# SAEMIX, an R version of the SAEM algorithm for parameter estimation in nonlinear mixed effect models

Emmanuelle Comets[1,2] & Audrey Lavenu[2] & Marc Lavielle[3]

[1] *CIC 0203, U. Rennes-I, CHU Pontchaillou, 35000 Rennes, France*
[2] *IAME, UMR 1137, INSERM, 75018 Paris, France; Univ Paris Diderot, Sorbonne Paris Cité, 75018 Paris, France*
[3] *Inria Saclay, team Popix, U. Paris Sud Mathematics, France*

**Résumé.**
Les modèles non-linéaires à effets mixtes sont de plus en plus répandus pour l'analyse de données longitudinales. Ces données sont recueillies dans les études de pharmacocinétique et de pharmacodynamie lors du développement de médicaments, mais sont plus généralement disponibles dans les essais cliniques où est mesurée au cours du temps l'évolution de marqueurs de l'efficacité ou de la toxicité clinique. La disponibilité croissante de ce type de données a conduit au développement de nouvelles méthodes d'estimation. Dans ce travail, nous présentons la librairie SAEMIX développée pour le logiciel statistique `R`, et qui met en œuvre l'algorithme SAEM (Stochastic Approximation Expectation Maximization), un algorithme EM où l'étape E inclut des simulations dans la loi individuelle des paramètres. L'efficacité par rapport aux algorithmes reposant sur des linéarisations du modèle a été démontrée théoriquement et en pratique.

Cette librairie permet d'estimer les paramètres d'un modèle non-linéaire à effets mixtes écrit sous forme analytique. La vraisemblance, estimée par linéarisation, par approximation stochastique ou numérique, peut être utilisée pour comparer des modèles entre eux à l'aide du test du rapport de vraisemblance ou de critères statistiques. L'algorithme fournit également des estimations de l'incertitude des paramètres, et de nombreux graphes de diagnostic peuvent être tracés grâce à une fonction `plot` adaptée aux objets résultant de l'analyse. La librairie a été programmée avec les classes S4 de `R` pour une utilisation facile.

**Mots-clés.** Modèles non-linéaires à effets mixtes, estimation de paramètres, algorithme SAEM, R, librairie R, pharmacocinétique, pharmacodynamie, données longitudinales.

**Abstract.** Nonlinear mixed effect models are being increasingly used and expanded to handle longitudinal data collected in clinical trials, prompting the development of new estimation methods. The Stochastic Approximation Expectation Maximization (SAEM) algorithm has proven very efficient, quickly converging to the maximum likelihood estimators and performing better than linearisation-based algorithms. It has been implemented

in the Monolix software which has enjoyed increasingly widespread use over the last few years, more recently in the Statistics toolbox of Matlab (nlmefitsa.m), and is also available in NONMEM version 7.

In the present paper, we describe the SAEMIX package, which implements SAEM in the R statistical software. This package performs parameter estimation in non-linear mixed effect models, providing estimates of uncertainty for each parameter. The likelihood can be computed using model linearisation or through stochastic or numeric approximations. Models can then be compared through likelihood ratio tests or statistical criteria. The SAEMIX package was programmed using the S4 class system to enhance user-friendliness, with a wide range of diagnostic plots that can help model assessment.

**Keywords.** Nonlinear mixed effect models, parameter estimation, SAEM algorithm, R, R package, pharmacokinetics, pharmacodynamics, longitudinal data.

## Introduction

The use of modelling and simulation in clinical drug development is now well established. Regardless of whether a single outcome is considered at the end of the study, clinical trials often collect longitudinal data, with each subject providing several measurements during the course of the study. Longitudinal data is a staple in particular of pharmacokinetic (PK) and pharmacodynamic (PD) studies, which are a required part of a new drug application file. Nonlinear mixed effect models can help to characterise and to understand many complex nonlinear biological processes, such as biomarkers or surrogate endpoints, and are crucial in describing and quantifying the mechanisms of drug action and the different sources of variation, e.g., the interindividual variability [1].

