Original article

EM-REML estimation
of covariance parameters
in Gaussian mixed models
for longitudinal data analysis

Jean-Louis FOULLEY[a*], Florence JAFFRÉZIC[b],
Christèle ROBERT-GRANIÉ[a]

[a] Station de génétique quantitative et appliquée,
Institut national de la recherche agronomique,
78352 Jouy-en-Josas Cedex, France
[b] Institute of Cell, Animal and Population Biology
The University of Edinburgh
Edinburgh EH9 3JT, UK

Abstract – This paper presents procedures for implementing the EM algorithm to compute REML estimates of variance covariance components in Gaussian mixed models for longitudinal data analysis. The class of models considered includes random coefficient factors, stationary time processes and measurement errors. The EM algorithm allows separation of the computations pertaining to parameters involved in the random coefficient factors from those pertaining to the time processes and errors. The procedures are illustrated with Pothoff and Roy's data example on growth measurements taken on 11 girls and 16 boys at four ages. Several variants and extensions are discussed.

EM algorithm / REML / mixed models / random regression / longitudinal data

Résumé – Estimation EM-REML des paramètres de covariance en modèles mixtes gaussiens en vue de l'analyse de données longitudinales. Cet article présente des procédés permettant de mettre en œuvre l'algorithme EM en vue du calcul d'estimations REML des composantes de variance covariance en modèles mixtes gaussiens d'analyse de données longitudinales. La classe de modèles considérée concerne les coefficients aléatoires, les processus temporels stationnaires et les erreurs de mesure. L'algorithme EM permet de dissocier formellement les calculs relatifs aux paramètres des coefficients aléatoires de ceux impliqués dans les processus et la résiduelle. Ces méthodes sont illustrées par un exemple provenant de Pothoff et Roy

---

* Correspondence and reprints
E-mail: foulley@jouy.inra.fr

sur des mesures de croissance prises sur 11 filles et 16 garçons à quatre âges différents. On discute enfin plusieurs variantes et extensions de cette méthode.

algorithme EM / REML / modèles mixtes / régression aléatoire / données longitudinales

1. INTRODUCTION

There has been a great deal of interest in longitudinal data analysis among biometricians over the last decade: see e.g., the comprehensive synthesis of both theoretical and applied aspects given in Diggle et al. [4] textbook. Since the pioneer work of Laird and Ware [13] and of Diggle [3], random effects models [17] have been the cornerstone of statistical analysis used in biometry for this kind of data. In fact, as well illustrated in the quantitative genetics and animal breeding areas, practitioners have for a long time restricted their attention to the most extreme versions of such models viz. to the so called intercept or repeatability model with a constant intra-class correlation, and to the multiple trait approach involving an unspecified variance covariance structure.

Harville [9] first advocated the use of autoregressive random effects to the animal breeding community for analysing lactation records from different parities. These ideas were later used by Wade and Quaas [33] and Wade et al. [34] to estimate correlation among lactation yields produced over different time periods within herds and by Schaeffer and Dekkers [28] to analyse daily milk records.

As well explained in Diggle et al. [3], potentially interesting models must include three sources of variation: (i) between subjects, (ii) between times within a subject and (iii) measurement errors. Covariance parameters of such models are usually estimated by maximum likelihood procedures based on second order algorithms. The objective of this study is to propose EM-REML procedures [1, 21] for estimating these parameters especially for those involved in the serial correlation structure (ii).

The paper is organized as follows. Section 2 describes the model structure and Section 3 the EM implementation. A numerical example based on growth measurements will illustrate these procedures in Section 4, and some elements of discussion and conclusion are given in Section 5.

2. MODEL STRUCTURE

Let y_{ij} be the jth measurement ($j = 1, 2, \ldots, n_i$) recorded on the ith individual $i = 1, 2, \ldots, I$ at time t_{ij}. The class of models considered here can be written as follows:

$$y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \varepsilon_{ij} \tag{1}$$

where $\mathbf{x}'_{ij}\boldsymbol{\beta}$ represents the systematic component expressed as a linear combination of p explanatory variables (row vector $\mathbf{x}'_{ij}$) with unknown linear coefficients (vector $\boldsymbol{\beta}$), and ε_{ij} is the random component.

As in [3], ε_{ij} is decomposed as the sum of three elements:

$$\varepsilon_{ij} = \sum_{k=1}^{K} z_{ijk} u_{ik} + w_i(t_{ij}) + e_{ij}. \tag{2}$$

The first term represents the additive effect of K random regression factors u_{ik} on covariable information z_{ijk} (usually a $(k-1)$th power of time) and which are specific to each ith individual. The second term $w_i(t_{ij})$ corresponds to the contribution of a stationary Gaussian time process, and the third term e_{ij} is the so-called measurement error.

