SAMBA: a Novel Method for Fast Automatic Model Building in Nonlinear Mixed-Effects Models

Mélanie Prague, Marc Lavielle

SAMBA: a Novel Method for Fast Automatic Model Building in Nonlinear Mixed-Effects Models

Mélanie Prague [1]

University of Bordeaux, Inria Bordeaux Sud-Ouest, Inserm, Bordeaux Population Health Research Center, SISTM Team, UMR 1219, F-33000 Bordeaux, France

Melanie.Prague@inria.fr

Marc Lavielle

Inria & CMAP, Ecole Polytechnique, CNRS, Institut Polytechnique de Paris, France.

Marc.Lavielle@inria.fr

Conflict of interest Marc Lavielle is chief scientist of Lixoft, the company that develops and distributes the Monolix Suite. The other author declared no competing interests for this work.

Funding information None to disclose.

Keyword: Nonlinear models; mixed-effects model; Population PKPD; Modeling; Covariate model selection; Stochastic algorithm.

[1] Corresponding author: ISPED - Université de Bordeaux - Bureau 23, 146 Rue Léo Saignat, 33070 Bordeaux Cedex; +33557574531

Abstract

The success of correctly identifying all the components of a nonlinear mixed-effects model is far from straightforward: it is a question of finding the best structural model, determining the type of relationship between covariates and individual parameters, detecting possible correlations between random effects, or also modeling residual errors. We present the SAMBA (Stochastic Approximation for Model Building Algorithm) procedure and show how this algorithm can be used to speed up this process of model building by identifying at each step how best to improve some of the model components. The principle of this algorithm basically consists in 'learning something' about the 'best model', even when a 'poor model' is used to fit the data. A comparison study of the SAMBA procedure with SCM and COSSAC show similar performances on several real data examples but with a much-reduced computing time. This algorithm is now implemented in Monolix and in the R package *Rsmlx*.

Introduction

Construction of a complex (nonlinear) mixed-effects model [14] is a challenging process which requires confirmed expertise, advanced statistical methods, the use of sophisticated software tools, but above all time and patience. Indeed, the success of correctly identifying all the components of the model is far from straightforward: it is a question of finding the best structural model, determining the type of relationship between covariates and individual parameters, detecting possible correlations between random effects, or also modeling residual errors. Our goal is to accelerate and optimize this process of model building by identifying at each step how best to improve some of the model components.

The procedure for constructing a model is usually iterative: one adjusts a first model to the data, and diagnosis plots and statistical tests allow to detect possible miss-specifications in the proposed model. A new model must then be proposed to correct these defects and improve the predictive abilities of the model. Most of common approaches consist in stepwise procedures consisting in testing the addition of variable forward and their elimination backward alternatively and progressing through the choice of model using a criterion derived from the log-likelihood (LL). A widely used approach is SCM (Stepwise Covariate Modeling) [12], which consists in an exhaustive search in the covariates space. Each covariate addition or deletion is tested in turn selecting models at each step leading to the best adjustment according to the objective criterion. Approaches such as WAM (Wald Approximation Method) [13] and COSSAC (COnditional Sampling use for Stepwise Approach based on Correlation tests) [2] are less computationally intensive as they use respectively a likelihood ratio test and a correlation test to move in the covariates space, which allow to test less models. All these methods are nevertheless computationally intensive as they require to re-estimate the model parameters and the likelihood many times. In particular, these methods are very sensitive to 'the curse of dimensionality' when the number of covariates to test on parameters is large.

The GAM (Generalized additive model) method [9, 17] is computationally appealing as it does not require as many models fitting. Indeed, it is based on a regression on the empirical Bayes estimates (EBEs). The EBEs are the modes of the conditional distributions of the individual parameters. In other words, they are the most likely value of the individual parameters, given the estimated population parameters and the data. These estimates are known to be misleading and prone to shrinkage when data are sparse [19]. An efficient method which can correct the bias caused by the shrinkage of the EBEs have been recently proposed for covariate analysis [21, 22]. In this article, we propose to develop similar method which relies on the use of random samples from the conditional distribution of each individual parameters instead of EBEs. Indeed, the random sample of the posterior distribution has been shown to correctly control the type-I error when performing tests to detect miss-specifications in the model [16].

As for most of the model building procedures, the objective of SAMBA (Stochastic Approximation for Model Building Algorithm) is to find a model that minimizes some information criterion, such as AIC, BIC or BICc, the corrected BIC [6]. The main principle of SAMBA is to use the results obtained with a wrong model to learn the right model. Then, SAMBA is an iterative procedure where a new model is used at each iteration of the algorithm. The values of the population parameters of the model are found by maximum likelihood estimation and then, the individual parameters are sampled from the conditional distribution defined under this estimated model. These simulated individual parameters combined with the observed data can now be used to select a new statistical model. It is important to underline that, as most of the iterative procedures for non-convex optimization, SAMBA does not pretend to be capable of always finding the global minimum of the used criterion, but it always allows to find very quickly a very good solution.

Two contributions mainly constitute the content of this article. First, we describe the novel algorithm called SAMBA for fast automatic model building in nonlinear mixed-effects models (Section 1). And, second we benchmark its performances compared to reference methods SCM and COSSAC in real world examples (Section 2). Section 3 concludes.

Methods

Model description

Let $y_i = (y_{ij}, 1 \leq j \leq n_i)$ be the vector of observations for subject i, where $1 \leq i \leq N$. The model that describes the observations y_i is assumed to be a parametric probabilistic model that depends on a vector of L (individual) parameters $\psi_i = (\psi_{i1}, \ldots, \psi_{Li})$. In a population framework, the vector of parameters ψ_i is assumed to be drawn from a population distribution $\mathrm{p}(\psi_i)$. Then, defining a model $\mathcal{M}$ consists in defining a joint probability distribution for the observations $y = (y_1, \ldots, y_N)$ and for the individual parameters $\psi = (\psi_1, \ldots, \psi_N)$. For sake of notation simplicity, we focus on models for continuous longitudinal data. However, extension to models for discrete data and time to event data is straightforward.

Let y_{ij}, the observation obtained from subject i at time t_{ij} be described as:

$$u(y_{ij}) = u(f(t_{ij}, \psi_i)) + g(t_{ij}, \psi_i, \xi)\varepsilon_{ij} \;\;,\; 1 \leq i \leq N \;\;,\;\; 1 \leq j \leq n_i. \qquad (1)$$

The structural model f is a fundamental component of the model since it defines the individual predictions of the observed kinetics for a given set of parameters. The residual errors (ε_{ij}) are assumed to be standardized Gaussian random variables (mean zero and variance 1). The residual error model is represented by function g in model (1) and may depends on some additional parameter ξ. Finally, one can use the function u to transform the observations, assuming for instance that they are log-normally distributed. In the following, we will assume u to be the identity.

We assume a linear model for the individual parameters (up to some transformation h):

$$h(\psi_i) = h(\psi_{\rm pop}) + \beta\, c_i + \eta_i \ \ , \ 1 \leq i \leq N, \tag{2}$$

where $\eta_i \sim \mathcal{N}(0, \Omega)$ is a vector of random effects and where c_i is a vector of individual covariates used to explain part of the variability of the ψ_i's. $\psi_{\rm pop}$ and β are fixed effects. The joint model of y and ψ then depends on a set of parameters $\theta = (\psi_{\rm pop}, \beta, \Omega, \xi)$.

