Extension of the SAEM algorithm for nonlinear mixed models with 2 levels of random effects

XAVIÈRE PANHARD*
INSERM U738 and UFR de Médecine, University Paris 7, 16 rue Huchard, 75018 Paris, France
xaviere.panhard@bichat.inserm.fr

ADELINE SAMSON
Laboratoire MAP5, UMR CNRS 8145, University Paris 5, France

SUMMARY
This article focuses on parameter estimation of multilevel nonlinear mixed-effects models (MNLMEMs). These models are used to analyze data presenting multiple hierarchical levels of grouping (cluster data, clinical trials with several observation periods, ...). The variability of the individual parameters of the regression function is thus decomposed as a between-subject variability and higher levels of variability (e.g. within-subject variability). We propose maximum likelihood estimates of parameters of those MNLMEMs with 2 levels of random effects, using an extension of the stochastic approximation version of expectation–maximization (SAEM)–Monte Carlo Markov chain algorithm. The extended SAEM algorithm is split into an explicit direct expectation–maximization (EM) algorithm and a stochastic EM part. Compared to the original algorithm, additional sufficient statistics have to be approximated by relying on the conditional distribution of the second level of random effects. This estimation method is evaluated on pharmacokinetic crossover simulated trials, mimicking theophylline concentration data. Results obtained on those data sets with either the SAEM algorithm or the first-order conditional estimates (FOCE) algorithm (implemented in the nlme function of R software) are compared: biases and root mean square errors of almost all the SAEM estimates are smaller than the FOCE ones. Finally, we apply the extended SAEM algorithm to analyze the pharmacokinetic interaction of tenofovir on atazanavir, a novel protease inhibitor, from the Agence Nationale de Recherche sur le Sida 107-Puzzle 2 study. A significant decrease of the area under the curve of atazanavir is found in patients receiving both treatments.

Keywords: Bioequivalence trials; Crossover trial; Multilevel nonlinear mixed-effects models; Multiple periods; SAEM algorithm.

1. INTRODUCTION
The use of nonlinear mixed-effects models (NLMEs) to model longitudinal data is increasing in several fields such as agronomy, forestry, clinical trials, population pharmacokinetics (PK), and pharmacodynam-

*To whom correspondence should be addressed.

© The Author 2008. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permissions@oxfordjournals.org.
ics. In some settings, data can present multiple hierarchical levels of grouping, leading to multiple nested levels of variability. For instance, we may study patients who are grouped in medical services that are themselves grouped into hospitals. In this article, we consider multilevel nonlinear mixed-effects models (MNLMEMs) for such data. MNLMEMs represent a natural extension of models with a single level of variability, and they have recently been subject to a great deal of attention in statistical literature and many application areas including population PK.

In NLMEMs with only one level of variability, often corresponding to between-subject variability, the analysis results in the estimation of the fixed effects parameters and the between-subject variability of the parameters, also called intersubject variability. When there is more than one level of grouping, the higher levels of variability can be estimated. In the specific case where the second level of grouping corresponds to multiple periods of measurement, this variability is called within-subject variability (or intrasubject variability or interoccasion variability) and corresponds to the variation of the individual parameters across the different study periods or units. In the context of PK, Karlsson and Sheiner (1993) demonstrate the importance of modeling this second level of variability in 2-level NLMEMs.

The parameter estimation of NLMEMs is not trivial because the likelihood of NLMEMs cannot be expressed in a closed form due to the nonlinearity of the regression function in the random effects. Therefore, several estimation methods have been proposed. The first-order conditional estimates (FOCE) algorithm performs a first-order linearization of the regression function with respect to the random effects (Beal and Sheiner, 1982; Lindstrom and Bates, 1990). The implementation of the FOCE algorithm in the NONMEM software and in the nlme function of Splus and R enables the estimation of both between-subject and within-subject variability. In our experience, the main drawback of this method is that it does not always converge when one estimates simultaneously the between-subject and the within-subject variability in several parameters. Furthermore, the theoretical basis of this linearization-based method is weak. For instance, Vonesh (1996) and Ge and others (2004) give examples of specific designs resulting in inconsistent estimates, such as when the number of observations per subject does not increase faster than the number of subjects or when the variability of random effects is too large.

Several estimation methods have been proposed as alternatives to linearization algorithms. A common method to handle numerical integration is the adaptative Gaussian quadrature (AGQ) method. An estimation algorithm of NLMEM parameters based on this classical AGQ method has been proposed by Pinheiro and Bates (1995) and is implemented in the SAS procedure NLMIXED (Wolfinger, 1993). However, the AGQ method requires a sufficiently large number of quadrature points implying an often slow convergence with very high computational time. Furthermore, 2-level NLMEMs can be implemented in the NLMIXED procedure, but convergence is difficult to obtain in practice (Jaffrezic and others, 2006). Improvements upon this method are thus needed. The second alternative to linearization is the use of the expectation–maximization (EM) algorithm (Dempster and others, 1977) in order to estimate models with missing or nonobserved data such as random effects. Because of the nonlinearity of the model, stochastic versions of the EM algorithm have been proposed. Wei and Tanner (1990), Walker (1996), and Wu (2004) propose Monte Carlo EM (MCEM) algorithms, with a Monte Carlo approximation of the E-step. However, the MCEM algorithm may have computational problems (i.e. slow or even nonconvergence). As an alternative to address computational problems, a stochastic approximation version of expectation–maximization (SAEM) has been proposed in Delyon and others (1999) and Kuhn and Lavielle (2005a), which requires the simulation of only one realization of the missing data for each iteration, substantially reducing the computation time. Kuhn and Lavielle (2005b) propose to combine the SAEM algorithm with a Monte Carlo Markov chain (MCMC) procedure adapted to the NLMEMs and prove that the resulting estimates are convergent and consistent.

