\left(\log k(t) - \log k^\star\right)dt + \gamma dW(t),$$

where W is a standard one dimensional Wiener process. This results in the following diffusion system:

$$dX(t) = b(X(t))dt + \gamma(X(t))dW(t),$$

where

$$X(t) = \begin{pmatrix} A(t) \\ \log k(t) \end{pmatrix}; \quad b(x) = \begin{pmatrix} -x_1 e^{x_2} \\ -\alpha(x_2 - \log k^\star) \end{pmatrix}$$

$$\gamma(x) = \begin{pmatrix} 0 & 0 \\ 0 & \gamma \end{pmatrix}.$$

Note that in this specific example, the Jacobian matrix of the drift function has a simple form:

$$B(x) = \begin{pmatrix} -\exp x_2 & -x_1 \exp x_2 \\ 0 & -\alpha \end{pmatrix}.$$

Example 2 (Oral with first-order absorption and linear elimination)**.** Oral PK model with first-order absorption and linear elimination is widely used to describe the time-course of a drug orally administered through a unique compartment of the human body. The drug is administrated in a depot compartment, absorbed by the central compartment with absorption rate k_a and eliminated with elimination rate k_e. Such model is described by the following system of ODEs:

$$\frac{d}{dt}\begin{pmatrix} A_d(t) \\ A_c(t) \end{pmatrix} = \begin{pmatrix} -k_a & 0 \\ k_a & -k_e \end{pmatrix}\begin{pmatrix} A_d(t) \\ A_c(t) \end{pmatrix}, \tag{14}$$

where $A_d(t)$ and $A_c(t)$ respectively represent the amounts of drug at time t in the depot compartment and in the central compartment. Assume now that the constant of elimination is driven by a stochastic process solution of this stochastic differential equation

$$dk_e(t) = -\alpha(k_e - k_e^\star)dt + \gamma\sqrt{k_e(t)}dW(t),$$

where W is a standard one dimensional Wiener process, then (14) becomes:

$$dX(t) = b(X(t))dt + \gamma(X(t))dW(t).$$

Here,

$$X(t) = \begin{pmatrix} A_d(t) \\ A_c(t) \\ k_e(t) \end{pmatrix}; \quad b(x) = \begin{pmatrix} -k_a x_1 \\ k_a x_1 - x_3 x_2 \\ -\alpha(x_3 - k_e^\star) \end{pmatrix}$$

$$\gamma(x) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \gamma\sqrt{x_3} \end{pmatrix}.$$

In both examples, the diffusion model can be easily extended in a population approach by defining the system's parameters as Gaussian random variables.

4.2 Simulation study

We now investigate the properties of the maximum likelihood estimates obtained with the SAEM algorithm combined with the extended Kalman filter through a short simulation study. The model used for the simulations is a mixed-effects bolus model with an elimination rate constant defined as a stochastic process (Example 1), with discrete-time observations consisting of the log-concentration of drug, up to a Gaussian noise:

Table 2. Results of the Monte Carlo simulation study for $N = 30$: for each parameter of the model, $n = 25$ and $n = 50$, the table displays the mean estimate, the standard deviation of the estimates and the mean estimated standard error

		$N = 30, n = 25$			$N = 30, n = 50$		
Parameter	θ_0	$\bar{\theta}$	sd	$\bar{se}$	$\bar{\theta}$	sd	$\bar{se}$
$k^\star$	0.2	0.199	0.011	0.011	0.202	0.011	0.011
V	0.5	0.502	0.020	0.019	0.500	0.018	0.018
γ	0.2	0.177	0.026	0.022	0.188	0.021	0.018
σ	0.1	0.102	0.007	0.003	0.101	0.002	0.002
ω_k	0.3	0.296	0.045	0.040	0.296	0.039	0.039
ω_V	0.2	0.198	0.027	0.028	0.192	0.027	0.026

$$\begin{cases} dA_i(t) = -k_i(t)A_i(t)dt, \\ d\log k_i(t) = -\alpha_i(\log k_i(t) - \log k_i^\star)dt + \gamma_i dW_i(t), \\ y_{ij} = \log(A_i(t_{ij})/V_i) + \xi_{ij}, \\ \xi_{ij} \underset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2), \quad j = 1, \ldots, n, \; i = 1, \ldots, N. \end{cases}$$

