EM algorithm coupled with particle filter for maximum likelihood parameter estimation of stochastic differential mixed-effects models

Sophie Donnet, Adeline Samson

EM algorithm coupled with particle filter for maximum likelihood parameter estimation of stochastic differential mixed-effects models

Running headline : SDE mixed-effects models

Sophie Donnet[1] and Adeline Samson[2*]

[1] Laboratoire Ceremade, Université Paris Dauphine, France
[2] UMR CNRS 8145, Laboratoire MAP5, Université Paris Descartes, France

*To whom correspondence should be addressed

Abstract

Biological processes measured repeatedly among a series of individuals are standardly analyzed by mixed models. They use common regression function and error model, both depending on individual random parameters. Regression functions defined by parametric Stochastic Differential Equations (SDEs) model adequately biological processes. This results in a mixed-effects model defined by an SDE. We focus on the parameter maximum likelihood estimation of this model. As the likelihood is not explicit, we propose the use of a stochastic version of the Expectation-Maximization algorithm combined with the Particle Markov Chain Monte Carlo method. We prove the convergence of the proposed algorithm towards the maximum likelihood estimator. We illustrate the performance of this estimation method on simulated datasets. We consider two examples: the first one is based on an Ornstein-Uhlenbeck process with two random parameters and an additive error model; the second one is based on a time-inhomogeneous SDE (Gompertz SDE) with a stochastic volatility error model and three random parameters. We highlight the superiority of our estimator over an estimator based on the EM algorithm coupled with standard MCMC algorithms.

Key Words: SAEM algorithm, Mixed models, Stochastic Differential Equations, Particle Filter, PMCMC algorithm, SMC algorithm, Stochastic volatility, Time-inhomogeneous SDE.

1 Introduction

Biological processes are usually measured repeatedly among a collection of individuals or experimental animals. It is reasonable to assume that all experimental subject responses follow the same biological model, but model parameters vary among individuals. These parameters are thus considered as individual non-observed random data. Furthermore, the biological processes are often measured with noise, due to the experimental device. The parametric statistical approach commonly used to analyze these longitudinal data is mixed-effects model methodology (Pinheiro and Bates, 2000). These models have the capacity to discriminate between inter-subjects variability (the variance of the individual regression parameters) and residual variability (measurement noise). The choice of the regression function depends on the biological context. Dynamical biological processes are frequently described by deterministic models defined as

systems of ordinary differential equations. However, these functions may not capture the exact process, as responses for some individuals may display local "random" fluctuations. These phenomena are not due to error measurements but are induced by an underlying biological process that is still unknown or unexplained today. For example, for a neuron in a cell tissue receiving information from a large number of other neurons, it has been observed that neuronal response (membrane potential) is highly variable, "random" and irregular, even under stable conditions. Thus we may assume that there is a random component in the physiological model, that leads to the class of stochastic differential equations (SDEs). Hence, this paper deals with parameter estimation for a mixed model defined by an SDE process (called SDE mixed model).

Though complex, parametric inference for discretely observed general SDEs has been widely investigated. When the transition probability of the diffusion process is explicitly known, Dacunha-Castelle and Florens-Zmirou (1986) propose a consistent maximum likelihood estimator. In the general case, Genon-Catalot and Jacod (1993) and Kessler (1997) propose estimators based on the minimization of suitable contrasts. For fixed sampling interval, Bibby and Sørensen (1995) propose martingale estimating functions. In a biological context, Ditlevsen *et al.* (2005) propose an estimation method based on simulation. Picchini *et al.* (2008) propose estimators based on the Hermite expansion of the transition densities. When combining the case of discrete and noisy observations, parameter estimation is a more delicate statistical problem. It is classical to estimate the unobserved signal by using filtering techniques. Favetto and Samson (2010) propose to compute the exact likelihood using the Kalman filter of a discretely, partially and noisily observed bi-dimensional Ornstein-Uhlenbeck process. They also propose to combine the well-known EM algorithm (Dempster *et al.*, 1977) with the Kalman filter. However, their approach is limited to the case of linear SDEs and Gaussian innovations. Moreover, the extension of these methods to mixed SDEs is not straightforward.

Indeed, parametric estimation of mixed-effects models is complex. The likelihood of mixed models is intractable in a closed form when the regression function is nonlinear. Several estimation methods have been proposed to avoid this difficulty when the regression function is deterministic (Davidian and Giltinan, 1995; Wolfinger, 1993; Pinheiro and Bates, 2000; Kuhn and Lavielle, 2005). Among them, Kuhn and Lavielle (2005) propose to couple the SAEM algorithm, a stochastic version of the EM algorithm, with a Markov Chain Monte Carlo (MCMC) procedure to simulate the unobserved random individual parameters.

When the regression function is a diffusion process, mixed model estimation is made more complicated by the difficulties of SDE estimation. Parametric estimation of SDE with random parameters (without measurement noise) has been studied by Ditlevsen and De Gaetano (2005) and Picchini *et al.* (2010). From a Bayesian point of view, Oravecz *et al.* (2009) study the estimation of an Ornstein-Uhlenbeck process with random parameters. SDE mixed model estimation, including measurement noise modeling, is more complex and hence it has received little attention. Overgaard *et al.* (2005) and Tornøe *et al.* (2005) propose estimators based on an extended Kalman filter, but the algorithm convergence is not proved. Donnet and Samson (2008) propose a maximum likelihood estimator based on the SAEM algorithm, coupled with an MCMC algorithm. They prove the convergence of their algorithm. Donnet *et al.* (2010) propose a general MCMC algorithm for a Bayesian estimation of mixed SDEs model. However, MCMC samplers do not take advantage of the temporal behavior of the hidden dynamic process, especially of its Markovian property. In this context, MCMC samplers typically consist of building an MCMC kernel updating each time component of the hidden process, individually. This leads to slow mixing algorithms that deteriorate the properties of the estimation algorithms (frequentist or Bayesian ones).

In this paper, we propose to combine the SAEM algorithm with a particle filter approach as an alternative to MCMC techniques. Particle filters, also called Sequential Monte Carlo (SMC) techniques, represent a promising alternative to classical filters such as the Kalman filter or the Hamilton-Kitagawa filter and to MCMC samplers. They can deal with nonlinear models and non-Gaussian random variables. In contrast to Hidden Markov Model filters, which work on a state space discretized to a fixed grid, particle filters focus sequentially on the higher density regions of the state space. Different particle filters are proposed in the literature. Doucet *et al.* (2001) provide a state of the art on these methods, discussing both applications and theoretical convergence of the algorithms. SMC methods have been proposed as a powerful tool to estimate the parameters of nonlinear state space models (see Kantas *et al.*, 2009, for an overview). Recently, Andrieu *et al.* (2010) have proposed Particle Markov Chain Monte Carlo (PMCMC) methods, which combine MCMC and SMC techniques, and take advantage of the strength of its two components. To our knowledge, PMCMC methods have been used for parameter estimation only in a Bayesian context and for models without random individual parameters. We propose to focus on maximum likelihood estimation for SDE mixed models by combining

a PMCMC algorithm with the SAEM algorithm. We prove the convergence of the SAEM-PMCMC algorithm towards the maximum likelihood estimator. The proof is based on convergence results of the Robbins-Monro type stochastic approximation scheme (Delyon *et al.*, 1999) and the convergence result of the PMCMC algorithm (Andrieu *et al.*, 2010). Our algorithm SAEM-PMCMC is applied to two simulated examples, an Ornstein-Uhlenbeck mixed SDE model observed with additive noise, and a stochastic volatility mixed model with a time-inhomogeneous SDE. The SAEM-PMCMC algorithm is also compared with the SAEM-MCMC algorithm proposed by Donnet and Samson (2008). We show the increase in accuracy for the SAEM-PMCMC estimates.
The paper is organized as follows. In Section 2, we present the SDE mixed model. In Section 3, the estimation method is detailed and the convergence of the algorithm is proved. Section 4 shows numerical results based on simulated data. In Section 5 we conclude and discuss the advantages and limitations of our approach.

2 Mixed model defined by stochastic differential equation

2.1 Model and notations

Let $\mathbf{y} = (y_{ij})_{1\leq i\leq n, 0\leq j\leq J_i}$ denote the data, where y_{ij} is the noisy measurement of the observed process for individual i at time $t_{ij} \geq t_{i0}$, for $i = 1, \ldots, n$, $j = 0, \ldots, J_i$. Let $\mathbf{y}_{i,0:J_i} = (y_{i0}, \ldots, y_{iJ_i})$ denote the data vector of subject i. We consider that the y_{ij}'s are governed by a mixed-effects model based on a stochastic differential equation defined as follows:

$$
\begin{aligned}
y_{ij} &= g(X_{it_{ij}}, \varepsilon_{ij}), \quad \varepsilon_{ij} \sim_{i.i.d.} \mathcal{N}(0, \sigma^2), && (1)\\
dX_{it} &= a(X_{it}, t, \phi_i)dt + b(X_{it}, t, \phi_i, \gamma)dB_{it}, && (2)\\
X_{it_{i0}}|\phi_i &\sim \pi_0(\cdot|\phi_i), && (3)\\
\phi_i &\sim_{i.i.d.} \mathcal{N}(\mu, \Omega). && (4)
\end{aligned}
$$

In equation (1), $(\varepsilon_{ij})_{i,j}$ is a sequence of i.i.d Gaussian random variables of variance σ^2 representing the measurement errors and g is the error model function, assumed to be known. Several functions g are classically considered. For example, the function $g(x, \varepsilon) = x + \varepsilon$ leads to an additive error model. Mul-

tiplicative error models or stochastic volatility models can be considered with $g(x,\varepsilon) = x(1+\varepsilon)$ or $g(x,\varepsilon) = f(x)\exp(\varepsilon)$ with a known function f.
For each individual i, $(X_{it_{ij}})_{0\leq j\leq J_i}$ are the values at discrete times $(t_{ij})_{0\leq j\leq J_i}$ of the continuous process $(X_{it})_{t\geq t_{i0}} = (X_{it}(\phi_i))_{t\geq t_{i0}}$ defined by equations (2) and (3). This process is proper to each individual through the individual parameters $\phi_i \in \mathbb{R}^p$ involved in the drift and volatility functions a and b and an individual Brownian motion $(B_{it})_{t\geq t_{i0}}$. Functions a and b – which may be functions of time, leading to time-inhomogeneous SDEs – are common to the n subjects. They are assumed to be sufficiently regular to ensure a unique weak solution (Oksendal, 2007). Besides, b depends on an unknown additional volatility parameter γ. The (B_{it}) and ϕ_k are assumed to be mutually independent for all $1 \leq i,k \leq n$. The initial condition $X_{it_{i0}}$ of process $(X_{it})_{t\geq t_{i0}}$, conditionally to the individual parameter ϕ_i is random with distribution $\pi_0(\cdot|\phi_i)$. This distribution can be a point mass distribution, resulting in a deterministic initial condition conditionally to the individual random parameters. In the following, $\mathbf{X}_{i,0:J_i} = (X_{it_{i0}},\ldots,X_{it_{iJ_i}})$ denotes the vector of the i-th process realization at observation times (t_{ij}) for $i = 1,\ldots,n$ and $\mathbf{X} = (\mathbf{X}_1,\ldots,\mathbf{X}_n)$ denotes the whole vector of processes at observation times, for all individuals.
The individual parameters ϕ_i are assumed to be normally distributed with mean p-vector μ and covariance-matrix Ω. We denote $\mathbf{\Phi} = (\phi_1,\ldots,\phi_n)$ the vector of all individual parameter vectors.

***Remark* 1.** *In such models, three fundamentally different types of noise have to be distinguished: the inter-subject variability Ω representing the variance of the individual parameters ϕ_i, the dynamic noise γ, reflecting the random fluctuations of the real process around the corresponding theoretical dynamic model, and the measurement noise σ, representing the uncorrelated part of the residual variability associated with assay and sampling errors for instance.*

Note that it is straightforward to extend this setting to a multidimensionnal process (X_t) and a multidimensionnal observation $\mathbf{y}$.
Notice that considering the discrete realizations $\mathbf{X}_{i,0:J_i} = (X_{it_{i0}},\ldots,X_{it_{iJ_i}})$ of each individual process, our model belongs to the wide framework of state space models (with random parameters), the latent processes $\mathbf{X}_{i,0:J_i}$ being Markovian. We detail this property in Section 4.