To date, none of the EM-based algorithms are directly applicable to the case of MNLMEMs. The objective of this paper is to extend the SAEM algorithm to MNLMEMs with 2 levels of variability: both E- and M-steps need to be adapted to integrate higher levels of random effects. We also propose estimates
of the likelihood and of the Fisher information matrix. We evaluate this algorithm on a simulated PK example, more precisely a 2-period 1-sequence crossover trial mimicking theophylline concentration data (Pinheiro and Bates, 1995). We also apply the SAEM algorithm to the PK interaction of 2 HIV molecules (tenofovir and atazanavir) from a PK substudy of the Agence Nationale de Recherche sur le Sida (ANRS) 107-Puzzle 2 trial.

Section 2 sets out the model and notation. Section 3 describes the SAEM algorithm. Section 4 reports the simulation study and its results. In Section 5, we study the PK interaction of tenofovir on atazanavir in HIV patients. Section 6 is a conclusion—discussion.

2. Models

We denote by $y_{ijk}$ the observation in unit $k$ ($k = 1, \ldots, K$) for subject $i$ ($i = 1, \ldots, n$) at time $t_{ijk}$ ($j = 1, \ldots, n_{ik}$). For instance, the different units can be the different periods in the case of PK trials or the different parents in the case of genetic analyses. We assume 2 known nonlinear functions $f$ and $g$ such that the 2-level NLMEM linking observations to sampling times can be written as

$$y_{ijk} = f(t_{ijk}, \phi_{ik}) + g(t_{ijk}, \phi_{ik})\varepsilon_{ijk},$$

where $\phi_{ik}$ is the $p$-vector of the parameters of subject $i$ for unit $k$ and $\varepsilon_{ijk}$ is the measurement error. We assume that the errors $\varepsilon_{ijk}$ are mutually independent, the individual parameters $\phi_{ik}$ are random vectors, and for each unit $k$, $\phi_{ik}$ can be decomposed as

$$\phi_{ik} = \mu + \beta_k + b_i + c_{ik},$$

$$b_i \sim N(0, \Omega),$$

$$c_{ik} \sim N(0, \Psi),$$

where $\mu + \beta_k$ is the mean value for unit $k$, $b_i$ is the random effect of size $p$ of subject $i$, and $c_{ik}$ is the random effect of size $p$ of subject $i$ and unit $k$. To ensure the identifiability of the parameters, we assume that $\beta_1 = 0$, that is, $\mu$ is the mean of the first unit and $\beta_k$ represents the difference (or effect) of the $k$th unit in comparison to this first unit. The random effects $(b_i)$ and $(c_{ik})$ are assumed to be mutually independent. The total variance of the parameters is thus broken into a between-subject variance $\Omega$ and a within-subject variance $\Psi$. Finally, the individual parameters vector $\phi_i = (\phi_{i1}, \ldots, \phi_{iK})$ of subject $i$ is distributed with a Gaussian distribution with mean vector $(\mu, \mu + \beta_2, \ldots, \mu + \beta_K)$ and a $pK \times pK$ covariance matrix $\Gamma$ equal to

$$\Gamma = \begin{pmatrix}
\Omega + \Psi & \Omega & \ldots & \Omega \\
\Omega & \Omega + \Psi & \ldots & \Omega \\
\vdots & \vdots & \ddots & \vdots \\
\Omega & \Omega & \ldots & \Omega + \Psi
\end{pmatrix}.$$  

Let $\theta = (\mu, \beta, \Omega, \Psi, \sigma^2)$, the vector of all the parameters of the model, where $\beta = (\beta_1, \ldots, \beta_K)$. The aim of this paper is to propose an estimation of $\theta \in \Theta$ by maximizing the likelihood of the observations $y = (y_{ijk})_{ijk}$.

We write $\tilde{b}_i := \mu + b_i$. The likelihood of $y$ can be written as

$$p(y; \theta) = \int p(y, \phi, \tilde{b}; \theta)d(\phi, \tilde{b}),$$
where \( p(y, \phi, \tilde{b}; \theta) \) is the likelihood of the complete data \((y, \phi, \tilde{b})\), with \( \phi = (\phi_{ik})_{i=1,\ldots,n;k=1,\ldots,K} \) and \( \tilde{b} = (\tilde{b}_1, \ldots, \tilde{b}_n) \). Because of the nonlinearity of the regression function \( f \) with respect to the random effects \( \phi_{ik} \), the likelihood has no closed form. Therefore, the maximization of the likelihood is a complex problem. We propose to use a stochastic version of the EM algorithm, which is presented in detail in Section 3.

### 3. Estimation algorithm

The EM algorithm introduced by Dempster and others (1977) is a classical approach to estimating parameters of models with nonobserved or incomplete data. In 2-level NLMEs, the unobserved data are the individual parameters \((\phi, \tilde{b})\) and the complete data are \((y, \phi, \tilde{b})\). Let denote \( L_c(y, \phi, \tilde{b}; \theta) = \log p(y, \phi, \tilde{b}; \theta) \) the log-likelihood of the complete data. The principle of the iterative EM algorithm is to maximize the function \( Q(\theta | \theta') = E(L_c(y, \phi, \tilde{b}; \theta) | y; \theta') \) where the expectation is the conditional expectation under the posterior distribution \( p(\phi, \tilde{b} | y; \theta') \). Each iteration of the EM algorithm is computed through 2 steps: the expectation step (E-step) and the maximization step (M-step). At the \( \ell \)th iteration of the algorithm, the E-step is the evaluation of \( Q(\theta | \hat{\theta}_\ell) \), while the M-step updates \( \hat{\theta}_\ell \) by maximizing \( Q(\theta | \hat{\theta}_\ell) \).