Selecting a model described by equations (1) and (2) consists for the modeler in selecting: *(i)* the structural model f, *(ii)* the transformation of the individual parameters h, *(iii)* the residual error model g, *(iv)* the list of covariates that have an impact on individual parameters, and *(v)* the structure of the variance-covariance matrix of the random effects in the linear model Ω. The selection of the two first items is problem-specific, and their selection is out of the scope of this article. We will therefore assume in this article that f and h are given. The SAMBA procedure proposes solutions to address the selection of the three other components of the model.

The SAMBA procedure

Automatic model building is a difficult task since it is generally not possible to fit and compare all possible models. Moreover, it is necessary to define what is the 'best model' among all the possible models. A classical approach consists in searching for the model

$\mathcal{M}_*$ that minimizes a criterion such as the penalized likelihood [8, 11]:

$$\mathcal{M}_* = \mathrm{argmin}_{\mathcal{M}} \left\{ \min_\theta \left(-2\log\left(\mathcal{L}_{\mathcal{M}}(\theta; y)\right)\right) + \mathrm{pen}(\mathcal{M}) \right\}. \tag{3}$$

The objective of this approach is to find a model that best fits the data (by minimizing $-2LL$) while being as simple as possible (it is the role of $\mathrm{pen}(\mathcal{M})$ to favor models with few parameters). When the space of possible models is large, an exhaustive search is clearly impossible, and an efficient minimization strategy must be implemented. It is precisely for this purpose that SAMBA was developed: to obtain very quickly the "best" model $\mathcal{M}_*$, or a model with an objective criterion value very close to that of $\mathcal{M}_*$.

SAMBA is an iterative procedure alternating three steps. Assume that model $\mathcal{M}_k$ was obtained at iteration k of the algorithm. We first compute $\theta^{(k)}$, the maximum likelihood estimate of θ for model $\mathcal{M}_k$. We then generate a set of individual parameters $\psi^{(k)}$ from the conditional distribution of individual parameters $p_{\mathcal{M}_k}(\psi \,|\, y; \theta^{(k)})$. The selection step finally consists in building a new model $\mathcal{M}_{k+1}$ using the *complete data* $(y; \psi^{(k)})$ and minimizing the *complete penalized criterion*:

$$\mathcal{M}_{k+1} = \mathrm{argmin}_{\mathcal{M}} \left\{ \min_\theta \left(-2\log\left(\mathcal{L}_{\mathcal{M}}(\theta; y, \psi^{(k)})\right)\right) + \mathrm{pen}(\mathcal{M}) \right\}. \tag{4}$$

As already mentioned, the statistical model to be built consists of a covariate model, a correlation model, and a residual error model. Then, the selection of model $\mathcal{M}_{k+1}$ is composed of three model selection procedures: the selection of the covariate model $\mathcal{M}_{k+1}^{COV}$, the selection of the correlation model $\mathcal{M}_{k+1}^{CORR}$ and the selection of the error model $\mathcal{M}_{k+1}^{ERR}$. Note that not all these components are necessarily selected: some may have been set arbitrarily because of existing knowledge. By noticing that $\mathcal{L}_{\mathcal{M}}(\theta; y, \psi^{(k)})$=$\mathcal{L}_{\mathcal{M}}(\theta | y, \psi^{(k)})$

$\mathcal{L}_{\mathcal{M}}(y, \psi^{(k)})$, it appears that the problem of selecting the error model is independent from the problem of selecting the covariate and correlation models. Figure 1 provides a flowchart of the complete procedure. Let's now take a closer look at what each step of the model selection process consists of.

The covariate model selection $\boldsymbol{\mathcal{M}_{k+1}^{COV}}$. The sample $\psi^{(k)}$ has been generated conditionally to the data y and the model $\mathcal{M}_k$. For the ℓ-th parameter, we build a linear model between $\psi_\ell^{(k)}$ and covariates c such as in Equation 2:

$$h_\ell(\psi_{i\ell}^{(k)}) = h_\ell(\psi_{\text{pop},\ell}) + \beta_\ell \, c_i + \eta_{i\ell}^{(k)} \;\;, \; 1 \leq i \leq N, \; 1 \leq \ell \leq L, \tag{5}$$

with h_ℓ the transformation associated to the ℓ-th parameter and where $\eta_{i\ell}^{(k)}$ is supposed normally distributed with mean zero and variance ω_ℓ^2. We define $\theta_\ell = (\psi_{\text{pop},\ell}, \beta_\ell, \omega_\ell^2)$. Best covariate model for parameter ℓ, denoted $\mathcal{M}_{k+1}^{COV_\ell}$, is selected as being the one minimizing a penalized criterion:

$$\mathcal{M}_{k+1}^{COV_\ell} = \text{argmin}_{\mathcal{M}} \left\{ \text{min}_{\theta_\ell} \left(-2 \log \left(\mathcal{L}_{\mathcal{M}}(\theta_\ell; \psi_\ell^{(k)}) \right)) + \text{pen}^{COV}(\mathcal{M}) \right) \right\}.$$

We denote n_β the number of non-null elements in β_ℓ for model $\mathcal{M}$. The penalization depends on the criterion selected for optimization: if AIC then $\text{pen}^{COV}(\mathcal{M}) = 2n_\beta$, if BIC or BICc then $\text{pen}^{COV}(\mathcal{M}) = \log(N)n_\beta$. Equation (5) tells us that the covariate selection problem has become here a classical problem of variable selection in a linear model [7]. This problem is much more easily tractable than the original one. The overall best covariate model combines the best model for each parameters such that $\mathcal{M}_{k+1}^{COV} = \{\mathcal{M}_{k+1}^{COV_1}, \ldots \mathcal{M}_{k+1}^{COV_L}\}$.

In the implemented version of package *Rsmlx* (R speaks Monolix), two different strategies are implemented depending on the dimension of the selection problem. If the number d of available covariates is less than 11, an exhaustive search is performed over all the 2^d possible covariate models for each parameter. Otherwise, the stepwise variable selection procedure implemented in the function *stepAIC* from package *MASS* is used. It consists of iteratively adding and removing covariates in stepwise manner to lower the objective criterion.

The correlation model selection $\boldsymbol{\mathcal{M}_{k+1}^{CORR}}$. Using the selected covariate model $\mathcal{M}_{k+1}^{COV}$ and the sample of individual parameters $\psi_i^{(k)}$, it is possible to extract the vector of individual random effects $\eta_i^{(k)} = (\eta_{i\ell}^{(k)}, \ell = 1, \ldots, L)$ from Equation 5. Assuming that $\eta_i^{(k)} \sim \mathcal{N}(0, \Omega)$ where Ω is a block diagonal matrix, the problem of correlation model selection consists in selecting the block structure of Ω. We then select the correlation model denoted $\mathcal{M}_{k+1}^{CORR}$ by minimizing a penalized criterion:

$$\mathcal{M}_{k+1}^{CORR} = \mathrm{argmin}_{\mathcal{M}} \left\{ \min_{\Omega} \left(-2 \log \left(\mathcal{L}_{\mathcal{M}}(\Omega; \eta_i^{(k)}) \right) \right) + \mathrm{pen}^{CORR}(\mathcal{M}) \right\}.$$

We denote n_Ω the number of non-zero elements in the upper triangular part of the matrix Ω. The penalization depends on the criterion selected for global optimization: if AIC then $\mathrm{pen}^{CORR}(\mathcal{M}) = 2n_\Omega$, if BIC or BICc then $\mathrm{pen}^{CORR}(\mathcal{M}) = \log(N)n_\Omega$.