The initial conditions are $A_i(0) = D_i$ and $k_i(0) = k_i^\star$, where D_i denotes the dose of drug injected at time $t = 0$. V_i stands for the apparent volume of distribution of the medication. In the present simulation study, we set

$$\begin{aligned} \alpha_i &= \alpha, \\ \gamma_i &= \gamma, \\ \log k_i^\star &= \log k^\star + \eta_{k,i}, \\ \log V_i &= \log V + \eta_{V,i}, \end{aligned}$$

for all $i = 1, \ldots, N$. We consider that the random variables $\eta_{k,i}$ and $\eta_{V,i}$ are independent zero-mean Gaussian variables with respective variance ω_k^2 and ω_V^2. Moreover, for reasons of identifiability in the model, we assume that parameter α is known. For the present study, we set $\alpha = 1$. Therefore the vector of population parameters to be estimated is $\theta = (\gamma, k^\star, V, \sigma, \omega_k, \omega_V)$.

100 datasets are simulated with different numbers of subjects ($N = 30$ and $N = 100$) and different numbers of observations per subject ($n = 25$ and $n = 50$). For a n given, we take equally spaced observation times t_{ij} between 0 and 24 hours. The parameter values used for the simulations are displayed in Tables 1 and 2. For each simulated dataset, the population parameters and the standard errors of $\hat{\theta}$ are estimated with SAEM according to the procedures described in Section 3. The algorithm is initialized with values randomly chosen in a neighborhood of the true parameter values θ_0. More precisely, we choose $\theta^{(0)}$ drawn from $\theta_0 \mathcal{U}_{[0,c]}$ with c a constant. The EKF and the model were implemented in Matlab and integrated in the MONOLIX software.

For $m = 1, 2, \ldots, 100$, let $\hat{\theta}_m$ be the estimated vector of population parameters obtained with the mth simulated dataset and let $\widehat{se}_m$ be their respective estimated standard errors. For each design, we have computed the mean estimated parameter:

$$\bar{\theta} = \frac{1}{100} \sum_{m=1}^{100} \hat{\theta}_m,$$

the parameter standard deviation

$$sd = \sqrt{\frac{1}{100} \sum_{m=1}^{100} (\hat{\theta}_m - \bar{\theta})^2},$$

and the mean estimated standard error:

$$\bar{se} = \frac{1}{100} \sum_{m=1}^{100} \widehat{se}_m.$$

The distribution of the relative estimation errors $100 \times \frac{\hat{\theta}_m - \theta_0}{\theta_0}$ is displayed Figure 2.

The extended SAEM algorithm gives satisfactory results overall. It supplies very accurate estimations for the parameters in every design tested through the present study. The population parameter estimates show very little bias, since the absolute relative bias is globally less than 5%, except for parameter γ. The volatility parameter is indeed 11.35% underestimated from the smallest simulated datasets ($N = 30, n = 25$), but the bias decreases when the sample size increases. The standard errors of the estimates are quite small, whatever the values for N and n, and increasing the number of simulated subjects reduces the empirical variance of the estimates. We estimate the Fisher Information Matrix accurately. For every parameter, we notice from Tables 1 and 2 that $\bar{se}$ is close to the standard deviation of the estimates. Finally, convergence is obtained in less than 250 iterations, which only requires about 4′30 on a laptop (processor Intel(R) Core(TM) i7-2760QM 2.40GHz) for a single simulated dataset including $N = 30$ subjects and $n = 25$ observations per subject, and 14′ in largest datasets ($N = 100$, $n = 50$).

4.3 Model diagnostics

The aim of this section is to illustrate with a numerical example the behavior of the Visual Predictive Check (VPC) when the model is correct and when it is uncorrect. We consider two models:

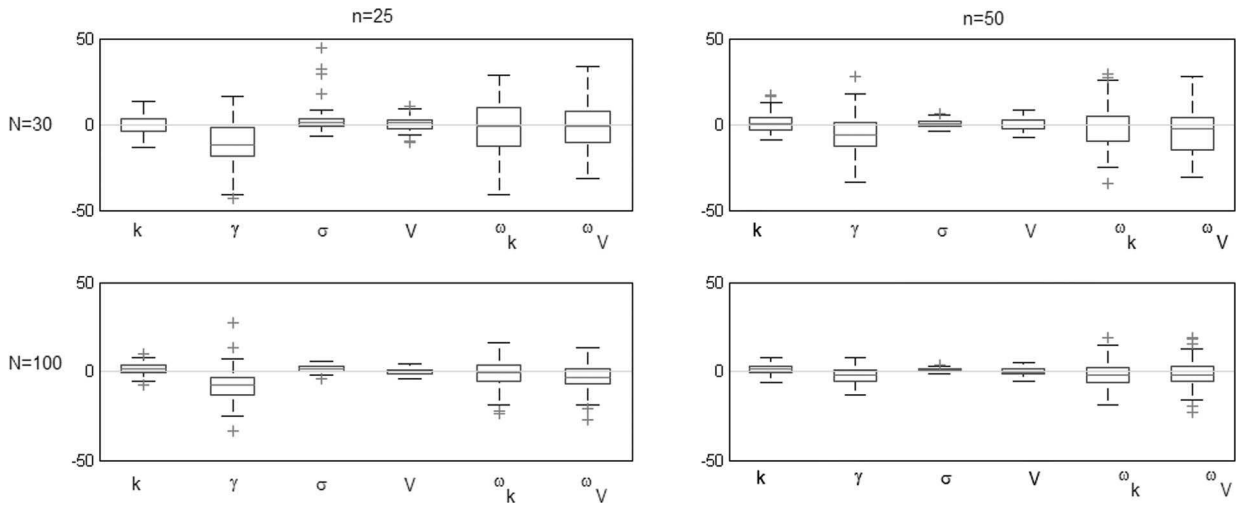

Figure 2. Results of the Monte Carlo simulation study: distribution of the relative estimation errors for the different designs for (N, n).

Table 3. Results of the Monte Carlo simulation study for $N = 100$: for each parameter of the model, $n = 25$ and $n = 50$, the table displays the mean estimate, the standard deviation of the estimates and the mean estimated standard error

		$N = 100, n = 25$			$N = 100, n = 100$		
Parameter	θ_0	$\bar{\theta}$	sd	$\bar{se}$	$\bar{\theta}$	sd	$\bar{se}$
$k^\star$	0.2	0.203	0.006	0.006	0.202	0.006	0.006
V	0.5	0.500	0.009	0.010	0.499	0.011	0.010
γ	0.2	0.184	0.018	0.012	0.195	0.010	0.010
σ	0.1	0.101	0.002	0.002	0.101	0.001	0.001
ω_k	0.3	0.296	0.021	0.021	0.294	0.021	0.021
ω_V	0.2	0.194	0.015	0.015	0.197	0.016	0.015

- Model M_A is a bolus model with a linear elimination defined by a stochastic process

$$dA(t) = -k(t)A(t)dt,$$
$$d\log k(t) = -\alpha(\log k(t) - \log k^\star)dt + \gamma dW(t).$$

- Model M_B is a bolus model with a non linear and non stochastic elimination

$$dA(t) = -\frac{V_m}{V\,K_m + A(t)}A(t)dt.$$

Two datasets with 100 individuals each were simulated with the same design used previously for Example 1:

- Dataset A was simulated using model M_A incorporating inter-individual variability on $k^\star$ and V by defining the elimination and the volume of subject i as $\log k_i^\star = \log k^\star + \eta_{k^\star,i}$ and $\log V_i = \log V + \eta_{V,i}$ where $\eta_{k^\star,i}$ and $\eta_{V,i}$ are centered Gaussian random variables with respective variance $\omega^2_{k^\star}$ and ω^2_V.
- Dataset B was simulated using model M_B incorporating inter-individual variability on V_m, K_m and V: for any subject $i = 1, \ldots, N$, we set $\log V_{m,i} = \log V_m + \eta_{V_m,i}$, $\log K_{m,i} = \log K_m + \eta_{K_m,i}$ and $\log V_i = \log V + \eta_{V,i}$ where $\eta_{V_m,i}$, $\eta_{K_m,i}$ and $\eta_{V,i}$ are centered Gaussian random variables with respective variance $\omega^2_{V_m}$, $\omega^2_{K_m}$ and ω^2_V.