2.2 Likelihood function

Let $\theta = (\mu, \Omega, \gamma, \sigma)$ be the parameter vector which belongs to some open subset Θ of the Euclidean space $\mathbb{R}^q$ with q the number of unknown parameters. Our objective is to propose a maximum likelihood estimation of θ. We consider that the likelihood function is well-defined. More precisely, the distribution of X_{it} given ϕ_i and $X_{is} = x_s$, $s < t$, has a strictly positive density with respect to the Lebesgue measure, which is denoted

$$x \mapsto p(x, t-s, s|x_s, \phi_i; \theta) > 0. \tag{5}$$

Note that this density may depend on time s for a time-inhomogeneous SDE. This allows us to write the likelihood of model (1-4) as

$$p(\mathbf{y}; \theta) = \prod_{i=1}^{n} \int p(\mathbf{y}_{i,0:J_i}|\mathbf{X}_{i,0:J_i}; \sigma) p(\mathbf{X}_{i,0:J_i}|\phi_i; \gamma) p(\phi_i; \mu, \Omega) d\phi_i d\mathbf{X}_{i,0:J_i}, \tag{6}$$

where $p(y|x;\sigma)$, $p(x|\phi;\gamma)$ and $p(\phi;\mu,\Omega)$ are density functions of the observations given the diffusion process, the diffusion process given the individual parameters and the individual parameters, respectively. By independence of the measurement errors (ε_{ij}), we have $p(\mathbf{y}_{i,0:J}|\mathbf{X}_{i,0:J_i};\sigma) = \prod_{j=0}^{J_i} p(y_{ij}|X_{it_{ij}};\sigma)$ where $p(y_{ij}|X_{it_{ij}};\sigma)$ is the density function associated to the error model (1). The Markovian property of the diffusion process (X_{it}) implies

$$p(\mathbf{X}_{i,0:J_i}|\phi_i;\theta) = \pi_0(X_{it_{i0}}|\phi_i) \prod_{j=1}^{J_i} p(X_{it_{ij}}, \Delta_{ij}, t_{ij-1}|X_{it_{ij-1}}, \phi_i; \gamma),$$

where $\Delta_{ij} = t_{ij} - t_{ij-1}$ is the step size between two observations.
The integral (6) has generally no-closed form. The only case with an explicit likelihood is a mixed SDE model defined by a Wiener process with drift

$$dX_{it} = \phi_i dt + \gamma dB_{it},$$

and an additive Gaussian error model $g(x, \varepsilon) = x + \varepsilon$. In this case, as $X_t|\phi$ is Gaussian with expectation depending linearly on ϕ, the vector data $\mathbf{y}$ is also Gaussian. In the other cases, including linear SDEs, the integral (6) has no-closed form. For example, the case of the linear Ornstein-Uhlenbeck mixed model is detailed in Section 4.

3 Estimation method

Maximizing (6) with respect to θ yields the corresponding Maximum Likelihood Estimate (MLE) $\hat{\theta}$ of the mixed SDE model (1-4). As the likelihood is not explicit, it has to be numerically evaluated or maximized. This requires to know the transition densities (5). Thus we consider the following assumption:

(M0) The SDE defined by (2) is assumed to have an explicit transition density $p(x, t-s, s|x_s, \phi_i; \theta)$.

The extension to general SDEs will be investigated in future works.
The estimation is complex because the n random parameters ϕ_i and random trajectories ($\mathbf{X}_{i,0:J_i}$) are not observed. This statistical problem can be viewed as an incomplete data model. The observable vector $\mathbf{y}$ is thus considered as part of a so-called complete vector $(\mathbf{y}, \mathbf{X}, \mathbf{\Phi})$.

3.1 Estimation algorithm (SAEM algorithm)

The Expectation-Maximization (EM) algorithm (Dempster *et al.*, 1977) is useful in situations where the direct maximization of the marginal likelihood $\theta \rightarrow p(\mathbf{y}; \theta)$ is more complex than the maximization of the conditional expectation of the complete likelihood

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

where $p(\mathbf{y}, \mathbf{X}, \mathbf{\Phi}; \theta)$ is the likelihood of the complete data $(\mathbf{y}, \mathbf{X}, \mathbf{\Phi})$. The EM algorithm is an iterative procedure: at the m-th iteration, given the current value $\widehat{\theta}_{m-1}$ of the parameters, the E-step is the evaluation of $Q_m(\theta) = Q(\theta \,|\, \widehat{\theta}_{m-1})$ while the M-step updates $\widehat{\theta}_{m-1}$ by maximizing $Q_m(\theta)$. For cases where the E-step has no closed form, Delyon *et al.* (1999) propose the Stochastic Approximation EM algorithm (SAEM) and replace the E-step by a stochastic approximation of $Q_m(\theta)$. The E-step is thus divided into a simulation step (S-step) of the non-observed data $(\mathbf{X}_m, \mathbf{\Phi}_m)$ with the conditional distribution $p(\mathbf{X}, \mathbf{\Phi}\,|\mathbf{y}; \widehat{\theta}_{m-1})$ and a stochastic approximation step (SA-step):

$$Q_m(\theta) = Q_{m-1}(\theta) + \alpha_m \left[\log p\left(\mathbf{y}, \mathbf{X}_m, \mathbf{\Phi}_m; \widehat{\theta}_{m-1}\right) - Q_{m-1}(\theta)\right],$$

where $(\alpha_m)_{m\in\mathbb{N}}$ is a sequence of positive numbers decreasing to zero. The advantage of the SAEM algorithm is the low-level dependence on the initialization

$\widehat{\theta}_0$, due to the stochastic approximation of the E-step.
To fulfill the convergence conditions of the SAEM algorithm (Delyon *et al.*, 1999), we consider the exponential case. More precisely, we assume:

(M1) The parameter space Θ is an open subset of $\mathbb{R}^q$. The complete likelihood $p(\mathbf{y}, \mathbf{X}, \mathbf{\Phi})$ belongs to the exponential family i.e.

$$\log p(\mathbf{y}, \mathbf{X}, \mathbf{\Phi}; \theta) = -\psi(\theta) + \langle S(\mathbf{y}, \mathbf{X}, \mathbf{\Phi}), \nu(\theta) \rangle \,,$$

where ψ and ν are two functions of θ, $S(\mathbf{y}, \mathbf{X}, \mathbf{\Phi})$ is known as the minimal sufficient statistics of the complete model, taking its value in a subset $\mathcal{S}$ of $\mathbb{R}^d$ and $\langle \cdot, \cdot \rangle$ is the scalar product on $\mathbb{R}^d$.

In this case, the SA-step of the SAEM algorithm reduces to the approximation of $E\left[S(\mathbf{y}, \mathbf{X}, \mathbf{\Phi})|\mathbf{y}; \theta'\right]$.
At step S, a simulation under the conditional distribution $p(\mathbf{X}, \mathbf{\Phi}|\mathbf{y}; \widehat{\theta}_{m-1})$ is required. However, this distribution is likely to be a complex distribution, resulting in the impossibility of a direct simulation of the non-observed data $(\mathbf{X}, \mathbf{\Phi})$. For such cases, Kuhn and Lavielle (2005) suggest to realize the simulation step via a Markov Chain Monte Carlo (MCMC) scheme, resulting in the SAEM-MCMC algorithm. The S-step consists in constructing a Markov chain with $p(\mathbf{X}, \mathbf{\Phi}|\mathbf{y}; \widehat{\theta}_{m-1})$ as ergodic distribution at the m-th iteration. However, standard MCMC algorithms have reached their limits in high dimensional contexts and we propose to replace them by the Particular MCMC algorithm proposed by Andrieu *et al.* (2010). We detail the S-step in the following subsection.

3.2 Simulation of the latent variables $(\mathbf{X}, \Phi)$ conditionally to the observations y

Conditionally to the observations $\mathbf{y}$, the individuals are independent and the simulation of each $(\mathbf{X}_{i,0:J_i}, \phi_i)$ conditionally to $\mathbf{y}_{i,0:J_I}$ can be performed separately. To ease the reading, in this subsection, we focus on the simulation of the non observed variables $(\mathbf{X}_{i,0:J_i}, \phi_i)$ of a single individual. *Therefore, we omit the index i* in the following. Moreover, we denote σ and γ the current values of the parameters and omit the index m of the SAEM algorithm iteration.

3.2.1 Standard Markov Chain Monte Carlo algorithm

To generate realizations of the distribution of interest $p(\mathbf{X}_{0:J}, \phi|\mathbf{y}_{0:J}; \gamma, \sigma)$, MCMC algorithms are routinely applied in the literature. These methods alternately simulate the trajectory $\mathbf{X}_{0:J}$ under the distribution $p(\mathbf{X}_{0:J}|\mathbf{y}_{0:J}, \phi; \gamma, \sigma)$ and the parameter ϕ under the distribution $p(\phi|\mathbf{y}_{0:J}, \mathbf{X}_{0:J}; \gamma, \sigma)$. Even if we assume that the transition density (5) is explicit, these two conditional distributions are likely not to be explicit.
If the conditional distributions $p(\mathbf{X}_{0:J}|\mathbf{y}_{0:J}, \phi; \gamma, \sigma)$ and $p(\phi|\mathbf{y}_{0:J}, \mathbf{X}_{0:J}; \gamma, \sigma)$ are easily simulated then the algorithm amounts to a Gibbs algorithm (see for instance Example 1 in Section 4). On the contrary, difficulties arise if these distributions are not explicit or cannot be simulated easily. In that case, we resort to Metropolis Hastings algorithms for each component of ϕ and for each time component of the vector $\mathbf{X}_{0:J}$ depending on a proposal distribution. The construction of the proposal distribution has been widely studied in the literature (see *e.g.* Gilks *et al.*, 1996). We describe a standard random walk Metropolis-Hastings algorithm for Example 2 in Section 4. However, it is well known that the efficiency of these methods is not satisfactory in high dimensional contexts. Furthermore, it does not take advantage of the temporal behavior of the hidden process (X_t). Hidden states are typically updated each time component X_j individually and iteratively. This leads to a slow mixing MCMC algorithm, inducing poor convergence properties for the SAEM-MCMC algorithm.

3.2.2 Particle Markov Chain Monte Carlo algorithm

Andrieu *et al.* (2010) recently proposed an innovative simulation scheme, the PMCMC algorithm, which takes advantage of the strength of MCMC and Sequential Monte Carlo (SMC) algorithms. More precisely, they suggest using this second class of stochastic methods to design efficient high dimensional proposal distributions for MCMC algorithms. In the following, we first recall the SMC algorithm and its inclusion in PMCMC as proposed by Andrieu *et al.* (2010). Then we expose its inclusion in the SAEM algorithm.

