Eric Houn gla Adjakossa^{1,2} / Norbert Mahouton Hounkonnou¹ / Grégory Nuel²

Computationally Stable Estimation Procedure for the Multivariate Linear Mixed-Effect Model and Application to Malaria Public Health Problem

¹ International Chair in Mathematical Physics and Applications (ICMPA-UNESCO Chair), Université d'Abomey-Calavi, Cotonou, Benin, E-mail: ericadjakossah@gmail.com, hounkonnou@yahoo.fr

² Laboratoire de Probabilités, Statistique et Modélisation (UMR 8001), Sorbonne Université, Paris, France, E-mail: ericadjakossah@gmail.com, Gregory.Nuel@math.cnrs.fr

Abstract:

In this paper, we provide the ML (Maximum Likelihood) and the REML (REstricted ML) criteria for consistently estimating multivariate linear mixed-effects models with arbitrary correlation structure between the random effects across dimensions, but independent (and possibly heteroscedastic) residuals. By factorizing the random effects covariance matrix, we provide an explicit expression of the profiled deviance through a reparameterization of the model. This strategy can be viewed as the generalization of the estimation procedure used by Douglas Bates and his co-authors in the context of the fitting of one-dimensional linear mixed-effects models. Beside its robustness regarding the starting points, the approach enables a numerically consistent estimate of the random effects covariance matrix while classical alternatives such as the EM algorithm are usually non-consistent. In a simulation study, we compare the estimates obtained from the present method with the EM algorithm-based estimates. We finally apply the method to a study of an immune response to Malaria in Benin.

Keywords: multivariate linear mixed-effects model, profiled deviance, ML and REML criteria, numerical consistency

DOI: 10.1515/ijb-2017-0076

Received: October 3, 2017; **Revised:** February 7, 2019; **Accepted:** May 5, 2019

1 Introduction

The linear mixed-effects model [1, 1–5] has become a popular tool for analyzing univariate multilevel data which arise in many areas (biology, medicine, economy, etc.) thanks to its flexibility to model the correlation contained in these data, and the availability of reliable and efficient software packages for fitting it [6–9]. Univariate multilevel data are referred to observations (or measurements) of a single variable of interest on several levels. For instance, schools grouped by village, and villages grouped at the city-level. In this context, multiple variables of interest measured at multiple levels characterize multivariate multilevel data. Examples include exam or test scores recorded for students across time and multi-dimensional subjects (e.g. maths, biology, French, etc.). In some situations where attention is focused on the analysis of multivariate multilevel data, some research questions can only be answered in a joint analysis of two or more variables, not separately [10]. A multivariate extension of a single response variable-based linear mixed-effects model is indeed having increasing popularity as a flexible tool for the analysis of multivariate multilevel data [11–14].

For example, in the context of the malaria immune study, fourteen antibodies against malaria have been measured on children exposed to malaria-infected mosquitoes in certain Benin republic villages from West Africa. One of the objectives of the malaria study was to evaluate the effect of the malaria infection on the child's immune against the disease. The antigens which characterize the child's immune status interact together, and it makes more sense to evaluate the effect of the malaria infection on all these antigens, simultaneously. The only valid way of doing this evaluation is to analyze the characteristics of the joint distribution of these antigens, conditional on the malaria infection. In [15] the authors pointed out the drawback of performing a separate

Eric Houn gla Adjakossa is the corresponding author.

© 2019 Walter de Gruyter GmbH, Berlin/Boston.

This content is free.

analysis instead. The malaria study is further described below and used as an illustration of the method at the end of the paper.

For the one-dimensional linear mixed-effects model, many methods for obtaining the estimates of the fixed and the random effects have been proposed in the literature. These methods include Henderson's mixed model equations [16], approaches proposed by [17] as well as techniques based on two-stage regression. Searle et al. [18] and [19] gave more details of the latter. Concerning the variance parameters estimation in linear mixed-effects model, the discussed methods in the literature include the ANOVA method for balanced data which uses the expected mean squares approach [20, 21]. For unbalanced data, [22] proposed the minimum norm quadratic estimation (MINQUE) method, where the resulting estimates are translation invariant under unbiased quadratic forms of the observations. Lee and Nelder [23] suggested another method of estimating variance parameters using extended quasi-likelihood, meaning gamma-log generalized linear models. Gumedze and Dunne [24] presented more details on the parameters' estimation methods in the linear mixed-effects model. The Maximum Likelihood (ML) and the Restricted Maximum Likelihood (REML) methods are the most popular estimation methods in the linear mixed-effects model [25]. The main attraction of these methods is the ability to handle a much wider class of variance models than simple variance components [24].

In the multivariate linear mixed-effects model, ML and REML estimates are frequently obtained through iterative schemes such as the EM algorithm [14, 26–29]. This avoids the difficulties related to the direct calculation of the parameters' likelihood, without increasing the implementation burden. Despite the existence of valid theorems proving the asymptotic convergence of the iterative sequences of these algorithms [27], the practical reality is often more complex. Verbeke et al. [10] explained that the main problem in fitting multivariate mixed-effects models is the loss of numerical stability due to the increasing of the dimension of the response variable. Indeed, the increase of the dimension of the response variable obviously leads to that of the dimension of the parameters vector of the model. Adjakossa et al. [15] illustrate through simulations studies that the random effects covariance matrix obtained by the EM algorithm has the worst Mean Square Error (MSE) among the parameters. In the same paper, the authors also gave a direct estimate of the random effects covariance matrix.

In this paper, we focus on the multivariate linear mixed-effects model with arbitrary correlation structure between the Gaussian random effects across dimensions, but independent (and possibly heteroscedastic) Gaussian residuals. In order to gain in numerical stability along with a better MSE for the variance parameter estimates, inspiring from [6] approach in the one-dimensional case, we optimize the profiled deviance which is obtained by partially maximizing the likelihood in the fixed effect parameters. This choice is justified by the fact that the fixed-effects vector is generally greater in dimension than the vector of variance parameters. Furthermore, avoiding a direct optimization on this vector of fixed-effects may decrease the complexity of the resulting optimization problem. Finally, the profiled deviance is a function of a small number of parameters that is easily optimized by any box-constrained optimizer, since some of the variance parameters are constrained. Our approach consists in directly calculating the likelihood of the model parameters after a reparametrization. This likelihood is used to obtain ML or REML estimates through the provided profiled deviance or the REML criterion, respectively. This strategy may explain the high quality of the estimates for the fixed effects parameters and the random effects variance parameters as well as the residual variance parameters. This approach may be viewed as the multi-dimensional generalization of the lme4 R package [6].

2 Multivariate linear mixed-effects model

Throughout the paper we adopt the following classical notations: $|\cdot|$ denotes the determinant of a matrix, $\|\cdot\|$ is the Euclidean norm, and $^\top$ is the matrix transpose symbol.

For the sake of simplicity, we focus on the bivariate case ($d=2$) in most of the paper, but the generalization to higher dimensions ($d>2$) is straightforward. Thus, in dimension 2, the model is the following:

$$\begin{cases} y_1 = X_1\beta_1 + Z_1\gamma_1 + \varepsilon_1 \\ y_2 = X_2\beta_2 + Z_2\gamma_2 + \varepsilon_2 \end{cases} \quad (1)$$

where

$$\gamma = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{\Gamma} = \begin{pmatrix} \Gamma_1 & \Gamma_{12} \\ \Gamma_{12}^\top & \Gamma_2 \end{pmatrix}\right), \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \sigma_1^2 \mathbf{I}_N & 0 \\ 0 & \sigma_2^2 \mathbf{I}_N \end{pmatrix}\right). \quad (2)$$

For $k \in \{1, 2\}$, $\beta_k \in \mathbb{R}^{p_k \times 1}$ and $\gamma_k \in \mathbb{R}^{q_k \times 1}$ denote respectively the fixed effect and the random effect vectors of covariates, while $\varepsilon_k \in \mathbb{R}^{N \times 1}$ is the marginal residual component in the dimension k of the model. Moreover, $y_k \in \mathbb{R}^{N \times 1}$ is the response variable in dimension k , $X_k \in \mathbb{R}^{N \times p_k}$ is a matrix of covariates and $Z_k \in \mathbb{R}^{N \times q_k}$ the random-effect design matrix. We denote $p = p_1 + p_2$ and $q = q_1 + q_2$.

