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

Sophie Donnet

Laboratoire Ceremade, Université Paris Dauphine, France

Adeline Samson

UMR CNRS 8145, Laboratoire MAP5, Université Paris Descartes, France

Abstract. Biological processes measured repeatedly among a series of individuals are standardly analyzed by mixed models. These biological processes can be adequately modeled by parametric Stochastic Differential Equations (SDEs). We focus on the parametric maximum likelihood estimation of this mixed-effects model defined by SDE. As the likelihood is not explicit, we propose a stochastic version of the Expectation-Maximization algorithm combined with the Particle Markov Chain Monte Carlo method. When the transition density of the SDE is explicit, we prove the convergence of the SAEM-PMCMC algorithm towards the maximum likelihood estimator. Two simulated examples are considered: an Ornstein-Uhlenbeck process with two random parameters and a time-inhomogeneous SDE (Gompertz SDE) with a stochastic volatility error model and three random parameters. When the transition density is unknown, we prove the convergence of a different version of the algorithm based on the Euler approximation of the SDE towards the maximum likelihood estimator.

Keywords. Mixed models, Stochastic Differential Equations, SAEM algorithm, Particle Filter, PMCMC, Stochastic volatility, Time-inhomogeneous SDE

1. Introduction

Biological processes are usually measured repeatedly among a collection of individuals or experimental animals. The parametric statistical approach commonly used to discriminate between the inter-subjects variability (the variance of the individual regression parameters) and the residual variability (measurement noise) is the mixed-effects model methodology (Pinheiro and Bates, 2000). The choice of the regression function depends on the context. Dynamical processes are frequently described by deterministic models defined as ordinary differential equations. However, these functions may not capture the exact process, since 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. The perturbation of deterministic models by a random component leads to the class of stochastic differential equations (SDEs). In this paper we develop a parametric estimation method for mixed model defined by an SDE process, called SDE mixed model. Note that SDE mixed models can be viewed as an extension of state space models with random parameters.

Except in the linear case, the likelihood of standard mixed models is intractable in a closed form, making the estimation a hard task. Davidian and Giltinan (1995), Wolfinger (1993),

Pinheiro and Bates (2000) and Kuhn and Lavielle (2005) develop various methods for determinitic regression function.

When the regression term is the realization of a diffusion process, the estimation is made more complicated by the difficulties deriving from the SDEs. Parametric estimation of SDE with random parameters (without measurement noise) has been studied by Ditlevsen and De Gaetano (2005), Picchini et al. (2010) or Oravecz et al. (2009) in a Bayesian context. SDE mixed model estimation, including measurement noise modeling, is more complex and has received little attention. Overgaard et al. (2005) and Tornøe et al. (2005) build estimators based on an extended Kalman filter but without proving the convergence of the algorithm. Donnet and Samson (2008) propose to use a stochastic version of the Expectation-Maximisation algorithm. However, the method involves Markov Chain Monte Carlo (MCMC) samplers which have proved their slow mixing properties in the context of state space models with random parameters.

For state space models with fixed known parameters, the Sequential Monte Carlo (SMC) algorithms have demonstrated their efficiency. However, these techniques are difficult to extent to the case of unknown parameters (Casarin and Marin, 2009). Recently, Andrieu et al. (2010) developed a powerful algorithm combining MCMC and SMC, namely the Particle Markov Chain Monte Carlo (PMCMC), taking the advantage of the strength of its two components. Besides, its convergence toward the exact distribution of interest is established.

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. When the transition density of the SDE is explicit, we prove the convergence of the SAEM-PMCMC algorithm towards the maximum likelihood estimator. Two simulated examples are considered, an Ornstein-Uhlenbeck mixed SDE model observed with additive noise and a stochastic volatility mixed model with a time-inhomogeneous SDE, showing the increase in accuracy for the SAEM-PMCMC estimates compared with the SAEM-MCMC algorithm (Donnet and Samson, 2008). When the transition density is unknown, we prove the convergence of a different version of the algorithm towards the maximum likelihood estimator.

The paper is organized as follows. In Section 2, we present the SDE mixed model. In Section 3, an estimation method is proposed when the transition density of the SDE is explicit. Section 4 shows numerical results based on simulated data in this case. In Section 5, we extend the approach to general SDEs and prove the convergence of the proposed algorithm. In Section 6 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 $y_{i,0:J_i} = (y_{i0}, \ldots, y_{iJ_i})$ 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:

$$y_{ij} = g(X_{ij}, \varepsilon_{ij}), \quad \varepsilon_{ij} \sim_{i.i.d.} \mathcal{N}(0, \sigma^2), \tag{1}$$

$$dX_{it} = a(X_{it}, t, \phi_i)dt + b(X_{it}, t, \phi_i, \gamma)dB_{it}, \quad X_{it_{i0}}|\phi_i \sim \pi_0(\cdot|\phi_i) \tag{2}$$

$$\phi_i \sim_{i.i.d.} p(\phi; \beta). \tag{3}$$

In equation (1), $(\varepsilon_{ij})_{i,j}$ are i.i.d Gaussian random variables of variance σ^2 representing the measurement errors and g is the known error model function. The function $g(x, \varepsilon) = x + \varepsilon$ leads to an additive error model. Multiplicative 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, $(X_{ij})_{0 \le j \le J_i}$ are the values at discrete times $(t_{ij})_{0 \le j \le J_i}$ of the continuous process $(X_{it})_{t \ge t_{i0}} = (X_{it}(\phi_i))_{t \ge t_{i0}}$ defined by equation (2). 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 $(B_{it})_{t \ge t_{i0}}$. 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 (with linear growth) to ensure a unique strong 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 $1 \le i, k \le n$. The initial condition X_{i0} of process $(X_{it})_{t \ge t_{i0}}$, conditionally to the individual parameter ϕ_i is random with distribution $\pi_0(\cdot|\phi_i)$. In the following, $X_{i,0:J_i} = (X_{i0}, \ldots, X_{iJ_i})$ denotes the vector of the i-th process realization at observation times (t_{ij}) for $i = 1, \ldots, n$ and $\mathbf{X} = (X_1, \ldots, X_n)$ denotes the whole vector of processes at observation times, for all individuals. The individual parameters ϕ_i are assumed to be distributed with a density $p(\phi; \beta)$ depending on parameter β. We denote $\mathbf{\Phi} = (\phi_1, \ldots, \phi_n)$ the vector of all individual parameter vectors.
For simplicity purpose, we restrict to a same number of observations per subject: $J_i = J$ for all i. The extension to the general case is straightforward.

2.2. Likelihood function

Let $\theta = (\beta, \gamma, \sigma)$ be the parameter vector of interest 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. The likelihood function is well-defined under the assumption of existence of a strong solution to (2). We denote $x \mapsto p(x, t-s, s|x_s, \phi_i; \theta)$ the density of X_{it} given ϕ_i and $X_{is} = x_s$, $s < t$ with respect to the Lebesgue measure. This allows us to write the likelihood of model (1-3) as

$$p(\mathbf{y}; \theta) = \prod_{i=1}^{n} \int p(y_{i,0:J}|X_{i,0:J}; \sigma) p(X_{i,0:J}|\phi_i; \gamma) p(\phi_i; \beta) d\phi_i dX_{i,0:J}, \tag{4}$$

where $p(y|x; \sigma)$, $p(x|\phi; \gamma)$ and $p(\phi; \beta)$ 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 $p(y_{i,0:J}|X_{i,0:J}; \sigma) = \prod_{j=0}^{J} p(y_{ij}|X_{ij}; \sigma)$ where $p(y_{ij}|X_{ij}; \sigma)$ is the density function associated to the error model (1). The Markovian property of the diffusion process (X_{it}) implies

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

where $\Delta_{ij} = t_{ij} - t_{ij-1}$. Except for a mixed Wiener process with drift and additive noise, the integral (4) has no-closed form.

Maximizing (4) with respect to θ yields the corresponding Maximum Likelihood Estimate (MLE) $\hat{\theta}$ of the SDE mixed model (1-3). As the likelihood (4) is not explicit, it has to be numerically evaluated or maximized. This requires eitheir to know the transition densities explicitely (see Section 3) or to approximate it by an Euler scheme (see Section 5).

3. SDE with explicit transition density: estimation method

In this section, we assume that:

(M0) the SDE defined by (2) has an explicit transition density $p(x, t-s, s|x_s, \phi_i; \theta)$.