Sequential Monte Carlo algorithm

The aim of the SMC algorithm is to obtain a set of K particles $(\mathbf{X}_{0:J}^{(k)})_{k=1\dots K}$ and weights $(W_{0:J}^{(k)})_{k=1\dots K}$ approximating the conditional distribution $p(\mathbf{X}_{0:J}|$ $\mathbf{y}_{0:J}, \phi;\ \gamma, \sigma)$. The SMC algorithm is a stochastic version of the filtering procedure (see Doucet *et al.*, 2001, for a complete review). This algorithm is

an iterative method supplying at step j, for $j = 0, \ldots, J$, the distribution $p(X_j|\mathbf{y}_{0:j}, \phi; \gamma, \sigma)$. It relies on a proposal distribution $q(X_j|y_j, X_{j-1}, \phi; \gamma, \sigma)$. To avoid the problem of degeneracy of the particles, a resampling step is introduced resulting in the following algorithm:

***Algorithm* 1 (SMC algorithm)**
Approximation of the conditional distribution $p(\mathbf{X}_{0:J}|\mathbf{y}_{0:J}, \phi; \gamma, \sigma)$

- At time $j = 0$: $\forall\, k = 1, \ldots, K$

  (1) sample $X_0^{(k)} \sim \pi_0(\cdot|\phi)$

  (2) compute and normalize the weights:

$$w_0\left(X_0^{(k)}\right) = p\left(y_0, X_0^{(k)}\,|\phi; \gamma, \sigma\right), \quad W_0\left(X_0^{(k)}\right) = \frac{w_0\left(X_0^{(k)}\right)}{\sum_{k=1}^K w_0\left(X_0^{(k)}\right)}$$

- At time $j = 1, \ldots, J$: $\forall\, k = 1, \ldots, K$

  (0) resample the particles, i.e. sample the indices $A_{j-1}^{(k)}$ *with a multinomial distribution such that*

$$P(A_{j-1}^{(k)} = l) = W_{j-1}(\mathbf{X}_{0:j-1}^l), \quad \forall\, l = 1, \ldots, K$$

  (1) sample $X_j^{(k)} \sim q\left(\cdot|y_j, X_{j-1}^{A_{j-1}^{(k)}}, \phi; \gamma, \sigma\right)$ *and set* $\mathbf{X}_{0:j}^{(k)} = (\mathbf{X}_{0:j-1}^{A_{j-1}^{(k)}}, X_j^{(k)})$

  (2) compute and normalize the weights:

$$\begin{aligned} w_j\left(\mathbf{X}_{0:j}^{(k)}\right) &= \frac{p\left(\mathbf{y}_{0:j}, \mathbf{X}_{0:j}^{(k)}|\phi; \gamma, \sigma\right)}{p\left(\mathbf{y}_{0:j-1}, \mathbf{X}_{0:j-1}^{A_{j-1}^{(k)}}|\phi; \gamma, \sigma\right) q\left(X_j^{(k)}|y_j, X_{j-1}^{A_{j-1}^{(k)}}, \phi; \gamma, \sigma\right)} \\ W_j(\mathbf{X}_{0:j}^{(k)}) &= \frac{w_j\left(\mathbf{X}_{0:j}^{(k)}\right)}{\sum_{k=1}^K w_j\left(\mathbf{X}_{0:j}^{(k)}\right)} \end{aligned}$$

Choice of the proposal distribution $q(X_j|y_j, X_{j-1}, \phi; \gamma, \sigma)$
As in any importance sampling method, the choice of the proposal distribution q is crucial to ensure good convergence properties. If possible, the conditional

posterior

$$\begin{aligned} q(X_j|y_j, X_{j-1}, \phi; \gamma, \sigma) &= p(X_j|y_j, X_{j-1}, \phi; \gamma, \sigma) \\ &\propto p(y_j|X_j; \sigma) p(X_j|X_{j-1}, \phi; \gamma), \end{aligned}$$

is the best proposal distribution. If its simulation is intractable, $q(X_j|y_j, X_{j-1}, \phi; \gamma, \sigma)$ should be chosen as close as possible to this ideal distribution $p(X_j|y_j, X_{j-1}, \phi; \gamma, \sigma)$. In Example 2 of Section 4, we propose a Gaussian approximation of the posterior conditional distribution. An optimal importance density q is discussed in Pitt and Shephard (1999) but in many cases the deriving importance weights $\{w_j^{(k)}\}_{k=1}^K$ do not have an explicit expression. Another option can be the predictive distribution $p(X_j|X_{j-1}, \phi; \gamma, \sigma)$ as proposal distribution. This non-optimal choice can work well in particular contexts if the dimension of latent variables is not too large (Andrieu *et al.*, 2010). However, when combined with an estimation procedure, this choice may be rather suboptimal, as illustrated in Example 1 of Section 4.

Output of the SMC algorithm

The SMC algorithm supplies a set of K particles and weights approximating the conditional distribution $p(\mathbf{X}_{0:J}|\mathbf{y}_{0:J}, \phi; \gamma, \sigma)$. In general, we are interested in one realization of the trajectory $\mathbf{X}_{0:J}$ under the conditional distribution $p(\mathbf{X}_{0:J}|\mathbf{y}_{0:J}, \phi; \gamma, \sigma)$. This is directly achieved by randomly choosing one particle among the K particles with weights $(W_J(\mathbf{X}_{0:J}^{(k)}))_{k=1\ldots K}$.

In the following, we denote by "a run of the SMC algorithm" the generation of one realization under the approximate SMC distribution.

Approximation of the conditional likelihood $p(y_j|\mathbf{y}_{0:j-1}, \phi; \gamma, \sigma)$

As described in Andrieu *et al.* (2010), the algorithm also provides an estimation of the conditional likelihood $p(y_j|\mathbf{y}_{0:j-1}, \phi; \gamma, \sigma)$ by:

$$\begin{aligned} \widehat{p}(y_0|\phi; \gamma, \sigma) &= \tfrac{1}{K} \textstyle\sum_{k=1}^K w_0 \left(X_0^{(k)}\right), \\ \widehat{p}(y_j|\mathbf{y}_{0:j-1}, \phi; \gamma, \sigma) &= \tfrac{1}{K} \textstyle\sum_{k=1}^K w_j \left(\mathbf{X}_{0:j}^{(k)}\right) \quad \forall\, j = 1, \ldots, J. \end{aligned} \tag{7}$$

As the quantity $p(\mathbf{y}_{0:J}|\phi; \gamma, \sigma)$ can be decomposed into the product $p(\mathbf{y}_{0:J}|\phi; \gamma, \sigma) = p(y_0|\phi; \gamma, \sigma) \prod_{j=1}^J p(y_j|\mathbf{y}_{0:j-1}, \phi; \gamma, \sigma)$, we estimate it by

$$\widehat{p}(\mathbf{y}_{0:J}|\phi; \gamma, \sigma) = \widehat{p}(y_0|\phi; \gamma, \sigma) \prod_{j=1}^J \widehat{p}(y_j|\mathbf{y}_{0:j-1}, \phi; \gamma, \sigma). \tag{8}$$

***Remark* 2.** *We are able to estimate the conditional likelihood distribution for a fixed ϕ. But we underline that this conditional likelihood differs from the likelihood (6) since the individual parameter ϕ has still not been integrated.*

Particle Markov Chain Monte Carlo algorithm

The PMCMC algorithm, proposed by Andrieu *et al.* (2010), is an MCMC algorithm which jointly updates ϕ and $\mathbf{X}_{0:J}$ conditionally to $\mathbf{y}_{0:J}$. A new candidate $(\mathbf{X}^c_{0:J}, \phi^c)$ is generated with a proposal distribution of the type:

$$q(\mathbf{X}^c_{0:J}, \phi^c | \mathbf{X}_{0:J}, \phi; \theta) = q(\phi^c|\phi) p(\mathbf{X}^c_{0:J} | \mathbf{y}_{0:J}, \phi^c; \gamma, \sigma).$$

where $q(\phi^c|\phi)$ is the proposal on ϕ to be designed: ϕ is a low dimensional parameter and so does not raise specific difficulties. For instance, a random walk proposal can be considered (see Section 4). The proposal distribution on the trajectories $p(\mathbf{X}^c_{0:J}|\mathbf{y}_{0:J}, \phi^c; \gamma, \sigma)$ is (approximately) sampled through the SMC algorithm presented above.
Consequently, the acceptation probability can be written as follows:

$$\begin{aligned}
\rho(\mathbf{X}^c_{0:J}, \phi^c | \mathbf{X}_{0:J}, \phi) &= \min\left\{1, \frac{p(\mathbf{X}^c_{0:J}, \phi^c | \mathbf{y}_{0:J}; \gamma, \sigma)}{p(\mathbf{X}_{0:J}, \phi | \mathbf{y}_{0:J}; \gamma, \sigma)} \frac{q(\phi|\phi^c) p(\mathbf{X}_{0:J}|\mathbf{y}_{0:J}, \phi; \gamma, \sigma)}{q(\phi^c|\phi) p(\mathbf{X}^c_{0:J}|\mathbf{y}_{0:J}, \phi^c; \gamma, \sigma)}\right\} \\
&= \min\left\{1, \frac{q(\phi|\phi^c)}{q(\phi^c|\phi)} \frac{p(\mathbf{y}_{0:J}|\phi^c; \gamma, \sigma) p(\phi^c; \mu, \Omega)}{p(\mathbf{y}_{0:J}|\phi; \gamma, \sigma) p(\phi; \mu, \Omega)}\right\},
\end{aligned}$$

which can be approximated through the SMC algorithm by:

$$\widehat{\rho}(\mathbf{X}^c_{0:J}, \phi^c | \mathbf{X}_{0:J}, \phi) = \min\left\{1, \frac{q(\phi|\phi^c)}{q(\phi^c|\phi)} \frac{\widehat{p}(\mathbf{y}_{0:J}|\phi^c; \gamma, \sigma) p(\phi^c; \mu, \Omega)}{\widehat{p}(\mathbf{y}_{0:J}|\phi; \gamma, \sigma) p(\phi, \mu, \Omega)}\right\}. \qquad (9)$$

The N iterations of the PMCMC algorithm, performed at the S-step of each iteration of the SAEM algorithm and for each individual i, are finally:

***Algorithm* 2** (**PMCMC algorithm**)
Approximation of the conditional distribution $p(\mathbf{X}_{0:J}, \phi|\mathbf{y}_{0:J}; \gamma, \sigma)$

- Initialization *: starting from an arbitrary $\phi(0)$, generate $\mathbf{X}_{0:J}(0)$ by a run of SMC algorithm targeting $p(\mathbf{X}_{0:J}|\mathbf{y}_{0:J}, \phi(0); \gamma, \sigma)$*
- At iteration $\ell = 1, \ldots, N$
  1. *Sample a candidate $\phi^c \sim q(\cdot|\phi(\ell-1))$*

2. *By a run of the SMC algorithm, generate* $\mathbf{X}_{0:J}^c$ *targeting* $p(\cdot|\mathbf{y}_{0:J}, \phi^c; \gamma, \sigma)$ *and compute* $\widehat{p}(\mathbf{y}_{0:J}|\phi^c; \gamma, \sigma)$.

3. *Set* $(\mathbf{X}_{0:J}(\ell), \phi(\ell)) = (\mathbf{X}_{0:J}^c, \phi^c)$ *with probability* $\widehat{\rho}\,(\mathbf{X}_{0:J}^c, \phi^c|\mathbf{X}_{0:J}(\ell-1), \phi(\ell-1))$. *If the candidate is not accepted, then set* $(\mathbf{X}_{0:J}(\ell), \phi(\ell)) = (\mathbf{X}_{0:J}(\ell-1), \phi(\ell-1))$.

We denote $\widehat{p}^N(\mathbf{X}_{0:J}, \phi|\mathbf{y}_{0:J}; \gamma, \sigma)$ the approximated distribution of $p(\mathbf{X}_{0:J}, \phi|\mathbf{y}_{0:J}; \gamma, \sigma)$ obtained by N iterations of the PMCMC algorithm. Andrieu *et al.* (2010) prove the following convergence result for the PMCMC algorithm under two additional assumptions.

(PMCMC1) For any θ, for any ϕ, we define the two sets $S_j^\phi = \{x_{0:j} : p(x_{0:j}|\mathbf{y}_{0:j}, \phi; \theta) > 0\}$ and $Q_j^\phi = \{x_{0:j} : p(x_{0:j-1}|\mathbf{y}_{0:j-1}, \phi; \theta)q(x_j|x_{0:j-1}, y_j, \phi; \theta) > 0\}$. We assume that $S_j^\phi \subseteq Q_j^\phi$ for $j = 0, \ldots, J$.

(PMCMC2) The Metropolis-Hasting sampler of target density $p(\phi|\mathbf{y}_{0:J}; \theta)$ with proposal density $q(\phi^c|\phi)$ is irreducible and aperiodic.