The variance-covariance matrices of γ_1 and γ_2 are respectively $\Gamma_1 \in \mathbb{R}^{q_1 \times q_1}$ and $\Gamma_2 \in \mathbb{R}^{q_2 \times q_2}$, and must be positive semidefinite. It is then convenient to express the model in terms of the relative covariance factors $\Lambda_{\theta_1} \in \mathbb{R}^{q_1 \times q_1}$ and $\Lambda_{\theta_2} \in \mathbb{R}^{q_2 \times q_2}$ such that

$$\Gamma_1 = \sigma_1^2 \Lambda_{\theta_1} \Lambda_{\theta_1}^\top \quad \text{and} \quad \Gamma_2 = \sigma_2^2 \Lambda_{\theta_2} \Lambda_{\theta_2}^\top. \quad (3)$$

In eq. (3), σ_1^2 and σ_2^2 are the same marginal residual variances used in the model eq. (2). Using the variance-component parameters θ_1 and θ_2 , the marginal random effects γ_1 and γ_2 are expressed as

$$\gamma_1 = \Lambda_{\theta_1} u_1, \quad \gamma_2 = \Lambda_{\theta_2} u_2,$$

where

$$\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \sim \mathcal{N}(\mathbf{0}, \Sigma_{\mathbf{u}}), \quad \text{with} \quad \Sigma_{\mathbf{u}} = \begin{pmatrix} \sigma_1^2 \mathbf{I}_{q_1} & \sigma_1 \sigma_2 \boldsymbol{\rho} \\ \sigma_1 \sigma_2 \boldsymbol{\rho}^\top & \sigma_2^2 \mathbf{I}_{q_2} \end{pmatrix}, \quad (4)$$

and $\boldsymbol{\rho} = \text{diag}(\rho, \dots, \rho)$. For example, if we suppose that we have one intercept (I) and one slope (S) in each dimension, ρ is expressed as

$$\rho = \begin{pmatrix} \text{corr}(\gamma_1^I, \gamma_2^I) & \text{corr}(\gamma_1^I, \gamma_2^S) \\ \text{corr}(\gamma_1^S, \gamma_2^I) & \text{corr}(\gamma_1^S, \gamma_2^S) \end{pmatrix},$$

where γ_k^I and γ_k^S are the random intercept and slope, respectively, in the k th dimension of the model. The bivariate linear mixed-effects model can therefore be expressed as

$$\begin{cases} y_1 = X_1 \beta_1 + Z_1 \Lambda_{\theta_1} u_1 + \varepsilon_1 \\ y_2 = X_2 \beta_2 + Z_2 \Lambda_{\theta_2} u_2 + \varepsilon_2 \end{cases} \quad (5)$$

with

$$\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \sim \mathcal{N}(\mathbf{0}, \Sigma_{\mathbf{u}}), \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \sigma_1^2 \mathbf{I}_N & 0 \\ 0 & \sigma_2^2 \mathbf{I}_N \end{pmatrix}\right), \quad (6)$$

where $\mathbf{I}_N$ denotes the dimension N identity matrix. From now on, our objective is to estimate the parameter $\Theta = (\beta_1, \beta_2, \sigma_1^2, \sigma_2^2, \theta_1, \theta_2, \rho)$.

2.1 ML criterion

The objective of this section is to express the log-likelihood of the model defined in eqs. (5) and (6) given an observation $\mathbf{y}^\top = (y_1^\top, y_2^\top)$.

If we denote by $\mathbf{u}$ the random variable of the random effects $\mathbf{u}$, the conditional expectation $\mu_{\mathbf{u}|\mathbf{y}=\mathbf{y}}$ of $\mathbf{u}$ given that $\mathbf{y} = \mathbf{y}$ can be obtained jointly with the fixed effect estimate $\hat{\beta}_{\theta, \rho, \sigma}$ conditionally to $\theta = (\theta_1, \theta_2)$, ρ , and $\sigma = (\sigma_1^2, \sigma_2^2)$ with the following lemma:

Lemma 2.1.

Denoting

$$Y_\sigma = \left(\sqrt{\sigma_2^2} y_1^\top, \sqrt{\sigma_1^2} y_2^\top \right)^\top, \quad X_\sigma = \begin{pmatrix} \sqrt{\sigma_2^2} X_1 & \mathbf{0} \\ \mathbf{0} & \sqrt{\sigma_1^2} X_2 \end{pmatrix}, \quad Z_{\sigma\theta} = \begin{pmatrix} \sqrt{\sigma_2^2} Z_1 \Lambda_{\theta_1} & \mathbf{0} \\ \mathbf{0} & \sqrt{\sigma_1^2} Z_2 \Lambda_{\theta_2} \end{pmatrix},$$

$\hat{\beta}_{\theta, \rho, \sigma}$ and $\mu_{\mathbf{u}|\mathbf{y}=\mathbf{y}}$ satisfy

$$\begin{pmatrix} X_\sigma^\top X_\sigma & X_\sigma^\top Z_{\sigma\theta} \\ Z_{\sigma\theta}^\top X_\sigma & Z_{\sigma\theta}^\top Z_{\sigma\theta} + \sqrt{\sigma_1^2 \sigma_2^2} \Sigma_{\mathbf{u}}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\beta}_{\theta, \rho, \sigma} \\ \mu_{\mathbf{u}|\mathbf{y}=\mathbf{y}} \end{pmatrix} = \begin{pmatrix} X_\sigma^\top \\ Z_{\sigma\theta}^\top \end{pmatrix} Y_\sigma. \quad (7)$$

By using Lemma 2.1 the log-likelihood of Θ given $\mathbf{y}$ is expressed in the following theorem.

Theorem 2.1.

Assume that $\hat{\beta}_{\theta, \rho, \sigma}$ and $\mu_{\mathbf{u}|\mathbf{y}=\mathbf{y}}$ satisfy eq. (7). By using the notations of Lemma 2.1, let us denote by $r(\hat{\beta}_{\theta, \rho, \sigma}, \mu_{\mathbf{u}|\mathbf{y}=\mathbf{y}})$ the penalized residual sum of squares (knowing θ , ρ and σ) such that

$$r(\hat{\beta}_{\theta, \rho, \sigma}, \mu_{\mathbf{u}|\mathbf{y}=\mathbf{y}}) = \left\| Y_\sigma - X_\sigma \hat{\beta}_{\theta, \rho, \sigma} - Z_{\sigma\theta} \mu_{\mathbf{u}|\mathbf{y}=\mathbf{y}} \right\|^2 + \sigma_1^2 \sigma_2^2 \mu_{\mathbf{u}|\mathbf{y}=\mathbf{y}}^\top \Sigma_{\mathbf{u}}^{-1} \mu_{\mathbf{u}|\mathbf{y}=\mathbf{y}},$$

and $L_{\theta, \rho, \sigma}$ and R_X two matrices satisfying the following Cholesky decomposition, respectively:

$$L_{\theta, \rho, \sigma} L_{\theta, \rho, \sigma}^\top = Z_{\sigma\theta}^\top Z_{\sigma\theta} + \sqrt{\sigma_1^2 \sigma_2^2} \Sigma_{\mathbf{u}}^{-1},$$

and

$$\begin{pmatrix} X_{\sigma}^\top X_{\sigma} & X_{\sigma}^\top Z_{\sigma\theta} \\ Z_{\sigma\theta}^\top X_{\sigma} & L_{\theta, \rho, \sigma} L_{\theta, \rho, \sigma}^\top \end{pmatrix} = \begin{pmatrix} R_X & \mathbf{0} \\ R_{ZX} & L_{\theta, \rho, \sigma}^\top \end{pmatrix}^\top \begin{pmatrix} R_X & \mathbf{0} \\ R_{ZX} & L_{\theta, \rho, \sigma}^\top \end{pmatrix}.$$

Then, the log-likelihood of Θ can be expressed as:

$$\begin{aligned} \ell(\beta, \theta, \rho, \sigma | \mathbf{y}) = & - \frac{r(\hat{\beta}_{\theta, \rho, \sigma}, \mu_{\mathbf{y}|\mathbf{y}=\mathbf{y}}) + \|R_X(\beta - \hat{\beta}_{\theta, \rho, \sigma})\|^2}{2\sigma_1^2 \sigma_2^2} - \frac{N-q}{2} \log(\sigma_1^2 \sigma_2^2) \\ & - \frac{1}{2} \log(|\Sigma_{\mathbf{u}}|) - \frac{1}{2} \log(|L_{\theta, \rho, \sigma}|^2). \end{aligned} \quad (8)$$

The ML criterion expressed in eq. (8) is a differentiable convex function of β , and $\partial \ell(\beta, \theta, \rho, \sigma | \mathbf{y}) / \partial \beta = 0$ implies that $\beta = \hat{\beta}_{\theta, \rho, \sigma}$. Then the ML estimate of Θ can be obtained through the following corollary.

Corollary 2.1.

Consider the log-likelihood defined in eq. (8) and let $D_{\text{ML}}(\theta, \rho, \sigma | \mathbf{y})$ be the ML variance parameters profiled deviance, given by

$$\begin{aligned} D_{\text{ML}}(\theta, \rho, \sigma | \mathbf{y}) &= -2\ell(\hat{\beta}_{\theta, \rho, \sigma}, \theta, \rho, \sigma | \mathbf{y}) \\ &= \frac{r(\hat{\beta}_{\theta, \rho, \sigma}, \mu_{\mathbf{y}|\mathbf{y}=\mathbf{y}})}{\sigma_1^2 \sigma_2^2} + (N-q) \log(\sigma_1^2 \sigma_2^2) + \log(|\Sigma_{\mathbf{u}}|) + \log(|L_{\theta, \rho, \sigma}|^2), \end{aligned}$$

the ML estimators $\hat{\beta}, \hat{\theta}, \hat{\rho}, \hat{\sigma}$ of β, θ, ρ and σ are obtained by

$$(\hat{\theta}, \hat{\rho}, \hat{\sigma}) = \arg \min_{\theta, \rho, \sigma} D_{\text{ML}}(\theta, \rho, \sigma | \mathbf{y}) \quad \text{and} \quad \hat{\beta} = \hat{\beta}_{\hat{\theta}, \hat{\rho}, \hat{\sigma}}. \quad (9)$$

The procedure that we use to derive the ML estimate of Θ components is summarized in Corollary 2.1 through eq. (9). First, the ML estimate of the fixed-effects vector is analytically computed while supposing that all the variance parameters are known. This leads to $\hat{\beta}_{\theta, \rho, \sigma}$ which satisfies eq. (7). Then, plugging $\hat{\beta}_{\theta, \rho, \sigma}$ into the full likelihood yields the profiled deviance $D_{\text{ML}}(\theta, \rho, \sigma | \mathbf{y})$ which is a function of variance parameters only. The numerical task consisting in optimizing the deviance $D_{\text{ML}}(\theta, \rho, \sigma | \mathbf{y})$ is easier than optimizing the full likelihood $\ell(\beta, \theta, \rho, \sigma | \mathbf{y})$. Finally, we get $\hat{\beta} = \hat{\beta}_{\hat{\theta}, \hat{\rho}, \hat{\sigma}}$ once $\hat{\theta}, \hat{\rho}$ and $\hat{\sigma}$ are obtained.