Even with this assumption, the estimation is complex because the n random parameters ϕ_i and random trajectories $(X_{i,0:J})$ are unobserved. 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 $\theta \to p(\mathbf{y}; \theta)$ 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):

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

where $(\alpha_m)_{m\in\mathbb{N}}$ is a sequence of positive numbers decreasing to zero.
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 included in a compact set 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 $E[S(\mathbf{y}, \mathbf{X}, \mathbf{\Phi})|\mathbf{y}; \theta']$. 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 $(\mathbf{X}, \mathbf{\Phi})$. 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. This S-step by a Markov kernel is detailed in the following subsection.

3.2. Simulation of the latent variables $(\mathbf{X}, \mathbf{\Phi})$ *given* $\mathbf{y}$

We know deal with the construction of a Markov Chain with ergodic distribution $p(\mathbf{X}, \mathbf{\Phi}|\mathbf{y}; \widehat{\theta}_{m-1})$. First note that conditionally to the observations $\mathbf{y}$, the individuals are independent. As a consequence, the simulation of each $(X_{i,0:J}, \phi_i)$ conditionally to $y_{0:J}$ can be performed separately. To ease the reading, in this subsection, we focus on the simulation of the non observed variables $(X_{i,0:J}, \phi_i)$ of a single individual. *Therefore, we omit the index i* in the following. Moreover, we denote $\theta = (\beta, \gamma, \sigma)$ the current values of the parameters and omit the index m of the SAEM algorithm iteration.

Standard version of the MCMC alternately simulates $X_{0:J}$ under the distribution $p(X_{0:J}|y_{0:J}, \phi; \gamma, \sigma)$ and ϕ under the distribution $p(\phi|y_{0:J}, X_{0:J}; \gamma, \sigma)$. If the conditional distributions $p(X_{0:J}|y_{0:J}, \phi; \gamma, \sigma)$ and $p(\phi|y_{0:J}, X_{0:J}; \gamma, \sigma)$ cannot be simulated easily, we resort to Metropolis Hastings algorithms for each component of ϕ and each time component of $X_{0:J}$ (Donnet and Samson, 2008). However, standard MCMC algorithms have reached their limits in high dimensional context: they do not exploit the Markovian structure of the data and have proved slow mixing properties. We propose to replace them by the Particle Markov Chain Monte Carlo (PMCMC) algorithm proposed by Andrieu et al. (2010). The idea of the PMCMC algorithm has been first proposed by Beaumont (2003), formalized by Andrieu and Roberts (2009) and developped in the context of state-space models by Andrieu et al. (2010).

Let us consider an ideal Metropolis-Hastings algorithm updating conjointly ϕ and $X_{0:J}$ conditionally to $y_{0:J}$. A new candidate $(X^c_{0:J}, \phi^c)$ would be generated with a proposal distribution:

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

and accepted with probability:

$$\rho(X^c_{0:J}, \phi^c|X_{0:J}, \phi) = \min\left\{1, \frac{q(\phi|\phi^c)}{q(\phi^c|\phi)} \frac{p(y_{0:J}|\phi^c; \gamma, \sigma)p(\phi^c; \beta)}{p(y_{0:J}|\phi; \gamma, \sigma)p(\phi; \beta)}\right\},$$

However, because of the complexity of the model, we are not able to simulate exactly under the conditional distribution $p(X_{0:J}|y_{0:J}, \phi; \gamma, \sigma)$ and the marginal likelihood $p(y_{0:J}|\phi; \gamma, \sigma)$ has no closed form. These two points can be tackled by an approximation through a particle filter, also called Sequential Monte Carlo (SMC) algorithm. The SMC algorithm produces a set of K particles $(X^{(k)}_{0:J})_{k=1\ldots K}$ and respective weights $(W^{(k)}_{0:J})_{k=1\ldots K}$ approximating the conditional distribution $p(X_{0:J}|y_{0:J}, \phi; \gamma, \sigma)$ by an empirical measure

$$\Psi^K_J = \sum_{k=1}^{K} W^{(k)}_{0:J} \delta_{X^{(k)}_{0:J}}$$

where δ_0 is any probability measure on $\mathbb{R}^{J+1}$.

Algorithm 1 (**SMC algorithm**).

- Time $j = 0$: *sample* $X_0^{(k)} \sim \pi_0(\cdot|\phi)\ \forall\, k = 1, \ldots, K$ *and compute the weights:*

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

- Time $j = 1, \ldots, J$: *sample* K *iid variables* $X_{0:j-1}^{'(k)}$ *according to the distribution* $\Psi_{j-1}^K = \sum_{k=1}^K W_{0:j-1}^{(k)} \delta_{X_{0:j-1}^{(k)}}$. *Then for each* $k = 1, \ldots, K$, *sample* $X_j^{(k)} \sim q\left(\cdot|y_j, X_{0:j-1}^{'(k)}, \phi; \gamma, \sigma\right)$ *and set* $X_{0:j}^{(k)} = (X_{0:j-1}^{'(k)}, X_j^{(k)})$. *Finally compute and normalize the weights:*

$$w_j\left(X_{0:j}^{(k)}\right) = \frac{p\left(y_j, X_j^{(k)}|y_{0:j-1}, X_{0:j-1}^{'(k)}\phi; \gamma, \sigma\right)}{q\left(X_j^{(k)}|y_j, X_{0:j-1}^{'(k)}, \phi; \gamma, \sigma\right)}, \quad W_j(X_{0:j}^{(k)}) = \frac{w_j\left(X_{0:j}^{(k)}\right)}{\sum_{k=1}^K w_j\left(X_{0:j}^{(k)}\right)}$$

The simulation of one trajectory $X_{0:J}$ (called a "run of SMC algorithm") under the approximation of $p(X_{0:J}|y_{0:J}, \phi; \gamma, \sigma)$ is directly achieved by randomly choosing one particle among the K particles with weights $(W_{0:J}^{(k)})_{k=1\ldots K}$. Besides, the marginal distribution $p(y_{0:J}|\phi; \gamma, \sigma)$ can be estimated through the weights

$$\widehat{p}^K(y_{0:J}|\phi; \gamma, \sigma) = \frac{1}{K}\sum_{k=1}^K w_0\left(X_0^{(k)}\right) \prod_{j=1}^J \left(\frac{1}{K}\sum_{k=1}^K w_j\left(X_{0:j}^{(k)}\right)\right). \tag{5}$$

As a consequence, Andrieu et al. (2010) propose the PMCMC algorithm:

Algorithm 2 (**PMCMC algorithm**).

- Initialization *: starting from* $\phi(0)$, *generate* $X_{0:J}(0)$ *by a run of SMC algorithm – with* K *particles– targeting* $p(X_{0:J}|y_{0:J}, \phi(0); \gamma, \sigma)$ *and estimate* $p(y_{0:J}|\phi(0); \gamma, \sigma)$ *by* $\widehat{p}^K(y_{0:J}|\phi(0); \gamma, \sigma)$
- At iteration $\ell = 1, \ldots, N$
  - *(a) Sample a candidate* $\phi^c \sim q(\cdot|\phi(\ell-1))$
  - *(b) By a run of SMC algorithm with* K *particles, generate* $X_{0:J}^c$ *targeting* $p(\cdot|y_{0:J}, \phi^c; \gamma, \sigma)$ *and compute* $\widehat{p}^K(y_{0:J}|\phi^c; \gamma, \sigma)$ *estimating* $p(y_{0:J}|\phi^c; \gamma, \sigma)$
  - *(c) Set* $(X_{0:J}(\ell), \phi(\ell)) = (X_{0:J}^c, \phi^c)$ *and* $\widehat{p}^K(y_{0:J}|\phi(\ell); \gamma, \sigma) = \widehat{p}^K(y_{0:J}|\phi^c; \gamma, \sigma)$ *with probability*

$$\widehat{\rho}^K(X_{0:J}^c, \phi^c|X_{0:J}(\ell-1), \phi(\ell-1)) = \min\left\{1, \frac{q(\phi(\ell-1)|\phi^c)}{q(\phi^c|\phi(\ell-1))}\frac{\widehat{p}^K(y_{0:J}|\phi^c; \gamma, \sigma)p(\phi^c; \beta)}{\widehat{p}^K(y_{0:J}|\phi(\ell-1); \gamma, \sigma)p(\phi(\ell-1); \beta)}\right\}$$

If the candidate is not accepted, then set $(X_{0:J}(\ell), \phi(\ell)) = (X_{0:J}(\ell-1), \phi(\ell-1))$ *and* $\widehat{p}^K(y_{0:J}|\phi(\ell); \gamma, \sigma) = \widehat{p}^K(y_{0:J}|\phi(\ell-1); \gamma, \sigma)$

Remark 1. *The proposal distributions* $q(\phi^c|\phi(\ell))$ *and* $q(X_j|y_j, X_{j-1}, \phi^c; \gamma, \sigma)$ *used in the SMC algorithm play a crucial role to ensure good mixing properties of the PMCMC algorithm. They are discussed with more details for the two simulated examples (Section 4).*