We then fitted model M_A to both datasets. We only show in Figure 3 the fits obtained with the first individual of each dataset (we obtain very similar fits for the 100 subjects). We obtain very nice fits in both cases, which means that the diffusion is able to capture some smooth component of the intra-individual variability around the solution of the ODE model which is obtained by assuming that k remains constant over time (i.e. $k(t) = k^\star, t \geq 0$) in model M_A. These fluctuations are described by a stochastic process in example A and by a deterministic function in example B (the deviation between the linear and the non linear elimination functions). Looking at only one individual fit does not allow to evaluate if the dynamical system is misspecified or not. We also displayed in the same figure the estimated functions $k(t)$ obtained with the (extended) Kalman smoother for these two individuals. Both series seem to randomly fluctuate around a typical value $k^\star$, but it is impossible to decide whether these fluctuations are purely stochastic, or if

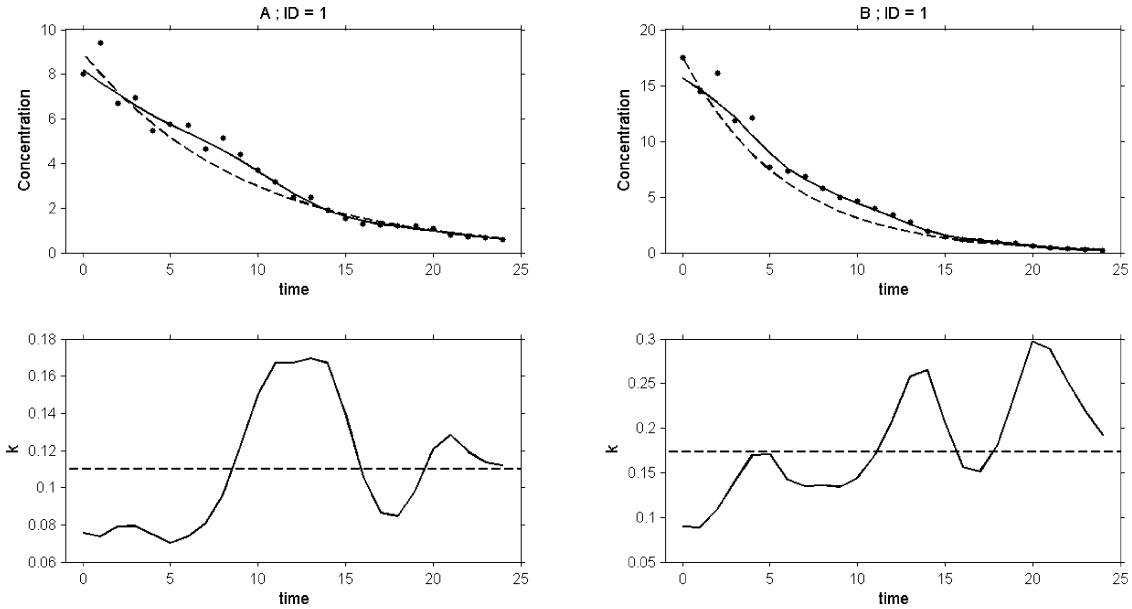

Figure 3. Top: observed and estimated concentrations obtained with two different models from two different patients, the dots are the observed concentrations, the solid line is the estimated concentration given by the SDE based model and the dashed line the estimated concentration given by the ODE based model; bottom: estimated elimination rate processes obtained with these two different models from these two individuals. Left: results obtained with the first individual of dataset A; right: results obtained with the first individual of dataset B.

there is a deterministic trend. Assessment of the dynamical system then requires combining information from several individuals. Assessment of the statistical model will obviously require information from several individuals as well.

Figure 4 displays the VPCs for the observations (y_{ij}) and the estimated elimination processes $(k(t_{ij}))$ for datasets A and B. On one hand, we see that the observed percentiles obtained from dataset A are mostly within the 90% prediction intervals, both for (y_{ij}) and $(k(t_{ij}))$. Model is hence potentially likely to have generated data A. On the other hand, there is a strong discrepancy between dataset B and model M_A. We can reject this model with high confidence. It can be then interesting to use the VPCs to improve the model. We see in this example that the elimination rate process $(k(t))$ does not randomly fluctuate around a typical value $k^\star$, but rather seems to increase with time, or equivalently to decrease with the concentration. Such a behavior may indicate a saturable elimination as described with the Michaelis Menten elimination of model B.