2.2 REML criterion

The REML criterion [4] is obtained by integrating the likelihood of the model (given $\mathbf{y}$) with respect to the fixed-effects β :

$$\mathcal{L}(\theta, \rho, \sigma | \mathbf{y}) = \int \exp \ell(\beta, \theta, \rho, \sigma | \mathbf{y}) d\beta.$$

In eq. (8), we recognize the expression of a multivariate Gaussian distribution density in β (up to a normalization factor), and we establish that:

Theorem 2.2.

Consider the log-likelihood defined in eq. (8). By using the notations of Theorem 2.1, the log-REML profiled deviance defined by $D_{\text{REML}}(\theta, \rho, \sigma | \mathbf{y}) = -2\mathcal{L}(\theta, \rho, \sigma | \mathbf{y})$ is given by

$$\begin{aligned} D_{\text{REML}}(\theta, \rho, \sigma | \mathbf{y}) &= \frac{r(\hat{\beta}_{\theta, \rho, \sigma}, \mu_{\mathbf{y}|\mathbf{y}=\mathbf{y}})}{\sigma_1^2 \sigma_2^2} + (N-p-q) \log(\sigma_1^2 \sigma_2^2) + \log(|\Sigma_{\mathbf{u}}|) \\ &+ \log(|L_{\theta, \rho, \sigma}|^2) + \log(|R_X|^2), \end{aligned} \quad (10)$$

and the REML estimators $\hat{\beta}_R, \hat{\theta}_R, \hat{\rho}_R, \hat{\sigma}_R$ of β, θ, ρ and σ are characterized by

$$(\hat{\theta}_R, \hat{\rho}_R, \hat{\sigma}_R) = \arg \min_{\theta, \rho, \sigma} D_{\text{REML}}(\theta, \rho, \sigma | \mathbf{y}) \quad \text{and} \quad \hat{\beta}_R = \arg \max_{\beta} \ell(\beta, \hat{\theta}_R, \hat{\rho}_R, \hat{\sigma}_R | \mathbf{y}). \quad (11)$$

The ML and the REML deviance optimization tasks can be solved by using any box-constrained optimizer. In this paper, we use for example the optimization function *nlmminb* available in the R software.

3 Simulation studies

In this section, the computational stability of the procedure is illustrated through simulation studies, and the present estimation procedure is compared with the estimation procedure based on the EM algorithm. For the sake of simplicity, the simulation studies are performed using simulated bivariate longitudinal data sets. In the following paragraph, we explain how we chose the parameters that have been used to simulate the “working” data sets.

The working data sets. We suppose that we are following up a sample of subjects where the goal is to evaluate how the increase of the weight and the height of the individuals of this population are jointly explained by the sex, the score of nutrition (Nscore), and the age. Each subject has been seen n_i (i indicating the subject) times in the course of the survey where at each time his/her height, his/her weight and his/her nutrition score are measured. We randomly affect sex, age and nutrition score using uniform distributions: balanced sex ratio, $\mathcal{U}([18, 37])$ for the age, and $\mathcal{U}([20, 50])$ for the nutrition score. All the computations in this paper are done using the R software. The model under which we simulate the data for individual $i = 1, \dots, n$ is the following:

$$\begin{aligned} \text{weight}_i &= (1_{n_i}, \text{sex}_i, \text{Nscore}_i, \text{age}_i) \beta_1 + (1_{n_i}, \text{Nscore}_i) \gamma_{1i} + \varepsilon_{1i}, \\ \text{height}_i &= (1_{n_i}, \text{sex}_i, \text{Nscore}_i, \text{age}_i) \beta_2 + (1_{n_i}, \text{Nscore}_i) \gamma_{2i} + \varepsilon_{2i}, \end{aligned} \quad (12)$$

where

$$\gamma_i = \begin{pmatrix} \gamma_{1i} \\ \gamma_{2i} \end{pmatrix} \sim \mathcal{N}(\mathbf{0}, \bar{\Gamma}), \varepsilon_{1i} \sim \mathcal{N}(0, \sigma_1^2 \mathbf{I}_{n_i}), \varepsilon_{2i} \sim \mathcal{N}(0, \sigma_2^2 \mathbf{I}_{n_i}), \gamma_i \perp \varepsilon_{1i} \perp \varepsilon_{2i}. \quad (13)$$

In eq. (12), weight_i is a vector (of length n_i) that contains the successive weights for the subject i . The random effect related to the dependent variable *weight* or *height* is a vector composed of one random intercept and one random slope in the direction of the covariate *Nscore*. The total number of observations is denoted by N ($= n_1 + n_2 + \dots$). The parameters in the model described by eqs. (12) and (13) are $\beta_1, \beta_2, \bar{\Gamma}, \sigma_1$ and σ_2 .

The true values of $\beta_1, \beta_2, \sigma_1$ and σ_2 are reported in the first column of Table 2. The matrix $\bar{\Gamma}$ was chosen such that it is positive definite with the following form:

$$\bar{\Gamma} = \begin{pmatrix} \eta_1^2 & \rho_{\eta} \eta_1 \eta_2 & \rho_{\eta} \tau_1 & \rho_{\eta} \tau_2 \\ \rho_{\eta} \eta_1 \eta_2 & \eta_2^2 & \rho_{\eta} \tau_2 & \rho_{\eta} \tau_1 \\ \rho_{\eta} \tau_1 & \rho_{\eta} \tau_2 & \tau_1^2 & \rho_{\tau} \tau_1 \tau_2 \\ \rho_{\eta} \tau_2 & \rho_{\eta} \tau_1 & \rho_{\tau} \tau_1 \tau_2 & \tau_2^2 \end{pmatrix}.$$

In order to have an almost strong correlation between the marginal random effects, we set $\rho = 0.8$ and randomly chose all other parameters involved in the expression of $\bar{\Gamma}$. The matrix $\bar{\Gamma}$ that resulted from this procedure is

$$\bar{\Gamma} = \begin{pmatrix} 27.77 & 18.80 & 41.70 & 4.93 \\ 18.80 & 36.00 & 47.47 & 5.62 \\ 41.70 & 47.47 & 97.81 & 8.91 \\ 4.93 & 5.62 & 8.91 & 1.37 \end{pmatrix}. \quad (14)$$

3.1 Estimates' performances

One practical way to show the numerical consistency of an estimator is by computing its Mean Square Error (MSE). If the MSE of an estimator is asymptotically zero, this estimator converges in probability, and is then consistent. We simulated datasets with gradually increasing sample sizes where $n \in \{50, 60, 100, 300\}$ and $N \in \{600, 1000, 3000\}$. In order to account for the unbalanced nature of the longitudinal data we observe in practice, we used a multinomial distribution of the N observations among the n individuals under the constraint that $n_i \geq 2$ for all individual i (using the rejection method). For each combination of (n, N) we performed 100 replications, and derive empirical mean and 95 % CI (confidence interval) for the estimators' MSE.

One should note that the chosen simulation design is challenging since it usually requires $n_i/n \rightarrow \infty$ for reaching the asymptotic state [30]. Indeed, the condition $n_i/n \rightarrow \infty$ implies that $n_i > n$ which is seldom satisfied in practical studies. In our simulations, we do not have $n_i \geq n$ in general but the parameters nevertheless are generally well estimated as shown in Table 1. This illustrates not only the good practical properties of the parameter estimators but also the improved computational stability of our procedure.

Furthermore, it requires not only $n \rightarrow \infty$ and $N \rightarrow \infty$ but also $N/n \rightarrow \infty$ for reaching the asymptotic state [30]. This means that it requires a sufficient total number of observations, a sufficient number of levels for the

grouping factor and a sufficient number of observations for each level of the grouping factor. For example, in Table 1, when the total number of observations is $N = 1000$, the MSE of $\bar{\Gamma}$ is better for $n = 100$, 0.47 [0.03;1.29], than for $n = 300$, 0.69 [0.06;1.94]. Observing Table 1, it is clear that as the number of observations increases along with a reasonable number of subjects, the MSE tends to 0. The estimation procedure discussed in this paper may therefore be named Numerically Consistent Estimation (procedure) for the Multivariate Linear Mixed-Effects model (NCEmlme).

Table 1: MSE in the simulation study. Mean Square Error of estimators with 95 % CI estimated on 100 replications for values of $n \in \{50, 60, 100, 300\}$ and $N \in \{600, 1000, 3000\}$.