In the implemented version of package *Rsmlx*, we limit the size of the block-structure that can be considered at each iteration. For $\mathcal{M}_1$, no correlation can be added and a diagonal matrix is used for Ω ; for $\mathcal{M}_2$ only blocks of size two are considered. At iteration k for selection of model $\mathcal{M}_{k+1}^{CORR}$, block size cannot be larger than $k + 1$, leading to no more than $(k - 1)k/2$ non-zero covariance terms in Ω.

The error model selection $\mathcal{M}_{k+1}^{ERR}$. For a given set of simulated individual parameters $(\psi_i^{(k)}, 1 \leq i \leq N)$, the residual errors can easily be computed:

$$e_{ij}^{(k)} = y_{ij} - f(t_{ij}, \psi_i^{(k)}),\ 1 \leq i \leq N\ \ ,\ \ 1 \leq j \leq n_i.$$

We then fit several error models with standard deviation of the form $g(t_{ij}, \psi_i^{(k)}, \xi)$ for $e_{ij}^{(k)}$ and select the one minimizing a penalized criterion:

$$\mathcal{M}_{k+1}^{ERR} = \text{argmin}_{\mathcal{M}} \left\{ \min_{\xi} \left(-2\log\left(\mathcal{L}_{\mathcal{M}}(\xi; e_{ij}^{(k)})\right)\right) + \text{pen}^{ERR}(\mathcal{M}) \right) \right\}.$$

We denote n_ξ the length of ξ, i.e. the number of parameters in model $\mathcal{M}$. The penalization depends on the criterion selected for global optimization: if AIC then $\text{pen}^{ERR}(\mathcal{M}) = 2n_\xi$, if BIC then $\text{pen}^{ERR}(\mathcal{M}) = \log(N)n_\xi$, and if BICc then $\text{pen}(\mathcal{M}) = \log(n_{\text{tot}})n_\xi$ where n_{tot} is the total number of observations, including below the limit of quantification (BLQ) data.

In the implemented version of package *Rsmlx*, five error models (provided by function g in Equation (1)) are tested by default: *constant* ($g_x(t_{ij}, \psi_i^{(k)}, \xi) = \xi$), *proportional* ($g_x(t_{ij}, \psi_i^{(k)}, \xi) = \xi f(t_{ij}, \psi_i)$), *combined*$_1$ ($g_x(t_{ij}, \psi_i^{(k)}, \xi) = \xi_1 + \xi_2 f(t_{ij}, \psi_i)$), *combined*$_2$ ($g_x(t_{ij}, \psi_i^{(k)}, \xi) = \sqrt{\xi_1^2 + \xi_2^2 f(t_{ij}, \psi_i)}$) or *exponential* in which a constant error model is fitted to the $\log(y)$ using the transformation $u = \log$ in Equation 1. Note that it is currently not possible to perform the selection on a restricted number of error models, but such a feature could be easily implemented.

Stopping rule procedure At each iteration k of the algorithm, we combine $\mathcal{M}_{k+1}^{COV}$, $\mathcal{M}_{k+1}^{CORR}$ and $\mathcal{M}_{k+1}^{ERR}$ to get the new selected model $\mathcal{M}_{k+1}$ which is passed forward on to the next estimation-simulation run. It is important to select the covariate model before the correlation model. On the other hand, the error model can be updated before or after the other two components of the model. The algorithm stops when $\mathcal{M}_k$ is strictly identical to $\mathcal{M}_{k+1}$ for all components and the last model is the selected one.

Remark In the above, $\psi_i^{(k)}$ represents a single realization of the conditional distribution $p_{\mathcal{M}_k}(\psi_i|y, \theta^{(k)})$ for each $i = 1, \dots N$. Instead of considering only one realization of this distribution, we could use a sample of size R $(\psi_{i\ell,r}^{(k)}, 1 \leq r \leq R)$. If so, the linear covariate model described in Equation (5) rewrites:

$$\overline{h_\ell(\psi_{\ell,i}^{(k)})} = h_\ell(\psi_{\ell,\text{pop}}) + \beta_\ell \, c_i + \overline{\eta_{\ell,i}^{(k)}} \ , \ 1 \leq i \leq N, \ 1 \leq \ell \leq L,$$

where:

$$\overline{h_\ell(\psi_{\ell,i}^{(k)})} = \frac{1}{R} \sum_{r=1}^{R} h_\ell \left(\psi_{i\ell,r}^{(k)} \right).$$

Procedures for covariate model selection and correlation model selection remains the same, but using now $(\overline{\psi_{i\ell}^{(k)}})$ and $(\overline{\eta_{i\ell}^{(k)}})$ at iteration k. On the other hand, the R series of residual errors $(e_{ij,r}^{(k)})$ are used for selecting the residual error model.

Results

Step-by-step example of the SAMBA procedure

To illustrate how SAMBA works in practice, we will describe step-by-step the complete procedure on the example of remifentanil [18]. We use here the SAMBA implementation in function *buildmlx* of the R package *Rsmlx*, using the default settings.

The remifentanil data. The dataset is composed of 65 healthy adults who have received remifentanil IV infusion at a constant infusion rate between 1 and 8 $\mu g.kg^{-1}.min^{-1}$ for 4 to 20 minutes. Time and rate of infusion are known for each individual. The pharmacokinetics data consists in the plasma concentration of remifentanil, which is measured during and after infusion for a total of 19 to 53 observations by patients, totalling 2057 observations. A total of 6 covariates are available: one qualitative covariate, the sex (SEX) and five continuous covariates: the age (AGE), the height (HT), the weight (WT), the lean body mass (LBM) and the body surface area (BSA). All the latter are normalized and log-transformed for the analysis. In the following, we adopt the notation $\mathrm{logAGE} = \log(\mathrm{AGE}/\mathrm{AGE}_{\mathrm{pop}})$, where AGE_{pop} is a typical value to normalize on, e.g. the mean value of age in population.

The model. The PK model for IV infusion has a central compartment (volume $V1$), two peripheral compartments (volume $V2$ and $V3$, inter-compartmental clearances $Q2$ and $Q3$), and a linear elimination (Cl). Log-normal distributions are used for the six individual parameters. The $2^6 = 64$ possible covariate models will be considered for each of the six individual parameters. Note that if we had to test all possible models, we would have had to test 64^6 combinations, which would have made the problem intractable.

SAMBA iterations. We start the SAMBA procedure with a model $\mathcal{M}_0$ without any covariate on all parameters, with no correlation between random effects and the so-called *combined*$_1$ error model. Figure 2 illustrates the selection steps on this specific example. One can notice that the BICc, which has been chosen as target criterion, decreases from 7186 for $\mathcal{M}_0$, to 6985 for $\mathcal{M}_1$, 6957 for $\mathcal{M}_2$, and 6903 for $\mathcal{M}_3$, which is finally selected as the best model for this example.