4.4 A real data example

We have then used the methodology proposed in Section 3 to analyze the well-known theophyllin data (see e.g. [24]). These data come from twelve subjects who received a single oral dose of 320 mg of theophyllin, a drug prescribed for the treatment of asthma. For each subject, the concentration of theophyllin in plasma is measured through ten consecutive blood samples taken around 15 minutes, 30 minutes, 1, 2, 3.5, 5, 7, 9, 12 and 24 hours after the dose administration. These data are classically described by the one compartment model with first order absorption and first order elimination based on the ODE system (14) presented in *Example 2* (Model 1). We would rather assume that the elimination rate of theophyllin is a diffusion process. We then consider the following mixed-effects diffusion model (Model 2):

$$\begin{aligned} dA_{d,i}(t) &= -k_{a,i}A_{d,i}(t)dt, \\ dA_{c,i}(t) &= \left(k_{a,i}A_{d,i}(t) - k_{e,i}(t)A_{c,i}(t)\right)dt, \\ d\log k_{e,i}(t) &= -\alpha_i(\log k_{e,i}(t) - \log k^\star_{e,i})dt + \gamma_i dW_i(t), \\ y_{ij} &= A_{c,i}(t_{ij})/V_i + \xi_{ij}, \\ \xi_{ij} &\underset{i.i.d.}{\sim} \mathcal{N}(0,\sigma^2). \end{aligned} \tag{15}$$

Only 12 patients with 10 observations per patient does not allow to estimate properly all the parameters of the model with their inter individual variability. It is therefore necessary to assume that some parameters do not vary and/or are fixed. Here, we assume that parameters α_i and γ_i are identical for the twelve subjects: $\alpha_i = \alpha$ and $\gamma_i = \gamma$, $i = 1, \ldots, N$.

We set $\log k_{a,i} = \log k_a + \eta_{1i}$, $\log k^\star_{e,i} = \log k^\star_e + \eta_{2i}$ and $\log V_i = \log V + \eta_{3i}$, where $(\eta_{1i}, \eta_{2i}, \eta_{3i})' \underset{i.i.d.}{\sim} \mathcal{N}(0, \Omega)$ and Ω diagonal. The estimated population parameters and the estimated standard errors are reported in Table 4. We see that it was not possible to estimate correctly the parameters of the diffusion α and γ which converged to a very small