Parameter	n	$N = 600$	$N = 1000$	$N = 3000$
β_1	50	2.43 (0.11–7.11)	1.89 (0.22–4.89)	1.02 (0.07–2.44)
	60	2.57 (0.14–7.87)	2.13 (0.26–5.55)	0.77 (0.10–2.09)
	100	2.27 (0.16–5.61)	1.55 (0.14–5.17)	0.71 (0.04–1.85)
	300	3.16 (0.14–10.54)	1.70 (0.10–4.55)	0.51 (0.04–1.36)
β_2	50	5.50 (0.09–16.35)	3.26 (0.02–11.64)	2.06 (0.09–5.98)
	60	5.06 (0.12–15.09)	3.22 (0.02–10.24)	1.78 (0.10–5.62)
	100	4.33 (0.03–13.17)	2.37 (0.02–6.89)	1.06 (0.02–3.60)
	300	4.58 (0.18–15.92)	2.43 (0.05–7.39)	0.90 (0.04–2.88)
σ_1	50	0.03 (0.00–0.14)	0.02 (0.00–0.09)	0.00 (0.00–0.02)
	60	0.04 (0.00–0.11)	0.02 (0.00–0.08)	0.00 (0.00–0.02)
	100	0.03 (0.00–0.13)	0.01 (0.00–0.06)	0.00 (0.00–0.01)
	300	0.05 (0.00–0.23)	0.03 (0.00–0.13)	0.00 (0.00–0.02)
σ_2	50	0.04 (0.00–0.15)	0.03 (0.00–0.08)	0.00 (0.00–0.02)
	60	0.04 (0.00–0.18)	0.03 (0.00–0.12)	0.01 (0.00–0.03)
	100	0.06 (0.00–0.18)	0.03 (0.00–0.11)	0.01 (0.00–0.03)
	300	0.13 (0.00–0.45)	0.04 (0.00–0.15)	0.01 (0.00–0.05)
$\bar{\Gamma}$	50	0.90 (0.12–2.41)	0.62 (0.06–1.41)	0.45 (0.04–1.14)
	60	1.07 (0.06–2.64)	0.68 (0.08–1.98)	0.25 (0.03–0.66)
	100	0.90 (0.05–2.40)	0.47 (0.03–1.29)	0.21 (0.02–0.71)
	300	1.45 (0.19–4.39)	0.69 (0.06–1.94)	0.17 (0.02–0.57)

3.2 Comparison with EM-based estimates

In this section, we compare the estimation procedure based on the EM algorithm with NCEmlme. For both algorithms, we consider two initialization schemes: naive or advised. For the naive initialization the starting parameter is randomly chosen (without specific control), and for the advised one the starting values are those obtained by fitting separately each dimension of the bivariate model. For each replication, the starting values are the same for both the NCEmlme and EM-based algorithms. For each algorithm, the number of iterations required for convergence are reported too. Our methodology consists in simulating one hundred longitudinal datasets of size ($N = 3000$, $n = 300$) and fitting the model to each of these datasets using the EM-based algorithm and the NCEmlme, respectively. This allows to compute both the 95 % CI and the empirical mean of the one hundred estimates in each case (naive and advised starting values). The results are presented in Table 2 and Table 3.

Table 2 contains the empirical means of the estimates with their 95 % CI, and the minimum, the maximum and the average numbers of iterations. Table 3 contains the empirical relative error of the estimators with their 95 % CI. The results show that in the case of naive initialization, the NCEmlme estimators outperform the EM estimators. For example, the component of β_1 which is 14.00 is well estimated by NCEmlme, 14.02 [13.27 – 14.45] with an empirical relative error of 0.02 [0.00;0.04], but poorly estimated by EM, -2.05 [-4.70 ;0.40] with an empirical relative error of 1.14 [1.01;1.32]. In the case of the advised initialization, both NCEmlme and EM algorithms perform well, but NCEmlme converge faster (64 iterations in average) than EM (169 iterations in average). The number of iterations required by NCEmlme with advised initialization ranges from 48 to 89 compared to 56 to 103 for the naive initialization. The NCEmlme with naive initialization therefore needs more iterations than NCEmlme with advised initialization to converge. This is expected and may be explained by

the fact that the advised initialization values contain some information from the data of interest, and the naive starting points do not.

Table 2: Comparison of EM-based and NCEmlme estimates in the simulation study. EM-based estimates compared with NCEmlme estimates on the same data sets. Empirical estimates with their 95 % CI and the number of iterations required for convergence.

Parameter	Value	Naive initialization				Advised initialization			
		NCEmlme		EM		NCEmlme		EM	
		Emp. Mean	95 % CI	Emp. Mean	95 % CI	Emp. Mean	95 % CI	Emp. Mean	95 % CI
β_1	50.67 -4.80 14.00 2.70	50.79 -5.00 14.02 2.70	49.14 – 52.11 -8.39 – -3.66 13.27 – 14.45 2.66 – 2.72	13.47 -4.79 -2.05 2.69	-76.70 – 43.37 -8.08 – -3.42 -4.70 – -0.40 2.66 – 2.72	50.80 -5.02 14.02 2.70	49.15 – 52.12 -8.39 – -3.66 13.28 – 14.45 2.66 – 2.72	50.78 -4.98 14.02 2.70	49.09 – 52.01 -8.38 – -3.65 13.27 – 14.45 2.66 – 2.72
β_2	13.20 -2.80 27.00 1.70	13.65 -2.80 27.00 1.68	11.79 – 15.06 -4.74 – -0.43 26.87 – 27.10 1.64 – 1.71	-84.47 -2.75 0.90 1.68	-114.28 – -50.63 -4.90 – 0.21 -1.62 – 2.68 1.64 – 1.71	13.65 -2.81 27.00 1.68	11.79 – 15.07 -4.79 – -0.43 26.87 – 27.10 1.64 – 1.71	13.68 -2.85 27.00 1.68	11.81 – 15.14 -4.80 – -0.51 26.87 – 27.10 1.64 – 1.71
σ_1	5.80	5.79	5.62 – 5.92	5.78	5.64 – 5.98	5.78	5.64 – 5.92	5.79	5.65 – 5.94
σ_2	7.60	7.61	7.34 – 7.74	7.59	7.33 – 7.73	7.61	7.34 – 7.73	7.63	7.36 – 7.73
Nbr. of iteration	Min	56	-	63	-	48	-	14	-
	Mean	71	-	109	-	64	-	169	-
	Max	103	-	157	-	89	-	645	-

Table 3: Comparison of EM-based and NCEmlme relative errors on the estimates in the simulation study. EM-based estimates compared with NCEmlme estimates on the same data sets. Empirical relative error of the estimates with their 95 % CI.

Parameter	Value	Naive initialization				Advised initialization			
		NCEmlme		EM		NCEmlme		EM	
		R. Error	95 % CI	R. Error	95 % CI	R. Error	95 % CI	R. Error	95 % CI
β_1	50.67	0.01	0.00 – 0.03	0.73	0.01 – 2.18	0.01	0.00 – 0.03	0.01	0.00 – 0.03
	–4.80	0.21	0.02 – 0.32	0.21	0.00 – 0.32	0.21	0.02 – 0.32	0.21	0.00 – 0.33
	14.00	0.02	0.00 – 0.04	1.14	1.01 – 1.32	0.02	0.00 – 0.04	0.02	0.00 – 0.04
	2.70	0.00	0.00 – 0.01	0.00	0.00 – 0.01	0.00	0.00 – 0.01	0.00	0.00 – 0.01
β_2	13.20	0.07	0.00 – 0.14	7.39	4.19 – 9.40	0.07	0.00 – 0.14	0.07	0.00 – 0.14
	–2.80	0.43	0.00 – 0.84	0.43	0.02 – 1.07	0.43	0.00 – 0.84	0.43	0.00 – 0.81
	27.00	0.00	0.00 – 0.00	0.96	0.88 – 1.02	0.00	0.00 – 0.00	0.00	0.00 – 0.00
	1.70	0.01	0.00 – 0.02	0.01	0.00 – 0.02	0.01	0.00 – 0.02	0.01	0.00 – 0.03
σ_1	5.80	0.01	0.00 – 0.02	0.01	0.00 – 0.03	0.01	0.00 – 0.02	0.01	0.00 – 0.03
σ_2	7.60	0.01	0.00 – 0.02	0.01	0.00 – 0.04	0.01	0.00 – 0.02	0.01	0.00 – 0.02

The empirical mean of the random effects covariance matrix, $\bar{\Gamma}$, is well estimated with advised initializations:

$$\bar{\Gamma}_{\text{adv}}^{\text{NCEmlme}} = \begin{pmatrix} 25.74 & 16.34 & 35.75 & 4.50 \\ 16.34 & 34.83 & 43.97 & 5.37 \\ 35.75 & 43.97 & 82.44 & 8.43 \\ 4.50 & 5.37 & 8.43 & 1.32 \end{pmatrix}, \quad \text{with} \quad \sigma_{\Gamma_{\text{adv}}}^{\text{NCEmlme}} = \begin{pmatrix} 5.73 & 2.74 & 4.45 & 0.61 \\ 2.74 & 2.59 & 3.96 & 0.44 \\ 4.45 & 3.96 & 12.07 & 0.80 \\ 0.61 & 0.44 & 0.80 & 0.11 \end{pmatrix},$$

and

$$\bar{\Gamma}_{\text{adv}}^{\text{EM}} = \begin{pmatrix} 23.66 & 16.45 & 32.60 & 4.45 \\ 16.45 & 34.82 & 43.39 & 5.39 \\ 32.61 & 43.39 & 75.16 & 8.66 \\ 4.45 & 5.39 & 8.65 & 1.31 \end{pmatrix}, \quad \text{with} \quad \sigma_{\Gamma_{\text{adv}}}^{\text{EM}} = \begin{pmatrix} 7.49 & 2.71 & 5.29 & 0.76 \\ 2.71 & 2.59 & 3.71 & 0.45 \\ 5.30 & 3.71 & 13.36 & 0.80 \\ 0.76 & 0.45 & 0.80 & 0.11 \end{pmatrix}.$$

The matrices $\sigma_{\Gamma_{\text{adv}}}^{\text{NCEmlme}}$ and $\sigma_{\Gamma_{\text{adv}}}^{\text{EM}}$ contain the standard deviations of the entries of $\bar{\Gamma}_{\text{adv}}^{\text{NCEmlme}}$ and $\bar{\Gamma}_{\text{adv}}^{\text{EM}}$, respectively. It seems that the empirical standard deviation of the entries of $\bar{\Gamma}$ with the greater magnitude are bigger with EM than with NCEmlme. For example, the standard deviation of $\bar{\Gamma}_{11} = 27.77$ is 7.49 for EM, but 5.73 for NCEmlme. The same remark is valid for the standard deviations of $\bar{\Gamma}_{31}$, $\bar{\Gamma}_{32}$ and $\bar{\Gamma}_{33}$, comparing EM and NCEmlme. This may be explained by the fact that the NCEmlme procedure is more numerically stable than the EM procedure. In the case of naive initializations, NCEmlme provides a well estimated empirical mean of $\bar{\Gamma}$, when the estimated $\bar{\Gamma}$ provided by EM is very bad (we choose not to show it here).