- **Run 0 (BICc=7185.8) + Iteration 1:** Model $\mathcal{M}_0$ is fitted to data and individual parameters are sampled conditionally on the data and this model. Each of the 64 possible linear covariate models is fitted to each individual parameters and the one with lowest BICc is selected. Let's take the example of Cl: the three best models include 1) an effect of logAGE and logWT (BICc=-55.0), 2) an effect of logAGE and logLBM (BICc=-56.1), and 3) an effect of logAGE and logBSA (BICc=-57.5). The latter is chosen as the best model for parameter Cl as it provides the lowest BICc ($\mathcal{M}_1^{COV,Cl}$). Altogether, for all parameters, the best covariate model ($\mathcal{M}_1^{COV}$) includes logAGE on all parameters, logBSA on Cl, and logLBM on $V1$ and $V2$. No correlation is added to the model since no correlation is allowed at first iteration. Then, $\mathcal{M}_1^{CORR}$ is a diagonal variance-covariance matrix for the random effects. Among the tested error models, the three best ones are *proportional* (BICc=5815.2), *combined*$_1$ (BICc=5811.2) and *combined*$_2$ (BICc=5807.0) which is selected for $\mathcal{M}_1^{ERR}$. These covariate, correlation and error models are then passed on to run 1: $\mathcal{M}_1 = \{\{\mathcal{M}_1^{COV,Cl}, \mathcal{M}_1^{COV,Q2}, \mathcal{M}_1^{COV,Q3}, \mathcal{M}_1^{COV,V1}, \mathcal{M}_1^{COV,V2},$ $\mathcal{M}_1^{COV,V3}\}, \mathcal{M}_1^{CORR}, \mathcal{M}_1^{ERR}\})$.

- **Run 1 (BICc=6984.9) + Iteration 2:** Model $\mathcal{M}_1$ is fitted to the data and individual parameters are sampled. Again, the three best model for each covariate are provided. Best covariate model for includes logAGE on all parameters except $V1$, logBSA on Cl, logLBM on $V1$, and SEX on $V2$ ($\mathcal{M}_2^{COV}$). Block-structured correlation with blocks up to size 2 are compared (i.e. up to one correlation term). The best three models are with a correlation between parameters Cl and $V2$ (BICc=1082.9), between parameters Cl and $Q2$ (BICc=1093.8), and between parameters $V2$ and $Q2$ (BICc=1072.0). The latter correlation model is selected for $\mathcal{M}_2^{CORR}$. Residual error model *combined*$_2$ remains the best one ($\mathcal{M}_2^{ERR}$). These covariate, correlation and error models are then passed on to run 2.

- **Run 2 (BICc=6956.9) + Iteration 3:** Model $\mathcal{M}_2$ is fitted to data and individual parameters are sampled. Best covariate model includes logAGE on all parameters except $V1$, logBSA on Cl, logLBM on $V1$ and $V2$ ($\mathcal{M}_3^{COV}$). Block-structured correlation with blocks up to size 3 are compared (i.e. up to three correlation terms), a correlation block is selected between Cl, $Q2$ and $V2$ ($\mathcal{M}_3^{CORR}$). Residual error model *combined*$_2$ remains the best one ($\mathcal{M}_3^{ERR}$). These covariate, correlation and error models are then passed on to run 3.

- **Run 3 (BICc=6903.4) + Iteration 4:** Model $\mathcal{M}_3$ is fitted to data and individual parameters are sampled. Of note, regarding the correlation model selection, block-structured correlation with blocks up to size 4 are compared (i.e. up to six correlation terms). During this iteration, the same model as the one in the previous iteration is selected ($\mathcal{M}_4 = \mathcal{M}_3$) resulting in the stopping of the procedure. Model $\mathcal{M}_3$ is therefore the final model selected with the SAMBA procedure.

Converging toward a global optimal model. Even if the selected criterion decreases at each iteration, there is no guarantee that SAMBA converges toward a global minimum of this criterion. The quality and the robustness of the convergence of SAMBA can then be assessed by running SAMBA several times from different starting models. In particular, a good practice is to: 1/ launch SAMBA from several initial models 2/ compare the best models found (if there is not only one) in term of objective criterion (e.g. BICc) and 3/ make a thorough analysis and interpretation of the nearby models in order to choose the most relevant one for a given application. Regarding the choice of the starting model, similarly to the EM and SAEM algorithms, there is no optimal choice [3, 4]. We recommend to test in priority the following three starting models: 1/ an empty model, 2/ (when possible) a complete model, and 3/ a model (or models) that make sense for the biological application. Note that this robustness assessment is standard for all non-convex optimization algorithms and should also be performed for SCM and COSSAC in routine.

Performances on real examples, comparison with the SCM and COSSAC procedures

To assess the performances of the SAMBA procedure compared to SCM and COSSAC procedures, we replicate the illustration provided in [2]. We applied the three routines to a collection of 10 representative datasets, including pharmacokinetics, pharmacodynamics, and disease models. Of note, the SCM method for variable selection used here is exactly the same as the one implemented in PsN (Pearl Speaks NONMEM), differences lie in the algorithms used to estimate the parameters of a model and to calculate the likelihood. We restricted the SAMBA procedure to the covariate model selection as correlation and error model selection are not implemented in COSSAC and SCM. The results can be found in Table 1.

Because the datasets are real data illustrations, there is no ”true” model. It is only possible to compare them in term of BIC. Out of 10 examples, the same best model was proposed by the three procedures in four examples. In two examples, the best model selected by SAMBA was better in term of BICc than with SCM and COSSAC (Theophylline Ext. Rel. and Warfarin PK/PD). In three other examples the model with lowest BICc was not selected by SAMBA. However, the difference in BICc was respectively smaller than 6 in comparison with SCM procedure and 4.2 in comparison with COSSAC procedure. We insist on the fact that a difference in BICc does not necessarily have any biological meaning. This is an arbitrary criterion that allows to quantify the goodness of fit with respect to the sparsity of the model chosen. We thus argue that the three procedures lead to rather similar models which all constitute very good starting points for the modeler to build a model based on biological hypothesis. Finally, in only one example discussed below, the difference in BICc was larger than 10 points of BICc both compared to SCM and COSSAC procedures.

Regarding the Cholesterol dataset, we run again the SAMBA procedure starting from a full model in which all covariates are supposed to have an effect on all parameters. The new model selected by SAMBA is the full model with an effect of logAGE on ($Chol0$, $slope$) and SEX on ($Chol0$, $slope$) is much closer in term of BICc than the one selected starting from an empty model (ΔBICc $=-2$). We can finally notice with this example that it is sometimes possible to improve the convergence of SAMBA by improving the convergence of SAEM. Indeed, using 10 Markov chains instead of only 1, SAMBA also finds the model selected by SCM and COSSAC. Finding the optimal settings that minimize computation time while maximizing the probability of finding the best model is an extremely difficult problem that remains open. We can claim that the default settings used in Rsmlx and Monolix give very good results in most cases, but not in all cases with absolute certainty.

In terms of computational effort, it is important to note that the SAMBA procedure completes the model building process in much less runs, hence much less CPU time than SCM and COSSAC. In the considered problems, the number of runs and the CPU computation time are equivalent since the other computation times are negligible in the order of a few seconds. Actually, the computation times are 6 to 149 smaller than for SCM and 2 to 11 times smaller than for COSSAC. Note that the number of evaluations required by SAMBA is always lower or equal to the number of evaluations performed by COSSAC and SCM.

Simulation study

Data generation and analysis. We simulated data from a 1-compartment pharmacokinetic model. The model has three population parameters $ka_{\text{pop}} = 1$, $V_{\text{pop}} = 10$ and $Cl_{\text{pop}} = 2$. All individual parameters are log-normally distributed around the population parameters ($\omega_{ka} = 0.2$, $\omega_V = 0.3$ and $\omega_{Cl} = 0.3$. We simulated five individual covariates $(C_1, C_2, C_3, C_4, C_5)$ from standard normal distributions. The covariate model is such that there only exist linear relationships between $\log(V)$ and C_1 ($\beta_{V,1} = 0.2$), $\log(Cl)$ and C_1 ($\beta_{Cl,1} = -0.2$), $\log(Cl)$ and C_2 ($\beta_{Cl,2} = 0.3$). The correlation model is such that there exists a linear correlation between η_V and η_{Cl} ($\rho_{V,Cl} = 0.6$). Finally, the error model is a $combined_2$ model with $a = 2$ and $b = 0.1$. A clinical trial could then be simulated by generating PK data from this model for 100 individuals and 11 time-points (0.25, 0.5, 1, 2, 5, 8, 12, 16, 20, 24, 30). In order to evaluate the properties of SAMBA by Monte Carlo, we simulated 100 replicates of the same trial and built the model for each replicate using SAMBA as implemented in *Rsmlx* and Monolix for minimizing BICc. The initial model didn't include any covariate-parameter relationship and any correlation between random effect. The initial residual error model was a $combined_1$ model. The R code used for this Monte-Carlo study is available as Supplementary material.