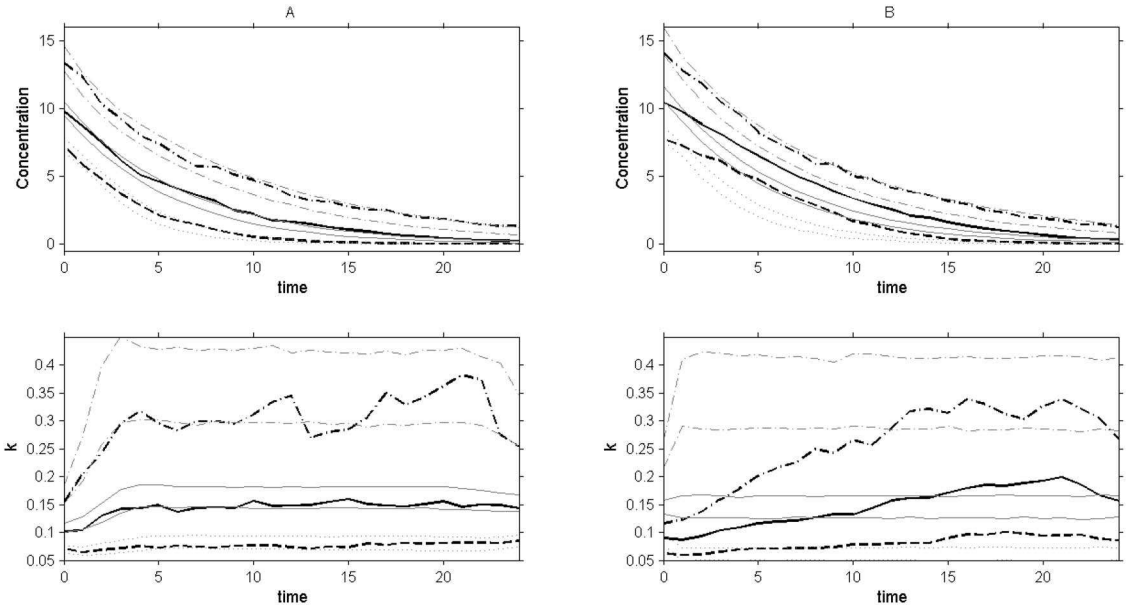

Figure 4. Visual predictive checks obtained with dataset A (left) and dataset B (right). Top: VPC for the observed concentration; bottom: VPC for the estimated elimination rate process. Black: observed 10th, 50th and 90th percentiles; grey: 90% prediction intervals. Dashed line: 10th percentile; solid line: 50th percentile; dashdot line: 90th percentile.

Table 4. Estimation of the population parameters for the theophyllin data with the mixed-effects diffusion model (15). The table displays the parameter estimates and their estimated standard errors obtained with two different estimates. No prior information is used in Model 1 and the MLE is computed. A prior information for α, γ and σ is used in Model 2 and the MAP is computed for these three parameters

	Model 1		Model 2	
Parameter	Estimate	Stand. Err.	Estimate	Stand. Err.
$k_e^\star$	0.088	0.005	0.103	0.008
α	2.94	NaN	0.522 (MAP)	0.290
γ	0.0003	NaN	0.416 (MAP)	0.180
k_a	1.31	0.27	1.14	0.280
σ	0.168	0.013	0.057 (MAP)	0.002
V	0.457	0.024	0.399	0.024
ω_{k_e}	0.158	0.067	0.085	0.059
ω_{k_a}	0.678	0.170	0.837	0.180
ω_V	0.149	0.042	0.171	0.042

value. The solution obtained by maximizing the likelihood without any constraint is therefore the ODE model without any stochastic behavior. We then introduced some prior information on α, γ and σ to force the solution to contain a stochastic component. Log-normal distributions were used, the prior values and the standard deviations were respectively (0.5, 0.05), (0.4, 0.04) and (0.04, 0.01) for α, γ and σ. We then computed the Maximum a posteriori (MAP) estimate of the population parameters by maximizing the conditional distribution. The MAP estimate reduces to a penalized MLE for parameters with prior and to standard MLE for parameters without prior.

The predicted curves for three subjects obtained either with Model 1 or with Model 2 are displayed in Figure 5. The predicted curves under both models are satisfactory for the three subjects, but we see that the SDE based model leads to a much better fit of the concentrations data since the model enables to capture smooth random variations of the concentration kinetics. Assuming that the elimination rate of theophyllin is driven by a stochastic differential equation seems to be a plausible assumption. Nevertheless, the three estimated elimination processes also displayed Figure 5 show the same behavior (a peak around $t = 5h$). This "non-stochastic" behavior could indicate a misspecification of the structural model. Given these figures, a modeler should then consider other PK models compatible with this behavior of the elimination.

