NIMROD: A Program for Inference via Normal Approximation of the Posterior in Models with Random effects based on Ordinary Differential Equations.

Mélanie Prague[a,b,*], Daniel Commenges[a,b], Jérémie Guedj[c], Julia Drylewicz[d], Rodolphe Thiébaut[a,b]

[a] *University of Bordeaux, ISPED, F-33000 Bordeaux, France*
[b] *INSERM, Centre INSERM U897-Épidémiologie-Biostatistique, Bordeaux, France*
[c] *Theoretical Biology and Biophysics, Los Alamos National Laboratory, NM87545, USA*
[d] *Department of Immunology, University Medical Center Utrecht, The Netherlands*

Abstract

Models based on ordinary differential equations (ODE) are a widespread tool to describe dynamical system. In biomedical sciences, data within each subject can be sparse but information is often gained from between-subjects variability. This makes natural the use of mixed effect models to estimate population parameters. Although maximum likelihood based approaches are a valuable option, both numerical and identifiability issues favour a Bayesian approach which can incorporate prior knowledge in a flexible way. However the combination of difficulties coming from the ODE system and from the presence of random effects raises a major numerical challenge. A normal approximation of the posterior can be obtained by computing the maximum of the posterior distribution (MAP). Here we present the NIMROD (Normal approximation Inference in Models with Random effects based on Ordinary Differential equations), a program devoted to the MAP estimation in ODE

*melanie.prague@isped.u-bordeaux2.fr

models. We describe the specific implemented features such as a convergence criterion called the relative distance to maximum (RDM) and an approximation of the leave-one-out cross validation. We illustrate the approach and the program by analysing a pharmacokinetics example, first in simulations, then on data from the PUZZLE clinical trial in HIV infected patients.
Keywords: Bayes; Convergence criteria; Maximum likelihood; Maximum *a posteriori*; HIV; Non-linear Mixed Effects model; Ordinary Differential Equations ; Pharmacokinetics; Program; Optimization.

---

1. Introduction

In biomedical sciences, the increasing body of data makes crucial the development of adequate models that naturally capture the interactions between biomarkers. When the interactions are dynamic and may change with time or concentrations, they can be naturally captured by models based on a system of Ordinary Differential Equations (ODE). As these models not only aim at mimicking the data but also at understanding the underlying biological mechanisms, they are often called mechanistic or semi-mechanistic models. The use of ODE is particularly popular in pharmacometrics to describe the drug pharmacokinetics (PK) in plasma (Wakefield and Racine-Poon, 1995) or to analyse viral kinetics during treatment. For instance, the seminal papers of Perelson et al. (1996) contributed to a better understanding of HIV pathogenesis by estimating the daily viral production of HIV in infected individuals.

In the clinical setting, studies are made on a relatively large number of patients, but technical and ethical reasons often limit the number of sam-

pling measurements that can be done within each patient. In this context, non-linear mixed effect models are a better way than subject-by-subject analysis to handle the available information and to estimate the mean and the variance of the model parameters in the study samples (Verbeke and Molenberghs, 2009). Inference for these models can be done by either frequentist or Bayesian approaches. Many algorithms have been suggested for frequentist approaches. A large body of literature suggests that approaches based on an "exact" (more exactly, the numerical) likelihood provide better results than methods based on approximation of the likelihood (Ding and Wu, 2001). Among methods based on an exact likelihood, those based on Gaussian quadratures tend to be the most precise but also the most time consuming (Plan et al., 2012). We recently suggested an algorithm based on Gaussian quadratures specially adapted for biological models defined by ODE (Guedj et al., 2007a), but its use was limited by the lack of an interface user.

On the other hand the large number of parameters in these mechanistic models often results in a lack of identifiability (Guedj et al., 2007b), which motivates the use of a Bayesian approach. However non-linear mixed effect models may become numerically challenging when the underlying biological model is defined by non-linear ODE which may not have an analytical solution. In this situation the conventional MCMC algorithm (such as WinBugs) for computing the marginal posterior distributions of the parameters is time-consuming and may even fail in complicated cases (Drylewicz et al., 2012a). Other Bayesian methods for approximation of the posterior distributions, such as the Integrated Nested Laplace Approximation (INLA) method (Rue

et al., 2009) cannot be applied here because ODE do not have an analytical solution.

An easier way to use Bayesian approaches is to assume a normal approximation of the posterior (NAP). This is justified asymptotically by the Bernstein-Von Mises theorem (Van der Vaart, 2000). Numerically, this amounts to compute the maximum of the posterior distribution (MAP) while the variance matrix is approximated by the inverse of the Hessian of minus the logarithm of the posterior. This approach was shown to work well in situation where MCMC may fail Drylewicz et al. (2012a).

Here we detail the implementation and the improvement of this algorithm called NIMROD (Normal approximation Inference in Models with Random effects based on Ordinary Differential equations) which can be used for either maximum likelihood or Bayesian inference by a normal approximation of the posteriors. We show some new statistical and numerical features of the algorithm, such as the properties of the stopping criterion and the optimization of the algorithm for parallel computing, respectively. The paper is organized as follows. In Section 2 a statistical model based on an ODE system in a general form is described. Section 3 presents the algorithm and its statistical properties. In Section 4, the NIMROD program is introduced in a simplified user-manual style. Section 5 presents the analysis of simulated data and of data coming from a clinical trial. Section 6 concludes.

2. Model and inference

2.1. General model for the system

Let us consider an ODE model for a population of n subjects. For subject i, with $i = 1 \dots n$, this can be written:

$$\begin{cases} \frac{dX^i(t)}{dt} = f(X^i(t), \xi^i(t)) \\ X^i(0) = h(\xi^i(t)) \end{cases}$$

where $\xi^i(t) = (\xi_1^i(t), ..., \xi_p^i(t))$ is a vector of p individual parameters which appear naturally in the ODE system and have a biological interpretation; $X^i(t) = (X_1^i(t), \dots, X_K^i(t))$ is the vector of the K state variables (or compartments). We let $X(t, \xi^i(t)) = X^i(t)$ to underline that $\xi_i(t)$ completely determines the trajectories $X^i(t)$. We assume that f and h are possibly non-linear functions, twice differentiable with respect to $\xi^i(t)$.