We now show that the function \( Q \) can be reduced in the case of a MNLMEM. First, note that as \( p(\tilde{b} | y, \phi; \theta) = p(\tilde{b} | \phi; \theta) \), by application of Bayes’ theorem, we have

\[
p(\phi, \tilde{b} | y; \theta) = p(\tilde{b} | \phi; \theta) p(\phi | y; \theta) = \prod_{i=1}^{n} p(\tilde{b}_i | \phi_i; \theta) p(\phi_i | y_i; \theta).
\]

(3.1)

Second, through the linearity of the individual parameters model in (2.1), the posterior distribution \( p(\tilde{b}_i | \phi_i; \theta) \) of the \( i \)th subject is explicit: \( p(\tilde{b}_i | \phi_i; \theta) \) is a Gaussian distribution \( \mathcal{N}(m(\phi_i, \theta), V(\theta)) \) with mean and variance equal to

\[
m(\phi_i, \theta) = V(\theta) \left( \Psi^{-1} \sum_{k=1}^{K} (\phi_{ik} - \beta_k) + \Omega^{-1} \mu \right),
\]

\[
V(\theta) = (\Omega^{-1} + K \Psi^{-1})^{-1}.
\]

(3.2)

The factorization in (3.1) allows the function \( Q \) to be rewritten as

\[
Q(\theta | \theta') = \int \left( \int L_c(y, \phi, \tilde{b}; \theta) p(\tilde{b} | \phi; \theta') d\tilde{b} \right) p(\phi | y; \theta') d\phi.
\]

Because of the explicit posterior distribution of random effects \( \tilde{b} \) given in (3.2), the computation of this conditional expectation can be split into 2 parts: the computation of the integral with respect to the posterior distribution of \( \tilde{b} \), which has an analytical form, and the computation of the integral with respect to the posterior distribution of \( \phi \), which has no analytical form. Therefore, the EM algorithm is split into an explicit direct EM algorithm for the computation of the first integral and the use of a stochastic version of the EM algorithm for the computation of the second integral.

We now consider the explicit computation of the first integral, denoted by

\[
R(y, \phi, \theta, \theta') = \int L_c(y, \phi, \tilde{b}; \theta) p(\tilde{b} | \phi; \theta') d\tilde{b}.
\]
This integral has an analytical form, and the complete log-likelihood \( L_c(y, \phi, \tilde{b}; \theta) \) is

\[
L_c(y, \phi, \tilde{b}; \theta) = -\frac{1}{2} \sum_{k=1}^{K} \sum_{i=1}^{n} \sum_{j=1}^{n_k} \log(2\pi \sigma^2 g^2(t_{ijk}, \phi_{ik})) - \frac{1}{2} \sum_{k=1}^{K} \sum_{i=1}^{n} \sum_{j=1}^{n_k} \frac{\left(y_{ijk} - f(t_{ijk}, \phi_{ik})\right)^2}{\sigma^2 g^2(t_{ijk}, \phi_{ik})}
\]

\[
- \frac{nK}{2} \log(2\pi \det \Psi) - \frac{1}{2} \sum_{k=1}^{K} \sum_{i=1}^{n} (\phi_{ik} - \tilde{b}_i - \beta_k)^T \Psi^{-1}(\phi_{ik} - \tilde{b}_i - \beta_k)
\]

\[
- \frac{n}{2} \log(2\pi \det \Omega) - \frac{1}{2} \sum_{i=1}^{n} (\tilde{b}_i - \mu)^T \Omega^{-1}(\tilde{b}_i - \mu).
\]

As the posterior distribution \( p(\tilde{b}|\phi; \theta) \) is known (3.2), \( R(y, \phi, \theta, \theta') \) is equal to

\[
R(y, \phi, \theta, \theta') = -\frac{1}{2} \sum_{i,j,k} \log(2\pi \sigma^2 g^2(t_{ijk}, \phi_{ik})) - \frac{1}{2} \sum_{i,j,k} \frac{(y_{ijk} - f(t_{ijk}, \phi_{ik}))^2}{\sigma^2 g^2(t_{ijk}, \phi_{ik})}
\]

\[
- \frac{nK}{2} \log(2\pi \det \Psi) - \frac{nK}{2} \text{tr}(\Psi^{-1} V(\theta'))
\]

\[
- \frac{1}{2} \sum_{i,k} \text{tr}(\Psi^{-1}(\phi_{ik} - m(\phi_i, \theta') - \beta_k)(\phi_{ik} - m(\phi_i, \theta') - \beta_k)^T)
\]

(3.3)

\[
- \frac{n}{2} \log(2\pi \det \Omega) - \frac{n}{2} \text{tr}(\Omega^{-1} V(\theta'))
\]

\[
- \frac{1}{2} \sum_{i=1}^{n} \text{tr}(\Omega^{-1}(m(\phi_i, \theta') - \mu)(m(\phi_i, \theta') - \mu)^T),
\]

where \( \text{tr} \) denotes the trace of matrix.