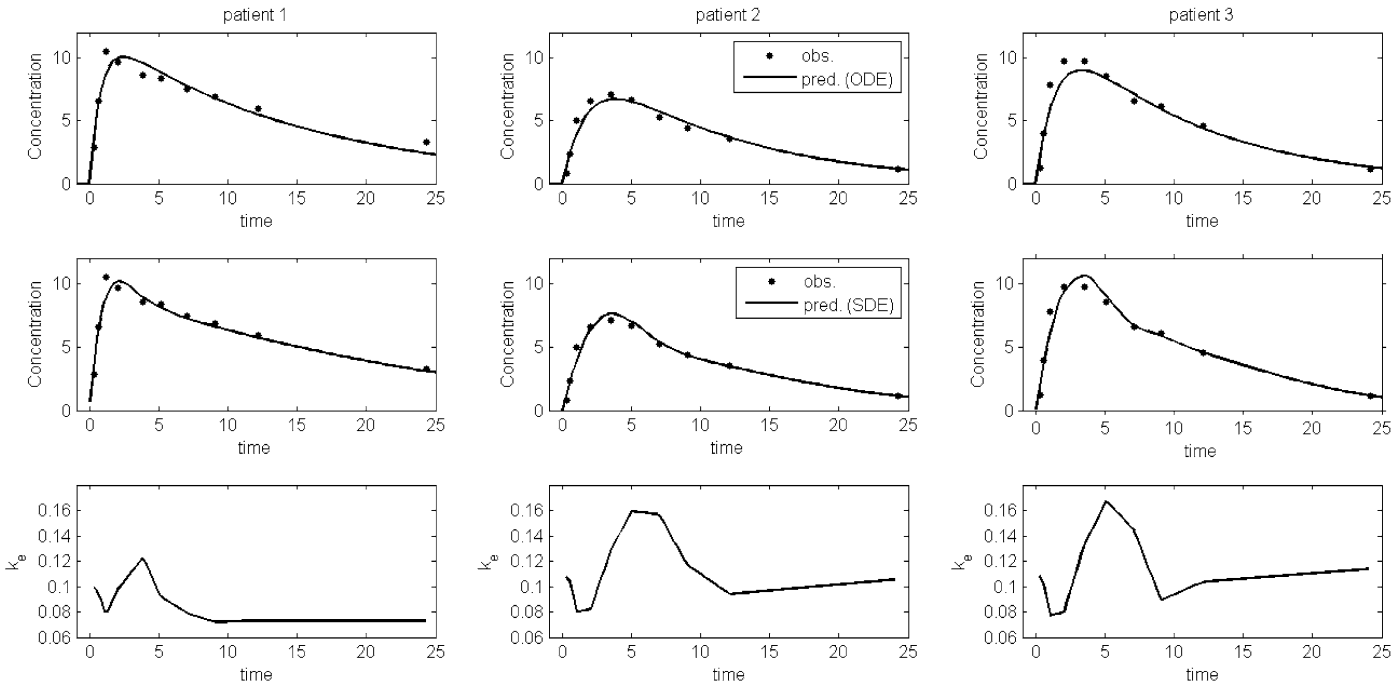

Figure 5. Top: observed plasmatic concentrations of theophyllin for three subjects and their predicted concentration profiles given by the ODE based model (Model 1): the unobserved drug's elimination rate is assumed to keep constant over time; middle: observed concentrations for the same three subjects and their predicted concentration profiles given by the SDE based model (Model 2): the unobserved drug's elimination rate is assumed to be a diffusion process; bottom: predicted evolution of the elimination rate of theophyllin over time given by the SDE based model.

5. DISCUSSION

This paper proposes a new inference methodology for mixed-effects diffusion models, observed at discrete time points up to a Gaussian noise. In particular, we suggest a specific extension of the SAEM algorithm combined with the extended Kalman filter to estimate the population parameters in these models. The particularity of this new version of SAEM is that the simulation step reduces to the simulation of the unknown individual parameters, considering the diffusion process as a nuisance parameter. This approach leads to much smaller computational times than other versions of SAEM for mixed-effects diffusion models previously proposed in the literature, particularly the combination between SAEM and particle filters [11]. It should be noted however, that contrary to the particle filters based method, our approach does not give exact results since the EKF provides approximation of the marginal density of the data.

The performances of the new estimation method are illustrated with a brief Monte-Carlo simulation study based on a stochastic extension of PK bolus model. The simulation globally shows satisfying performances of the new algorithm. The SAEM algorithm reveals very little bias of the estimates, even in quite small datasets, and converges in a very few iterations. Mixed-effects diffusion models have direct applications, especially for dynamical systems with linear transfers, like in pharmacokinetics. Our approach is to define the transfer rates as diffusion processes to model the perturbations observed on the systems rather than constants as it is often done in practice.

We mainly focused ourselves in this paper to the problem of estimating the population parameters, the individual parameters and the state variables of the system. We have also shown that Visual Predictive Checks is a valuable tool for a model diagnostic. A more complete methodology for model assessment deserves to be developed, including statistical tests to distinguish a stochastic and a deterministic behavior of the dynamical system.

Received 16 August 2012

REFERENCES

[1] Allassonniere, S., Kuhn, E. and Trouvé, A. (2010). Construction of Bayesian deformable models via a stochastic approximation algorithm: A convergence study. *Bernoulli* **16** 641–678. MR2730643

[2] Andrieu, C., Doucet, A. and Holenstein, R. (2010). Particle markov chain Monte Carlo. *Journal of the Royal Statistical Society: Series B* **72** 269–342. MR2758115

[3] Bergstrand, M., Hooker, A., Wallin, J. E. and Karlsson, M. O. (2011). Prediction-corrected visual predictive checks for diagnosing nonlinear mixed-effects models. *AAPS J.* **13** 143–151.