$$\bar{\Gamma}_{\text{naive}}^{\text{NCEmlme}} = \begin{pmatrix} 24.76 & 16.18 & 34.43 & 4.47 \\ 16.18 & 34.77 & 43.91 & 5.35 \\ 34.43 & 43.91 & 80.67 & 8.40 \\ 4.47 & 5.35 & 8.40 & 1.32 \end{pmatrix}, \quad \text{with} \quad \sigma_{\Gamma_{\text{naive}}}^{\text{NCEmlme}} = \begin{pmatrix} 7.58 & 2.67 & 6.76 & 0.61 \\ 2.67 & 2.54 & 3.91 & 0.42 \\ 6.76 & 3.91 & 13.37 & 0.80 \\ 0.61 & 0.42 & 0.80 & 0.11 \end{pmatrix}.$$

$\bar{\Gamma}_{\text{naive}}^{\text{NCEmlme}}$ compared to $\bar{\Gamma}_{\text{adv}}^{\text{NCEmlme}}$ and $\sigma_{\Gamma_{\text{naive}}}^{\text{NCEmlme}}$ compared to $\sigma_{\Gamma_{\text{adv}}}^{\text{NCEmlme}}$ show a slight difference which reveals a tiny sensibility of NCEmlme to the starting values. This may be corrected by doing more than one evaluation of the model's deviance.

For all the simulation studies, we use the ML deviance criterion and have minimized it using the *nlminb* function under the R software. Thus, the estimates obtained are from the ML estimators. In this paper, we do not provide an application of REML estimators.

4 Application on malaria dataset

4.1 Data description

The data that we analyze here come from a study which was conducted in nine villages (Avamé centre, Gbédjougou, Houngo, Anavié, Dohinoko, Gbétaga, Tori Cada Centre, Zébè and Zoungoudo) of Tori Bossito area (Southern Benin), where *P.falciparum* is the commonest species in the study area (95%) [31] from June 2007 to January 2010. The aim of this study was to evaluate the determinants of malaria incidence in the first months of life of children in Benin.

Mothers ($n = 620$) were enrolled at delivery and their newborns were actively followed up during the first year of life. One questionnaire was conducted to gather information on women's characteristics (age, parity, use of Intermittent Preventive Treatment during pregnancy (IPTp) and bed net possession) and during their pregnancy. After delivery, thick and thin placental blood smears were examined to detect placental infection defined by the presence of asexual forms of *P.falciparum*. Maternal peripheral blood as well as cord blood were collected. At birth, newborn's weight and length were measured and gestational age was estimated.

During the follow-up of newborns, axillary temperature was measured weekly. In case of temperature higher than 37.5°C , mothers were told to bring their children to the health center where a questionnaire was filled out. A rapid diagnostic test (RDT) for malaria was performed and a thick blood smear (TBS) made. Symptomatic malaria cases, defined as fever ($> 37.5^\circ\text{C}$) with TBS and/or RDT positive, were treated with an artemisinin-based combination. Systematically, TBS were made every month to detect asymptomatic infections. Every three months, venous blood was sampled to quantify the level of antibody against malaria promised candidate vaccine antigens. Finally, the environmental risk of exposure to malaria was modeled for each child, derived from a statistical predictive model based on climatic, entomological parameters, and characteristics of children's immediate surroundings. Also every 3 months (at 3, 6, 9, 12, 15, 18 months 130 of age), infant blood samples were collected.

Concerning the antibody quantification, two recombinant *P.falciparum* antigens were used to perform IgG subclass (IgG1 and IgG3) antibody. Recombinant antigens MSP2 (3D7 and FC27) were from La Trobe University [32, 33]. GLURP-R0 (amino acids 25-514, F32 strain) and GLURP-R2 (amino acids 706-1178, 140 F32 strain) were also expressed. The antibodies were quantified in plasma at different times and ADAMSEL FLp039 software (<http://www.malariaearesearch.eu/content/software>) was used to analyze automatically the ELISA optical density (OD) leading to antibody concentrations in ($\mu\text{g/mL}$).

In this paper, we use some of the data and we rename the proteins used in the study, for reasons of the protection of these data. We thus name A1, A2, B and C the proteins that we consider, which are related to the antigens IgG1 and IgG3 as mentioned above. The information contained in the multivariate longitudinal dataset of malaria is described in Table 4, where Y denotes an antigen which is one of the following:

$$\text{IgG1_A1, IgG3_A1, IgG1_A2, IgG3_A2, IgG1_B, IgG3_B, IgG1_C, IgG3_C} \quad (15)$$

Table 4: Variables of the Malaria dataset.

N°	Variable	Description
1	id	Child ID
2	conc.Y	concentration of Y
3	conc_CO.Y	Measured concentration of Y in the umbilical cord blood
4	conc_M3.Y	Predicted concentration of Y in the child's peripheral blood at 3 months
5	ap	Placental apposition
6	hb	Hemoglobin level
7	inf_trim	Number of malaria infections in the previous 3 months
8	pred_trim	Quarterly average number of mosquitoes child is exposed to
9	nutri_trim	Quarterly average nutrition scores

4.2 Data analysis

The aim of the data analysis is to evaluate the effect of the malaria infection on the child's immune acquisition (against malaria). Since the antigens which characterize the child's immune status interact together in the human body, we analyze the characteristics of the joint distribution of these antigens, conditionally on the malaria infection and other factors of interest. The dependent variables are then provided by conc.Y (Table 4) which describes the level of the antigen Y in the children at 3, 6, 9, 12, 15 and 18 months. All other variables in Table 4 are covariates. We then have $d = 8$ dependent variables which describe the longitudinal profile (in the child) of the proteins listed in eq. (15).

To illustrate the stability of our approach, we are fitting here a bivariate model (see eqs. (16) and (17)) to the data, with IgG1_A1 and IgG1_A2 as dependent variables:

$$\begin{aligned} \text{conc.IgG1_A1} &= (1, \text{ap}, \text{conc_CO.IgG1_A1}, \text{conc_M3.IgG1_A1}, \text{hb}, \text{inf_trim}, \\ &\quad \text{pred_trim}, \text{nutri_trim})\beta_1 + (1, \text{inf_trim})\gamma_1 + \varepsilon_1 \\ \text{conc.IgG1_A2} &= (1, \text{ap}, \text{conc_CO.IgG1_A2}, \text{conc_M3.IgG1_A2}, \text{hb}, \text{inf_trim}, \\ &\quad \text{pred_trim}, \text{nutri_trim})\beta_2 + (1, \text{inf_trim})\gamma_2 + \varepsilon_2, \end{aligned} \quad (16)$$

with

$$\boldsymbol{\gamma} = (\gamma_1^\top, \gamma_2^\top)^\top \sim N(\mathbf{0}, \mathbf{\Gamma}), \quad \boldsymbol{\varepsilon} = (\varepsilon_1^\top, \varepsilon_2^\top)^\top \sim N\left(\mathbf{0}, \begin{pmatrix} \sigma_1^2 \mathbf{I} & 0 \\ 0 & \sigma_2^2 \mathbf{I} \end{pmatrix}\right). \quad (17)$$

Our strategy is to (1) fit the model to the data by running the NCEmlme algorithm using 25 different naive starting points and (2) retain the estimates related to the best likelihood (the minimum of the 25 deviances) as the true parameters and compute the estimators' MSE using the 24 others estimates. This may allow to evaluate how much the NCEmlme algorithm is sensitive to the starting points. The results are contained in Table 5. Based on these results, the influence of the starting points on the NCEmlme algorithm is very low (see the MSE in

Table 5). The estimated random effects covariance matrix is

$$\mathbf{\Gamma} = \begin{pmatrix} 1.27 & -0.53 & 1.57 & -0.71 \\ -0.53 & 0.49 & -0.71 & 0.65 \\ 1.57 & -0.71 & 1.93 & -0.82 \\ -0.71 & 0.65 & -0.82 & 0.77 \end{pmatrix} \quad (18)$$

with an MSE of 0.0095.

Table 5: Bivariate and univariate linear mixed-effects model fitted to the malaria proteins. The parameters are estimated under the bivariate and the univariate models, respectively. The MSE are obtained for the bivariate model using 25 different naive starting points.

Covariates	Response variables					
	conc.IgG1_A1			conc.IgG1_A2		
	Joint estim.	MSE	Indep. estim.	Joint estim.	MSE	Indep. estim.
Intercept	0.648	9.05×10^{-5}	0.451	0.032	3.18×10^{-5}	-1.794
ap	-0.130	1.06×10^{-5}	-0.110	-0.038	1.04×10^{-6}	-0.324
conc_CO.IgG1_A1	0.127	1.68×10^{-6}	0.174	-	-	-
conc_M3.IgG1_A1	0.151	9.85×10^{-6}	0.130	-	-	-
conc_CO.IgG1_A2	-	-	-	0.092	6.44×10^{-7}	0.044
conc_M3.IgG1_A2	-	-	-	0.117	2.22×10^{-7}	0.137
hb	-0.166	3.22×10^{-7}	-0.154	-0.245	1.35×10^{-7}	-0.335
inf_trim	0.398	1.89×10^{-6}	0.373	0.785	5.09×10^{-7}	0.709
pred_trim	-0.002	5.25×10^{-8}	0.001	0.017	1.49×10^{-8}	0.019
nutri_trim	0.072	5.81×10^{-6}	0.046	0.271	3.71×10^{-5}	0.106
σ_1 and σ_2	1.345	4.96×10^{-6}	1.426	1.495	2.42×10^{-5}	1.672

In this kind of study, it is often useful to characterize the subjects (here, the children) by constructing their profiles. This is done through the computation of the expected random-effects vector (i.e. $\mathbb{E}[\gamma|y]$). We compute here the expected random-effects vector for each child under the joint model but also under the separate models (i.e. two independent models since the joint model is bivariate). The results are shown in Figure 1 and Figure 2. Children profiles change dramatically depending on whether one is in the joint model or in the independent models. This change is noticeable in the expected random intercepts but also in the expected random slopes. The discrepancy between the parameters estimated under the joint model and the separate models (see the columns labeled *Indep. estim.* and *Joint estim.*, respectively in Table 5) could be explained by the change noticed in the children profiles. This discrepancy is expected since the correlations between the dimensional random-effects are not null (see eq. (18)), and should be taken into account when characterizing the children profiles. These differences in the children profiles are crucial when analyzing such a highly important health dataset.