Therefore, \( Q \) is reduced to the computation of the second integral under the posterior distribution \( p(\phi|y; \theta) \) as follows:

\[
Q(\theta|\theta') = \int R(y, \phi, \theta, \theta') p(\phi|y; \theta') d\phi.
\]

(3.4)

Because of the nonlinearity of \( f \) with respect to \( \phi \), the posterior distribution \( p(\phi|y; \theta') \) has no closed form and the function \( Q \) defined by (3.4) is intractable. Thus, we propose to use the stochastic version SAEM of the EM algorithm proposed by Delyon and others (1999) which evaluates the integral \( Q \) by a stochastic approximation procedure.

Let us detail this SAEM algorithm in the case of 2-level NLMEMs. Let us note that the quantity \( R(y, \phi, \theta, \theta') \) belongs to the regular curved exponential family, that is, it can be written as

\[
R(y, \phi, \theta, \theta') = -\Lambda(\theta) + \langle S(y, \phi, \theta'), \Phi(\theta) \rangle,
\]

where \( \langle \cdot, \cdot \rangle \) is the scalar product, \( \Lambda \) and \( \Phi \) are twice continuously differentiable on \( \Theta \), and \( S(y, \phi, \theta') \) is the minimal sufficient statistic of the complete model. In this case, the \( Q \) function is reduced to

\[
Q(\theta|\theta') = -\Lambda(\theta) + \left( \int S(y, \phi, \theta') p(\phi|y; \theta') d\phi, \Phi(\theta) \right),
\]

and at the \( \ell \)th iteration, the SAEM algorithm proceeds as follows:
• Simulation step: simulation of the missing data \((\phi_i^{(\ell)})_i\) under the conditional distribution \(p(\phi|y; \hat{\theta}_\ell)\).

• Stochastic approximation step: computation of a stochastic approximation \(s_{\ell+1}\) of \(E[S(y, \phi, \hat{\theta}_\ell)| y; \hat{\theta}_\ell] = \int S(y, \phi, \hat{\theta}_\ell) p(\phi|y; \hat{\theta}_\ell) d\phi\), using \((\gamma_\ell)_{\ell \geq 0}\), a sequence of positive numbers decreasing to 0:

\[
s_{\ell+1} = s_\ell + \gamma_\ell (S(y, \phi^{(\ell)}, \hat{\theta}_\ell) - s_\ell).
\]

• Maximization step: update of the estimate \(\hat{\theta}_{\ell+1}\):

\[
\hat{\theta}_{\ell+1} = \arg \max_{\theta \in \Theta} (-\Lambda(\theta) + \langle s_{\ell+1}, \Phi(\theta) \rangle).
\]

The sufficient statistics of the complete model (3.3) evaluated during the SA-step of the SAEM algorithm are as follows:

\[
s_{1,i,\ell+1} = s_{1,i,\ell} + \gamma_{\ell} \left( \sum_{k=1}^{K} \phi_{ik}^{(\ell)} - s_{1,i,\ell} \right), \quad i = 1, \ldots, N;
\]

\[
s_{2,k,\ell+1} = s_{2,k,\ell} + \gamma_{\ell} \left( \sum_{i=1}^{n} \phi_{ik}^{(\ell)} - s_{2,k,\ell} \right), \quad k = 1, \ldots, K;
\]

\[
s_{3,\ell+1} = s_{3,\ell} + \gamma_{\ell} \left( \sum_{i=1}^{n} m(\phi_{i}^{(\ell)}, \hat{\theta}_\ell) t_{ik} m(\phi_{i}^{(\ell)}, \hat{\theta}_\ell) - s_{3,\ell} \right);
\]

\[
s_{4,\ell+1} = s_{4,\ell} + \gamma_{\ell} \left( \sum_{k=1}^{K} \sum_{i=1}^{n} (\phi_{ik}^{(\ell)} - m(\phi_{i}^{(\ell)}, \hat{\theta}_\ell)) t_{ijk} (\phi_{ik}^{(\ell)} - m(\phi_{i}^{(\ell)}, \hat{\theta}_\ell)) - s_{4,\ell} \right);
\]

\[
s_{5,\ell+1} = s_{5,\ell} + \gamma_{\ell} \left( \sum_{i,j,k} \left( \frac{y_{ijk} - f(t_{ijk}, \phi_{ik}^{(\ell)})}{g(t_{ijk}, \phi_{ik}^{(\ell)})} \right)^2 - s_{5,\ell} \right).
\]

The expression of the M-step is obtained by derivation of (3.3). The parameter estimates are as follows:

\[
\hat{\mu}_{\ell+1} = V(\hat{\theta}_\ell) \tilde{Q}_{\ell}^{-1} \left( \frac{1}{n} \sum_{i=1}^{n} s_{1,i,\ell+1} - \sum_{k=1}^{K} \tilde{p}_{k,\ell} \right) + V(\hat{\theta}_\ell) \tilde{\Omega}_{\ell}^{-1} \hat{\mu}_{\ell},
\]

\[
\hat{\beta}_{k,\ell+1} = \frac{s_{2,k,\ell+1}}{n} - \hat{\mu}_{\ell+1}, \quad \text{for } k = 2, \ldots, K,
\]

\[
\hat{\Omega}_{\ell+1} = V(\hat{\theta}_\ell) + \frac{s_{3,\ell+1}}{n} - (\hat{\mu}_{\ell+1})^T \hat{\mu}_{\ell+1},
\]

\[
\hat{Q}_{\ell+1} = V(\hat{\theta}_\ell) + \frac{s_{4,\ell+1}}{nK} - \frac{1}{K} \sum_{k=1}^{K} (\hat{\beta}_{k,\ell+1})^T \hat{\beta}_{k,\ell+1},
\]

\[
\hat{\sigma}^2_{\ell+1} = \frac{s_{5,\ell+1}}{\sum_{i=1}^{n} \sum_{k=1}^{K} n_{ik}}.
\]
The extension of the classic SAEM algorithm for single-level NLMEMs to 2-level NLMEMs is split into an explicit EM algorithm and a stochastic EM part. Furthermore, it requires the computation of 2 intermediate quantities (the conditional expectations $m(\phi_i, \theta)$ and variance $V(\theta)$ of the between-subject random effects parameters $b_i$) as well as 2 additional sufficient statistics ($S_3$ and $S_4$), functions of $m(\phi_i, \theta)$.