The most remarkable property of PMCMC is that the distribution of interest $p(X_{0:J}, \phi|y_{0:J}; \gamma, \sigma)$ is left invariant by the transition kernel, *whatever the number of particles* K, the ergodicity being reached under weak assumptions. More precisely, let Λ denote the auxiliary variables generated by the SMC algorithm, including the set of generated trajectories, the resampling indices and the indice of the particle randomly picked to obtain "a run of SMC algorithm". Andrieu and Roberts (2009) prove that *PMCMC is an exact MCMC algorithm* on $(X_{0:J}, \phi)$. Under general assumptions, the stationary distribution $\widetilde{\pi}^K(\Lambda, \phi)$ of the PMCMC algorithm is such that its marginalized distribution over the auxiliary variables Λ is exactly the distribution $p(X_{0:J}, \phi|y_{0:J}; \gamma, \sigma)$, independently on K: the PMCMC algorithm generates a sequence $(X_{0:J}(\ell), \phi(\ell))$ whose marginal distribution $\mathcal{L}^K(X_{0:J}(\ell), \phi(\ell)|y_{0:J}; \theta)$ is such that for all $\theta \in \Theta$ and for all $K > 0$,

$$||\mathcal{L}^K(X_{0:J}(\ell), \phi(\ell)|y_{0:J}; \theta) - p(X_{0:J}, \phi|y_{0:J}; \theta)||_{TV} \xrightarrow[\ell \to \infty]{} 0.$$

where $||\cdot||_{TV}$ is the total variation distance (Andrieu et al., 2010).

3.3. SAEM-PMCMC algorithm and convergence

We now combine algorithms PMCMC and SAEM:

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 by setting* $\phi_i(0)$ *as the expectation of* $p(\phi, \widehat{\beta}_{m-1})$
  - · *Run* N *iterations of the PMCMC algorithm with* K *particles at each iteration, targeting* $p(X_{i,0:J}, \phi_i|y_{i,0:J}; \widehat{\theta}_{m-1})$
- ∗ *Set* $\mathbf{X}^{(m)} = (X^{(m)}_{1,0:J}, \ldots, X^{(m)}_{n,0:J})$ *and* $\mathbf{\Phi}^{(m)} = (\phi^{(m)}_1, \ldots, \phi^{(m)}_n)$ *the simulated non observed variables*

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

$$s_m = s_{m-1} + \alpha_m \left[S(\mathbf{y}, \mathbf{X}^{(m)}, \mathbf{\Phi}^{(m)}) - s_{m-1}\right] \tag{6}$$

M-Step: *update of* $\widehat{\theta}_{m-1}$ *by* $\widehat{\theta}_m = \arg\max_\theta \left(-\psi(\theta) + \langle s_m, \nu(\theta)\rangle\right)$.

As the PMCMC algorithm can be viewed as a standard MCMC algorithm, the convergence of SAEM-PMCMC can be proved using Kuhn and Lavielle (2005) result. We recall the assumptions of Kuhn and Lavielle (2005).

(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 $\bar{s}(\theta) = \int S(\mathbf{y}, \mathbf{X}, \mathbf{\Phi}) p(\mathbf{X}, \mathbf{\Phi}|\mathbf{y}; \theta) d\mathbf{X} d\mathbf{\Phi}$ is continuously differentiable on Θ.

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

$$\partial_\theta \int p(\mathbf{y}, \mathbf{X}, \mathbf{\Phi}; \theta) d\mathbf{X} d\mathbf{\Phi} = \int \partial_\theta p(\mathbf{y}, \mathbf{X}, \mathbf{\Phi}; \theta) d\mathbf{X} d\mathbf{\Phi}.$$

(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 $S(\mathbf{y}, \mathbf{X}, \mathbf{\Phi})$.

(SAEM3) For all $\theta \in \Theta$, $\int ||S(\mathbf{y}, \mathbf{X}, \mathbf{\Phi})||^2 p(\mathbf{X}, \mathbf{\Phi}|\mathbf{y}; \theta) d\mathbf{X} d\mathbf{\Phi} < \infty$ and the function $\Gamma(\theta) = Cov_\theta(S(\mathbf{X}, \mathbf{\Phi}))$ is continuous.

(SAEM4) S is a bounded function.

(SAEM5) The transition probability Π_θ of the PMCMC algorithm is Lipschitz in θ and generates a uniformly ergodic chain. The Markov chain $(\mathbf{X}^{(m)}, \mathbf{\Phi}^{(m)})_{m \geq 0}$ takes its values in a compact subset.

We comment the different assumptions.
Assumption (M0) is a strong hypothesis, which is partially relaxed in Section 5.
Assumptions (M1-M5), (SAEM1-SAEM3) are standard and not restrictive.
Assumption (SAEM4) and the compacity assumption of (SAEM5) are the most restrictive and not really realistic. They could be relaxed using a principle of random boundaries presented in Allassonnière et al. (2009).
Assumption (SAEM5) is verified by the PMCMC algorithm depending on the proposal distributions. Indeed, the Lispchitz property holds if the complete likelihood is continuously derivable which is the case under (M2-M3) and if θ remains in a compact set, which is realistic in practice. About the uniform ergodicity, Andrieu et al. (2010) prove that if the noise density $p(y|X; \sigma)$ is bounded above

$$\sup_{y,X} p(y|X; \sigma) < M_\sigma, \tag{7}$$

PMCMC inherits the convergence properties of the corresponding ideal MCMC algorithm. For instance, if $q(\phi^c|\phi) = p(\phi^c; \beta)$ then the kernel of the ideal MCMC $q(X^c_{0:J}, \phi^c) = p(\phi^c; \beta) p(X^c_{0:J}|y_{0:J}, \phi^c; \gamma, \sigma)$ is independent. For this kernel, the ratio $\frac{p(X_{0:J}, \phi|y_{0:J})}{q(X_{0:J}, \phi)}$ is bounded if (7) holds, ensuring the uniform ergodicity (Tierney, 1994).

Theorem 1. *Assume that (M0-5), (SAEM1-5) hold. Under the assumption that* $\{s_m\}_{m\geq 0}$ *takes its values in a compact subset, the sequence* $\widehat{\theta}_m$ *supplied by the SAEM-PMCMC algorithm converges a.s. towards a (local) maximum of the log-likelihood* $\ell(\theta) = \log p(\mathbf{y};\theta)$.

Remark that despite the SMC approximation in the PMCMC algorithm, the fact that the marginal stationary distribution of MCMC is the exact conditional distribution $p(X, \Phi|\mathbf{y};\theta)$ is sufficient to prove the convergence of SAEM-PMCMC.

4. SDE with explicit transition density: simulation study

The respective performances of the SAEM implemented with a standard MCMC and of the SAEM combined with the PMCMC are compared to two models of various complexity: the Ornstein-Uhlenbeck process and the time-inhomogeneous Gompertz process.
In order to validate our results, we perform a large scale simulation study with various sets of parameters and designs (n, J). Moreover, we study the influence of the number of particles and of the proposals involved the SMC algorithm.
We compare the results using a biais and RMSE criteria. More precisely, for each condition (parameters, design, number of iterations) 100 datasets are generated. The corresponding estimate is obtained on each data set using both the SAEM-PMCMC algorithm and the SAEM-MCMC algorithm. 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 $\frac{1}{100}\sum_r(\widehat{\theta}_r - \theta)/\theta$ and relative root mean square error (RMSE) $\sqrt{\frac{1}{100}\sum_r(\widehat{\theta}_r - \theta)^2/\theta^2}$ for each component of θ are computed for both algorithms.
The SAEM algorithm requires initial value θ_0 and the choice of the sequence $(\alpha_m)_{m\geq 0}$. The initial values 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.

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

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

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

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

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

$$\begin{aligned} m_{iX} &= \left(\tau_i\kappa_i\left(1 - e^{-\frac{t_{i1}}{\tau_i}}\right), \ldots, \tau_i\kappa_i\left(1 - e^{-\frac{t_{iJ}}{\tau_i}}\right)\right)', \\ G_{iX} &= \left(\frac{\tau_i\gamma^2}{2}\left(1 - e^{-\frac{2\min(t_{ij}, t_{ik})}{\tau_i}}\right)e^{-\frac{|t_{ij} - t_{ik}|}{\tau_i}}\right)_{1\leq j,k\leq J}, \end{aligned} \tag{8}$$

where $'$ is the transposed vector. Altough this SDE is linear, the Gaussian transition density $p(X_{ij}, \Delta_{ij}, t_{ij-1}|X_{ij-1}, \phi_i; \theta)$ is a nonlinear function of ϕ_i. Thus, the likelihood has no closed form and the exact estimator of θ is unavailable.
This example is a toy example. First, we compare the performances of the SAEM algorithm implemented with a simple MCMC to those of the SAEM-PMCMC algorithm. Next, we compare the influence of the number of particles K and the choice of the proposal distribution $q(X_{t_{ij}}|\, X_{ij-1}, y_{ij}, \phi_i; \theta)$ in the SAEM-PMCMC.