Figure 1: Expected values of the random intercept/slope (in the malaria application) from the first dimension of the bivariate model. The red profiles are computed with the expected random intercepts and random slopes, using a univariate model that is equal to the first dimension (related to the dependent variable IgG1_A1 and fitted separately) of the model expressed by eqs. (16) and (17). The blue profiles correspond to the random effects computed using the joint bivariate model.

Figure 2: Expected values of the random intercept/slope (in the malaria application) from the second dimension of the bivariate model. The red profiles are computed with the expected random intercepts and random slopes, using a univariate model that is equal to the second dimension (related to the dependent variable IgG1_A2 and fitted separately) of the model expressed by eqs. (16) and (17). The blue profiles correspond to the random effects computed using the joint bivariate model.

5 Conclusion

In the context of multivariate linear mixed-effects models having homoscedastic dimensional residuals, we have suggested ML and REML estimation strategies by profiling the model's deviance and Cholesky factorizing of the random effect covariance matrix. This approach can be considered as the generalization of the approach used by [6] in the R software lme4 package. Through extensive simulation studies, we have illustrated that the present approach outperforms the traditional EM-based estimates and provides estimates that are numerically consistent for both fixed effects and variance components. Another interesting characteristic is its robustness relative to the initial value of the optimization procedure which can be randomly chosen without affecting the estimation results. These characteristics of our procedure show their improved computational stability in comparison with the traditional approaches. Furthermore, the optimization of the profiled ML or REML criterion can be easily and rapidly performed using any existing box-constrained optimizer in the R software. Further considerations of this approach may include heteroscedastic residuals as well as residuals correlated with the random effects, where the theoretical consistency of the resulting estimators should be demonstrated.

Appendix

A Proof of Theorem 2.1 and Lemma 2.1

Proof.

Denoting by $f_{\mathcal{X}}(\cdot)$ the density function of any random vector $\mathcal{X}$, the density function of $\mathcal{Y}$ is

$$f_{\mathcal{Y}}(\mathbf{y}) = \int_{\mathbb{R}^{q_1+q_2}} f_{\mathcal{Y},\mathcal{U}}(\mathbf{y},\mathbf{u})d\mathbf{u},$$

where $f_{\mathcal{Y},\mathcal{U}}(\mathbf{y},\mathbf{u})$ is the joint density function of $\mathcal{Y}$ and $\mathcal{U}$. The dependence between the coordinates of $\mathcal{Y}$ comes from $\mathcal{U}$, and by conditioning by $\mathcal{U}$, the density function $f_{\mathcal{Y},\mathcal{U}}(\mathbf{y}|\mathbf{u})$ can be factorized as: $f_{\mathcal{Y},\mathcal{U}}(\mathbf{y}|\mathbf{u}) = f_{\mathcal{Y}_1|\mathcal{U}_1}(y_1|u_1)f_{\mathcal{Y}_2|\mathcal{U}_2}(y_2|u_2)$. Thus,

$$\begin{aligned} f_{\mathcal{Y},\mathcal{U}}(\mathbf{y},\mathbf{u}) &= f_{\mathcal{Y}_1|\mathcal{U}_1}(y_1|u_1)f_{\mathcal{Y}_2|\mathcal{U}_2}(y_2|u_2)f_{\mathcal{U}}(\mathbf{u}) \\ &= (2\pi\sigma_1^2)^{-\frac{N}{2}}(2\pi\sigma_2^2)^{-\frac{N}{2}}(2\pi)^{-\frac{q_1+q_2}{2}}|\Sigma_{\mathbf{u}}|^{-\frac{1}{2}}\exp\left(-\frac{\|y_1 - X_1\beta_1 - Z_1\Lambda_{\theta_1}u_1\|^2}{2\sigma_1^2}\right. \\ &\quad \left.-\frac{\|y_2 - X_2\beta_2 - Z_2\Lambda_{\theta_2}u_2\|^2}{2\sigma_2^2} - \frac{1}{2}\mathbf{u}^\top\Sigma_{\mathbf{u}}^{-1}\mathbf{u}\right). \end{aligned}$$

Let us denote by $\tilde{\Sigma}$ a matrix such that $\Sigma_{\mathbf{u}}^{-1} = \tilde{\Sigma}^\top\tilde{\Sigma}$. Therefore, $\mathbf{u}^\top\Sigma_{\mathbf{u}}^{-1}\mathbf{u} = \|\tilde{\Sigma}\mathbf{u}\|^2$ and

$$\begin{aligned} &\frac{\|y_1 - X_1\beta_1 - Z_1\Lambda_{\theta_1}u_1\|^2}{\sigma_1^2} + \frac{\|y_2 - X_2\beta_2 - Z_2\Lambda_{\theta_2}u_2\|^2}{\sigma_2^2} + \mathbf{u}^\top\Sigma_{\mathbf{u}}^{-1}\mathbf{u} \\ &= \frac{\|\sqrt{\sigma_2^2}(y_1 - X_1\beta_1 - Z_1\Lambda_{\theta_1}u_1)\|^2 + \|\sqrt{\sigma_1^2}(y_2 - X_2\beta_2 - Z_2\Lambda_{\theta_2}u_2)\|^2 + \|\sqrt{\sigma_1^2\sigma_2^2}\tilde{\Sigma}\mathbf{u}\|^2}{\sigma_1^2\sigma_2^2} \\ &= (1/(\sigma_1^2\sigma_2^2))\left\|\begin{pmatrix} \sqrt{\sigma_2^2}y_1 \\ \sqrt{\sigma_1^2}y_2 \\ \mathbf{0}_{q_1+q_2} \end{pmatrix} - \begin{pmatrix} \sqrt{\sigma_2^2}X_1 & \mathbf{0}_{N,p_2} & \sqrt{\sigma_2^2}Z_1\Lambda_{\theta_1} & \mathbf{0}_{N,q_2} \\ \mathbf{0}_{N,p_1} & \sqrt{\sigma_1^2}X_2 & \mathbf{0}_{N,q_1} & \sqrt{\sigma_1^2}Z_2\Lambda_{\theta_2} \\ & \mathbf{0}_{q_1+q_2,p_1+p_2} & & \sqrt{\sigma_1^2\sigma_2^2}\tilde{\Sigma} \end{pmatrix} \begin{pmatrix} \beta \\ \mathbf{u} \end{pmatrix}\right\|^2 \\ &= (1/(\sigma_1^2\sigma_2^2))\left\|Y_\Lambda - Z_{X\Lambda} \begin{pmatrix} \beta \\ \mathbf{u} \end{pmatrix}\right\|^2 \end{aligned} \quad (19)$$

$$= g(\beta, \mathbf{u}, \theta, \rho, \sigma) / (\sigma_1^2\sigma_2^2), \quad (20)$$

where $\mathbf{0}_{q_1+q_2}$ is a zero vector of $\mathbb{R}^{q_1+q_2}$, and $\mathbf{0}_{N,p_1}, \mathbf{0}_{N,p_2}, \mathbf{0}_{N,q_1}, \mathbf{0}_{N,q_2}$ and $\mathbf{0}_{q_1+q_2,p_1+p_2}$ are zero matrices of dimension $N \times p_1, N \times p_1, N \times p_1, N \times q_2$ and $q_1 + q_2 \times p_1 + p_2$, respectively. In eq. (19),

$$Y_\Lambda = \begin{pmatrix} \sqrt{\sigma_2^2}y_1 \\ \sqrt{\sigma_1^2}y_2 \\ \mathbf{0}_{q_1+q_2} \end{pmatrix},$$

and

$$Z_{X\Lambda} = \begin{pmatrix} \sqrt{\sigma_2^2}X_1 & \mathbf{0}_{N,p_2} & \sqrt{\sigma_2^2}Z_1\Lambda_{\theta_1} & \mathbf{0}_{N,q_2} \\ \mathbf{0}_{N,p_1} & \sqrt{\sigma_1^2}X_2 & \mathbf{0}_{N,q_1} & \sqrt{\sigma_1^2}Z_2\Lambda_{\theta_2} \\ & \mathbf{0}_{q_1+q_2,p_1+p_2} & & \sqrt{\sigma_1^2\sigma_2^2}\tilde{\Sigma} \end{pmatrix}.$$