***Proposition* 1** (Andrieu *et al.* (2010))**.** *Under assumptions (PMCMC1-2), the PMCMC algorithm generates a sequence of distributions* $\hat{p}^N(\mathbf{X}, \Phi|\mathbf{y}; \theta)$ *such that for all* $\theta \in \Theta$,

$$||\hat{p}^N(\mathbf{X}, \Phi|\mathbf{y}; \theta) - p(\mathbf{X}, \Phi|\mathbf{y}; \theta)||_{TV} \xrightarrow[N\to\infty]{} 0.$$

where $||\cdot||_{TV}$ *is the total variation distance.*

We thus propose to use the PMCMC algorithm to simulate the non-observed data in the SAEM algorithm resulting into the following SAEM-PMCMC algorithm:

***Algorithm* 3** (SAEM-PMCMC algorithm)

- Iteration 0: *initialization of* $\widehat{\theta}_0$ *and* $s_0 = E\left[S(\mathbf{y}, \mathbf{X}, \mathbf{\Phi})|\mathbf{y}; \widehat{\theta}_0\right]$.
- Iteration $m = 1, \ldots, M$:

S-Step: $*$ *For each individual* i,

- *Initialize the PMCMC algorithm with* $\phi_i(0) = \widehat{\mu}_{m-1}$
- *Run* N_m *iterations of the PMCMC algorithm with* K *particles at each iteration, approximating* $p(\mathbf{X}_{i,0:J_i}, \phi_i|\mathbf{y}_{i,0:J_i}; \widehat{\theta}_{m-1})$ *with the approximate distribution* $\hat{p}^{N_m}(\mathbf{X}_{i,0:J_i}, \phi_i|\mathbf{y}_{i,0:J_i}; \widehat{\theta}_{m-1})$

* *Set* $\mathbf{X}_m = (\mathbf{X}_{1,0:J}(N_m), \ldots, \mathbf{X}_{n,0:J}(N_m))$ *and* $\Phi_m = (\phi_1(N_m), \ldots, \phi_n(N_m))$ *a realization of the non observed variables under the approximated distribution* $\hat{p}^{N_m}(\mathbf{X}, \Phi|\mathbf{y}; \widehat{\theta}_{m-1})$

SA-Step: *update of* s_{m-1} *using the stochastic approximation scheme:*

$$s_m = s_{m-1} + \alpha_m \left[S(\mathbf{y}, \mathbf{X}_m, \mathbf{\Phi}_m) - s_{m-1}\right] \tag{10}$$

M-Step: *update of* $\widehat{\theta}_m$ *by:*

$$\widehat{\theta}_m = \arg\max_{\theta} \left(-\psi(\theta) + \langle s_m, \nu(\theta)\rangle\right).$$

3.3 Convergence of the SAEM-PMCMC estimation algorithm

In the following, we set $\mathbf{Z} = (\mathbf{X}, \mathbf{\Phi})$. The convergence of the SAEM algorithm combined with the PMCMC is principally based on the work of Delyon *et al.* (1999). We first consider the additional regularity assumptions required for the convergence of the SAEM algorithm (Delyon *et al.*, 1999).

(M2) The functions $\psi(\theta)$ and $\nu(\theta)$ are twice continuously differentiable on Θ.

(M3) The function $\bar{s} : \Theta \longrightarrow \mathcal{S}$ defined as $\bar{s}(\theta) = \int S(y,z)p(z|y;\theta)dydz$ is continuously differentiable on Θ.

(M4) The function $\ell(\theta) = \log p(y, \theta)$ is continuously differentiable on Θ and

$$\partial_\theta \int p(y,z;\theta)dydz = \int \partial_\theta p(y,z;\theta)dydz.$$

(M5) Define $L : \mathcal{S} \times \Theta \longrightarrow \mathbb{R}$ as $L(s,\theta) = -\psi(\theta) + \langle s, \nu(\theta)\rangle$. There exists a function $\hat{\theta} : \mathcal{S} \longrightarrow \Theta$ such that

$$\forall \theta \in \Theta, \quad \forall s \in \mathcal{S}, \quad L(s, \hat{\theta}(s)) \geq L(s, \theta).$$

(SAEM1) The positive decreasing sequence of the stochastic approximation $(\alpha_m)_{m\geq 0}$ is such that $\sum_m \alpha_m = \infty$ and $\sum_m \alpha_m^2 < \infty$.

(SAEM2) $\ell : \Theta \to \mathbb{R}$ and $\hat{\theta} : \mathcal{S} \to \Theta$ are d times differentiable, where d is the dimension of $S(y,z)$.

(SAEM3) For any positive Borel function f

$$E(f(Z_{m+1})|\mathcal{F}_m) = \int f(z)p(z|y,\widehat{\theta}_m)dz,$$

where $\{\mathcal{F}_m\}$ is the increasing family of σ-algebra generated by the random variables $s_0, Z_1, Z_2, \ldots, Z_m$.

(SAEM4) For all $\theta \in \Theta$, $\int ||S(y,z)||^2 p(z|y;\theta)dz < \infty$ and the function $\Gamma(\theta) = Cov_\theta(S(\mathbf{Z}))$ is continuous.

Assumptions (M0-5) ensure the convergence of the EM algorithm. Assumptions (M0-5) and (SAEM1-4) with the fact that $(s_m)_{m\geq 0}$ takes its values in a compact subset of $\mathcal{S}$ ensure the convergence of the SAEM estimates to a stationary point of the observed likelihood when the simulation step is exact (Delyon *et al.*, 1999).

However the SAEM-PMCMC algorithm is more complex as the simulation step is not exact. Indeed, the stochastic approximation scheme differs from the classic SAEM stochastic approximation scheme by introducing a remainder term measuring the error induced by the simulation step. More precisely, we can decompose our stochastic approximation (10) into:

$$s_{m+1} = s_m + \alpha_m h(s_m) + \alpha_m e_m + \alpha_m r_m$$

with

$$\begin{aligned} h(s_m) &= E(S(y,z)|y,\hat{\theta}(s_m)) - s_m \\ e_m &= S(y,z^{(m)}) - E^{N_m}(S(y,z)|y,\hat{\theta}(s_m)) \\ r_m &= E^{N_m}(S(y,z)|y,\hat{\theta}(s_m)) - E(S(y,z)|y,\hat{\theta}(s_m)) \end{aligned}$$

where we denote $E(\cdot|y,\theta)$ the expectation under the exact distribution $p(z|y,\theta)$ and $E^{N_m}(\cdot|y,\theta)$ the expectation under the approximate distribution $\hat{p}^{N_m}(z|y,\theta)$ obtained with N_m iterations of the PMCMC algorithm. Consequently the proof of the SAEM-PMCMC convergence differs from SAEM-MCMC and is based on convergence results of the Robbins Monro type stochastic approximation procedure. We make the following assumptions:

(SAEM-PMCMC1) For all $\theta \in \Theta$ and for all N, the iteration number of the PMCMC algorithm, $\int ||S(y,z)||^2 \hat{p}^N(z|y;\theta)dz < \infty$

(SAEM-PMCMC2) The function $S(y,z)$ is uniformly bounded in y and z.

Theorem 1. *Assume that (M0-5), (SAEM1-2), (PMCMC1-2), (SAEM-PMCMC1-2) hold. Then, with probability 1,* $\lim_{m\to\infty} d(\widehat{\theta}_m, \mathcal{L}) = 0$ *where* $\mathcal{L} = \{\theta \in \Theta, \partial_\theta \ell(\theta) = 0\}$ *is the set of stationary points of the log-likelihood* $\ell(\theta) = \log p(y;\theta)$.

Proof Following Theorem 2 of Delyon *et al.* (1999) on the convergence of Robbins-Monro scheme, the convergence of SAEM-PMCMC is ensured if we prove the following assertions:

1. The function $V(s) = -\ell(\hat{\theta}(s))$ is such that for all $s \in \mathcal{S}$, $F(s) = \langle \partial_s V(s), h(s)\rangle \leq 0$ and such that the set $V(\{s, F(s) = 0\})$ is of zero measure.
2. $\lim_{m\to\infty} \sum_{k=1}^{m} \alpha_k e_k$ exists and is finite with probability 1.
3. $\lim_{m\to\infty} r_m = 0$ with probability 1.

Assertion 1 is proved by Lemma 2 of Delyon *et al.* (1999) under assumptions (M0-M5) and (SAEM2).
Assertion 2 is proved similarly as Theorem 5 of Delyon *et al.* (1999). Indeed, by construction of the PMCMC algorithm, due to the choice of PMCMC and its initialization, the equivalent of assumption (SAEM3) is checked for the approximate distribution $\hat{p}^N(\mathbf{Z}|\mathbf{y}, \widehat{\theta}_m)$. Indeed, the assumption of independence of the non-observed variables $\mathbf{Z}_1, \ldots, \mathbf{Z}_m$ given $\widehat{\theta}_0, \ldots, \widehat{\theta}_m$ is verified. As a consequence, we have: for any positive Borel function f, $E^{N_m}(f(Z_{m+1})|\mathcal{F}_m) = \int f(z)\hat{p}^{N_m}(z|\mathbf{y}, \widehat{\theta}_m)dz$. We can then prove that $\sum_{k=1}^{m} \alpha_k e_k$ is a martingale, bounded in L_2 under assumptions (M5) and (SAEM1)-(SAEM2).
To verify assertion 3, it is sufficient to prove that $E(r_m)$ and $Var(r_m)$ converge to 0. By assumption (SAEM-PMCMC2), there exists a constant $M > 0$ such that

$$|E(r_m)| \leq M\, ||\hat{p}^{N_m}(z|y;\hat{\theta}(s_m)) - p(z|y;\hat{\theta}(s_m))||_{TV}.$$

By (PMCMC1)-(PMCMC2) and Proposition 1, we deduce the convergence to 0 of $E(r_m)$ by choosing an increasing sequence N_m. Similarly, there exists a constant $C > 0$ such that

$$\begin{aligned} Var(r_m) &= E\left[\left\{\int (\hat{p}^{N_m}(z|y;\hat{\theta}(s_m)) - p(z|y;\hat{\theta}(s_m)))\,(S(y,z) - \right.\right. \\ &\qquad \left.\left. \int S(y,z)p(y)dy\right)dz\right\}^2\Bigg] \\ &\leq C\, E\left[||\hat{p}^{N_m}(z|y;\hat{\theta}(s_m)) - p(z|y;\hat{\theta}(s_m))||_{TV}^2\right] \end{aligned}$$

and we deduce the convergence of $Var(r_m)$ to 0. □

Just as for the SAEM algorithm, we can deduce the algorithm convergence to a (local) maximum of the likelihood.

***Corollary* 1**
Under assumptions of Theorem 1 and additional assumptions (LOC1)-(LOC3) proposed by Delyon et al. *(1999) on the regularity of the log-likelihood, the sequence $\widehat{\theta}_m$ converges with probability 1 to a (local) maximum of the likelihood.*

Proof Theorem 6 of Delyon *et al.* (1999) can be extended without difficulty to the SAEM-PMCMC algorithm. It proves that under assumptions of Theorem 1 and (LOC1), the sequence $\widehat{\theta}_m$ converges to a fixed point of the EM-mapping $T(\widehat{\theta}_m) = \hat{\theta}(s(\widehat{\theta}_m))$. Assumptions (LOC2-3), Lemma 3 of Delyon *et al.* (1999) and application of Brandière and Duflo (1996) imply that the sequence $\widehat{\theta}_m$ converges with probability 1 to a proper maximum of the likelihood. □
Assumption (SAEM-PMCMC2) is a strong hypothesis but it is only technical. This assumption could be relaxed using a principle of random boundaries presented in Allassonnière *et al.* (2009).