4.1.2. Simulation step of the SAEM algorithm

Whereas the conditional distribution $p(X_{i,1:J}|, y_{i,0:J}, \phi_i; \theta)$ is Gaussian, the joint distribution $p(X_{i,1:J}, \phi_i|, y_{i,0:J}, \phi_i; \theta)$ is not explicitly and we have to resort to a approximate simulation at the S-step.
A first solution is to implement a standard MCMC algorithm, alternatively simulating under the distributions $p(X_{i,1:J}|, y_{i,0:J}, \phi_i; \theta)$ and $p(\phi_i|, y_{i,0:J}, X_{i,1:J}; \theta)$ for each subject. The posterior distribution $p(X_{i,1:J}|, y_{i,0:J},\, \phi_i; \theta)$ is Gaussian with easily computable mean vector and variance matrix derived from (8). Similarly, the posterior distribution of κ_i is Gaussian with explicit mean and variance. On the other hand, the posterior distribution of τ_i is not explicit and we use a Metropolis-Hastings step with a random walk proposal.
If we consider implementing a PMCMC kernel at the S-step, we have to choose two proposals $q(X_{t_{ij}}|\, X_{ij-1}, y_{ij}, \phi_i; \theta)$ and $q(\cdot|\phi_i)$. In this particular linear example, the proposal $q(X_{ij}|\, X_{ij-1}, y_{ij}, \phi_i; \theta)$ can be the optimal proposal, the exact posterior density $p(X_{ij}|X_{ij-1}, y_{ij}, \phi_i; \theta)$, which minimizes the variance of the particle weights. Indeed, this distribution is explicit for the Ornstein-Uhlenbeck, Gaussian with conditional mean and variance easily computable.
As an alternative, we also consider the transition density $p(X_{ij}|X_{ij-1}, \phi_i; \theta)$ 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 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):

$$\begin{aligned} S_1(\mathbf{y}, X, \mathbf{\Phi}) &= \sum_{i=1}^{n} \phi_i, \quad S_2(\mathbf{y}, X, \mathbf{\Phi}) = \sum_{i=1}^{n} \phi_i \phi_i', \\ S_3(\mathbf{y}, X, \mathbf{\Phi}) &= \sum_{i=1}^{n} (y_i - X_i)'(y_i - X_i). \end{aligned}$$

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

$$\widehat{\mu}^{(m)} = \frac{1}{n} s_{1m} \quad \widehat{\Omega}^{(m)} = \frac{1}{n} s_{2m} - \frac{1}{n^2} s_{1m} s_{1m}' \quad \widehat{\sigma}^{(m)} = \sqrt{\frac{1}{n} s_{3m}}$$

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

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

The M step is thus $\widehat{\gamma}^{(m)} = \sqrt{\frac{1}{nJ} s_{4m}}$.

Convergence assumptions (M0-5), (SAEM1-2) are easily checked on this example. (SAEM3) is implied by (SAEM4). (SAEM4) is not theoretically verified even if in practice, this is not a problem. Finally, given our choice of proposals for PMCMC, (SAEM5) also holds, the compacity being verified in practice.

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 are 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 $\widehat{\sigma}_0 = 5$. The SAEM-MCMC algorithm is implemented with $N = 100$ MCMC iterations. Several values of N and K are used in the SAEM-PMCMC algorithm.

Figure 1 presents the convergence of the SAEM-PMCMC algorithm for one dataset simulated with the second set of parameters, $n = 40$ and $J = 20$. This illustrates the low dependence of the initialization of SAEM and the quick convergence in a small neighborhood of the maximum likelihood.

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

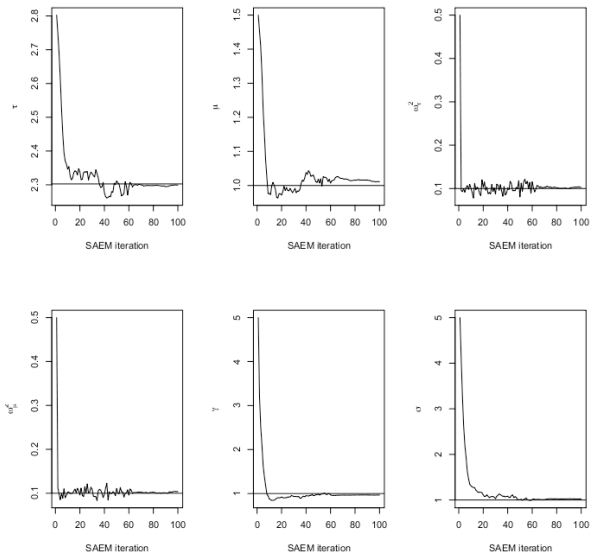

Figure 1. Convergence of the SAEM-PMCMC estimates $\widehat{\log(\tau)}_m$, $\widehat{\kappa}_m$, $\widehat{\omega_\tau}_m$, $\widehat{\omega_\kappa}_m$, $\widehat{\gamma}_m$ and $\widehat{\sigma}_m$ along the SAEM iterations for one dataset simulated with the Ornstein-Uhlenbeck mixed model, with the second set of parameters, $n = 40$ and $J = 20$. Horizontal lines represent the true values.

10%). Figure 2 shows the density of the estimators obtained by both algorithms with $n = 40, J = 20$ and the second set of parameters. 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.

In this particular linear model, the standard MCMC on $(\mathbf{X}, \mathbf{\Phi})$ is "ideal" as the distribution $p(X_{i,0:J}|y_{i,0:J}, \phi_i)$ can be sampled exactly. As a consequence, the SAEM-PMCMC can not be hopped to perform better. However, results show that the use of a particle approximation of $p(X_{i,0:J}|y_{i,0:J}, \phi_i)$ does not deteriorate the quality of estimation.

Besides, note that the design (n, J) of the study affects very little the estimation quality when γ and σ are small (first set of parameters). 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 ($n = 20, J = 40$).

In a second part, we study the influence of the particles number K in the SMC algorithm. 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 use $K = 25$, $K = 50$ and $K = 100$ particles and either $N = 50$, $N = 100$ or $N = 200$ PMCMC iterations, using the exact posterior distribution $q(X_{ij}|X_{ij-1}, y_{ij}, \phi_i; \theta) = p(X_{ij}|X_{ij-1}, y_{ij}, \phi_i; \theta)$. The results are almost identical. On the contrary, the choice of the proposal $q(X_{ij}|y_{ij}, X_{ij-1}, \phi; \gamma, \sigma)$ in the SMC algorithm may affect the result. Table 2 presents results obtained with the transition density $p(X_{ij}|X_{ij-1}, \phi; \gamma, \sigma)$ 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 par-

Table 1. Ornstein-Uhlenbeck mixed model: biais and RMSE (%) for $\widehat{\theta}$ 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
$n = 20$, $J = 40$							
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

ticles ($K = 1000$), but we observe the degeneracy of the algorithm on 90% of the simulated datasets. Morevoer, the bias and RMSE are greater than with the posterior conditional distribution $p(X_{ij}|y_{ij}, X_{ij-1}, \phi; \gamma, \sigma)$.

In conclusion, although the proposal q seems to be crucial in these simulations, the number of particles K has less influence on the convergence of the estimation algorithm. This is in concordance with the theoretical results proved by Andrieu et al. (2010).

However, as emphasized by Andrieu et al. (2010), increasing K can make arbitrarily small the probability of visiting unfavourable states by the Markov chain, which correspond to large values of the ratio target density to proposal density. But our simulations show that when choosing the "optimal" proposal density q, namely the exact posterior distribution, the ratio of the target density to the proposal density seems to have no larges values and thus the number of particles K has a very low influence. This low influence of K can also be explained by the fact that the SAEM only requires few iterations of PMCMC, without convergence of the Markov chain to the stationary distribution. The convergence is made over the iterations of SAEM.

Table 2. Ornstein-Uhlenbeck mixed model, $n = 20$, $J = 40$: biais and RMSE (%) for $\widehat{\theta}$ obtained by the SAEM-PMCMC algorithm on 100 simulated datasets. 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 $K = 25$, $K = 50$ or $K = 100$ particles, $N = 50$, $N = 100$ or $N = 200$ PMCMC iterations and prior proposal $q(X_{t_{ij}}|X_{t_{ij-1}}, y_{ij}, \phi_i; \theta) = p(X_{t_{ij}}|X_{t_{ij-1}}, \phi_i; \theta)$, $K = 1000$ particles and $N = 100$ iterations.