Reparametrization of the system allows us to take constraints into account: we introduce one-to-one functions $\psi_l(.)$, $l = 1 \dots p$ and defined transformed parameters $\tilde{\xi}_l^i(t) = \psi_l(\xi_l^i(t))$. For instance, biological parameters such as rates can be parametrized using a logarithmic transformation to ensure positivity, or parameters between 0 and 1 (such as probabilities or bioavailability) can be parametrized using a logistic transformation. A mixed effect model for the $\tilde{\xi}_l^i(t)$ allows introducing covariates and taking into account the between-subject variations. In this approach, random effects give each subject a different value for a subset $q \leq p$ of the biological parameters:

$$\tilde{\xi}_l^{\,i}(t) = \phi_l + \beta_l z_l^i(t) + u^i, \qquad u^i \sim \mathcal{N}\left(0, \sigma_l^2\right).$$

where ϕ_l is the intercept and z_l^i is a vector of n_e (possibly time-dependent) explanatory variables associated with the fixed effects of the l^{th} biological

parameter; β_l is a vector of regression coefficients; u^i is the individual vector of random effects with $\sigma_l^2 = 0$ for $l > q$.

In practice, we do not observe $X^i(t)$ directly, but we have discrete-time observations $Y^i(t_{ij})$ of some functions of $X^i(t)$. We assume that there are known link functions $g_m(.)$, $m = 1, \ldots, M$, leading to an additive measurement error model. For $i = 1, \ldots, n$, $m = 1, \ldots, M$ and $j = 1, \ldots T_i$ (the number of observation time for subject i), we observe:

$$Y_m^i(t_{ij}) = g_m(X^i(t_{ij})) + \epsilon_{ijm}, \qquad \epsilon_{ijm} \sim \mathcal{N}\left(0, \sigma_m^2\right).$$

This observation scheme can be complicated by left-censoring due to detection limits (Jacqmin-Gadda et al., 2000). We denote by θ the parameter vector to estimate of length $N = p + n_e + q + M$:

$$\theta = \left((\phi_l)_{l=1\ldots p}, (\beta_l)_{l=1\ldots n_e}, (\sigma_l)_{l=1\ldots q}, (\sigma_m)_{m=1\ldots M}\right).$$

2.2. Likelihood and NAP

The likelihood formula for the model and observations described in section 2.1 can be found in Guedj et al. (2007a). The individual likelihood given the random effects ($\mathcal{L}^{\theta}_{\mathcal{F}_i|u_i}$) is easily computed as a function of $g_m(X^i(t_{ij}))$ since the $Y_m^i(t_{ij})$ are independent Gaussian variables. However, the computation of the $g_m(X^i(t_{ij}))$'s requires to solve the ODE system. This is done by using the Livermore solver DLSODE (Hindmarsh, 1983), specially adapted for stiff systems using backward difference formula and gear type method BDF from Radhakrishnan et al. (1993). Then, the likelihood ($\mathcal{L}_i^{\theta}$) is computed by integrating over the random effects via the adaptive Gaussian quadrature (implemented as in Genz and Keister (1996)) as proposed by Pinheiro and Bates (2000).

Bayes therorem gives:

$$\log[P(\theta|Y)] = L(\theta) + \log[\pi(\theta)] + C,$$

where C is the normalization constant, $P(\theta|Y)$ the posterior distribution, $L(\theta)$ the log-likelihood and $\pi(\theta)$ the prior distribution. The normal approximation of the posterior is obtained by maximizing the penalized log-likelihood $L^P(\theta) = L(\theta) - J(\theta)$ (which does not involve the difficult normalization constant C) and computing its Hessian at the maximum value; here, $J(\theta)$ is a penalization term equal to $-\log[\pi(\theta)]$ up to a constant. If we assume normal independent priors for the fixed effects, half-Cauchy priors for the variances of the random effects, as recommended by Gelman (2006) with median parameter ${s^j}^2$ and conventional one-dimension Jeffreys-type improper priors for the variances of the measurement errors, the penalization function $J(\theta)$ can be written:

$$\begin{aligned} J(\theta) &= \sum_{j=1}^{p} \frac{[\phi_j - \mathrm{E}^0(\phi_j)]^2}{2\mathrm{var}^0(\phi_j)} + \sum_{j=1}^{n_e} \frac{[\beta_j - \mathrm{E}^0(\beta_j)]^2}{2\mathrm{var}^0(\beta_j)} \\ &\quad + \sum_{j=1}^{q} \log({\sigma^j}^2 + {s^j}^2) + \sum_{j=1}^{m} \log(\sigma_m), \end{aligned}$$

where E^0 and var^0 are the expectation and the variance under the prior.

2.3. Credible sets and Bayesian p-values

Parameter significance is often a major issue in biostatistics, for instance to conclude about treatment effects. In a frequentist approach, this is done by testing "$\beta = 0$" or "$\beta \leq 0$" (resp. "$\beta \geq 0$"). In a Bayesian approach, once we have the marginal posterior distributions of the parameters, it is

straightforward to compute $(1-\alpha)$-credible sets of highest probability density (HPD-sets, Berger (1985), part 4.3.2).

We implemented a Bayesian p-value as an analogue to the frequentist one. For the one-sided test "$\beta \leq 0$", it is simply $\mathbb{P}(\beta \leq 0|Y)$. For the two-sided Bayesian p-value testing "$\beta = 0$", called α_{min}, it is such that the largest HPD credible set not containing zero has probability $1 - \alpha_{min}$. This can be computed as 2 $\min[\mathbb{P}(\beta \leq 0|\mathrm{Y}), \mathbb{P}(\beta \geq 0|\mathrm{Y}))]$. If all priors are flat, this Bayesian p-value is consistent with the frequentist p-value obtained from a two-sided Wald test. Moreover, as credible sets and confidence intervals of the same coverage are asymptotically equivalent, this Bayesian p-value is also asymptotically consistent with the frequentist p-value for any prior.