4 Simulation study

Two models are considered in the simulation study: the Ornstein-Uhlenbeck process and the stochastic volatility time-inhomogeneous Gompertz process. For each mixed SDE model, 100 datasets are generated with n subjects and identical design with $J+1$ data points for all the individuals . Different sets of parameters and different values of n and J are used. The corresponding MLE is obtained on each data set using both the SAEM-PMCMC and the SAEM-MCMC algorithms. Let $\widehat{\theta}_r$ denote the estimate of θ obtained on the r-th simulated dataset, for $r = 1, \ldots, 100$ by the corresponding algorithm. The relative bias $\frac{1}{100}\sum_r \frac{(\widehat{\theta}_r - \theta)}{\theta}$ and relative root mean square error (RMSE) $\sqrt{\frac{1}{100}\sum_r \frac{(\widehat{\theta}_r - \theta)^2}{\theta^2}}$ for each component of θ are computed for both algorithms.
The implementation of the SAEM algorithm requires initial value θ_0 and the choice of the stochastic approximation sequence $(\alpha_m)_{m\geq 0}$. The initial values of the parameters are chosen arbitrarily as the convergence of the SAEM algorithm little depends on the initialization. The step of the stochastic approximation scheme is chosen as recommended by Kuhn and Lavielle (2005): $\alpha_m = 1$ during the first iterations $1 \leq m \leq M_1$, and $\alpha_m = \frac{1}{(m-M_1)^{0.8}}$ during the subsequent

ones. Indeed, the initial guess θ_0 might be far from the maximum likelihood value and the first iterations with $\alpha_m = 1$ allow the sequence of estimates to converge to a neighborhood of the maximum likelihood estimate. Subsequently, smaller step sizes during $M - M_1$ additional iterations ensure the almost sure convergence of the algorithm to the maximum likelihood estimate. We implement the SAEM algorithm with $M_1 = 60$ and $M = 100$ iterations. The PMCMC algorithm is implemented with $K = 50$ particles and $N = 100$ iterations. The MCMC algorithm is implemented with 50 iterations of the chain. We now detail the two examples.

4.1 Example 1: Ornstein-Uhlenbeck process

4.1.1 Ornstein-Uhlenbeck mixed model

The Ornstein-Uhlenbeck process has been widely used in neuronal modeling, biology, and finance (see *e.g.* Kloeden and Platen, 1992). Consider an SDE mixed model driven by the Ornstein-Uhlenbeck process and an additive error model

$$\begin{aligned} y_{ij} &= X_{it_{ij}} + \varepsilon_{ij}, \;\; \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2), \\ dX_{it} &= -\left(\frac{X_{it}}{\tau_i} - \kappa_i\right) dt + \gamma dB_{it}, \\ X_0 &= 0, \end{aligned}$$

where $\kappa_i \in \mathbb{R}$, $\tau_i > 0$. We set $\phi_i = (\log(\tau_i), \kappa_i)$ the vector of individual random parameters. We assume that

$$\log(\tau_i) \sim_{i.i.d.} \mathcal{N}(\log(\tau), \omega_\tau^2), \quad \kappa_i \sim_{i.i.d.} \mathcal{N}(\kappa, \omega_\kappa^2).$$

The parameter vector is $\theta = (\log \tau, \kappa, \omega_\tau, \omega_\kappa, \gamma, \sigma)$. This model can be easily discretized, resulting in a state-space model with random individual parameters:

$$\begin{aligned} y_{ij} &= X_{it_{ij}} + \varepsilon_{ij}, \;\; \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2), \\ X_{it_{ij}} &= X^i_{t_{ij-1}} e^{-\Delta_{ij}\tau_i} - \kappa_i \tau_i (1 - e^{-\Delta_{ij}\tau_i}) + \eta_{ij}, \\ \eta_{ij} &\sim \mathcal{N}\left(0, \gamma^2 \tau_i (1 - e^{-\Delta_{ij}\tau_i})\right). \end{aligned}$$

The vector $\mathbf{X}_i = (x_{it_{i1}}, \ldots, x_{it_{iJ}})$ conditional on ϕ_i is Gaussian with mean vector m_{iX} and covariance matrix G_{iX} equal to

$$
\begin{aligned}
m_{iX} &= \left(\tau_i\kappa_i\left(1-e^{-\frac{t_{i1}}{\tau_i}}\right), \ldots, \tau_i\kappa_i\left(1-e^{-\frac{t_{iJ}}{\tau_i}}\right)\right)', \qquad (11)\\
G_{iX} &= \left(\frac{\tau_i\gamma^2}{2}\left(1-e^{-\frac{2\min(t_{ij},t_{ik})}{\tau_i}}\right)e^{-\frac{|t_{ij}-t_{ik}|}{\tau_i}}\right)_{1\leq j,k,\leq J}, \qquad (12)
\end{aligned}
$$

where $'$ is the transposed vector. Remark that we do not consider X_{i0} in this calculus as the initial condition is assumed to be deterministic and equal to 0. The transition density $p(x_t, t-s, s|x_s, \phi)$ is Gaussian and equals to

$$
\begin{aligned}
\log p(x_{it_{ij}}, \Delta_{ij}, t_{ij-1}|x_{it_{ij-1}}, \phi_i; \theta) &= -\frac{1}{2}\log(\pi\gamma^2\tau_i(1-e^{-2\Delta_{ij}/\tau_i}))\\
&\quad -\frac{\left(x_{it_{ij}} - x_{it_{ij-1}}e^{-\Delta_{ij}\tau_i} - \kappa_i\tau_i(1-e^{-\Delta_{ij}\tau_i})\right)^2}{\gamma^2\tau_i(1-e^{-\Delta_{ij}\tau_i})}.
\end{aligned}
$$

It does not depend on time as the Ornstein-Uhlenbeck process is time-homogeneous. The conditional expectation and variance of y_{ij} are

$$
\begin{aligned}
\mathbb{E}(y_{ij}|X_0^i = 0, \phi_i) &= \tau_i\kappa_i(1-e^{-\frac{t_{ij}}{\tau_i}}),\\
\mathrm{Var}(y_{ij}|X_0^i = 0, \phi_i) &= \frac{\tau_i\gamma^2}{2}(1-e^{-\frac{t}{\tau_i}}) + \sigma^2.
\end{aligned}
$$

Even if this SDE is linear, the transition density $p(X_{it_{ij}}, \Delta_{ij}, t_{ij-1}|X_{t_{ij-1}}, \phi_i; \theta)$ is a nonlinear function of ϕ_i. Thus, the likelihood has no closed form. The exact estimator of θ is unavailable and we have to resort either to the SAEM-PMCMC or the SAEM-MCMC algorithms.

4.1.2 Simulation step of the SAEM algorithm

The S-step of the SAEM algorithm requires the simulation of the non-observed data under the distributions $p(\mathbf{X}_{i,1:J}, \phi_i|\mathbf{y}_{i,0:J}; \theta)$ for each subject. We detail this S-step for the SAEM-MCMC and SAEM-PMCMC algorithms.
For the S-step of the SAEM-MCMC algorithm, we have to simulate under the distributions $p(\mathbf{X}_{i,1:J}|, \mathbf{y}_{i,0:J}, \phi_i; \theta)$ and $p(\phi_i|, \mathbf{y}_{i,0:J}, \mathbf{X}_{i,1:J}; \theta)$ for each subject iteratively. As the Ornstein-Uhlenbeck is a linear Gaussian model and the error model is additive, the posterior distribution $p(\mathbf{X}_{i,1:J}|, \mathbf{y}_{i,0:J},\ \phi_i; \theta)$ is explicit,

Gaussian with mean vector $m_{i,X,post}$ and variance matrix $\Gamma_{i,X,post}$ equal to

$$\begin{aligned}\Gamma_{i,X,post} &= \left(\sigma^{-2} I_J + G_{iX}^{-1}\right)^{-1}, \\ m_{i,X,post} &= \Gamma_{i,X,post}\left(\sigma^{-2}\mathbf{y}_{i,1:J} + G_{iX}^{-1} m_{iX}\right).\end{aligned}$$

The conditional simulation of X is thus direct and exact. The posterior distribution of κ_i is explicit, Gaussian with mean $m_{i,\kappa,post}$ and variance $V_{i,\kappa,post}$ equal to

$$\begin{aligned}V_{i,\kappa,post} &= \left((H_i)' G_{iX}^{-1} H_i + \omega_\kappa^{-2}\right)^{-1}, \\ m_{i,\kappa,post} &= V_{i,\kappa,post}\left((H_i)' G_{iX}^{-1}\mathbf{X}_{i,1:J} + \kappa\,\omega_\kappa^{-2}\right),\end{aligned}$$

with $H_i = (\tau_i(1-e^{-t_{i1}/\tau_i}), \ldots, \tau_i(1-e^{-t_{iJ}/\tau_i}))'$. The posterior distribution of τ_i is not explicit. We use a Metropolis-Hastings algorithm with a random walk proposal to simulate with this distribution.

For the S-step of the SAEM-PMCMC algorithm, we have to choose two proposals $q(X_{t_{ij}} |\ X_{t_{ij-1}}, y_{ij}, \phi_i; \theta)$ and $q(\cdot|\phi_i)$ for each subject. The proposal $q(X_{t_{ij}} |\ X_{t_{ij-1}}, y_{ij}, \phi_i; \theta)$ of the SMC within the PMCMC can be the exact posterior density $p(X_{t_{ij}} | X_{t_{ij-1}},\ y_{ij}, \phi_i; \theta)$. Indeed, in the particular case of the Ornstein-Uhlenbeck, this density is explicit, Gaussian with conditional mean $m_{X_{ij},post}$ and variance $G_{X_{ij},post}$ equal to

$$\begin{aligned}G_{X_{ij},post} &= \left(\sigma^{-2} + \left(\frac{\tau_i\gamma^2}{2}(1-e^{-2\Delta_{ij}/\tau_i})\right)^{-1}\right)^{-1}, \\ m_{X_{ij},post} &= G_{X_{ij},post}\left(\sigma^{-2} y_{ij} + \left(\frac{\tau_i\gamma^2}{2}(1-e^{-2\Delta_{ij}/\tau_i})\right)^{-1}\left(e^{-\Delta_{ij}/\tau_i} X_{t_{ij-1}}\right.\right. \\ &\quad \left.\left. + \tau_i\kappa_i(1-e^{-\Delta_{ij}/\tau_i})\right)\right),\end{aligned}$$

As an alternative, we also consider the predictive distribution $p(X_{t_{ij}} | X_{t_{ij-1}}, \phi_i; \theta)$ as proposal. The proposal $q(\cdot|\phi)$ for the individual parameters ϕ within the PMCMC algorithm is a classical random walk on each component of the vector ϕ.

4.1.3 Maximization step of the SAEM algorithm

The SAEM algorithm is based on the computation of the sufficient statistics for the maximization step. The statistics for the parameters $\mu = (\log\tau, \kappa)$, $\Omega = diag(\omega_\tau^2, \omega_\kappa^2)$ and σ are the three classic ones for mixed models (see *e.g.*

Samson *et al.*, 2007):

$$S_1(\mathbf{y}, \mathbf{X}, \mathbf{\Phi}) = \sum_{i=1}^{n} \phi_i, \quad S_2(\mathbf{y}, \mathbf{X}, \mathbf{\Phi}) = \sum_{i=1}^{n} \phi_i \phi_i', \tag{13}$$

$$S_3(\mathbf{y}, \mathbf{X}, \mathbf{\Phi}) = \sum_{i=1}^{n} (y_i - X_i)'(y_i - X_i). \tag{14}$$

Let s_{1m}, s_{2m}, s_{3m} denote the corresponding stochastic approximated conditional expectations at iteration m of the SAEM algorithm. The M step for these three parameters reduces to

$$\widehat{\mu}^{(m)} = \frac{1}{n} s_{1m}, \tag{15}$$

$$\widehat{\Omega}^{(m)} = \frac{1}{n} s_{2m} - \frac{1}{n^2} s_{1m} s_{1m}', \tag{16}$$

$$\widehat{\sigma}^{(m)} = \sqrt{\frac{1}{n} s_{3m}}. \tag{17}$$

The sufficient statistic corresponding to the parameter γ depends on the SDE. For the Ornstein-Uhlenbeck, we have

$$S_4(\mathbf{y}, \mathbf{X}, \mathbf{\Phi}) = \sum_{i=1}^{n} \sum_{j=1}^{J} \left(X^i_{it_{ij}} - X^i_{it_{ij-1}} e^{-\Delta_{ij}\tau_i} - \kappa_i \tau_i (1 - e^{-\Delta_{ij}\tau_i}) \right)^2. \tag{18}$$

The M step is thus

$$\widehat{\gamma}^{(m)} = \sqrt{\frac{1}{nJ} s_{4m}}. \tag{19}$$

4.1.4 Simulation design and results

Two different designs are used for the simulations with equally spaced observation times: $n = 20$, $J = 40$, $\Delta = 0.5$ and $n = 40$, $J = 20$, $\Delta = 1$, $t_0 = 0$. Two sets of parameter values are used. The first set is $\log(\tau) = 0.6$, $\kappa = 1$, $\omega_\tau = 0.1$, $\omega_\kappa = 0.1$, $\gamma = 0.05$, $\sigma = 0.05$. The second set uses greater variances and is $\log(\tau) = \log(10)$, $\kappa = 1$, $\omega_\tau = 0.1$, $\omega_\kappa = 0.1$, $\gamma = 1$, $\sigma = 1$. For each design and each set of parameter values, one hundred datasets are simulated.