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, $K = 25$, $N = 100$						
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, $K = 50$, $N = 100$						
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, $K = 100$, $N = 100$						
Bias	0.04	0.24	-0.97	-4.11	6.86	-2.68
RMSE	0.48	0.29	1.24	1.29	1.15	0.51
SAEM-PMCMC with posterior proposal, $K = 50$, $N = 50$						
Bias	0.28	0.04	4.07	-11.61	10.06	-3.32
RMSE	0.49	0.30	1.40	1.55	1.38	0.55
SAEM-PMCMC with posterior proposal, $K = 50$, $N = 200$						
Bias	-0.06	0.31	-1.27	-2.68	6.10	-2.36
RMSE	0.48	0.30	1.16	1.30	1.15	0.49
SAEM-PMCMC with prior proposal, $K = 1000$, $N = 100$						
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

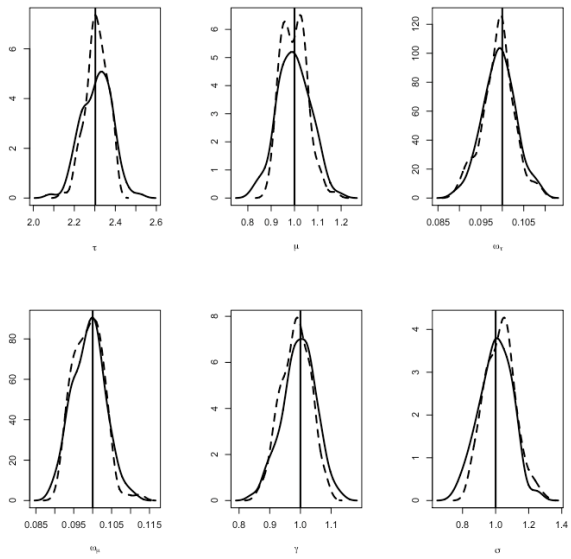

Figure 2. Ornstein-Uhlenbeck mixed model, $n = 40$, $J = 20$ and second set of parameters. Densities of the estimators $\widehat{\theta}$ obtained with the SAEM-PMCMC (plain line) and SAEM-MCMC algorithms (dashed line) on 100 simulated datasets, for each component of $(\log \tau, \kappa, \omega_\tau, \omega_\kappa, \gamma, \sigma)$. The true parameter value is the vertical line.

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:

$$\begin{aligned} y_{ij} &= X_{ij}(1+\varepsilon_{ij}),\ \varepsilon_{ij} \sim \mathcal{N}(0,\sigma^2), \\ dX_{it} &= B_i C_i e^{-C_i t} X_{it} dt + \gamma X_{it} dB_{it}, \quad X_{i0} = A_i e^{-B_i} \end{aligned} \tag{9}$$

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 initial conditional is a function of the random individual parameters ϕ_i. We assume that $\log A_i \sim_{i.i.d.} \mathcal{N}(\log A, \omega_A^2)$, $\log B_i \sim_{i.i.d.} \mathcal{N}(\log B, \omega_B)$, $\log C_i \sim_{i.i.d.} \mathcal{N}(\log C, \omega_C^2)$ and the parameter of interest is $\theta = (\log A, \log B, \log C, \omega_A, \omega_B, \omega_C, \gamma, \sigma)$. By the Itô formula, the conditional expectation and variance of $\log(X_{it})|\phi_i$, for $t \geq 0$, are

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

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

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

which is a nonlinear function of ϕ_i. As a consequence, the likelihood has no closed form and the exact estimator of θ is unavailable.

4.2.2. Simulation step of the SAEM algorithm

First, the S step of the SAEM algorithm is tackled with a standard MCMC algorithm. Due to the multiplicative structure of the observation noise, the distribution $p(X_{i,0:J}|y_{i,0:J},\phi_i,\theta)$ is no more explicit and we have to resort to a random walk proposal to simulate the trajectories $X_{i,0:J}$ conditionally to the observations $y_{i,0:J}$. The components of $X_{i,0:J}$ are updated time by time using a random walk proposal. A random walk proposal is also used to simulate the individual parameters ϕ_i from the conditional distribution.

Secondly, we implement the SAEM-PMCMC algorithm and the proposal $q(X_{ij}|X_{ij-1},y_{ij},\phi_i;\theta)$ has to be chosen. We propose to approximate the ideal proposal $p(X_{ij}|X_{ij-1},y_{ij},\phi_i;\theta)$ by a Gaussian distribution with mean and variance deduced from the true ones. More precisely, we consider the following proposal on $\log X_{t_{ij}}$

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

with

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

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. As the parameter γ appears both in the expectation and the variance of $\log(X_t)|\phi$, its estimator is not the same as in Example 1. The sufficient statistic corresponding to γ is

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

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.

Table 3. Stochastic volatility Gompertz mixed model: bias and RMSE (%) of $\widehat{\theta}$ 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.

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

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: $\log A = \log(3000)$, $\log B = \log(5)$, $\log C = \log(14)$, $\omega_A = 0.1$, $\omega_B = 0.1$, $\omega_C = 0.1$. Two levels of variability 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 $\widehat{\log A}_0 = 8.21$, $\widehat{\log B}_0 = 1.81$, $\widehat{\log C}_0 = 2.84$, $\widehat{\omega_A}_0 = 0.5$, $\widehat{\omega_B}_0 = 0.5$, $\widehat{\omega_C}_0 = 0.5$, $\widehat{\gamma}_0 = 1.2$, $\widehat{\sigma}_0 = 0.66$. The PMCMC is implemented with $N = 100$ iterations and $K = 50$ particles. The MCMC is implemented with $N = 100$ iterations.

Bias and RMSE (%) are presented in Table 3. 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

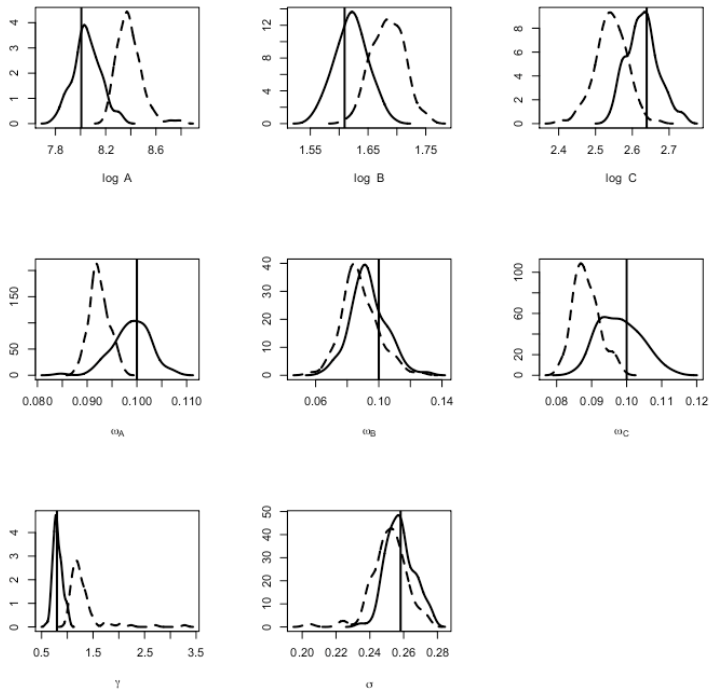

Figure 3. Stochastic volatility Gompertz mixed model, $n = 20$, $J = 40$ and second set of parameters. Densities of the estimators $\widehat{\theta}$ obtained with the SAEM-PMCMC (plain line) and SAEM-MCMC algorithms (dashed line) on 100 simulated datasets, for each component of $(\log A, \log B, \log C, \omega_A, \omega_B, \omega_C, \gamma, \sigma)$. The true parameter value is th vertical line.

also be observed in Figure 3 where we plot the densities of the parameters $\widehat{\theta}$ among the 100 simulated datasets for the second set of parameters with $n = 20$ and $J = 40$.
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. Theoretical results for general SDE requiring Euler approximation

The class of SDEs considered under assumption (M0) is limited. For a general SDE, we propose to approach the process by an Euler-Maruyama scheme, leading to an approximate model with an explicit (Gaussian) transition density.