[4] Bleakley, K. and Lavielle, M. (2011). Automatic data binning for improved visual diagnosis of pharmacometric models. *Journal of Pharmacokinetics and Pharmacodynamics* **38** 861–871.

[5] Cappé, O., Godsill, S. J. and Moulines, E. (2007). An overview of existing methods and recent advances in sequential Monte Carlo. *Proceedings of the IEEE* **95** 899–924.

[6] Delattre, M., Genon-Catalot, V. and Samson, A. (2013). Maximum likelihood estimation for stochastic differential equations with random effects. *Scandinavian Journal of Statistics*, to appear.

[7] Delyon, B., Lavielle, M. and Moulines, E. (1999). Convergence of a stochastic approximation version of the EM algorithm. *The Annals of Statistics* **27** 94–128. MR1701103

[8] Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society – Series B* **39** 1–38. MR0501537

[9] Ditlevsen, S. and Gaetano, A. D. (2005). Mixed effects in stochastic differential equation models. *REVSTAT Statistical Journal* **3** 137–153. MR2259358

[10] Donnet, S. and Samson, A. (2008). Parametric inference for mixed models defined by stochastic differential equations. *ESAIM: Probability and Statistics* **12** 196–218. MR2374638

[11] Donnet, S. and Samson, A. (2010). EM algorithm coupled with particle filter for maximum likelihood parameter estimation of stochastic differential mixed-effects models. *Technical report*. http://hal.archives-ouvertes.fr/hal-00519576/.

[12] Doucet, A., de Freitas, N. and Gordon, N. (2001). *Sequential Monte Carlo Methods in Practice*. Springer. MR1847783

[13] Faugeras, O., Touboul, J. and Cessac, B. (2009). A constructive mean field analysis of multi population neural networks with random synaptic weights and stochastic inputs. *Frontiers in Computational Neuroscience* **3** 1–28.

[14] Grewal, M. S. and Andrews, A. P. (2008). *Kalman Filtering: Theory and Practice Using MATLAB*, Third ed. Wiley.

[15] Klim, S., Mortensen, S. B., Kristensen, N. R., Overgaard, R. V. 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.

[16] Kuhn, E. and Lavielle, M. (2004). Coupling a stochastic approximation version of EM with an MCMC procedure. *ESAIM: Probability and Statistics* **8** 115–131. MR2085610

[17] Kuhn, E. and Lavielle, M. (2005). Maximum likelihood estimation in nonlinear mixed effects models. *Computational Statistics and Data Analysis* **49** 1020–1038. MR2143055

[18] Lixoft, (2013). *Monolix 4.2.1 User Guide*. http://www.lixoft.eu/monolix/documentation.

[19] Mazzoni, T. (2008). Computational aspects of continuous-discrete extended Kalman-filtering. *Computational Statistics* **23** 519–539. MR2448180

[20] Mbalawata, I. S., Särkkä, S. and Haario, H. (2012). Parameter estimation in stochastic differential equations with Markov chain Monte Carlo and non-linear Kalman filtering. *Computational Statistics* 1–29.

[21] 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.

[22] Picchini, U. and Ditlevsen, S. (2011). Practical estimation of high dimensional stochastic differential mixed-effects models. *Computational Statistics and Data Analysis* **55** 1426–1444. MR2741425

[23] Picchini, U., Gaetano, A. D. and Ditlevsen, S. (2010). Stochastic differential mixed-effects models. *Scandinavian Journal of Statistics* **37** 67–90. MR2675940

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

[25] Tornøe, C. W., Overgaard, R. V., Agersø, H., Nielsen, H. A., Madsen, H. and Jonsson, E. N. (2005). Stochastic differential equations in NONMEM: Implementation, application, and comparison with ordinary differential equations. *Pharmaceutical Research* **22** 1247–1258.

Maud Delattre
Inria Saclay & University Paris Sud Orsay & AgroParisTech
UMR 518 MIA, F-75005 Paris
France
E-mail address: maud.delattre@agroparistech.fr

Marc Lavielle
Inria Saclay & University Paris Sud Orsay
France
E-mail address: Marc.Lavielle@inria.fr