In the context of maximum likelihood, a dedicated software, called `NONMEM`, was developed in the 70's, which handles specific characteristics of pharmacologic data such as dosage regimen and other variables measured during treatment ([2]). The first algorithms implemented in this software relied on model linearisation to obtain an approximation of the likelihood, which cannot be computed easily in nonlinear mixed effect models, and the estimation algorithm operates through Newton-Raphson minimisation, an iterative algorithm using the gradient of the function to minimise. Linearisation-based methods however have statistical and practical shortcomings, and have namely been shown to increase the type I error of likelihood tests [3, 4]. Over the past decade, new and powerful estimation algorithms have therefore been proposed to estimate the parameters of these models. The Stochastic Approximation Expectation Maximization (SAEM) algorithm has proven very efficient, quickly converging to the maximum likelihood estimators [5] and performing better than linearisation-based algorithms [6]. It has been implemented in the Monolix software [7] which has enjoyed increasingly widespread use over the last few years, more recently in the Statistics toolbox of Matlab (nlmefitsa.m), and is also available in NONMEM version 7 [2].

The objective of the present package was to implement SAEM in the R software [8].

## Methods

Detailed and complete presentations of the nonlinear mixed effects model can be found in several reference textbooks, for instance [9]. We consider the following general nonlinear mixed effects model for continuous outputs:

$$y_{ij} = f(x_{ij}, \psi_i) + g(x_{ij}, \psi_i, \xi)\varepsilon_{ij} \quad , \; 1 \leq i \leq N \quad , \quad 1 \leq j \leq n_i \tag{1}$$

where $y_{ij}$ is the $j$th observation of subject $i$, $N$ is the number of subjects, $n_i$ is the number of observations of subject $i$; the regression variables, or design variables, $(x_{ij})$ are assumed to be known. We further assume that for subject $i$, the vector $\psi_i$ is a vector of $n_\psi$ individual parameters, and is a function of an unknown vector of fixed effects $\mu$, an unknown vector of normally distributed random effects $\eta_i$, and possibly of individual covariates $c_i$. The within-group errors $(\varepsilon_{ij})$ are supposed to be Gaussian random variables with mean zero and variance 1. Furthermore, we suppose that the $\varepsilon_{ij}$ and the $\eta_i$ are mutually independent. Different error models for $g$ can be used in SAEMIX.

The SAEM algorithm is used to obtain maximum likelihood estimates of the parameters of nonlinear mixed effects models without any linearisation of the model. The log-likelihood for nonlinear mixed effect models is analytically intractable since it requires integration over the unknown individual parameters. The SAEM algorithm uses an EM algorithm [10], where the unknown individual parameters are treated as missing data, and replaces the usual E-step with a stochastic approximation step [11]. The missing parameters are simulated at each iteration via a MCMC procedure, which can be used after the algorithm has converged to obtain the conditional modes, the conditional means and the conditional standard deviations of the individual parameters.

## Package overview

The library uses the S4 class system of R to provide a user-friendly input and output system through object-oriented programming.The main function in the SAEMIX package is the `saemix()` function, which estimates the population parameters of a nonlinear mixed effect model. This function requires two mandatory arguments, a saemixModel object defining the models (both structural and statistical), and the data, and a saemixData object created by the `saemixData()` function. An optional list of settings can be passed on as a list and can be used to set the directory in which to save data and graphs, as well as to tune the algorithm settings.

Functions like `summary` or `plot` have been developed and apply to fitted objects. The package provides summaries of the results, individual parameter estimates, standard errors (obtained using a linearised computation of the Fisher information matrix) Wald tests for fixed effects, and a number of diagnostic plots, including VPC plots and npde [12]. The log-likelihood can be computed by three methods: a linearisation of the model, an importance sampling procedure, or a Gaussian quadrature. The diagnostic graphs can

be tailored to the user's individual preferences by setting a number of options, and are easily exported to a file.