2.4. Model choice

Commenges et al. (2007) proposed an approximate cross-validation criterion that can be used to choose estimators based on penalized likelihood; see also Commenges et al. (2012) which generalizes this approach. Taking advantage of the fact that we maximize a penalized likelihood, we propose to use this criterion for model choice:

$$\mathrm{LCV_a} = -\mathrm{n}^{-1}[\mathrm{L}(\hat{\theta}) - \mathrm{Tr}(\mathrm{H_{L^P}^{-1}H_L})],$$

where H_{L} and H_{L^P} are the Hessians of minus the log-likelihood and the penalized log-likelihood respectively, taken in $\hat{\theta}$. The criterion is implemented in NIMROD with H_{L} and H_{L^P} approximated as indicated in section 3.1. The lower the criterion, the better the model. Differences of criteria between two models are especially meaningful: differences larger that 0.1 are considered

as "large" while differences lower than 0.001 are considered as negligible (Commenges et al., 2008).

3. Algorithm

3.1. Optimization procedure

Under the conditions of the Bernstein-Von Mises theorem, the posterior tends to a normal, thus, when n is large enough the penalized log-likelihood is close to a quadratic form in θ. The use of the Newton-Raphson method is justified as this method is based on a quadratic approximation of the surface to maximize. However with n not very large, the surface may be close to a quadric only in a region near the maximum and may not be convex everywhere. More robust methods such as the Levenberg-Marquardt algorithm (Marquardt, 1963) are called for. Moreover, in ODE mixed-effects models the likelihood is difficult to compute: we need to solve the ODE system and numerically compute multiple integrals; on the top of that, both Newton-Raphson and Levenberg-Marquardt algorithms require first and second derivatives of the objective function which also have to be numerically computed. To decrease the computation time we use an approximation of the second derivative of the log-likelihood leading to the Robust Variance Scoring (RVS) Algorithm (Commenges et al., 2006), an improved version of the BHHH algorithm (Berndt et al., 1974).

The individual score for a given value of the parameter θ, $(U_i(\theta))$ is calculated using Louis' formula (Louis, 1982) and the systems of sensitivity equations (given by $\left(\frac{df(X^i(t),\xi^i(t))}{d\xi_l^i(t)}\right)_{l=1\ldots p}$) as described in Guedj et al. (2007a). Then, the penalized scores $(U_i^P(\theta))$ are derived such that $U_i^P(\theta) =$

$U_i(\theta) - (1/n)\frac{\partial J(\theta)}{\partial \theta}$. The observed log-likelihood and scores are calculated as the sum over all the subjects: $L^P(\theta) = \sum_{i=1}^n L_i^P(\theta)$, $U(\theta) = \sum_{i=1}^n U_i(\theta)$, $U^P(\theta) = \sum_{i=1}^n U_i^P(\theta)$. The Hessian of $-L^P(\theta)$, denoted $H_{L^P}(\theta)$, is approximated by an estimator of the variance of L plus the second derivative of the penalization J:

$$G(\theta) = \sum_{i=1}^{n} U_i(\theta)U_i^T(\theta) - \frac{1}{n}U(\theta)U^T(\theta) + \frac{\partial^2 J(\theta)}{\partial \theta^2}.$$

For θ close to the true value and n large, $\sum_{i=}^n U_i(\theta)U_i^T(\theta) - \frac{1}{n}U(\theta)U^T(\theta)$ approximates $H_{L}(\theta)$ and $G(\theta)$ approximates $H_{L^P}(\theta)$, the Hessians of $-L$ and $-L^P$ respectively.

The RVS Algorithm with the calculations described before is implemented in NIMROD with an optional automatic switch to a classical Levenberg-Marquardt algorithm when the algorithm gets stuck, i.e. between two iteration steps there is neither log-likelihood change nor parameters movement whereas the main convergence criterion based on RDM (see section 3.2) is not met yet. Finally, a line search algorithm, such as in Potra and Shi (1995), is used to adjust the displacement when the log-likelihood is not improved at the first try. Despite of all these tricks, it may happen that the log-likelihood can not be improved, possibly due to too large errors in the computation of the scores. In that case, the displacement is done toward the highest proposed log-likelihood.

3.2. Convergence criteria

Any iterative algorithm must have a stopping rule; a good stopping rule stops the algorithm when the current parameter value is close enough to the argument of the maximum, $\hat{\theta}$. Stopping rules are often of the type: "convergence

criterion below a threshold". Two commonly used criteria assess the stability of the algorithm concerning the displacement in the parameter space $||\theta_{k+1} - \theta_k||$ and the variation of the objective function, here: $L^P(\theta_{k+1}) - L^P(\theta_k)$. If both of these criteria take very low values, this means that the algorithm does not move any more. This is what we expect if θ_k is very close to $\hat{\theta}$. These stopping rules are not really satisfactory. In the first place, it is not easy to fix good thresholds called respectively η_1 and η_2 (at least for the first criterion). More importantly, these stopping rules are not directly linked to the properties of the surface to maximize. If the algorithm does not move, this is not a proof that it has reached the maximum: it may just get stuck and be unable to find a good direction to improve the objective function.

If there is a unique maximum (which is justified by the Bernstein-Von Mises theorem), it is characterized by $U^P(\hat{\theta}) = 0$. Thus, a norm of the gradient will be a good candidate as convergence criterion. Commenges et al. (2006) proposed a criterion based on the metric G^{-1} that we call the Relative Distance to Maximum (RDM):

$$\mathrm{RDM}(\theta_\mathrm{k}) = \frac{\mathrm{U^P}(\theta_\mathrm{k})^\mathrm{T}\mathrm{G}^{-1}(\theta_\mathrm{k})\mathrm{U^P}(\theta_\mathrm{k})}{\mathrm{m}}$$

The name "RDM" can be justified by noting that near the maximum we have $U^P(\theta_k)^T G^{-1}(\theta_k) U^P(\theta_k) \approx (\theta_k - \hat{\theta})^T H_{L^P}(\theta_k)(\theta_k - \hat{\theta})$, that is a distance between θ_k and $\hat{\theta}$ (this comes from a Taylor expansion of $\frac{\partial L^P(\theta)}{\partial \theta}$ around $\hat{\theta}$). Using the same metric, by neglecting the effect of penalization, we have $\mathrm{E}_*[(\hat{\theta} - \theta_*)^T H_{L^P}(\theta_*)(\hat{\theta} - \theta_*)] \approx \mathrm{E}_*[(\hat{\theta} - \theta_*)^T I(\hat{\theta} - \theta_*)] = m$, where I is the information matrix. Thus, RDM is approximately the ratio of the distance of θ_k from $\hat{\theta}$ over the expectation of the distance between $\hat{\theta}$ and θ_*. It can also be interpreted as the ratio of the numerical error over the statistical

error. This criterion is asymptotically invariant near the maximum to any one-to-one transformation of the parameters.