Performances. Table 2 summarizes the results obtained for the covariate model selection. On the one hand, we can see that, for this particular example, SAMBA finds the 3 existing covariate-parameter relationships in 100% of the cases. On the other hand, very few spurious relationships are detected (less than 2%). Importantly, in all cases for which the final covariate model included more covariates than the true model M^*, the BICc of the selected model was lower than that of M^* (the differences ranging from 3 to 14.7 with *Rsmlx* and from 2.4 to 14.6 for Monolix). In other words, SAMBA always finds a covariate model as good or better than M^* in terms of BICc. Regarding the selection of the correlation model, the correct model was selected for all the replicates. Finally the correct error model was selected in 86% of the times with *Rsmlx* and 85% of the times with Monolix. Note that all the wrong selected error models were all $combined_1$ model (instead of $combined_2$) with a slightly larger BICc most of the time. Actually, these two models are quite similar and difficult to distinguish on the basis of a criterion like BICc. SAMBA then may get stuck in a local minimum in such a situation. Finally and importantly, the final selected models obtained with *Rsmlx* and Monolix are different in only 6% of cases. These small differences are due to small differences in the implementation of the algorithm (see the Discussion section for more details).

Discussion

This paper presents a novel model building procedure which offers covariate, correlation, and error model selection. It is fast as it requires only a limited number of runs of population parameter estimation and simulation compared to SCM and COSSAC. It allows to explore the space of models rapidly and provides to the modeler a very good model in term of the selection criterion. However, we insist on the fact that this procedure does not aim at replacing model building based on biological knowledge, which is in essence the strength of mechanistic modeling. Thus, it should not be blindly used and the best - potentially few best - models should be interpreted and compared.

SAMBA is an efficient algorithm for minimizing an objective function. In this article, we do not aim at evaluating the quality of the criterion used for model selection [5]. What is of interest here is the convergence of SAMBA. As it is also the case for SCM and COSSAC, SAMBA may not converge to the global minimum. This is particularly the case when the amount of data is too small compared to the complexity of the model to build. This phenomenon will be particularly critical when the number of covariates is high and/or when these are highly correlated. We then strongly encourage the user to build strategies to assess the robustness of the results. Extensions of the proposed algorithm are possible but are outside the scope of this paper and constitute a possible new research direction.

When there is a large number of available covariates, COSSAC and mainly SCM often fail in finding the best model in a reasonable time. In this case, SAMBA represents a particularly appealing approach since the covariate model selection is based on a stepwise variable selection procedure for linear models, which is known to handle high-dimension problems. While stepwise AIC/BIC are designed to obtain a sparse estimator that works well on the training set, other methods such as the lasso [20], where the penalty is chosen with cross validation, is designed to obtain the sparse linear model that minimize the prediction error. A lasso type approach [10] can sometimes present better performances than an approach based on an information criterion such as AIC or BIC, in particular when the number of covariates is very high. However, it should be noted that the choice of the penalty parameter by cross-validation can be complicated to implement and require a large number of runs. This type of method could be alternatively implemented in the covariate selection procedure and compared in further works. Note finally that it would be interesting to study the behavior of SAMBA using the EBEs (corrected as proposed in [21, 22]), rather than the individual simulated parameters, to build the covariate model.

The SAMBA procedure is implemented the R Package *Rsmlx* in the function *buildmlx* [15]. Minimal required input is a Monolix project used as initial model. Additional arguments can be used to enable specific features (all not listed): select the components of the model to optimize between the covariate, correlation, and error model, restrict the number of parameters or covariates to use, select a specific objective criterion, etc. *Rsmlx* is on CRAN and the R code can be modified to investigate any of the alternative implementations mentioned above for a specific problem. Note that the execution of *Rsmlx* requires the Monolix software, since it is only an algorithm combining tasks implemented in Monolix. The R codes allowing to replicate the analyses of this article are available in supplementary material. All the illustration datasets can be downloaded from the Supporting Information Appendix S2 of [2].

Finally, the SAMBA procedure is also implemented in the Monolix-GUI software starting from version 2019. Implementation is similar to the one in *Rsmlx* with two noteworthy differences. First, for the selection of covariates, a stepwise procedure is used even if the number of covariates d is small. Second, compiling differences exist between C++ and R. The full SAMBA procedure is available in the model building perspective, under a task called automatic statistical model building method. A single iteration of the SAMBA procedure is also proposed in the section Proposal in the tab Results after running a single estimation and simulation step for a model in Monolix [1].

Study Highlights

What is the current knowledge on the topic?

Existing model building methods for nonlinear mixed-effects models have high computational time, especially for selecting the covariate model.

What question did this study address?

The study describes the principle of the SAMBA (Stochastic Approximation for Model Building Algorithm) procedure which allow to build a covariate, a correlation, and an error model automatically and compares it with SCM (Stepwise Covariate Modeling) and COSSAC (COnditional Sampling use for Stepwise Approach) procedures.

What does this study add to our knowledge? SAMBA allows to select the best covariate model without having to fit the complete nonlinear mixed-effects model to the data for each possible covariate model. This study confirms that it is possible to obtain relevant information on the model we are looking for, even when another model is fitted to the data. This allows to drastically reduce the computation time with respect to other existing procedures while keeping the same performances. We also show that it is possible to perform correlation and error model selection in nonlinear mixed-effects models.

How might this change clinical pharmacology or translational science? This method will allow the practitioner to very quickly find a set of very good models in term of data fitting and parsimony, even when the number of parameters or the number of covariates available is large.

Author contribution

MP and ML wrote the manuscript, designed the research, performed the research and analyzed the data. ML is the main developer of the *Rsmlx* package.

Acknowledgements

This study has received funding from the Nipah virus project financed by the French Ministry of Higher Education, Research and Innovation.

References

[1] Monolix online documentation - Proposal tab description. `https://monolix.lixoft.com/tasks/proposal/`. Last Accessed: 2021-09-15.

[2] G. Ayral, J.-F. Si Abdallah, C. Magnard, and J. Chauvin. A novel method based on unbiased correlations tests for covariate selection in nonlinear mixed-effects

models–the COSSAC approach. *CPT: Pharmacometrics & Systems Pharmacology*, 2021. doi:10.1002/psp4.12612.

[3] J.-P. Baudry and G. Celeux. EM for mixtures. *Statistics and computing*, 25(4): 713–726, 2015. doi:10.1007/s11222-015-9561-x.

[4] C. Biernacki, G. Celeux, and G. Govaert. Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate gaussian mixture models. *Computational Statistics & Data Analysis*, 41(3-4):561–575, 2003. doi:10.1016/S0167-9473(02)00163-9.

[5] S. Buatois, S. Ueckert, N. Frey, S. Retout, and F. Mentré. Comparison of model averaging and model selection in dose finding trials analyzed by nonlinear mixed-effect models. *The AAPS journal*, 20(3):1–9, 2018. doi:10.1208/s12248-018-0205-x.