And in eq. (20), $g(\beta, \mathbf{u}, \theta, \rho, \sigma) = \left\|Y_\Lambda - Z_{X\Lambda} \begin{pmatrix} \beta \\ \mathbf{u} \end{pmatrix}\right\|^2$ is a squared norm or, more specifically, a residual sum-of-squares. This is a standard least squares problem:

$$\begin{pmatrix} \hat{\beta}_{\theta,\rho,\sigma} \\ \mu_{\mathcal{U}|\mathcal{Y}=\mathbf{y}} \end{pmatrix} = \arg \min_{\mathbf{u},\beta} g(\beta, \mathbf{u}, \theta, \rho, \sigma)$$

for which the solution satisfy the normal equation:

$$Z_{X\Lambda}^\top Z_{X\Lambda} \begin{pmatrix} \hat{\beta}_{\theta,\rho,\sigma} \\ \mu_{\mathcal{U}|\mathcal{Y}=\mathbf{y}} \end{pmatrix} = Z_{X\Lambda}^\top Y_\Lambda, \quad (21)$$

where

$$Z_{X\Lambda}^\top Z_{X\Lambda} = \begin{pmatrix} X_\sigma^\top X_\sigma & X_\sigma^\top Z_{\sigma\theta} \\ Z_{\sigma\theta}^\top X_\sigma & Z_{\sigma\theta}^\top Z_{\sigma\theta} + \sqrt{\sigma_1^2\sigma_2^2}\Sigma_{\mathbf{u}}^{-1} \end{pmatrix} \quad \text{and} \quad Z_{X\Lambda}^\top Y_\Lambda = \begin{pmatrix} X_\sigma^\top \\ Z_{\sigma\theta}^\top \end{pmatrix} Y_\sigma.$$

This concludes the proof of Lemma 2.1. In eq. (21), we behave as if θ, ρ and σ are known, and this justify the notation $\hat{\beta}_{\theta,\rho,\sigma}$ which means that the resulting estimator is a function of θ, ρ and σ .

By setting $p = p_1 + p_2$, $\dim(Z_{X\Lambda}) = (2N + q) \times (p + q)$ and $S = \text{Im}(Z_{X\Lambda})$ is a subspace of $\mathbb{R}^{2N+q}$. $Y_\Lambda \in \mathbb{R}^{2N+q}$ and is the orthogonal projection of Y_Λ on S . Then,

$$Z_{X\Lambda}u \perp \left[Y_\Lambda - Z_{X\Lambda} \begin{pmatrix} \hat{\beta}_{\theta,\rho,\sigma} \\ \mu_{\mathbf{u}|\mathbf{y}=\mathbf{y}} \end{pmatrix} \right], \forall u \in \mathbb{R}^{p+q}.$$

And $g(\beta, \mathbf{u}, \theta, \rho, \sigma)$ can then be rewritten as:

$$\begin{aligned} g(\beta, \mathbf{u}, \theta, \rho, \sigma) &= \left\| Y_\Lambda - Z_{X\Lambda} \begin{pmatrix} \beta \\ \mathbf{u} \end{pmatrix} + Z_{X\Lambda} \begin{pmatrix} \hat{\beta}_{\theta,\rho,\sigma} \\ \mu_{\mathbf{u}|\mathbf{y}=\mathbf{y}} \end{pmatrix} - Z_{X\Lambda} \begin{pmatrix} \hat{\beta}_{\theta,\rho,\sigma} \\ \mu_{\mathbf{u}|\mathbf{y}=\mathbf{y}} \end{pmatrix} \right\|^2 \\ &= \left\| Y_\Lambda - Z_{X\Lambda} \begin{pmatrix} \hat{\beta}_{\theta,\rho,\sigma} \\ \mu_{\mathbf{u}|\mathbf{y}=\mathbf{y}} \end{pmatrix} \right\|^2 + \left\| Z_{X\Lambda} \begin{pmatrix} \beta - \hat{\beta}_{\theta,\rho,\sigma} \\ \mathbf{u} - \mu_{\mathbf{u}|\mathbf{y}=\mathbf{y}} \end{pmatrix} \right\|^2 \\ &= \left\| Y_\Lambda - Z_{X\Lambda} \begin{pmatrix} \hat{\beta}_{\theta,\rho,\sigma} \\ \mu_{\mathbf{u}|\mathbf{y}=\mathbf{y}} \end{pmatrix} \right\|^2 + \begin{pmatrix} \beta - \hat{\beta}_{\theta,\rho,\sigma} \\ \mathbf{u} - \mu_{\mathbf{u}|\mathbf{y}=\mathbf{y}} \end{pmatrix}^\top Z_{X\Lambda}^\top Z_{X\Lambda} \begin{pmatrix} \beta - \hat{\beta}_{\theta,\rho,\sigma} \\ \mathbf{u} - \mu_{\mathbf{u}|\mathbf{y}=\mathbf{y}} \end{pmatrix}. \end{aligned}$$

$Z_{X\Lambda}^\top Z_{X\Lambda}$ can be Cholesky decomposed as

$$Z_{X\Lambda}^\top Z_{X\Lambda} = \begin{pmatrix} R_X & \mathbf{0} \\ R_{ZX} & L_{\theta,\rho,\sigma}^\top \end{pmatrix}^\top \begin{pmatrix} R_X & \mathbf{0} \\ R_{ZX} & L_{\theta,\rho,\sigma}^\top \end{pmatrix},$$

where

$$L_{\theta,\rho,\sigma} L_{\theta,\rho,\sigma}^\top = Z_{\sigma\theta}^\top Z_{\sigma\theta} + \sqrt{\sigma_1^2 \sigma_2^2 \Sigma_{\mathbf{u}}^{-1}}.$$

Thereafter,

$$\begin{aligned} g(\beta, \mathbf{u}, \theta, \rho, \sigma) &= \left\| Y_\sigma - X_\sigma \hat{\beta}_{\theta,\rho,\sigma} - Z_{\sigma\theta} \mu_{\mathbf{u}|\mathbf{y}=\mathbf{y}} \right\|^2 + \sigma_1^2 \sigma_2^2 \mu_{\mathbf{u}|\mathbf{y}=\mathbf{y}}^\top \Sigma_{\mathbf{u}}^{-1} \mu_{\mathbf{u}|\mathbf{y}=\mathbf{y}} \\ &\quad + \left\| R_X (\beta - \hat{\beta}_{\theta,\rho,\sigma}) \right\|^2 + \left\| R_{ZX} (\beta - \hat{\beta}_{\theta,\rho,\sigma}) + L_{\theta,\rho,\sigma}^\top (\mathbf{u} - \mu_{\mathbf{u}|\mathbf{y}=\mathbf{y}}) \right\|^2. \end{aligned}$$

By setting

$$r(\hat{\beta}_{\theta,\rho,\sigma}, \mu_{\mathbf{u}|\mathbf{y}=\mathbf{y}}) = \left\| Y_\sigma - X_\sigma \hat{\beta}_{\theta,\rho,\sigma} - Z_{\sigma\theta} \mu_{\mathbf{u}|\mathbf{y}=\mathbf{y}} \right\|^2 + \sigma_1^2 \sigma_2^2 \mu_{\mathbf{u}|\mathbf{y}=\mathbf{y}}^\top \Sigma_{\mathbf{u}}^{-1} \mu_{\mathbf{u}|\mathbf{y}=\mathbf{y}},$$

and returning to the calculation of $f_{\mathbf{y}}(\mathbf{y})$, we get

$$\begin{aligned} f_{\mathbf{y}}(\mathbf{y}) &= \frac{\int \exp \left[-\frac{r(\hat{\beta}_{\theta,\rho,\sigma}, \mu_{\mathbf{u}|\mathbf{y}=\mathbf{y}}) + \left\| R_X (\beta - \hat{\beta}_{\theta,\rho,\sigma}) \right\|^2 + \left\| R_{ZX} (\beta - \hat{\beta}_{\theta,\rho,\sigma}) + L_{\theta,\rho,\sigma}^\top (\mathbf{u} - \mu_{\mathbf{u}|\mathbf{y}=\mathbf{y}}) \right\|^2}{2\sigma_1^2 \sigma_2^2} \right] d\mathbf{u}}{(2\pi\sigma_1^2)^{N/2} (2\pi\sigma_2^2)^{N/2} (2\pi)^{q/2} |\Sigma_{\mathbf{u}}|^{1/2}} \\ &= \frac{\exp \left[-\frac{r(\hat{\beta}_{\theta,\rho,\sigma}, \mu_{\mathbf{u}|\mathbf{y}=\mathbf{y}}) + \left\| R_X (\beta - \hat{\beta}_{\theta,\rho,\sigma}) \right\|^2}{2\sigma_1^2 \sigma_2^2} \right]}{(2\pi\sigma_1^2)^{N/2} (2\pi\sigma_2^2)^{N/2} (2\pi)^{q/2} |\Sigma_{\mathbf{u}}|^{1/2}} \times \\ &\quad \int \exp \left[-\frac{\left\| R_{ZX} (\beta - \hat{\beta}_{\theta,\rho,\sigma}) + L_{\theta,\rho,\sigma}^\top (\mathbf{u} - \mu_{\mathbf{u}|\mathbf{y}=\mathbf{y}}) \right\|^2}{2\sigma_1^2 \sigma_2^2} \right] d\mathbf{u}. \end{aligned}$$