Thanks to its interpretation, the threshold (called η_3) can be chosen in a way that does not depend on the problem or on the data. It should clearly be lower than 1 and as close as possible to 0. RDM is an indicator of the number of significant digits that are obtained. See section 5.1 for indications about the choice of this threshold, 0.1 is a good default value. Anyway, we recommend to run the algorithm from several starting points to ensure reproducibility and assess the number of significant digits that have been obtained. For a secure stopping rule, we may combine the three criteria, although RDM is clearly the best criterion.

In the case where G is not inversible, RDM can not be computed and the program will stop only when the maximum number of iterations is reached; this case is a failure of convergence.

4. The NIMROD Program

4.1. Packaging

The program is written in Fortran 90 and comes with a Maple routine which allows the user to calculate derivations and sensitivity equations for ODE system. An open source code is available and can be modified with external accessible subroutines. The code is organized in two levels: a user level that has to be modified to specify a new model, and an optimisation part that can be used as a black box for new users. Nevertheless, it is freely modifiable if necessary.

4.2. Parallel computing

Depending on the number of random effects, number of subjects and number of parameters, the optimisation can take from few minutes to several days. Parallel computing offers the possibility of speeding the program. Thus, we took advantage of the fact that, in this algorithm, calculations presented in 3.1 are highly parallel: likelihood and scores are first calculated subject-by-subject, then aggregated. We implemented parallel computing over the subjects of the study. Thus, the program can efficiently use $n_p \leq n$ processors. Our experience shows that the speed-up approximately follows the power law of $n_p^{-0.85}$. As an example, treating a non-linear model of HIV dynamics similar to that used in Prague et al. (2012) with 3 random effects, 149 subjects and 17 parameters takes about 7h30. Using 149 processors for this problem allows to reduce the time to a value that can be approximated by the power law formula to be $7.5 \times 60 \times 149^{-0.85} = 6.4$ minutes, and this is the order of magnitude that is observed in practice.

Technically, MPI (Message passing Interface) is a parallel distributed computing standard consisting of a library of functions and macros for use in programs that exploit the existence of multiple processors (Pacheco, 1997). It is a freely-available internationally recognized standard. It can work on existing networks of workstations on processors with or without shared memory. Even if a super-computer is not available, one can still efficiently use the twelve cores that are available on standard workstations to obtain a speed-up of more than 7.

4.3. Requirements and abilities

The user must specify the number of parameters, the ODE system, the ODE sensitivity systems, the parameters transformations, the statistical model and the observational model. Structurally, there is no limitation in the number of states and random effects. However, if the model complexity is too high the program can either be time consuming or fail to converge. In our experience, 5 to 6 equations in the system can be handled without problem but the number of random effects must be limited to the case $l \leq 5$. Drylewicz et al. (2010a) proposed a forward selection of the random effects.

Penalization as described in section 2.2 can be set on all parameters or biological parameters only. We recommend to use the smallest penalization first and then move to a heavier one if necessary. The user can also easily implement his own penalizations. The program can handle censored data. A left-censoring threshold can be supplied, indicating that observations below this value are censored.

NIMROD can operate two ways: inference or predictions. Concerning inference there are several algorithm options: foremost, the user must set the optimization algorithm by choosing either the RVS only, or Marquardt only or a combination of both. Maximum number of iteration (default is 1000), starting points (default are priors means) and convergence threshold (default is $\eta_1 = \eta_2 = \eta_3 = 0.1$) have to be set. Predictions (see Web-Appendix C for details) with Parametric Empirical Bayes (Kass and Steffey, 1989) are automatically done after the inference with the stopping points as input values.

Other details on NIMROD implementation (subroutines, functions, fea-

tures), utilization (how to ...) and running methods (compilation and parallel computing advices) are available in a user manual coming with the source code.

5. Illustration in pharmacokinetics (PK)

NIMROD can be used in complex non-linear ODE systems without analytic solution. It was first developed for models of HIV dynamics (Prague et al., 2012; Drylewicz et al., 2010b, 2012b). In illustration, for the simulations we use a simple one-compartment PK model with absorption, and for the application a two-compartments model with absorption. In all these models, there were three random effects.

5.1. Simulation: One-compartment pharmacokinetics model with oral absorption

We consider a one-compartment PK model with oral absorption (Web-Appendix A, Figure 1 - notation from Rosenbaum (2011) p. 185). The amount of drug in the gastrointestinal compartment ($\mathrm{A_{GI}}$) is described by a first-order drug absorption equation (Equation 1a); the concentration of drug in the plasma ($\mathrm{C_P}$) at any time depends on the relative rate of drug absorption and elimination (Equation 1b):

$$\begin{cases} \frac{d\mathrm{A_{GI}}}{dt} &= -k_a \mathrm{A_{GI}}, & (a) \\ \frac{d\mathrm{C_P}}{dt} &= \frac{k_a}{V_0}\mathrm{A_{GI}} - \mathrm{k_e C_P}. & (b) \end{cases} \tag{1}$$

We neglected bio-availability (F) and salt factor (S), thus we took them equal to one. Initial values are given by $\mathrm{A_{GI}}(0) = \text{dose}$ and $\mathrm{C_P}(0) = 0$.