The SAEM-PMCMC and SAEM-MCMC algorithms have to be initialized. For the first set of parameters, we set $\widehat{\log(\tau)}_0 = 1.1$, $\widehat{\kappa}_0 = 1.5$, $\widehat{\omega_\tau}_0 = 0.5$, $\widehat{\omega_\kappa}_0 = 0.5$, $\widehat{\gamma}_0 = 0.25$ and $\widehat{\sigma}_0 = 0.25$, i.e. the initial standard deviations are 5 times greater than the true standard deviations. For the second set of parameters, we set $\widehat{\log(\tau)}_0 = \log(10) + 0.5$, $\widehat{\kappa}_0 = 1.5$, $\widehat{\omega_\tau}_0 = 0.5$, $\widehat{\omega_\kappa}_0 = 0.5$, $\widehat{\gamma}_0 = 5$ and$\widehat{\sigma}_0 = 5$.

Table 1 presents the bias and RMSE (%) of the SAEM-PMCMC and SAEM-MCMC algorithms obtained for the two designs and two sets of parameters. The results are almost identical for both algorithms and very satisfactory (bias less than 3% and RMSE less than 10%). The SAEM-PMCMC algorithm has slightly greater bias than SAEM-MCMC for the standard deviations of the random effects (ω_τ and ω_κ). On the contrary, the bias for γ and σ are always lower with SAEM-PMCMC than with SAEM-MCMC. The design (n, J) of the study affects very little the estimation quality when γ and σ are small (first set of parameters). On the contrary, when γ and σ are larger (second set of parameters), the RMSE are greater with fewer measures per subject ($n = 40, J = 20$) than with more ($n = 20, J = 40$). Figure 1 shows the density of the estimators obtained by both algorithms with $n = 40, J = 20$ and the second set of parameters. Densities are estimated non-parametrically with a Gaussian kernel. This illutrates again the small bias of SAEM-PMCMC for the parameters γ and σ.

The choice of the number of particles K in the SMC algorithm may seem crucial but it is not. Table 2 presents the bias and RMSE obtained on the 100 datasets simulated with the first set of parameters, $n = 40, J = 20$ and different implementations of the SAEM-PMCMC algorithm. We successively implement the PMCMC algorithm with $K = 25$, $K = 50$ and $K = 100$ particles using the exact posterior distribution $q(X_{t_{ij}}|X_{t_{ij-1}}, y_{ij}, \phi_i; \theta) = p(X_{t_{ij}}|X_{t_{ij-1}}, y_{ij}, \phi_i; \theta)$. The results are almost identical. On the contrary, the choice of the proposal $q(X_j|y_j, X_{j-1}, \phi; \gamma, \sigma)$ in the SMC algorithm may affect the result. Table 2 presents results obtained with the predictive distribution $p(X_j|X_{j-1}, \phi; \gamma, \sigma)$ as proposal q. The algorithm fails to converge due to the degeneracy of the particles. We present the results obtained by increasing significantly the number of particles ($K = 1000$). Even in this case, we observe the degeneracy of the algorithm on 90% of the simulated datasets. Morevoer, the bias and RMSE are significantly greater than with the posterior conditional distribution $p(X_j|y_j, X_{j-1}, \phi; \gamma, \sigma)$.

[Table 1 about here.]

[Table 2 about here.]

[Figure 1 about here.]

4.2 Example 2: stochastic volatility Gompertz model

4.2.1 Stochastic volatility Gompertz mixed model

The Gompertz model is a well-known growth model (see *e.g.* Jaffrézic and Foulley, 2006). Recently, Donnet *et al.* (2010) proposed a stochastic version of this model to take into account random fluctuations in growth process. The stochastic volatility Gompertz mixed model is the following one:

$$\begin{aligned} y_{ij} &= X_{it_{ij}}(1+\varepsilon_{ij}),\ \varepsilon_{ij} \sim \mathcal{N}(0,\sigma^2), \\ dX_{it} &= B_i C_i e^{-C_i t} X_{it} dt + \gamma X_{it} dB_{it}, \\ X_{i0} &= A_i e^{-B_i}, \end{aligned} \tag{20}$$

where $A_i > 0, B_i > 0, C_i > 0$. We set $\phi_i = (\log A_i, \log B_i, \log C_i)$ the vector of individual random parameters. The SDE (20) is a time-inhomogeneous SDE. The initial conditional is a function of the random individual parameters ϕ_i. We assume that

$$\begin{aligned} \log A_i &\sim_{i.i.d.} \mathcal{N}(\log A, \omega_A^2), \quad \log B_i \sim_{i.i.d.} \mathcal{N}(\log B, \omega_B), \\ \log C_i &\sim_{i.i.d.} \mathcal{N}(\log C, \omega_C^2), \end{aligned}$$

and the parameter of interest is $\theta = (\log A, \log B, \log C, \omega_A, \omega_B, \omega_C, \gamma, \sigma)$. The transition density of the process (X_t) is not explicit. However, by the Itô formula, the conditional expectation and variance of $\log(X_{it})|\phi_i$, for $t \geq 0$, are

$$\begin{aligned} \mathbb{E}(\log X_{it}|\phi_i) &= \log A_i - B_i e^{-C_i t} - \frac{1}{2}\gamma^2 t, \\ \mathrm{Var}(\log X_{it}|\phi_i) &= \gamma^2 t. \end{aligned}$$

The transition density of $(\log(X_t))$ is explicit, Gaussian and equals to

$$\log p(\log X_{it_{ij}}, \Delta_{ij}, t_{ij-1} | \log X_{it_{ij-1}}, \phi_i; \theta) = -\tfrac{1}{2}\log(2\pi\gamma^2\Delta_{ij}) \\ -\tfrac{1}{2}\frac{\left(\log X_{it_{ij}} - \log X_{it_{ij-1}} + B_i e^{-Ct_{ij-1}}(e^{-C_i\Delta_{ij}}-1) - \frac{1}{2}\gamma^2\Delta_{ij}\right)^2}{\gamma^2\Delta_{ij}}.$$

As the transition density $p(X_{it_{ij}}, \Delta_{ij}, t_{ij-1}|X_{it_{ij-1}}, \phi_i; \theta)$ is a nonlinear function of ϕ_i, the likelihood has no closed form and the exact estimator of θ is unavailable. We resort to the SAEM algorithm.

4.2.2 Simulation step of the SAEM algorithm

The S-step of the SAEM algorithm requires the simulation of the non-observed data under the distribution $p(\mathbf{X}_{i,0:J}, \phi_i|\mathbf{y}_{i,0:J};\theta)$ for each subject. We detail this step for the SAEM-MCMC and SAEM-PMCMC algorithms.

For the S step of the SAEM-MCMC algorithm, we implement a standard Metropolis-Hastings-within-Gibbs algorithm. More precisely, we use a random walk proposal to simulate the trajectories $\mathbf{X}_{i,0:J}$ conditionally to the observations $\mathbf{y}_{i,0:J}$, updating the components of $\mathbf{X}_{i,0:J}$ time by time. A random walk proposal is also used to simulate the individual parameters ϕ_i from the conditional distribution.

For the S step of the SAEM-PMCMC algorithm, the proposal $q(X_{it_{ij}}|X_{it_{ij-1}}, y_{ij}, \phi_i;\theta)$ has to be chosen. Contrary to Example 1, the exact posterior distribution $p(X_{it_{ij}}|X_{it_{ij-1}}, y_{ij}, \phi_i;\theta)$ is not explicit. We propose to approximate this distribution by a Gaussian distribution with mean and variance deduced from the true ones. More precisely, we consider the following proposal

$$q(\log X_{t_{ij}}|\log X_{it_{ij-1}}, y_{ij}, \phi_i;\theta) = \mathcal{N}(m_{X_{ij},post}, \Gamma_{X_{ij},post}),$$

on $\log X_{t_{ij}}$ with

$$\begin{aligned}
\Gamma_{X_{ij},post} &= \left(\sigma^{-2} + (\gamma^2\Delta_{ij})^{-1}\right)^{-1}, \\
m_{X_{ij},post} &= \Gamma_{X_{ij},post}\,\mu_{X_{ij},post}, \\
\mu_{X_{ij},post} &= \left(\frac{\log y_{ij}}{\sigma^2} + \frac{1}{\gamma^2\Delta_{ij}}\left(\log X_{it_{ij-1}} - B_i e^{-C_i t_{ij-1}}(e^{-C_i\Delta_{ij}} - 1) - \frac{\gamma^2\Delta_{ij}}{2}\right)\right).
\end{aligned}$$

and then take the exponential to obtain a candidate for $X_{t_{ij}}$. The proposal $q(\cdot|\phi)$ for the individual parameters ϕ within the PMCMC algorithms is a classical random walk on each component of the vector ϕ.

4.2.3 Maximization step of the SAEM algorithm

Within the SAEM algorithm, for parameters $\mu = (\log A, \log B, \log C)$, $\Omega = diag(\omega_A^2, \omega_B^2, \omega_C^2)$ and σ, the sufficient statistics and the M step are the same as in Example 1 (see equations (13)-(17)). In this particular model, the parameter γ appears both in the expectation and the variance of $\log(X_t)|\phi$ and thus its estimator is not the same as in Example 1. The sufficient statistic corresponding

to γ is

$$S_4(\mathbf{y}, \log \mathbf{X}, \Phi) = \sum_{i=1}^{n} \sum_{i=1}^{J} \Delta_{ij} \left(\log X_{it_{ij}} - \log X_{it_{ij-1}} + B_i e^{-C_i t_{ij-1}} (e^{-C_i \Delta_{ij}} - 1) \right)^2 .$$

Let s_{4m} denote the stochastic approximation of this sufficient statistic at iteration m of the SAEM algorithm. For the sake of simplicity, we assume that the step size Δ_{ij} is a constant Δ. The estimator $\widehat{\gamma}_m$ at iteration m is deduced by maximizing the complete likelihood:

$$\widehat{\gamma}_m = \sqrt{\frac{2}{\Delta} \left(-1 + \sqrt{1 + \frac{s_{4m}}{nJ}} \right)}$$

When Δ_{ij} is not a constant, the estimator is more complex but also explicit.

4.2.4 Simulation results

We propose to simulate two experimental designs with the same observation times for all individuals: $n = 20, J = 40, \Delta = 0.01$ and $n = 40, J = 20, \Delta = 0.02$ ($t_0 = 0$), respectively. As in Donnet *et al.* (2010), we use the following population parameters:

$$\begin{array}{ccc} \log A = \log(3000), & \log B = \log(5), & \log C = \log(14), \\ \omega_A = 0.1, & \omega_A = 0.1, & \omega_C = 0.1 \end{array}$$

We propose two levels of variability, i.e. γ and σ are taken respectively equal to $\left(0.4, 1/\sqrt{20}\right)$ and $\left(0.8, 1/\sqrt{15}\right)$ and referred to as first and second set of parameters thereafter. For each design and each set of parameters, 100 datasets are simulated and θ is estimated with the two algorithms (SAEM-PMCMC and SAEM-MCMC). The SAEM algorithm is initialized with $\widehat{\log A}_0 = 8.21, \widehat{\log B}_0 = 1.81, \widehat{\log C}_0 = 2.84, \widehat{\omega_A}_0 = 0.5, \widehat{\omega_B}_0 = 0.5, \widehat{\omega_C}_0 = 0.5, \widehat{\gamma}_0 = 1.2, \widehat{\gamma}_0 = 0.66$. Bias and RMSE (%) are presented in Table 3.