The M-step differs from the one of the classic SAEM for single-level NLMEMs, especially for the estimation of the variance matrix $\Omega_1$ and $\Psi_1$ which uses the additional quantity $V(\theta)$.

As proved by Delyon and others (1999) and Kuhn and Lavielle (2005a), the convergence of the SAEM algorithm is ensured under the following assumption.

**ASSUMPTION A1**

1. Functions $\Lambda$ and $\Phi$ are twice continuously differentiable on $\Theta$.
2. The log-likelihood $\log p(y; \theta)$ is $d$ times differentiable on $\Theta$, where $d$ is the dimension of $S(y, \phi, \theta')$.
3. The function
   \[
   \bar{s}(\theta, \theta') = \int S(y, \phi, \theta') p(\phi|y; \theta) \, d\phi
   \]
   is continuously differentiable on $\Theta$ with respect to its first variable.
4. For all $\ell$ in $\mathbb{N}$, $\gamma_\ell \in [0, 1]$, $\sum_{\ell=1}^{\infty} \gamma_\ell = \infty$ and $\sum_{\ell=1}^{\infty} \gamma_\ell^2 < \infty$.

For a convenient step-size sequence $\gamma_\ell$, Assumption (A1) is trivially checked in our model. A choice of step-size sequence $\gamma_\ell$ is presented in Section 4.

However, the simulation step of the SAEM algorithm, which consists in sampling $\phi$ in the posterior distribution $p(\phi|y; \theta)$, cannot be directly performed because the posterior distribution is only known up to a constant. In this case, Kuhn and Lavielle (2005b) propose to combine the SAEM algorithm with a MCMC procedure for the simulation step. This version of the SAEM–MCMC algorithm can be used for the estimation of MNLMEMs. The simulation step of the SAEM–MCMC algorithm performs the simulation of the missing data $\phi$ through a Markov chain which has $p(\phi|y; \theta)$ as unique stationary distribution. For subject $i$, by Bayes’ formula, this conditional distribution is proportional to

\[
p(\phi_i|y_i; \theta) \propto \prod_{k=1}^{K} p(y_{ik}|\phi_{ik}; \theta) p(\phi_i; \theta).
\]

To simulate this Markov chain, we use a Metropolis–Hastings algorithm.

As proved by Kuhn and Lavielle (2005a), the convergence of the SAEM–MCMC algorithm is ensured under Assumption (A1) and the following additional assumption.

**ASSUMPTION A2** For any $\theta$ in $\Theta$, the conditional distribution $p(\cdot|y; \theta)$ is the unique limiting distribution of a transition probability $\Pi_\theta$ with the following properties:

1. For any compact subset $V$ of $\Theta$, there exists a real constant $L$ such that for any $(\theta, \theta')$ in $V^2$,
   \[
   \sup_{(\phi, \phi') \in \mathcal{E}} |\Pi_\theta(\phi' | \phi) - \Pi_{\theta'}(\phi' | \phi)| \leq L \|\theta - \theta'\|.
   \]
2. The transition probability $\Pi_\theta$ supplies a uniformly ergodic chain whose invariant probability is the conditional distribution $p(\phi|y; \theta)$, that is,
   \[
   \exists K_\theta \in \mathbb{R}^+, \exists \rho_\theta \in ]0, 1[ \forall \ell \in \mathbb{N}, \|\Pi_\theta^\ell(\cdot|\phi) - p(\cdot, \cdot|y; \theta)\|_{TV} \leq C_\theta \rho_\theta^\ell,
   \]
where \( \| \cdot \|_{TV} \) is the total variation norm. Furthermore,
\[
C = \sup_{\theta \in \Theta} C_{\theta} < \infty \quad \text{and} \quad \rho = \sup_{\theta \in \Theta} \rho_{\theta} < 1.
\]

3. The function \( S \) is bound on \( \mathcal{E} \).

At iteration \( \ell \), the S-step of the SAEM–MCMC algorithm therefore consists of simulating \( \phi^{(\ell)} \) with the transition probability \( \Pi_{\phi_{\ell}}(\phi^{(\ell-1)}|d\phi^{(\ell)}) \).

Part (2) of Assumption (A2) is the most delicate to check and depends on the choice of the MCMC algorithm. This is detailed in the supplementary material available at Biostatistics online (http://www.biostatistics.oxfordjournals.org).

Because the SAEM algorithm is a stochastic algorithm, a deterministic convergence criterion is not appropriate. We recommend implementing the SAEM algorithm with a sufficiently large number of iterations and graphically checking the convergence by plotting the values of the SAEM estimates against iteration number, as illustrated in Section 4.

To perform statistical tests such as the Wald test or likelihood ratio test (LRT), we propose estimators of the Fisher information matrix and the likelihood, respectively. As the Fisher information matrix has no closed form in MNLMEMs, we propose to approximate it by the Fisher information matrix of the multilevel linear mixed model deduced from the MNLMEM after linearization of the function \( f \) around the conditional expectation of the individual parameters \( (E(\phi_i|y_i; \hat{\theta}), 1 \leq i \leq n) \). The computation of this linearized Fisher information matrix is direct and does not need any approximation.