Random effects were associated to all parameters but no covariate was introduced. We simulated data for 40 subjects with this one-compartment PK model for a unique initial dose of 100 mg by oral absorption. Parameter values are presented Table 1 (Rosenbaum (2011), chapter 9). Only the plasma concentration $\mathrm{C_P}$ was supposed observed at 10 time points (baseline, 0.5, 1, 1.5, 2, 3, 4, 6, 8 and 10 hours). The observational model for estimations, with a standard deviation measurement error of $\sigma_{\mathrm{C_P}} = 0.3$, was:

$$Y_1^i(t_{ij}) = (\mathrm{C_P^i(t_{ij})})^{0.25} + \epsilon_{\mathrm{ij1}}, \qquad \epsilon_{\mathrm{ij1}} \sim \mathcal{N}\left(0, \sigma^2_{\mathrm{C_P}}\right).$$

Concerning penalization, first, we chose weakly informative priors (Table 1) for inference on biological parameters. Convergence criteria were $\eta_1 = \eta_2 = 10^{-5}$ and $\eta_3 = 10^{-2}$. We ran the NIMROD algorithm with starting points taken at the mean of the priors. Convergence was reached after 23 iterations which took 97 seconds (parallel computing on 40 cores). However, starting the algorithm from different points led to two local maximums. Actually, in a one-compartment PK model with absorption, k_a and k_e are mathematically exchangeable without impacting the trajectories, this is called the *flip-flop* paradox (Godfrey et al., 1980). Thus, we defined well tight priors (less than one standard deviation far from the true value) and a tighter prior for k_a to reach practical identifiability. With the same running conditions, convergence was reached after 17 iterations which took 64 seconds (parallel computing on 40 cores). The obtained values were rather satisfactory, with on average two significant digits (see Table 1).

We evaluated our convergence criterion (RDM) behaviors during optimization (figure 1). The RDM decreases when the log-likelihood improves.

A zoom of figure 1 for the last iterations with a better scale and parameters trajectories over iterations are illustrated in Web-Appendix B.

To evaluate the link between the RDM and reproducibility, we randomly generated 1000 starting points at 2 standard deviations from the prior mean. We ran NIMROD with a unique convergence criterion: $\eta_3 < 1.0$. Convergence was reached for 97% of the starting points, otherwise program was stopped after 1500 iterations. The final RDM value is necessarily smaller than 1.0, but takes by chance different values. The final RDM value correlates with the final log-likelihood (see Figure 2a). At the stopping value, the variability of the log-likelihood values is smaller when the RDM is close to zero. We ran NIMROD again with $\eta_3 < 0.01$ as a convergence criterion from the same 1000 starting points sample. This led to 96.3% of convergence. Compared to the previous convergence criterion, we observed less variability in the final log-likelihood values (figure 2b).

5.2. Real data: Amprenavir in HIV infected patients

We used a one then a two-compartments with absorption PK model to fit data from the PUZZLE ANRS 104 clinical trial (Raguin et al., 2004). Plasma concentrations of Amprenavir (APV) were measured in 39 patients after multiple oral doses either at week 2 - day 15 or at week 6 - day 45. There is no left-censoring, all observations were greater than 40 ng/mL. Blood samples were drawn in the morning before the scheduled 600 mg drug intake BID (at time 0) and 0.5, 1, 1.5, 2, 3, 4, 6, 8 and 10 hours after the intake. Somehow, the design was the same as the one presented in section 5.1, but, the plasma concentration evaluation at $t = 0$ is complicated by multiple oral doses taken before the PK assay. Let us call τ (=12 hours)

the time between two doses and define the minimum steady-state plasma concentration. We assume that it is the baseline value for the patient:

$$\mathrm{C_P}(0) = \frac{\mathrm{k_a dose}\left(\frac{\mathrm{e}^{-\mathrm{k_e}\tau}}{1-\mathrm{e}^{-\mathrm{k_e}\tau}} - \frac{\mathrm{e}^{-\mathrm{k_a}\tau}}{1-\mathrm{e}^{-\mathrm{k_a}\tau}}\right)}{(\mathrm{k_a} - \mathrm{k_e})\mathrm{V_0}}.$$

We fitted the data using the one-compartment model described in section 5.1 with random effects on all parameters. For biological parameters we took informative priors arising from published studies (Sadler et al., 2001; Sale et al., 2002; Okusanya et al., 2007). Table 2 presents the chosen priors and the estimates. Random effect on k_e was not significant and was removed. We observed (figure 3 - left) a poor adjustment to the data. Moreover, $\sigma_{\tilde{k}_a}$ is very high; this gives a hint about the model misspecification. Thus, we implemented a two-compartments model with absorption; a peripheral compartment representing tissues, $\mathrm{C_T}$, was added (Equation 2). System graphical representation and minimum steady-state plasma and tissue concentrations, assumed as baseline for the patient, are in Web-Appendix A.

$$\begin{cases} \frac{d\mathrm{A_{GI}}}{dt} &= -k_a \mathrm{A_{GI}}, \\ \frac{d\mathrm{C_P}}{dt} &= \frac{k_a}{V_0}\mathrm{A_{GI}} - \mathrm{k_e C_P} - \mathrm{k_{PT} C_P} + \frac{\mathrm{k_{TP}}}{\mathrm{V_0}}\mathrm{C_T}, \\ \frac{d\mathrm{C_T}}{dt} &= \frac{k_{PT}}{V_T}\mathrm{C_P} - \mathrm{k_{TP} C_T}. \end{cases} \tag{2}$$

Fitting curves (figure 3 - right) look better than with the one-compartment model with absorption. Indeed, the $\mathrm{LCV_a}$ for the two-compartments model was lower than for the one-compartment model. We can then conclude that the two-compartments model is more suitable to explain the APV PK in human.

Finally, we tried introduced the subject's weight (w^i in kg) in the model as a covariate such as: $\tilde{V}_0 = \phi_{\tilde{V}_0} + \beta_w * w^i + u^i$. The NIMROD program

estimates $\phi_{\tilde{V}_0} = -3.02$ ($sd = 0.06$) and $\beta_w = 0.014$ ($sd = 0.006$) with a Bayesian p-value equals to 0.02. Thus, there is a significant effect of weight on V_0. This is in accordance with a decrease of about 0.1 in the $\mathrm{LCV_a} = 9.57$. This is consistent with previous studies (Sale et al., 2002) on APV.