## Application

We illustrate the use of the library with the well known PK dataset of theophylline. The dataset includes the concentration versus time data collected in 12 subjects given a single oral dose of theophylline, and for whom 11 blood samples were collected over a period of 24 h. We modelled this data using a one-compartment model with first-order absorption, parameterised as $k_a$, V, CL. The IIV was modelled using an exponential model with diagonal variance-covariance matrix, while the residual variability was modelled with a combined error model. Many diagnostic plots are available to evaluate convergence or model adequacy, such as individual plots. They can be accessed through options given to the `plot` function, and tailored to change specific features such as colours or axis scales.

The SAEMIX package comes with both in-line help and an extensive documentation. It provides several examples in addition to the theophylline PK example. Two examples are taken from agronomy studies, and describe the weight gain of cows and the yield of winter wheat in farms. The performance of SAEMIX was evaluated in a third example, on simulated data in a pharmacodynamic context, which showed a good performance of the method, with the ability to detect a treatment effect and precise parameter estimates. The final example in the documentation shows an application of SAEMIX to linear data, using height data collected in children over a period of two years.

## Conclusion

SAEMIX has some limitations in the models that can be set up. It is limited to structural models with closed-form solutions, although modifying the model function to call an ODE solver may be done. Covariates must enter the parameter model linearly (after one of the available transformations), and there is no automated treatment of categorical covariates, although this can be circumvented by users creating sets of binary covariates. In addition, for the moment SAEMIX can only handle one response

Within those limitations, the SAEMIX package, designed to estimate the parameters in non-linear mixed effect models, adds a powerful algorithm to the toolbox already present in `R`. As such, it offers a complement to more mainstream packages such as nlme [13] and lme4. It has many potential applications for the analysis of longitudinal data, which are encountered for instance in agronomy, chemistry, medical trials and of course pharmacokinetics and pharmacodynamics.

# References


[1] Ette, E. and Williams, P. (2007), *Pharmacometrics: the science of quantitative pharmacology*, Wiley-Interscience, Hoboken, New Jersey.

[2] Boeckmann, A., Sheiner, L. and Beal, S. (1994), *NONMEM Users Guide: Part V*, University of California, NONMEM Project Group, San Francisco.

[3] Comets, E. and Mentré, F. (2001), Evaluation of tests based on individual versus population modelling to compare dissolution curves, *J Biopharm Stat*, 11, 107–123.

[4] Bertrand, J., Comets, E. and Mentré, F. (2008), Comparison of model-based tests and selection strategies to detect genetic polymorphisms influencing pharmacokinetic parameters, *Journal of Biopharmaceutical Statistics*, 18, 1084–102.

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

[6] Girard, P. and Mentré, F. (2005), A comparison of estimation methods in nonlinear mixed effects models using a blind analysis, *$14^{th}$ meeting of the Population Approach Group in Europe, Pamplona, Spain.*

[7] Lavielle, M. (2013), *MONOLIX (MOdèles NOn LInéaires à effets miXtes)*, INRIA, Orsay, France, URL `http://www.lixoft.eu/wp-content/resources/docs/UsersGuide.pdf`.

[8] R Development Core Team (2006), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, URL `http://www.R-project.org`, ISBN 3-900051-07-0.

[9] Davidian, M. and Giltinan, D. (1995), *Nonlinear models for repeated measurement data*, Chapman & Hall, London.

[10] Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977), Maximum likelihood from incomplete data via the EM algorithm, *J. Roy. Statist. Soc. Ser. B*, 39, 1–38, with discussion.

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

[12] Brendel, K., Comets, E., Laffont, C., Laveille, C. and Mentré, F. (2006), Metrics for external model evaluation with an application to the population pharmacokinetics of gliclazide, *Pharmaceutical Research*, 23, 2036–49.

[13] Pinheiro, J. and Bates, D. (1995), Approximations to the log-likelihood function in the non-linear mixed-effect models, *Journal of Computational and Graphical Statistics*, 4, 12–35.