[Table 3 about here.]

Both algorithms give satisfactory results overall (bias and RSME $\leq 20\%$ except for γ). Nevertheless, we clearly see that the SAEM-PMCMC algorithm performs better than the SAEM-MCMC algorithm. Indeed, for the population means $(\log A, \log B, \log C)$, the population standard deviations $(\omega_A, \omega_B, \omega_C)$ and σ, the bias and RMSE are smaller for SAEM-PMCMC even if SAEM-MCMC results

are already small. The difference is obvious for γ, which is estimated with a large bias by SAEM-MCMC (up to a bias of 77%) while this is not the case for SAEM-PMCMC (bias $< 7\%$). This phenomenon is accentuated when the variabilities γ and σ increase (second set of parameters). The better performances of SAEM-PMCMC versus SAEM-MCMC can also be observed in Figure 5 where we plot the kernel estimated densities of the parameters $\widehat{\theta}$ among the 100 simualted datasets for the second set of parameters with $n = 20$ and $J = 40$.

[Figure 2 about here.]

The improved quality of estimation has a computational cost since it takes less than 8 minutes for the SAEM-MCMC and nearly 23 minutes for the SAEM-PMCMC algorithm to supply results. However, estimating precisely the population parameters is a key issue in population studies (such a genetic specification) and a bias can have real consequences on the conclusions.

5 Discussion

The stochastic differential mixed-effects models are quite widespread in the applied statistics field. However, in the absence of efficient and computationally reasonable estimation methods, a simplifying assumption is often made: either the observation noise or the volatility term are standardly neglected.
In this paper we present an EM algorithm combined with a Particular Monte Carlo Markov Chain method to estimate parameters in stochastic differential mixed-effects models including observation noise. We prove the convergence of the algorithm towards the maximum likelihood estimator. The convergence of the SAEM-MCMC algorithm is different and has been studied by Kuhn and Lavielle (2005). In the SAEM-MCMC, the simulation step is performed with an MCMC algorithm, implying a dependence between the simulated variables contrary to the SAEM algorithm (assumption SAEM3). Consequently, Kuhn and Lavielle (2005) prove the convergence of the SAEM-MCMC under the (strong) assumption of geometric ergodicity of the Markov Chain generated along the SAEM iterations. This cannot be applied to prove the convergence of the SAEM-PMCMC algorithm as Andrieu *et al.* (2010) prove that this assumption is not checked by the PMCMC algorithm we focus on. For the SAEM-PMCMC algorithm, we use the convergence results of more general stochastic approximation schemes, the SAEM algorithm being a Robbins-Monro

type procedure. Our proof is based on results proposed by Delyon *et al.* (1999) when the stochastic approximation scheme includes a remainder term, namely the error induced by the PMCMC algorithm in the SAEM-PMCMC algorithm. The PMCMC convergence results proposed by Andrieu *et al.* (2010) yield the convergence of the estimation algorithm.

The suggested method supplies accurate parameter estimation in a really moderate computational time on practical examples. On various simulated datasets, we illustrate the superiority of this method over the SAEM algorithm combined with a standard MCMC algorithm. This efficiency is due to the fact that the PMCMC algorithm takes advantage of the Markovian properties of the non-observed process.

A major advantage of this methodology is its automatic implementation. Indeed the generation of the non-observed process does not involve any tuning parameters, such as the size of the random move in the random walk Metropolis-Hastings algorithm. For example, the number of particles is not crucial, as illustrated in Example 1. In this paper we present a PMCMC algorithm where all the particles are re-sampled at each iteration. Many other resampling distributions have been proposed in the literature, trying to achieve an optimal procedure. West (1993) developed an effective method of adaptive importance sampling to address this issue. The procedure developed by Pitt and Shephard (2001) is similar in spirit and has real computational advantages. A stratified resampling is proposed in Kitawaga (1996). Each resampling distribution implies a specific update of the weights.

Several major assumptions are made in this paper. Assumption (M0) requires the existence of an explicit transition density for the non-observed random process. If this assumption is not fulfilled then a Euler approximation can be considered. In this context, the Markovian structure is still guaranted and the PMCMC strategy should prove its efficiency too. The relaxation of this assumption opens great perspectives but is beyond the scope of this paper. Assumption (M1) requires the exponential form of the complete likelihood which essentially contrains the structure of the noise observation and of the random effects. This hypothesis could be relaxed using another stochastic version of the EM algorithm such as that of Celeux and Diebolt (1992). Assumption (SAEM-PMCMC2) is a strong hypothesis but it is only technical. This assumption could be relaxed using a principle of random boundaries presented in Allassonnière *et al.* (2009). Additional assumptions are rather classical and not restrictive.

SDEs have been recently developed for medical applications in glucose dynamics

(Picchini *et al.*, 2008), in neuron potential dynamics (Höpfner, 2007), in pharmacokinetics (Ditlevsen *et al.*, 2005; Donnet and Samson, 2008), or in growth curve data (Donnet *et al.*, 2010). These applications of SDEs show that stochastic versions of physiological/biological models improve data fitting and stabilize parameter estimations. However, most of these applications are restricted to the analysis of single-individual data or ignore the measurement noise. The extension of this work to these real data analysis should be of great interest.

References

Allassonnière, S., Kuhn, E. and Trouvé, A. (2009). Construction of bayesian deformable models via stochastic approximation algorithm: A convergence study. *to appear in Bernoulli.* .

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

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

Brandière, O. and Duflo, M. (1996). Les algorithmes stochastiques contournent-ils les pièges? *Ann. Inst. H. Poincaré Probab. Statist.* **32**, 395–427.

Celeux, G. and Diebolt, J. (1992). A stochastic approximation type EM algorithm for the mixture problem. *Stochastics Stochastics Rep.* **41**, 119–134.

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

Davidian, M. and Giltinan, D. (1995). *Nonlinear models to repeated measurement data*. Chapman and Hall.

Delyon, B., Lavielle, M. and Moulines, E. (1999). Convergence of a stochastic approximation version of the EM algorithm. *Ann. Statist.* **27**, 94–128.

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. (2005). Stochastic vs. deterministic uptake of dodecanedioic acid by isolated rat livers. *Bull. Math. Biol.* **67**, 547–561.

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.

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

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

Doucet, A., de Freitas, N. and Gordon, N. (2001). An introduction to sequential Monte Carlo methods. In *Sequential Monte Carlo methods in practice*, Stat. Eng. Inf. Sci., pp. 3–14. Springer, New York.

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.

Genon-Catalot, V. and Jacod, J. (1993). On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Ann. Inst. H. Poincaré Probab. Statist.* **29**, 119–151.

Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (1996). *Markov chain Monte Carlo in practice*. Interdisciplinary Statistics. Chapman & Hall, London.

Höpfner, R. (2007). On a set of data for the membrane potential in a neuron. *Math. Biosci.* **207**, 275–301.

Jaffrézic, F. and Foulley, J. (2006). Modelling variances with random effects in non linear mixed models with an example in growth curve analysis. *XXIII International Biometric Conference, Montreal, 16-21 juillet* .

Kantas, N., Doucet, A., Singh, S. and Maciejowski, J. (2009). An overview of sequential Monte Carlo methods for parameter estimation in general state-space models. *Proceedings IFAC System Identification (SySid) Meeting* .

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

Kitawaga, G. (1996). Monte carlo filter and smoother for non-gaussian nonlinear state space models. *J. Comp. Graph. Statist.* pp. 1–25.

Kloeden, P. E. and Platen, E. (1992). *Numerical solution of stochastic differential equations*, vol. 23 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin.

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

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

Oravecz, Z., Tuerlinckx, F. and Vandekerckhove, J. (2009). A hierarchical Ornstein-Uhlenbeck model for continuous repeated measurement data. *Psychometrika* pp. 395–418.

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.

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

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

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

Pitt, M. and Shephard, N. (2001). Auxiliary variable based particle filters. In *Sequential Monte Carlo methods in practice*, Stat. Eng. Inf. Sci., pp. 273–293. Springer, New York.

Pitt, M. K. and Shephard, N. (1999). Filtering via simulation: auxiliary particle filters. *J. Amer. Statist. Assoc.* **94**, 590–599. ISSN 0162-1459.

Samson, A., Lavielle, M. and Mentré, F. (2007). The SAEM algorithm for group comparison tests in longitudinal data analysis based on non-linear mixed-effects model. *Stat. Med.* **26**, 4860–4875.

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.

West, M. (1993). Approximating posterior distributions by mixtures. *J. Roy. Statist. Soc. Ser. B* **55**, 409–422. ISSN 0035-9246.

Wolfinger, R. (1993). Laplace's approximation for nonlinear mixed models. *Biometrika* **80**, 791–795. ISSN 0006-3444.

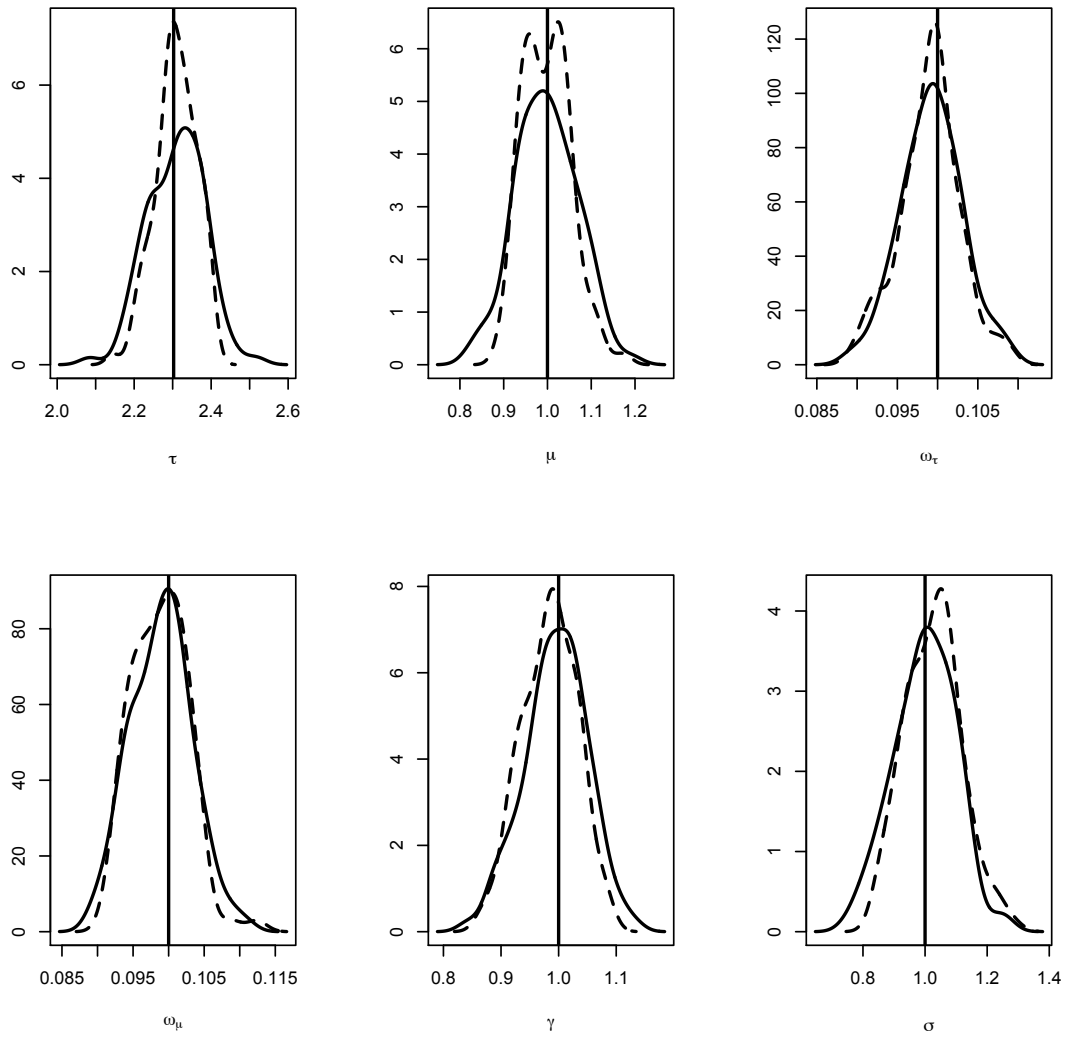

Figure 1: Ornstein-Uhlenbeck mixed model, $n = 40$, $J = 20$ and second set of parameters. Densities of the estimators $\hat{\theta}$ obtained with the SAEM-PMCMC algorithm (plain line) and the SAEM-MCMC algorithm (dashed line) on 100 simulated datasets, for each component of $(\log \tau, \kappa, \omega_\tau, \omega_\kappa, \gamma, \sigma)$. The true parameter value is in vertical line.