6. Conclusion

We have presented the NIMROD program which allows to make approximate Bayesian inference in models with random effects based on differential equations. The program bases inference on a normal approximation of the posteriors and a maximization procedure. NIMROD can also perform maximum likelihood inference if flat priors are specified. We have illustrated the inference by a simulation study and estimations on real data from a PK study. The program works when the ODE system has an analytical solution as well as for non-linear ODE systems as the one used in Prague et al. (2012).

Supplementary Materials

Web-Appendix referenced in section 4.3, 5.1 and 5.2 are available with this paper at the Computer Methods and Programs in Biomedecine website on Elsevier Library. The NIMROD program is available on request.

Acknowledgement

The authors would like to thank the investigators of the PUZZLE ANRS 104 trial, particularly G. Raguin (principal investigator), G. Chêne (methodologist). Parallel computing was used thanks to the computing facilities

MCIA (Mésocentre de Calcul Intensif Aquitain) of the Université de Bordeaux and of the Université de Pau et des Pays de l'Adour.

Reference

Berger, J., 1985. Statistical decision theory and Bayesian analysis. Springer.

Berndt, E., Hall, B., Hall, R., 1974. Estimation and inference in nonlinear structural models. Annals of Economic and Social Measurement 3 (4), 103–116.

Commenges, D., Jacqmin-Gadda, H., Proust, C., Guedj, J., 2006. A newton-like algorithm for likelihood maximization : The robust-variance scoring algorithm . Arxiv math/0610402.

Commenges, D., Joly, P., Gégout-Petit, A., Liquet, B., 2007. Choice between semi-parametric estimators of markov and non-markov multi-state models from coarsened observations. Scandinavian Journal of Statistics 34 (1), 33–52.

Commenges, D., Proust-Lima, C., Samieri, C., Liquet, B., 2012. A universal approximate cross-validation criterion and its asymptotic distribution. Arxiv preprint arXiv:1206.1753.

Commenges, D., Sayyareh, A., Letenneur, L., Guedj, J., Bar-Hen, A., 2008. Estimating a difference of kullback–leibler risks using a normalized difference of aic. The Annals of Applied Statistics 2 (3), 1123–1142.

Ding, A., Wu, H., 2001. Assessing antiviral potency of anti-hiv therapies in vivo by comparing viral decay rates in viral dynamic models. Biostatistics 2 (1), 13–29.

Drylewicz, J., Commenges, D., Thiébaut, R., 2010a. Score tests for exploring complex models: application to hiv dynamics models. Biometrical Journal 52 (1), 10–21.

Drylewicz, J., Commenges, D., Thiébaut, R., 2012a. Maximum a posteriori estimation in dynamical models of primary hiv infection. Statistical Communications in Infectious Diseases 4 (1), online.

Drylewicz, J., Commenges, D., Thiebaut, R., 2012b. Maximum a posteriori estimation in dynamical models of primary hiv infection. Statistical Communications in Infectious Diseases 4 (1).

Drylewicz, J., Guedj, J., Commenges, D., Thiébaut, R., 2010b. Modeling the dynamics of biomarkers during primary hiv infection taking into account the uncertainty of infection date. The Annals of Applied Statistics 4 (4), 1847–1870.

Gelman, A., 2006. Prior distributions for variance parameters in hierarchical models. Bayesian analysis 1 (3), 515–533.

Genz, A., Keister, B., 1996. Fully symmetric interpolatory rules for multiple integrals over infinite regions with gaussian weight. Journal of Computational and Applied Mathematics 71 (2), 299–309.

Godfrey, K., Jones, R., Brown, R., 1980. Identifiable pharmacokinetic models: The role of extra inputs and measurements. Journal of Pharmacokinetics and Pharmacodynamics 8 (6), 633–648.

Guedj, J., Thiébaut, R., Commenges, D., 2007a. Maximum likelihood estimation in dynamical models of HIV. Biometrics 63 (4), 1198–1206.

Guedj, J., Thiébaut, R., Commenges, D., 2007b. Practical identifiability of hiv dynamics models. Bulletin of mathematical biology 69 (8), 2493–2513.

Hindmarsh, A., 1983. Odepack: A systematized collection of ode solvers. IMACS ransactions on Scientific Computations 1 (1), 55–64.

Jacqmin-Gadda, H., Thiébaut, R., Chêne, G., Commenges, D., 2000. Analysis of left-censored longitudinal data with application to viral load in hiv infection. Biostatistics 1 (4), 355–368.

Kass, R., Steffey, D., 1989. Approximate bayesian inference in conditionally independent hierarchical models (parametric empirical bayes models). Journal of the American Statistical Association 84 (407), 717–726.

Louis, T., 1982. Finding the observed information matrix when using the em algorithm. Journal of the Royal Statistical Society. Series B (Methodological) 44 (2), 226–233.

Marquardt, D., 1963. An algorithm for least-squares estimation of nonlinear parameters. Journal of the society for Industrial and Applied Mathematics 11 (2), 431–441.

Okusanya, O., Forrest, A., DiFrancesco, R., Bilic, S., Rosenkranz, S., Para, M., Adams, E., Yarasheski, K., Reichman, R., Morse, G., et al., 2007. Compartmental pharmacokinetic analysis of oral amprenavir with secondary peaks. Antimicrobial agents and chemotherapy 51 (5), 1822–1826.

Pacheco, P., 1997. Parallel programming with MPI. Morgan Kaufmann.

Perelson, A., Neumann, A., Markowitz, M., Leonard, J., Ho, D., 1996. Hiv-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. Science 271 (5255), 1582–1586.

Pinheiro, J., Bates, D., 2000. Mixed-effects models in S and S-PLUS. Springer Verlag.

Plan, E., Maloney, A., Mentré, F., Karlsson, M., Bertrand, J., 2012. Performance comparison of various maximum likelihood nonlinear mixed-effects estimation methods for doseresponse models. The AAPS Journal 14 (3), 420–432.

Potra, F., Shi, Y., 1995. Efficient line search algorithm for unconstrained optimization. Journal of Optimization Theory and Applications 85 (3), 677–704.

Prague, M., Commenges, D., Drylewicz, J., Thiébaut, R., 2012. Treatment monitoring of hiv infected patients based on mechanistic models. Biometrics, in press.