The estimation of the likelihood of the MNLMEM is based on an importance sampling procedure, as proposed by Samson and others (2007) for NLMEMs. In this case, an estimate of the contribution \( p(y_i; \theta) \) of the individual \( i \) to the likelihood is
\[
\hat{p}(y_i; \theta) = \frac{1}{T} \sum_{t=1}^{T} \frac{p(y_i, \phi^{(t)}_i; \theta)}{q_i(\phi^{(t)}_i)},
\]
where \( (\phi^{(t)}_i)_{t=1,...,T} \) are simulated using the individual instrumental distribution \( q_i \). As an individual instrumental distribution \( q_i \), we propose a Gaussian approximation of the individual conditional posterior distribution \( p(\phi_i|y_i; \theta) \).

4. Simulation study: a PK example

4.1 Simulation settings

The objective of this simulation study is to illustrate the main statistical properties of the extended SAEM algorithm (bias, root mean square errors [RMSEs], group comparison tests) and to compare them to the FOCE algorithm.

We use the PK data of orally administered theophylline to define the population model for the simulation study. These data are classical ones in population PK, often used for software evaluation (Pinheiro and Bates, 2000). We assume that concentrations can be described by a 1-compartment model with first-order absorption and first-order elimination:
\[
f(t, \phi) = \frac{D K_a}{V K_a - Cl} \left( e^{-\frac{Q}{t}} - e^{-K_a t} \right),
\]
where \( D \) is the dose, \( V \) is the volume of distribution, \( K_a \) is the absorption rate constant, and \( Cl \) is the clearance of the drug elimination. These parameters are positive and distributed according to a lognormal...
distribution. Thus, $\phi$ has the following components: $\phi = (\log V, \log K_a, \log \text{AUC})$, with AUC = $D/Cl$ is the Area Under the time-concentration Curve. We assume identical sampling times for all subjects, hence $t_{ijk} = t_j$, for $j = 1, \ldots, J$. Additive Gaussian random effects are assumed for each parameter with a diagonal covariance matrix $\Omega$ and a diagonal covariance matrix $\Psi$. Let $\omega^2 = (\omega^2_V, \omega^2_{K_a}, \omega^2_{\text{AUC}})$ and $\psi^2 = (\psi^2_V, \psi^2_{K_a}, \psi^2_{\text{AUC}})$ denote the vector of the variances of the random effects. A combined error model is assumed by setting $g(t, \phi) = 1 + f(t, \phi)$.

We set the dose for all subjects to the value of 4 mg. For all the parameters, the values are those proposed by Panhard and Mentré (2005): $\log V = -0.73$, $\log K_a = 0.39$, $\log \text{AUC} = 4.61$, $\omega^2_V = 0.01$, $\omega^2_{K_a} = 0.04$, $\omega^2_{\text{AUC}} = 0.04$, $\psi^2_V = 0.0025$, $\psi^2_{K_a} = 0.01$, $\psi^2_{\text{AUC}} = 0.01$, and $\sigma^2 = 0.01$. We generate $n = 24$ and $n = 40$ total numbers of subjects with $J = 10$ blood samples per subject, taken at 15 min, 30 min, 1, 2, 3.5, 5, 7, 9, 12, and 24 h after dosing. The individual data of one simulated trial are displayed in Figure 1.

### 4.2 Evaluation of estimates

Our aim is to evaluate and compare the estimates produced by the extended SAEM algorithm with those produced by the nlme function of the R software. We fit the simulation model and compute the relative bias and RMSEs for each component of $\theta$ from 1000 replications of the 2 trials described below ($n = 24$ and $n = 40$ total number of subjects).

The simulation model includes a treatment effect on all components of $\theta$. We test the null hypothesis $\{\beta_{\text{log AUC}} = 0\}$ using the Wald test. We also fit the model where the treatment effect on log AUC is not estimated and test the same null hypothesis using the LRT.

The SAEM algorithm is implemented with 500 iterations. During the first 200 iterations, we use a constant step size $\gamma_0 = 1$. Then, during the last 300 iterations, the stochastic approximation scheme is implemented with a step size equal to $\gamma_\ell = 1/(\ell - 200)$ at iteration $\ell$ to satisfy (1) of Assumption (A1). The evolution of each SAEM parameter estimate is plotted against iteration number in Figure 2. In this example, the number of iterations has been chosen such that the convergence is clearly attained for all the model parameters.
Fig. 2. Evolution of the estimates, function of the iteration of SAEM algorithm.

The relative bias and RMSEs obtained on the 1000 data sets with $n = 24$ and $n = 40$ subjects are presented in Table 1.