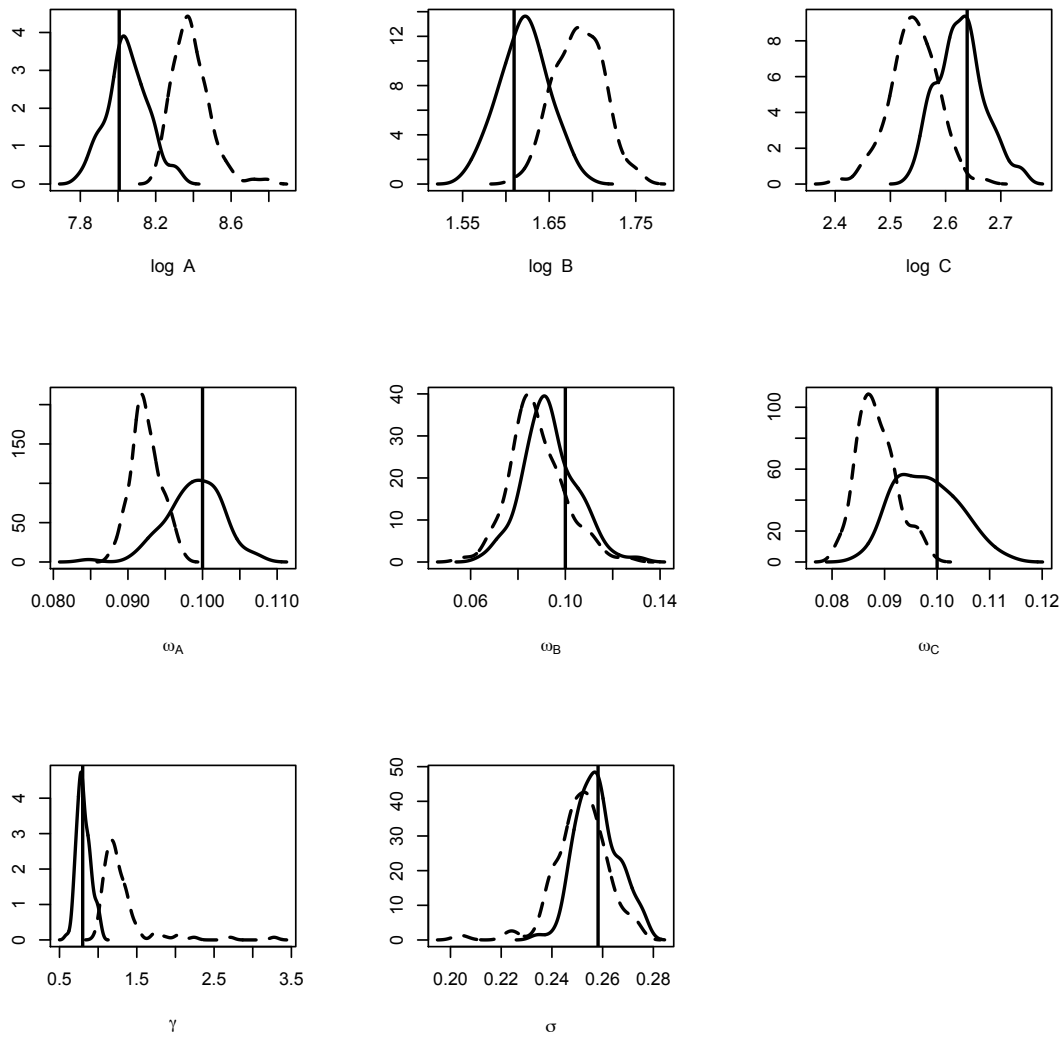

Figure 2: Stochastic volatility Gompertz mixed model, $n = 20$, $J = 40$ and second set of parameters. Densities of the estimators $\widehat{\theta}$ obtained with the SAEM-PMCMC algorithm (plain line) and the SAEM-MCMC algorithm (dashed line) on 100 simulated datasets, for each component of $(\log A, \log B, \log C, \omega_A, \omega_B, \omega_C, \gamma, \sigma)$. The true parameter value is in vertical line.

Parameters		$\log(\tau)$	κ	ω_τ	ω_κ	γ	σ
True value		0.62	1.00	0.10	0.10	0.05	0.05
$n = 20$, $J = 40$							
SAEM-PMCMC	Bias	0.38	0.05	-2.55	-3.29	-0.84	0.07
	RMSE	0.48	0.30	1.22	1.34	0.93	0.44
SAEM-MCMC	Bias	-0.02	0.29	-0.60	-2.38	3.34	-1.43
	RMSE	0.49	0.30	1.23	1.29	0.96	0.45
$n = 40$, $J = 20$							
SAEM-PMCMC	Bias	0.11	0.01	-2.75	-1.59	-0.73	0.50
	RMSE	0.33	0.20	0.82	0.99	0.94	0.52
SAEM-MCMC	Bias	-0.24	0.22	-1.46	-1.12	2.96	-1.37
	RMSE	0.34	0.21	0.81	1.02	0.96	0.55
True value		2.30	1.00	0.10	0.10	1.00	1.00
$n = 20$, $J = 40$							
SAEM-PMCMC	Bias	0.55	-0.63	-1.49	-1.92	0.15	-1.33
	RMSE	0.42	0.97	0.51	0.55	0.64	0.81
SAEM-MCMC	Bias	-0.41	1.24	-0.62	-1.25	0.44	-1.61
	RMSE	0.35	0.84	0.52	0.54	0.56	0.72
$n = 40$, $J = 20$							
SAEM-PMCMC	Bias	0.28	-0.18	-0.73	-1.04	-0.47	-0.03
	RMSE	3.19	7.03	3.91	4.38	5.45	9.81
SAEM-MCMC	Bias	0.26	-0.19	-0.99	-1.17	-1.59	2.64
	RMSE	2.32	5.39	3.89	4.05	4.96	9.42

Table 1: Ornstein-Uhlenbeck mixed model: biais and RMSE (%) for the estimation of θ obtained by the SAEM-PMCMC and SAEM-MCMC algorithms on 100 simulated datasets with two designs ($n = 20$, $J = 40$ and $n = 40$, $J = 20$) and two sets of parameters. PMCMC is implemented with $K = 50$ particles and the exact posterior proposal $q(X_{t_{ij}}|X_{t_{ij-1}}, y_{ij}, \phi_i; \theta) = p(X_{t_{ij}}|X_{t_{ij-1}}, y_{ij}, \phi_i; \theta)$.

Parameters	$\log(\tau)$	κ	ω_τ	ω_κ	γ	σ
True value	0.62	1.00	0.10	0.10	0.05	0.05
SAEM-PMCMC with posterior proposal, $K = 25$						
Bias	0.26	0.11	0.32	-4.84	0.64	-0.41
RMSE	0.49	0.30	1.28	1.30	1.04	0.45
SAEM-PMCMC with posterior proposal, $K = 50$						
Bias	0.38	0.05	-2.55	-3.29	-0.84	0.07
RMSE	0.48	0.30	1.22	1.34	0.93	0.44
SAEM-PMCMC with posterior proposal, $K = 100$						
Bias	0.29	0.10	-0.37	-4.35	-0.09	-0.30
RMSE	0.48	0.30	1.25	1.32	1.01	0.46
SAEM-PMCMC with prior proposal, $K = 1000$						
Bias	-0.63	0.66	-2.78	-4.73	15.42	-5.44
RMSE	0.53	0.34	1.43	1.77	2.17	0.83

Table 2: Ornstein-Uhlenbeck mixed model: biais and RMSE (%) for the estimation of θ obtained by the SAEM-PMCMC algorithm on 100 simulated datasets with $n = 20$, $J = 40$. The PMCMC algorithm is successively implemented with posterior proposal $q(X_{t_{ij}}|X_{t_{ij-1}}, y_{ij}, \phi_i; \theta) = p(X_{t_{ij}}|X_{t_{ij-1}}, y_{ij}, \phi_i; \theta)$ and $K = 25$, $K = 50$ and $K = 100$ particles, and prior proposal $q(X_{t_{ij}}|X_{t_{ij-1}}, y_{ij}, \phi_i; \theta) = p(X_{t_{ij}}|X_{t_{ij-1}}, \phi_i; \theta)$ and $K = 1000$ particles.

Parameters		$\log A$	$\log B$	$\log C$	ω_A	ω_B	ω_C	γ	σ
True value		8.01	1.61	2.64	0.10	0.10	0.10	0.40	0.22
$n = 20$, $J = 40$									
SAEM-PMCMC	Bias	0.01	-0.07	0.03	-0.69	-2.31	-1.79	2.77	-0.04
	RMSE	0.61	1.16	0.91	2.90	9.71	5.98	11.15	3.15
SAEM-MCMC	Bias	2.04	1.713	-1.99	-8.00	-5.00	-7.55	61.93	-2.56
	RMSE	2.16	2.07	2.19	8.15	10.87	9.14	63.78	4.30
$n = 40$, $J = 20$									
SAEM-PMCMC	Bias	0.05	-0.30	0.12	-1.92	-2.38	-1.41	6.79	-0.81
	RMSE	0.797	1.64	1.36	4.53	14.88	9.10	18.54	3.31
SAEM-MCMC	Bias	2.37	1.90	-1.98	-6.76	-4.58	-8.17	77.66	-2.00
	RMSE	2.52	2.43	2.33	7.20	14.80	10.73	79.74	4.01
True value		8.01	1.61	2.64	0.10	0.10	0.10	0.80	0.26
$n = 20$, $J = 40$									
SAEM-PMCMC	Bias	0.56	0.72	-0.44	-1.11	-6.56	-1.97	1.67	-0.15
	RMSE	1.47	1.90	1.67	3.90	13.10	6.33	11.20	3.20
SAEM-MCMC	Bias	4.76	4.64	-3.67	-7.58	-11.92	-11.49	62.90	-2.60
	RMSE	4.92	4.95	4.01	7.83	16.55	12.08	75.11	4.84
$n = 40$, $J = 20$									
SAEM-PMCMC	Bias	0.36	0.38	-0.47	-0.70	-2.16	-1.31	0.37	0.82
	RMSE	1.07	1.41	1.43	2.21	8.84	4.64	9.55	3.57
SAEM-MCMC	Biais	4.37	4.11	-3.60	-8.15	-8.05	-10.05	49.70	-4.33
	RMSE	4.46	4.31	3.81	8.26	11.38	10.55	52.36	5.64

Table 3: Stochastic volatility Gompertz mixed model: bias and RMSE (%) of the estimation of θ obtained by the SAEM-PMCMC and SAEM-MCMC algorithms, on the 100 datasets with two designs ($n = 20$, $J = 40$ and $n = 40$, $J = 20$) and two sets of parameters.