Radhakrishnan, K., Hindmarsh, A., Aeronautics, U. S. N., Administration, S., 1993. Description and use of LSODE, the Livermore solver for ordinary

differential equations. National Aeronautics and Space Administration, Office of Management, Scientific and Technical Information Program.

Raguin, G., Chêne, G., Morand-Joubert, L., Taburet, A., Droz, C., Le Tiec, C., Clavel, F., Girard, P., Group, P. . S., et al., 2004. Salvage therapy with amprenavir, lopinavir and ritonavir 200 mg/d or 400 mg/d in hiv-infected patients in virological failure. Antiviral therapy 9, 615–626.

Rosenbaum, S., 2011. Basic Pharmacokinetics and Pharmacodynamics: An Integrated Textbook and Computer Simulations. Wiley.

Rue, H., Martino, S., Chopin, N., 2009. Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. Journal of the royal statistical society: Series b (statistical methodology) 71 (2), 319–392.

Sadler, B., Piliero, P., Preston, S., Lloyd, P., Lou, Y., Stein, D., 2001. Pharmacokinetics and safety of amprenavir and ritonavir following multiple-dose, co-administration to healthy volunteers. Aids 15 (8), 1009.

Sale, M., Sadler, B., Stein, D., 2002. Pharmacokinetic modeling and simulations of interaction of amprenavir and ritonavir. Antimicrobial agents and chemotherapy 46 (3), 746–754.

Van der Vaart, A., 2000. Asymptotic statistics. Cambridge Univ Pr.

Verbeke, G., Molenberghs, G., 2009. Linear mixed models for longitudinal data. Springer Verlag.

Wakefield, J., Racine-Poon, A., 1995. An application of bayesian population pharmacokinetic/pharmacodynamic models to dose recommendation. Statistics in medicine 14 (9), 971–986.

Table 1: Inference in a one-compartment PK model with oral absorption: simulation values, mean and standard deviation of the priors and posteriors computed for the log-transformed parameters for one NIMROD run each. Units are m^3 and minutes. However, in natural scale, we have as simulation parameters: $k_e = 0.23\ h^{-1}$, $V_0 = 15\ L$ and $k_a = 0.6\ h^{-1}$.

Simulated Data		Large Priors		Tight Priors	
Parameters	Simulation	Priors	NIMROD	Priors	NIMROD
(in log)	value	Mean (sd.)	Mean (sd.)	Mean (sd.)	Mean (sd.)
$\tilde{k}_e$	-5.56	-6.0 (5.0)	-5.525 (0.07)	-6.0 (1.0)	-5.544 (0.07)
$\tilde{V}_0$	-4.19	-4.6 (5.0)	-4.176 (0.04)	-4.6 (1.0)	-4.158 (0.04)
$\tilde{k}_a$	-4.60	-4.5 (5.0)	-4.591 (0.06)	-4.5 (0.1)	-4.559 (0.05)
$\sigma_{\tilde{k_e}}$	0.25	no	0.251 (0.04)	0.5	0.249 (0.04)
$\sigma_{\tilde{V_0}}$	0.10	no	0.091 (0.05)	0.5	0.095 (0.05)
$\sigma_{\tilde{k_a}}$	0.20	no	0.220 (0.05)	0.5	0.219 (0.05)
$\sigma_{\mathrm{C_P}}$	0.30	no	0.280 (0.01)	$1/\sigma_{CP}$	0.280 (0.01)

Table 2: Inference in a one and two-compartments PK model with absorption for APV concentrations in HIV infected patients: priors and posteriors means and standard deviations. Convergence criteria were $\eta_1 = \eta_2 = 0.1$ and $\eta_3 = 0.3$. NIMROD reached convergence after 181 (resp. 301) iterations for the one-compartment PK model with oral absorption (resp. two-compartments) which took about 17 minutes (resp. 70 minutes).

Parameters	Priors	one-compartment w/abs.	two-compartments w/abs.
(in log)	Mean (sd.)	Mean (sd.)	Mean (sd.)
$\mathrm{LCV_a}$	-	11.78	9.65
$\tilde{k}_e$	-5.0 (3.0)	-6.249 (0.01)	-5.453 (0.01)
$\tilde{V}_0$	-3.0 (3.0)	-1.405 (0.04)	-2.051 (0.10)
$\tilde{k}_a$	-3.0 (3.0)	-1.912 (0.13)	-3.386 (0.10)
$\tilde{k}_{PT}$	-5.0 (2.0)	-	-4.941 (0.01)
$\tilde{k}_{TP}$	-7.0 (2.0)	-	-5.315 (0.01)
$\tilde{V}_T$	-1.0 (0.5)	-	-0.579 (0.41)
$\sigma_{\tilde{k}_e}$	0.5	-	0.191 (0.01)
$\sigma_{\tilde{V}_0}$	0.5	2.487 (0.10)	0.475 (0.05)
$\sigma_{\tilde{k}_a}$	1.0	0.454 (0.02)	0.679 (0.05)
$\sigma_{\mathrm{C_P}}$	$1/\sigma_{CP}$	0.659 (0.03)	0.516 (0.02)