[6] M. Delattre, M. Lavielle, and M.-A. Poursat. A note on BIC in mixed-effects models. *Electronic journal of statistics*, 8(1):456–475, 2014. doi:10.1214/14-EJS890.

[7] E. I. George. The variable selection problem. *Journal of the American Statistical Association*, 95(452):1304–1308, 2000. doi:10.1080/01621459.2000.10474336.

[8] P. J. Green. Penalized likelihood. *Encyclopedia of Statistical Sciences*, 2:578–586, 1998. doi:10.1002/9781118445112.stat01595.

[9] T. Hastie and R. Tibshirani. Generalized additive models: some applications. *Journal of the American Statistical Association*, 82(398):371–386, 1987. doi:10.1080/01621459.1987.10478440.

[10] T. Hastie, R. Tibshirani, and R. Tibshirani. Best subset, forward stepwise or lasso? analysis and recommendations based on extensive comparisons. *Statistical Science*, 35(4):579–592, 2020. doi:10.1214/19-STS733.

[11] G. James, D. Witten, T. Hastie, and R. Tibshirani. *An introduction to statistical learning*, volume 112. Springer, 2013. doi:10.1007/978-1-4614-7138-7.

[12] E. N. Jonsson and M. O. Karlsson. Automated covariate model building within NONMEM. *Pharmaceutical research*, 15(9):1463–1468, 1998. doi:10.1023/a:1011970125687.

[13] K. G. Kowalski and M. M. Hutmacher. Efficient screening of covariates in population models using Wald's approximation to the likelihood ratio test. *Journal of pharmacokinetics and pharmacodynamics*, 28(3):253–275, 2001. doi:10.1023/a:1011579109640.

[14] M. Lavielle. *Mixed-effects models for the population approach: models, tasks, methods and tools*. CRC press, 2014. doi:10.1201/b17203.

[15] M. Lavielle. *Rsmlx: R Speaks 'Monolix'*, 2021. URL `http://rsmlx.webpopix.org`. R package version 3.0.3.

[16] M. Lavielle and B. Ribba. Enhanced method for diagnosing pharmacometric models: random sampling from conditional distributions. *Pharmaceutical research*, 33 (12):2979–2988, 2016. doi:10.1007/s11095-016-2020-3.

[17] J. W. Mandema, D. Verotta, and L. B. Sheiner. Building population pharmacoki-

netic/pharmacodynamic models. I. models for covariate effects. *Journal of pharmacokinetics and biopharmaceutics*, 20(5):511–528, 1992. doi:10.1007/BF01061469.

[18] C. F. Minto, T. W. Schnider, T. D. Egan, E. Youngs, H. J. Lemmens, P. L. Gambus, V. Billard, J. F. Hoke, K. H. Moore, D. J. Hermann, et al. Influence of age and gender on the pharmacokinetics and pharmacodynamics of remifentanil: I. model development. *The Journal of the American Society of Anesthesiologists*, 86(1):10–23, 1997. doi:10.1097/00000542-199701000-00004.

[19] T. Nguyen, M.-S. Mouksassi, N. Holford, N. Al-Huniti, I. Freedman, A. C. Hooker, J. John, M. O. Karlsson, D. Mould, J. P. Ruixo, E. Plan, R. Savic, J. van Hasselt, B. Weber, C. Zhou, E. Comets, and F. Mentré for the Model Evaluation Group of the International Society of Pharmacometrics (ISoP) Best Practice Committee. Model evaluation of continuous data pharmacometric models: metrics and graphics. *CPT: pharmacometrics & systems pharmacology*, 6(2):87–109, 2017. doi:10.1002/psp4.12161.

[20] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996. doi:10.1111/j.2517-6161.1996.tb02080.x.

[21] M. Yuan, X. S. Xu, Y. Yang, J. Xu, X. Huang, F. Tao, L. Zhao, L. Zhang, and J. Pinheiro. A quick and accurate method for the estimation of covariate effects based on empirical bayes estimates in mixed-effects modeling: correction of bias due to shrinkage. *Statistical methods in medical research*, 28(12):3568–3578, 2019. doi:10.1177/0962280218812595.

[22] M. Yuan, Z. Zhu, Y. Yang, M. Zhao, K. Sasser, H. Hamadeh, J. Pinheiro, and X. S. Xu. Efficient algorithms for covariate analysis with dynamic data using nonlinear mixed-effects model. *Statistical Methods in Medical Research*, 30(1):233–243, 2021. doi:10.1177/0962280220949898.

Figures and Legends

Figure 1: Scheme of the SAMBA model building procedure.

Figure 2: Step-by-step SAMBA procedure on the remifentanil example with 6 covariates (SEX, logAGE, logBSA, logHT,logLBM, logWT) and 6 model parameters (Cl, $Q2$, $Q3$, $V1$, $V2$, $V3$). For each selection (covariate, correlation, error model), the three best models in term of BICc are displayed. Non selected models are in white, newly accepted models are in darker grey, and models which have been already accepted at previous run are in lighter grey.

Table 1: Comparison of the SAMBA procedure with the SCM and COSSAC procedure on 10 representative datasets.

Table 2: Performance of the SAMBA algorithm for the selection of the covariate model in a simulation study using a one-compartment PK model. 100 individuals with 11 observations each have been generated. True model $\mathcal{M}^*$ includes an effect of C_1 on V and Cl and an effect of C_2 on Cl. The percentages of times (over 100 replicates) each covariate-parameter relationship is selected in the final model are displayed. Implementation of SAMBA in *Rsmlx* and Monolix are compared.