By setting $\mathbf{v} = R_{ZX} (\beta - \hat{\beta}_{\theta,\rho,\sigma}) + L_{\theta,\rho,\sigma}^\top (\mathbf{u} - \mu_{\mathbf{u}|\mathbf{y}=\mathbf{y}})$, $d\mathbf{u} = \frac{1}{|L_{\theta,\rho,\sigma}|} d\mathbf{v}$ and

$$\begin{aligned} f_{\mathbf{y}}(\mathbf{y}) &= \frac{\exp \left[-\frac{r(\hat{\beta}_{\theta,\rho,\sigma}, \mu_{\mathbf{u}|\mathbf{y}=\mathbf{y}}) + \left\| R_X (\beta - \hat{\beta}_{\theta,\rho,\sigma}) \right\|^2}{2\sigma_1^2 \sigma_2^2} \right] (\sigma_1^2 \sigma_2^2)^{\frac{q}{2}}}{(2\pi\sigma_1^2)^{N/2} (2\pi\sigma_2^2)^{N/2} |\Sigma_{\mathbf{u}}|^{1/2} |L_{\theta,\rho,\sigma}|} \int \frac{1}{(2\pi\sigma_1^2 \sigma_2^2)^{\frac{q}{2}}} \exp \left[-\frac{\|\mathbf{v}\|^2}{2\sigma_1^2 \sigma_2^2} \right] d\mathbf{v} \\ &= \frac{\exp \left[-\frac{r(\hat{\beta}_{\theta,\rho,\sigma}, \mu_{\mathbf{u}|\mathbf{y}=\mathbf{y}}) + \left\| R_X (\beta - \hat{\beta}_{\theta,\rho,\sigma}) \right\|^2}{2\sigma_1^2 \sigma_2^2} \right] (\sigma_1^2 \sigma_2^2)^{\frac{q}{2}}}{(2\pi\sigma_1^2)^{N/2} (2\pi\sigma_2^2)^{N/2} |\Sigma_{\mathbf{u}}|^{1/2} |L_{\theta,\rho,\sigma}|}. \end{aligned} \quad (22)$$

The log-likelihood to be maximized can therefore be expressed as,

$$\ell(\beta, \theta, \rho, \sigma | \mathbf{y}) = -\frac{r(\hat{\beta}_{\theta,\rho,\sigma}, \mu_{\mathbf{u}|\mathbf{y}=\mathbf{y}}) + \left\| R_X (\beta - \hat{\beta}_{\theta,\rho,\sigma}) \right\|^2}{2\sigma_1^2 \sigma_2^2} - \frac{N-q}{2} \log(\sigma_1^2 \sigma_2^2) - \frac{1}{2} \log(|\Sigma_{\mathbf{u}}|) - \frac{1}{2} \log(|L_{\theta,\rho,\sigma}|^2). \quad \square$$

B Proof of Theorem 2.2

Proof.

Using the expression of $f_{\mathbf{y}}(\mathbf{y})$ in eq. (22), we get

$$\begin{aligned}\mathcal{L}(\sigma, \theta, \rho | \mathbf{y}) &= \int_{\mathbb{R}^p} f_{\mathbf{y}}(\mathbf{y}) d\boldsymbol{\beta} \\ &= \frac{\exp \left[-\frac{r(\hat{\boldsymbol{\beta}}_{\theta, \rho, \sigma} | \mathcal{H}(\mathbf{y} = \mathbf{y}))}{2\sigma_1^2 \sigma_2^2} \right] (\sigma_1^2 \sigma_2^2)^{\frac{q}{2}}}{(2\pi\sigma_1^2)^{N/2} (2\pi\sigma_2^2)^{N/2} |\Sigma_{\mathbf{u}}|^{1/2} |L_{\theta, \rho, \sigma}|} \int_{\mathbb{R}^p} \exp \left[-\frac{\|R_X(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_{\theta, \rho, \sigma})\|^2}{2\sigma_1^2 \sigma_2^2} \right] d\boldsymbol{\beta} \\ &= \frac{\exp \left[-\frac{r(\hat{\boldsymbol{\beta}}_{\theta, \rho, \sigma} | \mathcal{H}(\mathbf{y} = \mathbf{y}))}{2\sigma_1^2 \sigma_2^2} \right] (\sigma_1^2 \sigma_2^2)^{\frac{p+q-N}{2}}}{(2\pi)^{(2N-p)/2} |\Sigma_{\mathbf{u}}|^{1/2} |L_{\theta, \rho, \sigma}| |R_X|} \int_{\mathbb{R}^p} (2\pi\sigma_1^2 \sigma_2^2)^{-\frac{p}{2}} \exp \left[-\frac{\|\mathbf{t}\|^2}{2\sigma_1^2 \sigma_2^2} \right] d\mathbf{t},\end{aligned}$$

where $\mathbf{t} = \boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_{\theta, \rho, \sigma} \Rightarrow d\boldsymbol{\beta} = \frac{1}{|R_X|} d\mathbf{t}$ and $\int_{\mathbb{R}^p} (2\pi\sigma_1^2 \sigma_2^2)^{-\frac{p}{2}} \exp \left[-\frac{\|\mathbf{t}\|^2}{2\sigma_1^2 \sigma_2^2} \right] d\mathbf{t} = 1$. $\square$

References

- [1] Fitzmaurice GM, Laird NM, Ware JH. Applied longitudinal analysis, volume 998. New Jersey: John Wiley & Sons 2012
- [2] Hartley HO, Rao JN. Maximum-likelihood estimation for the mixed analysis of variance model. *Biometrika*. 1967;54:93–108.
- [3] Hedeker D, Gibbons RD. Longitudinal data analysis, volume 451. John Wiley & Sons 2006
- [4] Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics*. 1982;38:963–974
- [5] Verbeke G. Linear mixed models for longitudinal data. In: Linear mixed models in practice. New York: Springer. 1997;126:63–153
- [6] Bates D, Maechler M, Bolker B, Walker S, et al. lme4: Linear mixed-effects models using eigen and s4. R Package Version. 2014;7:1.
- [7] Halekoh U, Højsgaard S, Yan J. The r package geepack for generalized estimating equations. *J Stat Software*. 2006;15:1–11.
- [8] Littell R, Milliken G, Stroup W, Wolfinger R, Schabenberger O. Random coefficient models. SAS System for Mixed Models. Cary, NC: SAS Institute Inc, 1996:253–66.
- [9] Pinheiro J, Bates D, DebRoy S, Sarkar D, et al. Linear and nonlinear mixed effects models. R Package Version. 2007;3:57.
- [10] Verbeke G, Fieuws S, Molenberghs G, Davidian M. The analysis of multivariate longitudinal data: a review. *Stat Methods Med Res*. 2014;23:42–59.
- [11] Jensen KL, Spiild H, Toftum J. Implementation of multivariate linear mixed-effects models in the analysis of indoor climate performance experiments. *Int J Biometeorol*. 2012;56:129–36.
- [12] Sammel M, Lin X, Ryan L. Multivariate linear mixed models for multiple outcomes. *Stat Med*. 1999;18:2479–92.
- [13] Wang W-L, Fan T-H. Ecm-based maximum likelihood inference for multivariate linear mixed models with autoregressive errors. *Comput Stat Data Anal* 2010;54:1328–41.
- [14] Schafer JL, Yucel RM. Computational strategies for multivariate linear mixed-effects models with missing values. *J Comput Graphical Stat*. 2002;11:437–57.
- [15] Adjakossa EH, Sadissou I, Hounkonnou MN, Nuel G. Multivariate longitudinal analysis with bivariate correlation test. *PloS one*. 2016;11:e0159649.
- [16] Henderson CR. Estimation of genetic parameters. *Biometrics*. 1950;6:186–7.
- [17] Goldberger AS. Best linear unbiased prediction in the generalized linear regression model. *J Am Stat Assoc*. 1962;57:369–75.
- [18] Searle S, Casella G, McCulloch C. Variance components. New York, NY, USA: John Wiley & Sons, 1992.
- [19] Robinson GK. That blup is a good thing: the estimation of random effects. *Stat Sci*. 1991;6:15–32.
- [20] Searle SR. Linear models. New York: Wiley & Sons, 1971.
- [21] Searle SR. An overview of variance component estimation. *Metrika*. 1995;42:215–30.
- [22] Rao CR. Estimation of variance and covariance components minque theory. *J Multivariate Anal*. 1971;1:257–75.
- [23] Lee Y, Nelder J. Generalized linear models for the analysis of quality-improvement experiments. *Can J Stat*. 1998;26:95–105.
- [24] Gumedze F, Dunne T. Parameter estimation and inference in the linear mixed model. *Linear Algebra Appl*. 2011;435:1920–44.
- [25] Lindstrom MJ, Bates DM. Newton raphson and em algorithms for linear mixed-effects models for repeated-measures data. *J Am Stat Assoc*. 1988;83:1014–22.
- [26] An X, Yang Q, Bentler PM. A latent factor linear mixed model for high-dimensional longitudinal data analysis. *Stat Med*. 2013;32:4229–39.
- [27] Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the em algorithm. *J R Stat Soc: B (methodological)*. 1977;39:1–38.
- [28] Meng X-L, Rubin DB. Maximum likelihood estimation via the ecm algorithm: a general framework. *Biometrika*. 1993;80:267–78.
- [29] Shah A, Laird N, Schoenfeld D. A random-effects model for multiple characteristics with possibly missing data. *J Am Stat Assoc*. 1997;92:775–9.
- [30] Nie L. Convergence rate of mle in generalized linear and nonlinear mixed-effects models: theory and applications. *J Stat Plann Inference*. 2007;137:1787–804.

- [31] Djènontin A, Bio-Bangana S, Moiroux N, Henry M-C, Bousari O, Chabi J, et al. Culicidae diversity, malaria transmission and insecticide resistance alleles in malaria vectors in ouidah-kpomasse-tori district from benin (west africa): a pre-intervention study. *Parasit Vectors*. 2010;3:83.
- [32] Anders RF, Adda CG, Foley M, Norton RS. Recombinant protein vaccines against the asexual blood-stages of plasmodium falciparum. *Human Vaccines*. 2010;6:39–53.
- [33] McCarthy JS, Marjason J, Elliott S, Fahey P, Bang G, Malkin E, et al. A phase 1 trial of msp2-c1, a blood-stage malaria vaccine containing 2 isoforms of msp2 formulated with montanide (r) isa 720. *PLoS One*. 2011;6:e24413.