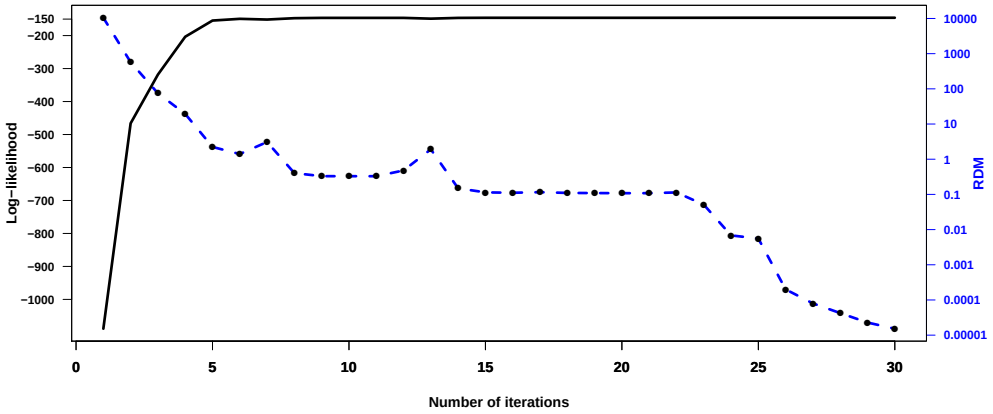
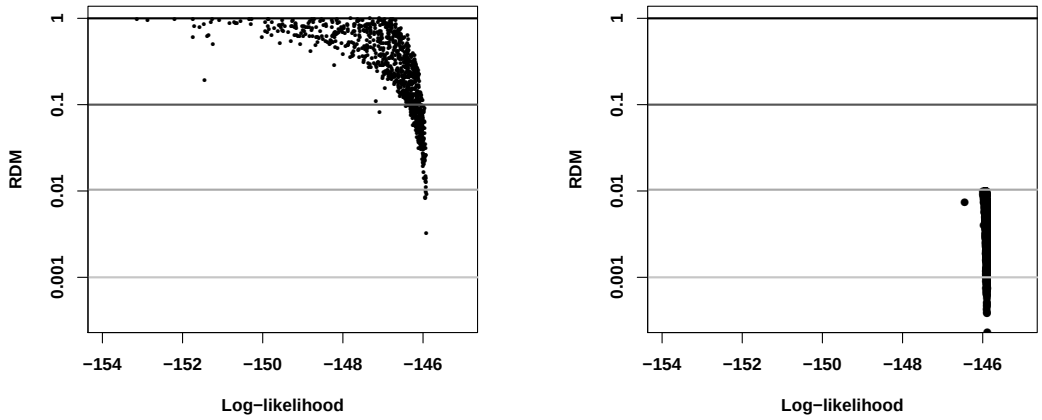


Figure 1: Trajectories of the RDM and the log-likelihood along iterations from a random starting point to convergence.



(a) $\eta_3 < 1$ (b) $\eta_3 < 0.01$

Figure 2: Link between the final RDM and the final log-likelihood at convergence stopping points for 1000 starting points with two values of convergence criterion η_3.

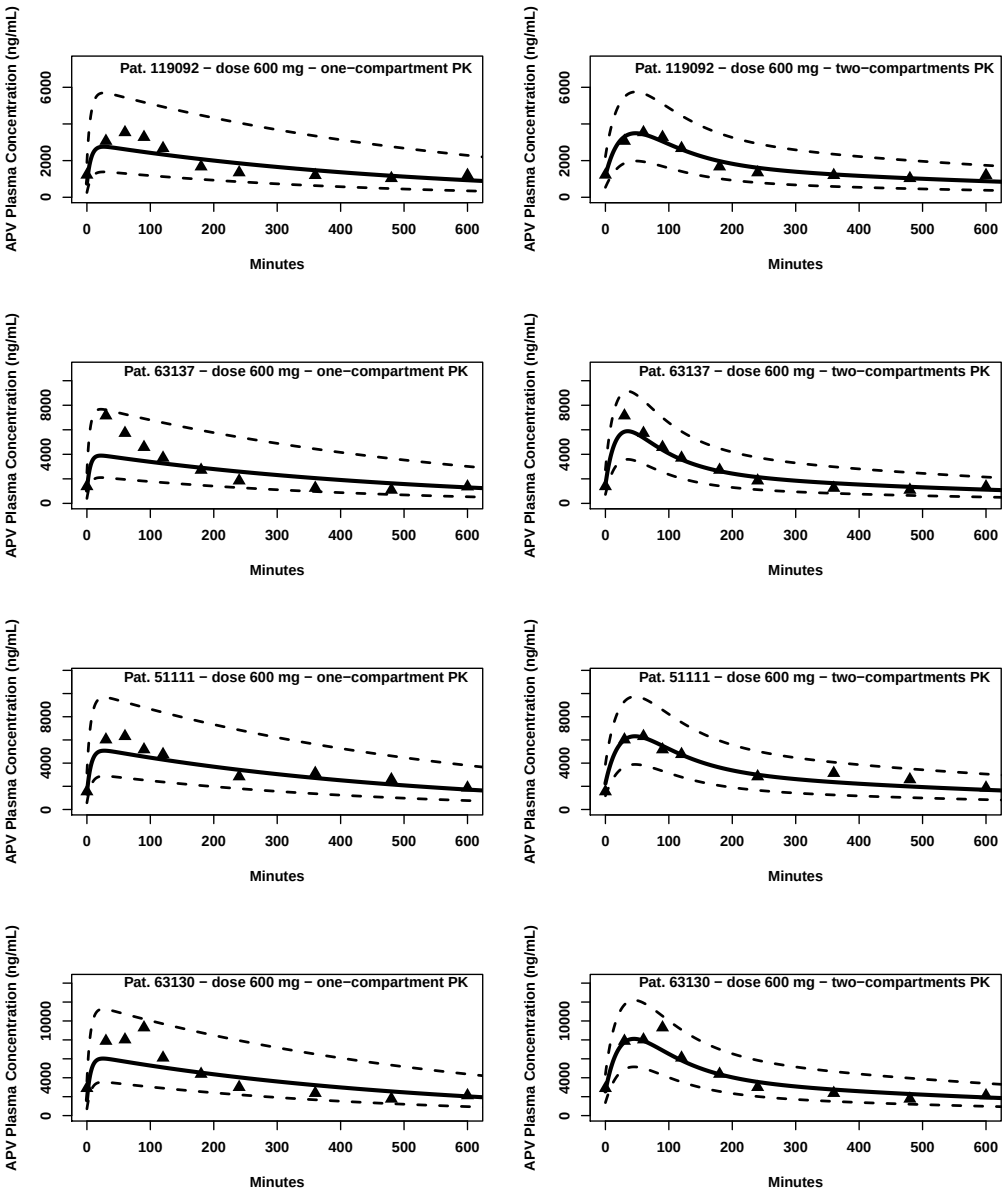


Figure 3: Examples of fits for APV PK dynamics in 4 HIV infected patients from PUZZLE Study. Left-side presents one-compartment PK models with absorption; Right-side is two-compartments PK models with absorption. Triangles are observations. Plain lines are the predicted plasma concentration. Dashed lines represent the "95% measurement error predictive interval" given by $\hat{Y} \pm 1.96\sigma_{\mathrm{C_P}}$.