5.1. Euler approximation

For the sake of simplicity, we now consider that $\forall i = 1 \dots n, \forall j = 0 \dots J$, $t_{ij} = j$ and that the observations times are regularly spaced with interval Δ. Let r be an integer ≥ 1. On each

interval time $[t_j, t_{j+1}[$, we define an approximate diffusion by the recursive Euler-Maruyama scheme of step size $1/r$ by : $X_{i0}(r) \sim \pi_0(\cdot|\phi_i)$ and

$$\begin{aligned} X_{ij}(0) &= X_{ij-1}(r), \\ X_{ij}(u+1) &= X_{ij}(u) + \frac{1}{r}\, a(X_{ij}(u), t_{j-1} + u/r, \phi_i) \\ &\quad + \sqrt{\frac{1}{r}}\, b(X_{ij}(u), t_{j-1} + u/r, \phi_i, \gamma) U_u \end{aligned} \tag{10}$$

where the variables U_u are iid unit centered Gaussian. Then $X_{ij}(r)$ is a realization of the approximate diffusion at time t_j. We consider the error model

$$y_{ij} = g(X_{ij}(r), \varepsilon_{ij}) \tag{11}$$

Equations (10), (11) and (3) define the so-called *approximate mixed model*. Under this model, the likelihood of the observations $\mathbf{y}$ is :

$$p_{(r)}(\mathbf{y}; \theta) = \prod_{i=1}^{n} \int p(y_{i,0:J}|X_{i,0:J}(r); \sigma) p_{(r)}(X_{i,0:J}(r)|\phi_i; \gamma) p(\phi_i; \beta) d\phi_i dX_{i,0:J}(r),$$

where $p_{(r)}(X_{i,0:J}(r)|\phi_i; \theta)$ is the density of the Euler approximation $X_{ij}(r)$ given ϕ_i.

5.2. SAEM-PMCMC on the approximate Euler model

A first approach to estimate θ is to use the algorithm SAEM-PMCMC presented in Section 3 on the approximate model. In that case, Theorem 1 implies the following corollary:

Corollary 1. *Assume that (M1-5), (SAEM1-5) hold for the approximate model. Under the assumption that $\{s_m\}_{m\geq 0}$ takes its values in a compact subset, the sequence $\{\widehat{\theta}_m\}_{m\geq 0}$ generated by the algorithm SAEM-PMCMC on the approximate Euler model converges to a (local) maximum of the likelihood $p_{(r)}(\mathbf{y}; \theta)$ of the approximate model.*

The main limit of this result is that the SAEM-PMCMC converges to the maximum of the Euler approximate likelihood $p_{(r)}(\mathbf{y}; \theta)$ instead of the maximum of the exact likelihood $p(\mathbf{y}; \theta)$. Donnet and Samson (2008) prove that there exists a constant C such that the total variation distance between $p_{(r)}(\mathbf{y}; \theta)$ and the exact likelihood $p(\mathbf{y}; \theta)$ can be bounded by

$$||p_{(r)}(\mathbf{y}; \theta) - p(\mathbf{y}; \theta)||_{TV} \leq \frac{C}{r}$$

This results says that the two likelihoods are close but this does not imply that the two maxima are close.

A solution would be to consider $r \to \infty$, so that the Euler approximation converges to the exact diffusion. This requires r to increase along the iterations of the SAEM algorithm, also implying the Euler approximate model and the likelihood to change with the iterations. The concept of stationary distribution of PMCMC is no more valid in this context. We could imagine a reversible jump approach for the simulation step of the SAEM algorithm. But even in this case, as the approximate likelihood changes with the iterations, it is not possible to prove the convergence of the SAEM algorithm.

From a practical point of view, the choice of r is sensitive because it adds intermediate latent variables to the model. A large value of r is prefered as the Euler approximate likelihood is closer to the exact likelihood. However, it is well known that a large volume of latent variables is difficult to handle within a MCMC or PMCMC approach. This has been discussed by Roberts and Stramer (2001) in a general diffusion context and by Donnet and Samson (2008) for SDE mixed models. Furthermore, as emphasized by Roberts and Stramer (2001), when $r \to \infty$, MCMC algorithms with r additional latent variables are unable to correctly estimate γ. Indeed, in that case, the density $p_{(r)}(x_j|x_{j-1}, \phi; \gamma)$ of the Euler approximate diffusion converges to the likelihood of the continuous path $(X_t)_{t\in[t_{j-1},t_j]}$ given by the Girsanov formula. It is well-known that it is not possible to estimate the volatility from this continuous path likelihood.

5.3. Exact theoretical convergence for a SAEM algorithm coupled with a SMC algorithm

To circumvent the problem of the estimation of γ, we restrict to the estimation of

$$\theta = (\beta, \sigma)$$

with γ assumed to be known. To estimate θ, we propose a theoretical algorithm coupling SAEM with a naive SMC. This algorithm is known to have bad numerical properties (this is the reason why we do not implement it) but we prove that it supplies a sequence $(\widehat{\theta}_m)_{m\in\mathbb{N}}$ converging a.s. towards a local maximum of the *exact likelihood* $\log p(\mathbf{y}; \theta)$. This theoretical result yields to a consequent improvement of Corollary 1.

Now, we present the SMC that we propose to use in the simulation step of the SAEM algorithm. In the literature, SMC algorithms have originally a filtering purpose to approximate the distribution $p(X_{0:J}|\phi, y_{0:J}; \theta)$. However, they have been rapidly extended to the case where the parameters ϕ are also sampled. In the following, we present a naive SMC extension targeting the distribution $p_{(r)}(X_{0:J}, \phi|y_{0:J}; \theta)$ of the approximate model.

Algorithm 4 (**Naive SMC algorithm with parameters sampling**).

- Time $j = 0$: $\forall\, k = 1, \ldots, K$, *sample* $\phi_0^{(k)} \sim p(\cdot; \beta)$ *and* $X_0^{(k)} \sim \pi_0(\cdot|\phi_0^{(k)})$ *and compute the weights:*
$$W_0\left(\phi_0^{(k)}, X_0^{(k)}\right) \propto p_{(r)}\left(y_0, X_0^{(k)}, \phi_0^{(k)}; \theta\right)$$

- Time $j = 1, \ldots, J$: *sample* K *iid variables* $(\phi_{j-1}^{'(k)}, X_{0:j-1}^{'(k)})$ *according to the law* $\Psi_{j-1}^K = \sum_{k=1}^K W_{0:j-1}^{(k)} \mathbf{1}_{\phi_{j-1}^{(k)}, X_{0:j-1}^{(k)}}$. *Then for each* $k = 1, \ldots, K$, *sample* $X_j^{(k)} \sim q_{(r)}\left(\cdot|y_j, X_{0:j-1}^{'(k)}, \phi_{j-1}^{'(k)}; \gamma, \sigma\right)$ *and set* $\phi_j^{(k)} = \phi_{j-1}^{'(k)}$ *and* $X_{0:j}^{(k)} = (X_{0:j-1}^{'(k)}, X_j^{(k)})$. *Finally compute and normalize the weights:*
$$W_j\left(\phi_j^{(k)}, X_{0:j}^{(k)}\right) \propto \frac{p_{(r)}\left(y_j, X_j^{(k)}|y_{0:j-1}, X_{0:j-1}^{'(k)}\phi_{j-1}^{'(k)}; \gamma, \sigma\right)}{q_{(r)}\left(X_j^{(k)}|y_j, X_{0:j-1}^{'(k)}, \phi_{j-1}^{'(k)}; \gamma, \sigma\right)},$$

We consider a SAEM algorithm in which the simulation step is performed by this naive SMC sampler. To ensure the convergence of this estimation algorithm towards the exact

maximum likelihood, we need the number of particles K and the step size $1/r$ of the Euler approximation to vary along the iterations of the SAEM algorithm. Thus K_m and r_m denote the number of particles and the step size of the Euler scheme at iteration m of the SAEM algorithm. We denote by $(\widehat{\theta}_m)_{m\in\mathbb{N}}$ the sequence of estimates supplied by this algorithm. Technical assumptions (E1-4) are required to prove the convergence of the naive SMC sampler and are presented in Appendix.

Theorem 2. *Assume that there exists a constant $c > 1$ such that K_m varies along the iterations:*

$$K_m = O\left(\log(m)^c\right)$$

and that $r_m = \min\{r \in \mathbb{N} \,|\, r \geq \sqrt{K_m}\}$. Assume that (M1-5), (SAEM1-4), (E1-4) hold for the exact model. 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 exact log-likelihood $\ell(\theta) = \log p(\mathbf{y}; \theta)$.