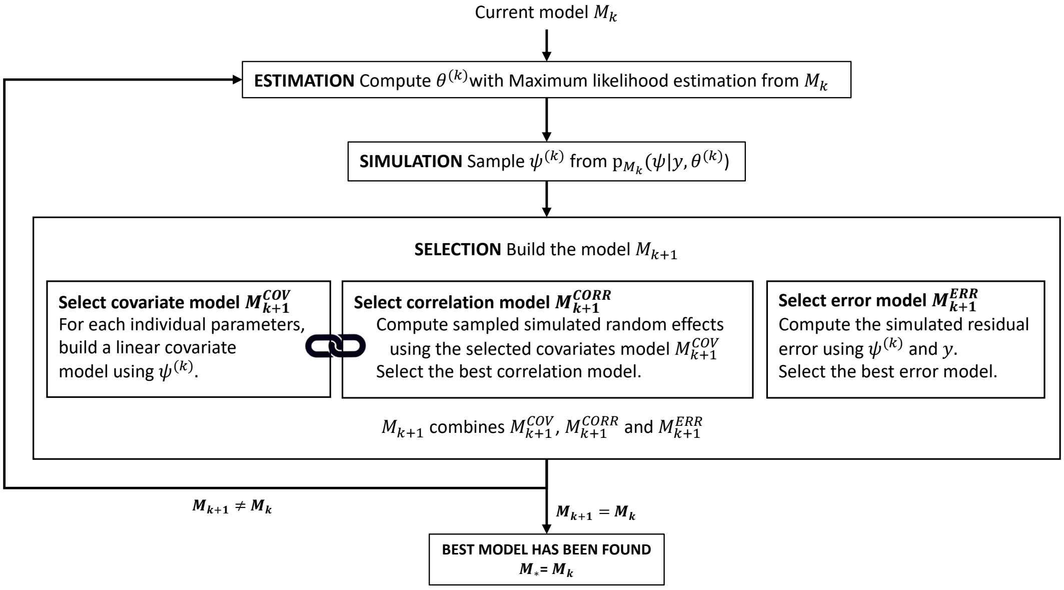
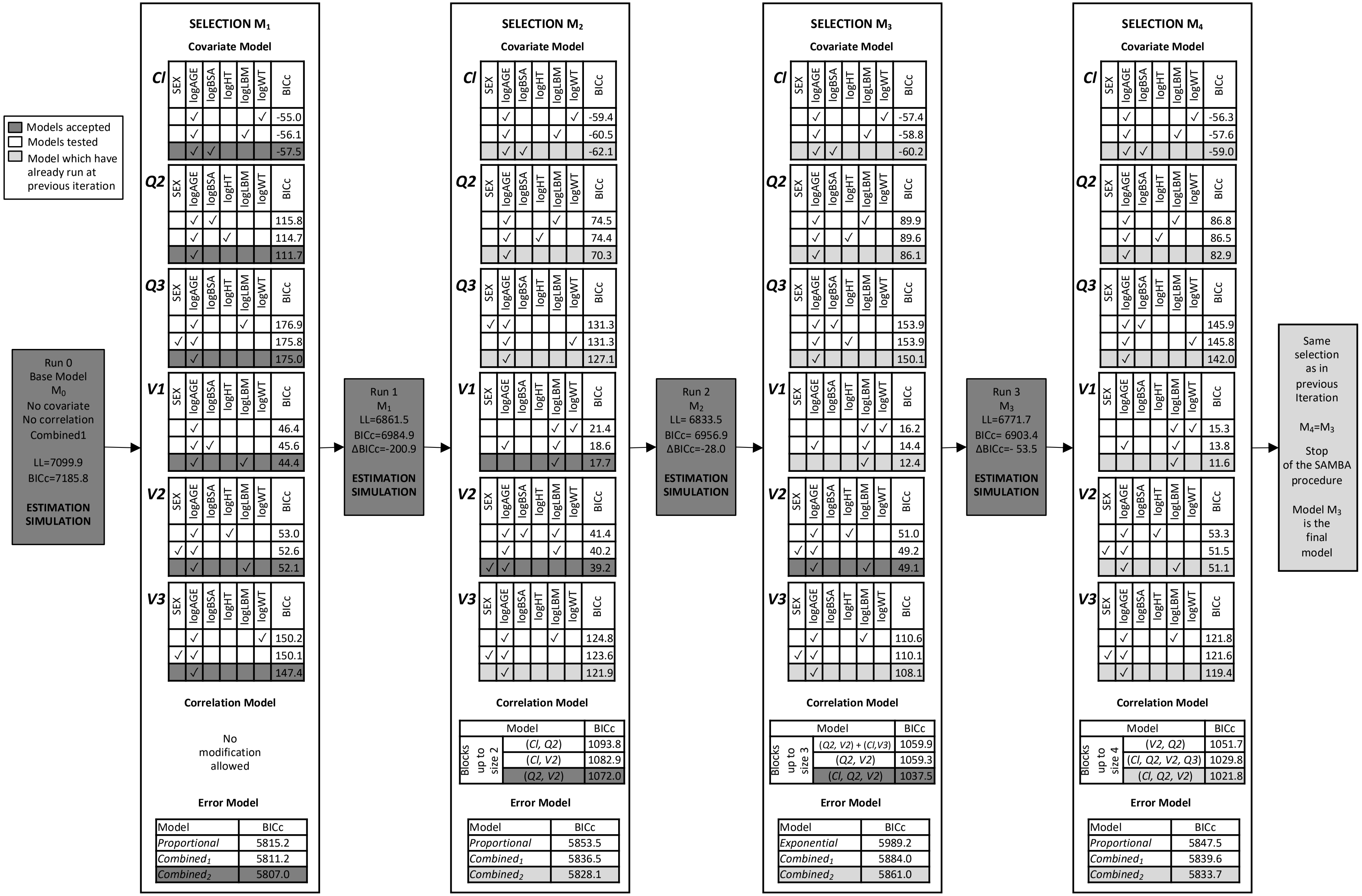

Current model M_k
ESTIMATION Compute $\theta^{(k)}$with Maximum likelihood estimation from M_k
SIMULATION Sample $\psi^{(k)}$ from $p_{M_k}(\psi|y,\theta^{(k)})$
SELECTION Build the model M_{k+1}
Select covariate model M_{k+1}^{COV}
For each individual parameters, build a linear covariate model using $\psi^{(k)}$.
Select correlation model M_{k+1}^{CORR}
Compute sampled simulated random effects using the selected covariates model M_{k+1}^{COV}
Select the best correlation model.
Select error model M_{k+1}^{ERR}
Compute the simulated residual error using $\psi^{(k)}$ and y.
Select the best error model.
M_{k+1} combines M_{k+1}^{COV}, M_{k+1}^{CORR} and M_{k+1}^{ERR}
$M_{k+1} \neq M_k$
$M_{k+1} = M_k$
BEST MODEL HAS BEEN FOUND
$M_* = M_k$

Figure 1: Legend Figure 1



SELECTION M$_1$

Covariate Model

Cl

SEX	logAGE	logBSA	logHT	logLBM	logWT	BICc
	✓				✓	-55.0
	✓			✓		-56.1
■	✓	✓				-57.5

Q2

SEX	logAGE	logBSA	logHT	logLBM	logWT	BICc
	✓	✓				115.8
	✓		✓			114.7
■	✓					111.7

Q3

SEX	logAGE	logBSA	logHT	logLBM	logWT	BICc
	✓			✓		176.9
✓	✓					175.8
■	✓					175.0

V1

SEX	logAGE	logBSA	logHT	logLBM	logWT	BICc
	✓					46.4
	✓	✓				45.6
■	✓			✓		44.4

V2

SEX	logAGE	logBSA	logHT	logLBM	logWT	BICc
	✓		✓			53.0
✓	✓					52.6
■	✓			✓		52.1

V3

SEX	logAGE	logBSA	logHT	logLBM	logWT	BICc
	✓				✓	150.2
✓	✓					150.1
■	✓					147.4

Correlation Model

No modification allowed

Error Model

Model	BICc
Proportional	5815.2
Combined$_1$	5811.2
■ *Combined$_2$*	5807.0

SELECTION M$_2$

Covariate Model

Cl

SEX	logAGE	logBSA	logHT	logLBM	logWT	BICc
	✓				✓	-59.4
	✓			✓		-60.5
▨	✓	✓				-62.1

Q2

SEX	logAGE	logBSA	logHT	logLBM	logWT	BICc
	✓			✓		74.5
	✓		✓			74.4
▨	✓					70.3

Q3

SEX	logAGE	logBSA	logHT	logLBM	logWT	BICc
✓	✓					131.3
	✓				✓	131.3
▨	✓					127.1

V1

SEX	logAGE	logBSA	logHT	logLBM	logWT	BICc
				✓	✓	21.4
	✓			✓		18.6
■				✓		17.7

V2

SEX	logAGE	logBSA	logHT	logLBM	logWT	BICc
	✓	✓		✓		41.4
	✓			✓		40.2
■ ✓	✓					39.2

V3

SEX	logAGE	logBSA	logHT	logLBM	logWT	BICc
	✓			✓		124.8
✓	✓					123.6
▨	✓					121.9

Correlation Model

	Model	BICc
Blocks up to size 2	*(Cl, Q2)*	1093.8
	(Cl, V2)	1082.9
	■ *(Q2, V2)*	1072.0