The bias and the RMSEs of the fixed effects ($\mu$) are small with the SAEM algorithm and almost half of those obtained with nlme (especially the RMSEs). The bias and RMSEs of the unit effect ($\beta$) are small
Table 1. Relative biases (%) and relative RMSEs (%) of the estimated parameters evaluated by the extended SAEM algorithm and the FOCE algorithm (nlme function) from 1000 simulated trials

<table>
<thead>
<tr>
<th>Parameter</th>
<th>n = 24 subjects</th>
<th>n = 40 subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias SAEM nlme</td>
<td>Bias SAEM nlme</td>
</tr>
<tr>
<td>V</td>
<td>0.01 0.53</td>
<td>−0.06 0.54</td>
</tr>
<tr>
<td>( k_a )</td>
<td>0.48 −1.48</td>
<td>14.4 24.4</td>
</tr>
<tr>
<td>AUC</td>
<td>−0.08 −0.20</td>
<td>1.0 1.5</td>
</tr>
<tr>
<td>( \beta_V )</td>
<td>−0.00 −0.01</td>
<td>3.6 3.6</td>
</tr>
<tr>
<td>( \beta_{k_a} )</td>
<td>−0.73 −0.76</td>
<td>14.2 18.8</td>
</tr>
<tr>
<td>( \beta_{AUC} )</td>
<td>0.02 −0.02</td>
<td>0.7 1.1</td>
</tr>
<tr>
<td>( \omega^2_V )</td>
<td>−5.13 −5.92</td>
<td>38.7 38.4</td>
</tr>
<tr>
<td>( \omega^2_{k_a} )</td>
<td>−3.99 −7.07</td>
<td>42.4 41.5</td>
</tr>
<tr>
<td>( \omega^2_{AUC} )</td>
<td>−4.88 −7.29</td>
<td>34.5 34.2</td>
</tr>
<tr>
<td>( \psi^2_V )</td>
<td>−8.67 −7.29</td>
<td>69.4 68.5</td>
</tr>
<tr>
<td>( \psi^2_{k_a} )</td>
<td>−10.94 −9.09</td>
<td>73.5 72.0</td>
</tr>
<tr>
<td>( \psi^2_{AUC} )</td>
<td>−5.37 −5.37</td>
<td>43.6 42.6</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>−0.33 −0.10</td>
<td>7.7 7.7</td>
</tr>
</tbody>
</table>

and on the same order with both methods. For the between-subject variability parameters (\( \Omega \)), the biases are reduced with SAEM, while the RMSEs are of the same order with both methods. For the within-subject variability parameters (\( \Psi \)), the bias and the RMSEs are satisfactory and on the same order with both methods. The bias and RMSE for \( \sigma^2 \) are small and satisfactory for both methods.

The type I errors of the Wald test and the LRT are evaluated on the same 1000 data sets. For \( n = 24 \), the type I errors are 6.0% and 6.5% for SAEM and nlme, respectively, for the Wald test and 4.6% and 5.6% for SAEM and nlme, respectively, for the LRT. For \( n = 40 \), the type I errors are 5.6% and 5.4% for SAEM and nlme, respectively, for the Wald test and 5.8% and 5.2% for SAEM and nlme, respectively, for the LRT.

5. Application to the population PK of atazanavir with tenofovir

5.1 Study population: ANRS 107-Puzzle 2 study

The Puzzle 2-ANRS 107 trial was a randomized open-label, multiple-dose study supported by the French ANRS with HIV-infected patients in treatment failure with their antiretroviral therapy. Patients were randomized to receive for the first 2 weeks either their unchanged treatment with protease inhibitors (PIs) and nucleoside reverse transcriptase inhibitors (NRTIs) (group 1) or unchanged treatment with NRTIs in combination with atazanavir (300 mg once a day) plus ritonavir (100 mg once a day) as a substitute for the failing PI therapy (group 2). From week 3 (day 15) to week 26, patients from either group switched to atazanavir (300 mg once a day) plus ritonavir (100 mg once a day) plus tenofovir disoproxil fumarate at 300 mg once a day and NRTIs selected according to the baseline reverse transcriptase genotype of the HIV isolated in each patient.

In this paper, we analyze concentration data obtained from 10 patients from group 2 who were included and measured at 1, 2, 3, 5, 8, and 24 h after administration of the drug during each treatment period. The objective of the substudy was to measure the pharmacokinetic parameters of atazanavir (administered with ritonavir) either before (day 14 [week 2]) or after (day 42 [week 6]) initiation of tenofovir DF in HIV-infected patients in order to detect pharmacokinetic interactions of tenofovir on atazanavir. Data of
this substudy were analyzed using a NLMEM by Panhard and others (2007) and a significant effect of the coadministration of tenofovir on the pharmacokinetic parameters of atazanavir was found using the nlme function of the Splus software.

The aim of the present analysis is to evaluate the effect of tenofovir on the PK parameters of atazanavir using the SAEM algorithm and the Wald test described in Section 4.

We use the 1-compartment model with zero-order absorption proposed by Panhard and others (2007) to describe atazanavir concentrations:

\[ f(t, \phi) = \frac{FD}{T_a Cl} \left( (1 - e^{-\frac{Cl}{V}T_a}) I_{t<T_a} + \frac{e^{-\frac{Cl}{V}T_a} I_{t<T_a} (1 - e^{-\frac{Cl}{V}(T_a - t)})}{1 - e^{-\frac{Cl}{V}T_a}} \right), \]

with \( F \) the bioavailability, \( V \) the volume of distribution of atazanavir, \( T_a \) the absorption duration, \( Cl \) the elimination clearance, and \( \tau \) the dosing interval (24 h until the PK visit). The vector of the logarithm of the identifiable parameters is \( \phi = (\log(V/F), \log(T_a), \log(AUC)) \). Data of both treatment periods are simultaneously analyzed using a NLMEM with 2 levels of variability (the between-patient and the within-patient variabilities) on each PK parameter. A treatment effect is also estimated for each PK parameter, and a homoscedastic error model is used.

5.2 Results

Concentrations versus time are displayed in Figure 3. The SAEM algorithm succeeds in the estimation of all the parameters. The resulting parameter estimates are displayed in Table 2. The SAEM algorithm estimates the AUC between-patient variability and the \( V/F \) and \( T_a \) within-patient variabilities to 0.48, 0.69, and 0.19, respectively. The 3 other variances are estimated to be 0. A significant effect of comedication with tenofovir is found on \( \log(AUC) \) \( (p=0.00015) \) with the Wald test based on the SAEM algorithm.