Theorem 2 is proved in Appendix 7.2. This theoretical result is noteworthy. Indeed, even if the SAEM-SMC algorithm is performed on the approximate Euler model, the algorithm converges to the maximum of the exact likelihood. This is due to the convergence of the Euler approximate SMC towards the exact filter distribution. This powerful result has first been proved by Del Moral et al. (2001). We propose an extension of their results to our SMC (see Lemma 1 in Appendix 7.1). Then, by generalizing the proof of convergence of the SAEM algorithm to an approximate simulation step, we are able to deduce the convergence of the estimates to the maximum of the exact likelihood.
The naive SMC algorithm 4 provides an asymptotically consistent estimate of the target distribution $p_{(r)}(X_{0:J}, \phi | y_{0:J}; \theta)$ under very weak assumptions but has proved bad properties in practice (in particular a degeneracy of the parameters due to the resampling step). It can be improved by including a MCMC step on the parameters ϕ (Doucet et al., 2001). Although these algorithms remains poorly efficient in practice (explaining the developement of the PMCMC algorithm for instance), they allow to prove this theorem.

6. 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 when the transition density is explicit. This proof is classical as the PMCMC acts as an exact marginal MCMC. On the contrary, when we consider SDE with unexplicit transition density, we prove the convergence of the SAEM-PMCMC algorithm to the maximum likelihood of an approximate model obtained by a Euler scheme. We are not able to prove the convergence of this algorithm to the exact maximum likelihood. But we propose the convergence of a SAEM-SMC algorithm to the exact maximum likelihood.
The suggested method supplies accurate parameter estimation in a really moderate computational time on practical examples. Moreover, it is not restricted to homogeneous time SDE. 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 involves less tuning parameters than a standard MCMC, 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.

Acknowledgments

Sophie Donnet has been financially supported by the french program ANR BigMC.

References

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

Andrieu, C., A. Doucet, and R. Holenstein (2010). Particle Markov chain Monte Carlo methods. *J. R. Statist. Soc. B* (72), 1–33.

Andrieu, C. and G. O. Roberts (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist. 37*(2), 697–725.

Beaumont, M. (2003). Estimation of population growth or decline in genetically monitored populations. *Genetics* (164), 1139–1160.

Casarin, R. and J.-M. Marin (2009). Online data processing: comparison of Bayesian regularized particle filters. *Electron. J. Stat. 3*, 239–258.

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

Del Moral, P., J. Jacod, and P. Protter (2001). The Monte-Carlo method for filtering with discrete-time observations. *Probab. Theory Related Fields 120*(3), 346–368.

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

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

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

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

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

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

Jaffrézic, F. and J. Foulley (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*.

Kitawaga, G. (1996). Monte carlo filter and smoother for non-gaussian nonlinear state space models. *J. Comp. Graph. Statist.* (5), 1–25.

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

Kuhn, E. and M. Lavielle (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*. Berlin-Heidelberg: Springer-Verlag.

Oravecz, Z., F. Tuerlinckx, and J. Vandekerckhove (2009). A hierarchical Ornstein-Uhlenbeck model for continuous repeated measurement data. *Psychometrika* (74), 395–418.

Overgaard, R., N. Jonsson, C. Tornøe, and H. Madsen (2005). Non-linear mixed-effects models with stochastic differential equations: Implementation of an estimation algorithm. *J Pharmacokinet. Pharmacodyn. 32*(1), 85–107.

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

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

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

Roberts, G. O. and O. Stramer (2001). On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm. *Biometrika 88*(3), 603–621.

Samson, A., M. Lavielle, and F. Mentré (2007). The SAEM algorithm for group comparison tests in longitudinal data analysis based on non-linear mixed-effects model. *Stat. Med. 26*(27), 4860–4875.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *Ann. Statist. 22*(4), 1701–1762.

Tornøe, C., R. Overgaard, H. Agersø, H. Nielsen, H. Madsen, and E. Jonsson (2005). Stochastic differential equations in NONMEM: implementation, application, and comparison with ordinary differential equations. *Pharm. Res. 22*(8), 1247–58.

West, M. (1993). Approximating posterior distributions by mixtures. *J. Roy. Statist. Soc. Ser. B 55*(2), 409–422.

Wolfinger, R. (1993). Laplace's approximation for nonlinear mixed models. *Biometrika 80*(4), 791–795.

7. Appendix

7.1. Convergence of the Euler approximate naive SMC

Let us introduce some notations. For any bounded Borel function f taking real values, we denote $\pi_{j\theta} f = E\left(f(X_{0:j}, \phi)|y_{0:j}; \theta\right)$ the conditional expectation under the exact smoothing distribution of the exact model $p(X_{0:j}, \phi|y_{0:j}; \theta)$ and $\Psi_{J,\theta}^{(K)} f = \sum_{k=1}^{K} W_{0:J}^{(k)} f(\phi^{(k)}, X_{0:J}^{(k)})$ the conditional expectation under the naive SMC distribution of the approximate model with K particles and a Euler step size $r = \min\{r \in \mathbb{N} \,|\, r \geq \sqrt{K}\}$.

Lemma 1. *Assume that*

- *E1. The functions a and b are two times differentiable with derivatives w.r.t x and ϕ all orders up to 2 uniformly bounded w.r.t ϕ*
- *E2. Θ is included in a compact set $\mathcal{C}$. Besides, the functions $\theta \mapsto \pi(\phi; \beta)$ and $\sigma \mapsto p(y|x; \sigma)$ are continuous on $\mathcal{C}$.*
- *E3. f is a function of $(\phi, X_{0:J})$ taking its values into $\mathbb{R}$, uniformly bounded*
- *E4. There exist κ_1 and κ_2 independent of σ such that $p(y|x, \sigma) \leq \kappa_1$ and such that*

$$\frac{p_{(r)}\left(y_j, X_j^{(k)} | \mathbf{y}_{0:j-1}, X_{0:j-1}^{'(k)} \phi_{j-1}^{'(k)}; \gamma, \sigma\right)}{q\left(X_j^{(k)} | y_j, X_{0:j-1}^{'(k)}, \phi_{j-1}^{'(k)}; \gamma, \sigma\right)} \leq \kappa_2$$

Then, for any $\varepsilon > 0$, for any $j = 1, \ldots, J$, there exist constants C_1 and C_2 independent of θ such that

$$\mathbb{P}\left(\left|\Psi_{j,\theta}^{(K)} f - \pi_{j,\theta} f\right| \geq \delta\right) \leq C_1 \exp\left(-K \frac{\delta^2}{C_2}\right)$$

Proof. Following Del Moral et al. (2001) proof, let us define the empirical measure $\Psi_{j,\theta}^{'(K)} f = \sum_{k=1}^{K} W_{0:j}^{(k)} f(\phi^{'(k)}, X_{0:j}^{'(k)})$ where $X_{0:j}^{'(k)}$ has been defined in algorithm 4. Set the σ-filed $\mathcal{G}$ generated by the variables $\phi_0^{(k)}, X_0^{(k)}, \ldots, \phi_j^{(k)}, X_{0:j}^{(k)}$. We have

$$\Psi_{j,\theta}^{'(K)} f - \Psi_{j,\theta}^{(K)} f = \frac{1}{K} \sum_{k=1}^{K} \eta_k - \mathbb{E}(\eta_k | \mathcal{G})$$

with $\eta_k = f(\phi_j^{'(k)}, X_{0:j}^{'(k)})$. By deviation inequalities, under assumption (E3), we obtain that there exist two constants such that

$$\mathbb{P}(|\Psi_{j,\theta}^{'(K)} f - \Psi_{j,\theta}^{(K)} f| \geq \delta) \leq A_1 \exp\left(-K \frac{A_2 \delta^2}{\|f\|^2}\right)$$

Next, we introduce the kernels H_j such that :

$$H_j f(\phi, X_{0:j-1}) = \int p(y_j|x_j)p(x_j|X_{0:j-1},\phi)p(\phi)f(\phi, X_{0:j-1}, x_j)d\phi dx_j$$