Error Model

Model	BICc
Proportional	5853.5
Combined$_1$	5836.5
▨ *Combined$_2$*	5828.1

SELECTION M$_3$

Covariate Model

Cl

SEX	logAGE	logBSA	logHT	logLBM	logWT	BICc
	✓				✓	-57.4
	✓			✓		-58.8
▨	✓	✓				-60.2

Q2

SEX	logAGE	logBSA	logHT	logLBM	logWT	BICc
	✓			✓		89.9
	✓		✓			89.6
▨	✓					86.1

Q3

SEX	logAGE	logBSA	logHT	logLBM	logWT	BICc
	✓	✓				153.9
	✓		✓			153.9
▨	✓					150.1

V1

SEX	logAGE	logBSA	logHT	logLBM	logWT	BICc
				✓	✓	16.2
	✓			✓		14.4
▨				✓		12.4

V2

SEX	logAGE	logBSA	logHT	logLBM	logWT	BICc
	✓		✓			51.0
✓	✓					49.2
■	✓			✓		49.1

V3

SEX	logAGE	logBSA	logHT	logLBM	logWT	BICc
	✓			✓		110.6
✓	✓					110.1
▨	✓					108.1

Correlation Model

	Model	BICc
Blocks up to size 3	*(Q2, V2)* + *(Cl,V3)*	1059.9
	(Q2, V2)	1059.3
	■ *(Cl, Q2, V2)*	1037.5

Error Model

Model	BICc
Exponential	5989.2
Combined$_1$	5884.0
▨ *Combined$_2$*	5861.0

SELECTION M$_4$

Covariate Model

Cl

SEX	logAGE	logBSA	logHT	logLBM	logWT	BICc
	✓				✓	-56.3
	✓			✓		-57.6
▨	✓	✓				-59.0

Q2

SEX	logAGE	logBSA	logHT	logLBM	logWT	BICc
	✓			✓		86.8
	✓		✓			86.5
▨	✓					82.9

Q3

SEX	logAGE	logBSA	logHT	logLBM	logWT	BICc
	✓	✓				145.9
	✓				✓	145.8
▨	✓					142.0

V1

SEX	logAGE	logBSA	logHT	logLBM	logWT	BICc
				✓	✓	15.3
	✓			✓		13.8
▨				✓		11.6

V2

SEX	logAGE	logBSA	logHT	logLBM	logWT	BICc
	✓		✓			53.3
✓	✓					51.5
▨	✓			✓		51.1

V3

SEX	logAGE	logBSA	logHT	logLBM	logWT	BICc
	✓			✓		121.8
✓	✓					121.6
▨	✓					119.4

Correlation Model

	Model	BICc
Blocks up to size 4	*(V2, Q2)*	1051.7
	(Cl, Q2, V2, Q3)	1029.8
	▨ *(Cl, Q2, V2)*	1021.8

Error Model

Model	BICc
Proportional	5847.5
Combined$_1$	5839.6
▨ *Combined$_2$*	5833.7

Figure 2: Legend Figure 2

		SCM		COSSAC		SAMBA		ΔBICc	
Dataset	Characteristics	#Runs[2]	Final Model[1]	#Runs[2]	Final Model[1]	#Runs[2]	Final Model[1]	SAMBA-SCM	SAMBA-COSSAC
Warfarin Linear PK	32 ind. - 247 obs. 4 param. - 3 cov. 4 re - 1 outcome	44	logtWt - V, Cl logtAge - C	4	Identical	2	Identical	0	0
Remifentanil Linear PK	65 ind. - 1992 obs. 6 param. - 6 cov. 4 re - 1 outcome	295	logLBM - $V1$ logAGE - Cl,$Q2$,$\boldsymbol{Q3}$,$V2$,$V3$ logBSA - Cl logHT - $\boldsymbol{V2}$	13	logLBM - $V1$,$\boldsymbol{V2}$ logAGE - Cl,$Q2$,$V2$,$V3$ logBSA - Cl SEX - $\boldsymbol{V3}$	4	logLBM - $V1$ logAGE - Cl,$Q2$,$\boldsymbol{Q3}$,$V2$,$V3$ logBSA - Cl SEX - $\boldsymbol{V2}$	0.8	0.5
Theophylline Linear PK	12 ind. - 20 obs. 3 param. - 2 cov. 4 re - 1 outcome	12	logtWEIGHT - $\boldsymbol{ka}$	4	Identical	2	Identical	0	0
Quinidine Sparse PK	136 ind. - 361 obs. 3 param. - 2 cov. 3 re - 1 outcome	22	none	11	Identical	1	Identical	0	0
Tobramycin Sparse PK	97 ind. - 322 obs. 3 param. - 2 cov. 2 re - 1 outcome	22	logCLCR - Cl logWT - $\boldsymbol{V}$	6	logCLCR - Cl logWT - $\boldsymbol{V}$	2	logCLCR - Cl logWT - $\boldsymbol{Cl}$	4.2	4.2
Theophylline Ext. Rel. Linear PK	18 ind. - 362 obs. 7 param. - 3 cov. 7 re - 1 outcome	98	logWT - $\boldsymbol{Tlag1, V}$	8	logWT - $\boldsymbol{Tlag1}$ logAGE - $\boldsymbol{ka2}$	6	logWT - $\boldsymbol{F, V}$ logAGE - $\boldsymbol{F}$ logHT - $\boldsymbol{ka1, ka2, Tlag1, diffTlag2}$	-11.7	-27
Warfarin PK/PD Joint	32 ind. - 247+232 obs. 8 param. - 3 cov. 8 re - 2 outcomes	92	logWT - Cl	10	logWT - Cl	2	logWT -Cl,$\boldsymbol{V}$ logAGE - $\boldsymbol{Cl, R0}$	-1.4	-1.4
Cholesterol Disease Progression	200 ind. - 1044 obs. 2 param. - 2 cov. 2 re - 1 outcome	12	logAGE - $Chol0$,$\boldsymbol{slope}$ SEX - $\boldsymbol{slope}$	5	logAGE - $Chol0$,$\boldsymbol{slope}$ SEX - $\boldsymbol{slope}$	2	logAGE - $Chol0$	13.5	13.5
Alzheimer Sparse PK	896 ind. - 3707 obs. 2 param. - 7 cov. 2 re - 1 outcome	73	APOE - $alpha$,$p0$ logAGE - $p0$,$\boldsymbol{alpha}$ logBMI - $\boldsymbol{alpha}$ logWT - $p0$	8	APOE - $alpha$,$p0$ logAGE - $p0$,$\boldsymbol{alpha}$ logBMI - $\boldsymbol{alpha}$ logWT - $p0$	2	APOE - $alpha$,$p0$ logAGE - $p0$ logWT - $p0$	6	1.5
Tranexamic PK	166 ind. - 817 obs. 4 param. - 10 cov. 4 re - 1 outcome	298	GROUP - Cl,$V2$ logBMI - Cl logCOCK - Cl logLBW - Q logWeight - V2	12	Identical	2	Identical	0	0

[1] Differences of variable selection between different methods are highlighted in bold.

[2] The number of runs is defined as the number of time the estimation and the simulation steps are performed (which is the most time-consuming).

Table 1: Legend Table 1

	Rsmlx			Monolix		
Covariates	ka	V	Cl	ka	V	Cl
C_1	2	100	100	2	100	100
C_2	0	1	100	0	1	100
C_3	1	2	1	2	2	1
C_4	0	3	4	0	3	4
C_5	0	1	1	1	2	1

Table 2: Legend Table 2.