Individual prediction curves for the 2 periods are overlaid on the concentration data in Figure 3. The goodness-of-fit plots (population and individual predicted concentrations versus observed concentrations; standardized residuals versus predicted concentrations and versus time) are judged satisfactory and are displayed in Figure 4.

6. Discussion

The main original element of this study is the development of the SAEM algorithm for 2-level NLMEMs. We have extended the SAEM algorithm developed by Kuhn and Lavielle (2005b) to the case of MNLMEMs with 2 levels of random effects. This algorithm will be implemented in the 3.1 version of the monolix software, freely available on the following Web site: http://www.monolix.org. The 2 levels of random effects are the between-subject variance and the within-subject (or between-unit) variance, with \( N \) subjects and \( K \) units, with no restriction on \( N \) or \( K \). We show that the SAEM algorithm is split into 2 parts: an explicit EM algorithm and a stochastic EM part. The integration of the term \( p(b|\phi; \theta) \) in the likelihood results in the derivation of 2 additional sufficient statistics compared to the original algorithm. Furthermore, it uses 2 intermediate quantities, the conditional expectations and variances of the between-subject random effects parameters \( b \). The addition of higher levels of variability would therefore require other extensions of the algorithm.

The convergence of the algorithm is monitored graphically, as shown in Figure 1. An automatic implementation of a stopping criterion to optimize both the number of iterations and the stochastic approximation step will be considered in future work.

The simulation study illustrates the accuracy of our approach. We show that the biases and RMSEs obtained by the extended SAEM algorithm are satisfactory for all parameters. The biases are reduced compared to those obtained with the FOCE algorithm implemented in the nlme function of the R software.
Furthermore, whereas the nlme implementation of the FOCE algorithm does not always converge with both between- and within-patient variability on all parameters, the extended SAEM algorithm does. Associated tests for a difference between the units achieve accurate type I error rates.

The analysis of the PK of atazanavir with tenofovir in the Puzzle 2-ANRS 107 trial also demonstrates the ability of the extended SAEM algorithm to detect treatment interaction on a real data set. When

### Table 2. Pharmacokinetic parameters of atazanavir (estimate and SE [%]) estimated with the SAEM

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Estimate</th>
<th>SE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\log(V/F)$ (L)</td>
<td>4.01</td>
<td>5.79</td>
</tr>
<tr>
<td>$\log(T_a)$ (h)</td>
<td>1.36</td>
<td>6.72</td>
</tr>
<tr>
<td>$\log(AUC)$ (ng mL$^{-1}$ h)</td>
<td>10.67</td>
<td>1.61</td>
</tr>
<tr>
<td>$\beta_{V/F}$</td>
<td>0.12</td>
<td>267.43</td>
</tr>
<tr>
<td>$\beta_{T_a}$</td>
<td>0.33</td>
<td>45.03</td>
</tr>
<tr>
<td>$\beta_{AUC}$</td>
<td>-0.38</td>
<td>25.31</td>
</tr>
<tr>
<td>$\omega_{V/F}$</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>$\omega_{T_a}$</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>$\omega_{AUC}$</td>
<td>0.48</td>
<td>25.48</td>
</tr>
<tr>
<td>$\psi_{V/F}$</td>
<td>0.55</td>
<td>28.30</td>
</tr>
<tr>
<td>$\psi_{T_a}$</td>
<td>0.16</td>
<td>35.76</td>
</tr>
<tr>
<td>$\psi_{AUC}$</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>$\sigma$ (ng mL$^{-1}$)</td>
<td>732.29</td>
<td>8.40</td>
</tr>
</tbody>
</table>

SE, standard error.
Fig. 4. Goodness-of-fit plots for atazanavir final population PK model: (a) population and (b) individual predicted concentrations (in ng/mL) versus observed concentrations (in ng/mL), standardized residuals versus (c) predicted concentrations (in ng/mL) and (d) time (in hours).

testing for an interaction of tenofovir on the PK of atazanavir, the impact of tenofovir on the absorption of atazanavir is confirmed; more precisely, a decrease of the AUC of atazanavir as shown by Panhard and others (2007) is found.

In addition to comparing our extension of SAEM to the FOCE algorithm, which is the most popular method in population PK, we tried to use the NLMIXED procedure of SAS implementing Gaussian quadrature. However, this procedure failed to estimate the parameters of such complex variance models, both on our simulated data and on the atazanavir data set.

The next ambitious development would be an extension of the algorithm to the case of MNLMEMs with more than 2 levels of random effects. However, it would be difficult to develop a general estimation method since this depends strongly on the relation (linear or not) of the different levels of random effects.

To conclude, the extended SAEM algorithm combines the statistical properties of an exact method together with a high computational efficiency. We thus recommend the use of this method in MNLMEMs.

ACKNOWLEDGMENTS

The authors are grateful to Marc Lavielle for his supportive advice and help. The authors would like to thank the principal investigator Dr Christophe Piketti, the pharmacology coordinator Dr Anne-Marie Taburet, and the scientific committee of the Puzzle 2-ANRS 107 trial for giving us access to the concentration measurements of the patients. Conflict of Interest: None declared.
Funding

The Agence Nationale de la Recherche (X.P.).

References


[Received 21 December 2007; first revision 19 February 2008; second revision 2 April 2008; accepted for publication 12 May 2008]