Thus we have

$$\pi_j f = \frac{\pi_{j-1}H_j f}{\pi_{j-1}H_j 1}$$

and the same definitions on the Euler approximate model, with kernels H_j^K by

$$H_j^K f(\phi, X_{0:j-1}) = \int p(y_j|x_j)p_{(r)}(x_j|X_{0:j-1},\phi)p(\phi)f(\phi, X_{0:j-1}, x_j)d\phi dx_j$$

such that $\pi_j^K f = \mathbb{E}_r(f(\phi, X_{0:j-1})|y_{0:j};\theta)$ the conditional expectation under the approximate model is

$$\pi_j^K f = \frac{\pi_{j-1}^K H_j^K f}{\pi_{j-1}^K H_j^K 1}$$

Then, considering the σ-filed $\mathcal{G}$ generated by the variables $\phi_0^{'(k)}, X_0^{'(k)}, \ldots, \phi_{j-1}^{'(k)}, X_{0:j-1}^{'(k)}$ we can write

$$W_{0:j}^{(k)} f - \Psi_j^{'K} H_j f \quad = \quad \frac{1}{K}\sum_{k=1}^{K} \eta_k - \mathbb{E}(\eta_k|\mathcal{G})$$

with $\eta_k = f(\phi_j^{(k)}, X_{0:j}^{(k)}) \frac{p_{(r)}\left(y_j, X_j^{(k)}|\mathbf{y}_{0:j-1}, X_{0:j-1}^{'(k)}\phi_{j-1}^{'(k)};\gamma,\sigma\right)}{q\left(X_j^{(k)}|y_j, X_{0:j-1}^{'(k)}, \phi_{j-1}^{'(k)};\gamma,\sigma\right)}$. We have to bound $|\eta_k|$ and $\mathbb{E}(|\eta_k|^2|\mathcal{G})$. Assumption (E4) implies that there exist two constants C_1, C_2 such that

$$|\eta_k| \leq C_1||f||, \quad \mathbb{E}(|\eta_k|^2|\mathcal{G}) \leq C_2||f||^2$$

Deviation inequalities imply that there exists a constant C such that

$$\mathbb{P}\left(\left|W_{0:j}^{(k)} f - \Psi_j^{'K} H_j f\right| \geq \delta\right) \leq 2e^{-K\frac{\delta^2}{C||f||^2}}$$

With similar arguments than Del Moral et al. (2001) and assumptions (E1-2), we prove by induction that there exist two constants A_1 and A_2 such that

$$\mathbb{P}(|\pi_{j\theta} f - \Psi_{j,\theta}^{(K)} f| \geq \delta) \leq A_1 \exp\left(-K\frac{\delta^2}{A_2\|f\|^2}\right)$$

where $A_1 = 3^{j+2}$, $A_2 = C(8\rho)^{j+1}$ with $C = 2 + 2\alpha$ and $\rho = 2\kappa^{J+1}/\varepsilon_{J+1}$, where α is the constant quantifying the error induced by the Euler scheme and $\varepsilon_{J+1} = H_J \ldots H_1 1$. With regularities assumptions and assumptions (E1-2), we can prove that the constants are independent of θ.

7.2. Proof of Theorem 2

PROOF. At iteration m, the simulation step provides $(\mathbf{X}^{(m)}, \mathbf{\Phi}^{(m)})$ simulated with an approximate distribution $\Psi_{J,\hat{\theta}_m}^{(K_m)} = \otimes_{i=1}^n \Psi_{i,J,\hat{\theta}_m}^{(K_m)}$ where $\Psi_{i,J,\hat{\theta}_m}^{(K_m)}$ is the empirical measure obtained by the naive SMC for individual i. The stochastic approximation step of the SAEM-SMC algorithm can be decomposed into:

$$\begin{aligned} s_{m+1} &= s_m + \alpha_m(S(\mathbf{y}, \mathbf{X}^{(m)}, \mathbf{\Phi}^{(m)}) - s_m) \\ &= s_m + \alpha_m h(s_m) + \alpha_m e_m + \alpha_m R_m \end{aligned}$$

with

$$\begin{aligned} h(s_m) &= E(S(\mathbf{y}, \mathbf{X}, \mathbf{\Phi})|\mathbf{y}, \hat{\theta}(s_m)) - s_m \\ e_m &= S(\mathbf{y}, \mathbf{X}^{(m)}, \mathbf{\Phi}^{(m)}) - \Psi_{J,\hat{\theta}(s_m)}^{(K_m)} S(\mathbf{y}, \mathbf{X}, \mathbf{\Phi}) \\ R_m &= \Psi_{J,\hat{\theta}(s_m)}^{(K_m)} S(\mathbf{y}, \mathbf{X}, \mathbf{\Phi}) - E(S(\mathbf{y}, \mathbf{X}, \mathbf{\Phi})|\mathbf{y}, \hat{\theta}(s_m)) \end{aligned}$$

Following Theorem 2 of Delyon et al. (1999) on the convergence of Robbins-Monro scheme, the convergence of SAEM-SMC to the set of stationary points of the exact log-likelihood $\ell(\theta)$ is ensured if we prove the following assertions:

(a) 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.
(b) $\lim_{m\to\infty} \sum_{k=1}^m \alpha_k e_k$ exists and is finite with probability 1.
(c) $\lim_{m\to\infty} R_m = 0$ with probability 1.

- To prove assertion 1, we first remark that, as γ is known and $\theta = (\beta, \sigma)$, we have

$$\begin{aligned} \partial_\theta \log p(\mathbf{y}, \mathbf{X}, \mathbf{\Phi}; \theta) &= \partial_\theta \log p_{(r)}(\mathbf{y}, \mathbf{X}, \mathbf{\Phi}; \theta) \\ \partial_\theta \log p(\mathbf{y}; \theta) &= \partial_\theta \log p_{(r)}(\mathbf{y}; \theta) \end{aligned}$$

  Thus under (M1-M5), (SAEM2), the functions $h(s) = \bar{s}(\hat{\theta}(s)) - s$ and $V(s) = -\ell(\hat{\theta}(s))$ verify

$$\langle \partial_s V(s), h(s) \rangle = -\langle \partial_s \ell(\hat{\theta}(s)), h(s) \rangle \leq 0$$

  With similar arguments than Lemma 2 of Delyon et al. (1999), we can also prove that the set $V(\{s, F(s) = 0\})$ is of zero measure.

- Assertion 2 is proved following Theorem 5 of Delyon et al. (1999). Note $\mathcal{F} = \{\mathcal{F}_m\}_{m\geq 0}$ the increasing family of σ-algebra generated by all the random variables used by the SMC algorithm until the mth iteration. By construction of the simulation step, the simulated variables $(\mathbf{X}^{(1)}, \mathbf{\Phi}^{(1)}), \ldots, (\mathbf{X}^{(m)}, \mathbf{\Phi}^{(m)})$ given $\widehat{\theta}_0, \ldots, \widehat{\theta}_m$ are independent. Consequently, we have that for any positive Borel function f,

$$E(f(\mathbf{X}^{(m+1)}, \mathbf{\Phi}^{(m+1)})|\mathcal{F}_m) = \Psi_{J,\theta}^{(K_m)} f$$

  We can then prove that $\sum_{k=1}^m \alpha_k e_k$ is a martingale, bounded in L_2 under assumptions (M5) and (SAEM1-4).

- We now prove the almost sure convergence of R_m (assertion 3). The term R_m is of dimension d. It converges a.s. i.f.f. each of its component converges a.s. As a

consequence, for the sake of simplicity, we suppose that d equals 1. Applying Borel-Cantelli lemma, it is sufficient to prove that $\sum_{m\geq 0} P(|R_m| \geq \epsilon) < \infty$ for all $\epsilon > 0$. We have

$$R_m = \Psi^{(K_m)}_{J,\hat{\theta}(s_m)} S - \pi_{J,\hat{\theta}(s_m)} S$$

where $\pi_{J,\theta} S = \mathbb{E}(S(\mathbf{X}, \mathbf{\Phi})|\mathbf{y}; \theta)$ First, by independence of the individual $i = 1 \dots n$, we can decompose S into a sum of n terms $S(\mathbf{X}, \mathbf{\Phi}) = \sum_{i=1}^{n} S_i(X_{i,0:J}, \phi_i)$. and write:

$$\begin{aligned}
&\mathbb{P}\left(\left|\Psi^{(K_m)}_{J,\hat{\theta}(s_m)} S - \pi_{J,\hat{\theta}(s_m)} S\right| \geq \delta\right) = \\
&\mathbb{P}\left(\left|\sum_{i=1}^{n} \Psi^{(K)}_{J,\hat{\theta}(s_m)} S_i - \sum_{i=1}^{n} \mathbb{E}(S_i(X_{i,0:J}, \phi_i)|y_{i,0:J}; \hat{\theta}(s_m))\right| \geq \delta\right) \\
&\leq \sum_{i=1}^{n} \mathbb{P}\left(\left|\Psi^{(K)}_{J,\hat{\theta}(s_m)} S_i - \pi_{i,J,\hat{\theta}(s_m)} S_i\right| \geq \frac{\delta}{n}\right)
\end{aligned}$$

with $\pi_{i,J,\hat{\theta}(s_m)} S_i = \mathbb{E}(S_i(X_{i,0:J}, \phi_i)|y_{i,0:J}; \theta)$. Now, we bound each term of the sum by Lemma 1 and under assumptions (SAEM4), (E1-E2). This implies the almost sure convergence of (R_m) towards 0.