By gathering the n_i measurements made on the ith individual such that $\mathbf{y}_i = \{y_{ij}\}$, $\boldsymbol{\varepsilon}_i = \{\varepsilon_{ij}\}$ and $\mathbf{X}_{i(n_i \times p)} = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{in_i})'$, (1) and (2) can be expressed in matrix notation as

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i, \tag{3}$$

and

$$\boldsymbol{\varepsilon}_i = \mathbf{Z}_i \mathbf{u}_i + \mathbf{W}_i + \mathbf{e}_i, \tag{4}$$

where $\mathbf{Z}_{i(n_i \times K)} = (\mathbf{z}_{i1}, \mathbf{z}_{i2}, \ldots, \mathbf{z}_{ij}, \ldots, \mathbf{z}_{in_i})'$, $\mathbf{z}_{ij(K \times 1)} = \{z_{ijk}\}$, $\mathbf{u}_{i(K \times 1)} = \{u_{ik}\}$ for $k = 1, 2, \ldots, K$, $\mathbf{W}_i = \{w_i(t_{ij})\}$, and $\mathbf{e}_i = \{e_{ij}\}$ for $j = 1, 2, \ldots, n_i$.

We will assume that $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \mathbf{V}_i)$ with

$$\mathbf{V}_i = \mathbf{Z}_i \mathbf{G} \mathbf{Z}_i' + \mathbf{R}_i \tag{5}$$

where $\mathbf{G}_{(K \times K)}$ is a symmetric positive definite matrix, which may alternatively be represented under its vector form $\mathbf{g} = \mathrm{vech}\mathbf{G}$. For instance, for a linear regression, $\mathbf{g} = (g_{00}, g_{01}, g_{11})'$ where g_{00} refers to the variance of the intercept, g_{11} to the variance of the linear regression coefficient and g_{01} to their covariance.

$\mathbf{R}_i$ in (5) has the following structure in the general case

$$\mathbf{R}_i = \sigma^2 \mathbf{H}_i + \sigma_e^2 \mathbf{I}_{n_i}, \tag{6}$$

where $\sigma_e^2 \mathbf{I}_{n_i} = \mathrm{var}(\mathbf{e}_i)$, and for stationary Gaussian simple processes, σ^2 is the variance of each $w_i(t_{ij})$ and $\mathbf{H}_i = \{h_{ij,ij'}\}$ the $(n_i \times n_i)$ correlation matrix among them such that $h_{ij,ij'} = f(\rho, d_{ij,ij'})$ can be written as a function f of a real positive number ρ and of the absolute time separation $d_{ij,ij'} = |t_{ij} - t_{ij'}|$ between measurements j and j' made on the individual i.

Classical examples of such functions are the power: $f(\rho, d) = \rho^d$; the exponential: $\exp(-d/\rho)$, and the Gaussian: $\exp(-d^2/\rho^2)$, functions. Notice that for equidistant intervals, these functions are equivalent and reduce to a first order autoregressive process (AR1).

$\mathbf{R}_i$ in (6) can be alternatively expressed in terms of ρ, σ^2 and of the ratio $\lambda = \sigma_e^2/\sigma^2$

$$\mathbf{R}_i = \sigma^2(\mathbf{H}_i + \lambda \mathbf{I}_{n_i}) = \sigma^2 \tilde{\mathbf{H}}_i. \tag{7}$$

This parameterisation via $\mathbf{r} = (\sigma^2, \rho, \lambda)'$ allows models to be addressed both with and without measurement error variance (or "nugget" in geostatistics).

3. EM IMPLEMENTATION

Let $\boldsymbol{\gamma} = (\mathbf{g}', \mathbf{r}')'$ be the $3 + K(K+1)/2$ parameter vector and $\mathbf{x} = (\mathbf{y}', \boldsymbol{\beta}', \mathbf{u}')'$ be the complete data vector where $\mathbf{y} = (\mathbf{y}_1', \mathbf{y}_2', \ldots, \mathbf{y}_i', \ldots, \mathbf{y}_I')'$ and $\mathbf{u} = (\mathbf{u}_1', \mathbf{u}_2' \ldots, \mathbf{u}_i', \ldots, \mathbf{u}_I')'$. Following Dempster et al. [1], the EM algorithm proceeds from the log-likelihood $L(\boldsymbol{\gamma}; \mathbf{x}) = \ln p(\mathbf{x}|\boldsymbol{\gamma})$ of $\mathbf{x}$ as a function of $\boldsymbol{\gamma}$. Here $L(\boldsymbol{\gamma}; \mathbf{x})$ can be decomposed as the sum of the log-likelihood of $\mathbf{u}$ as a function of $\mathbf{g}$ and of the log-likelihood of $\boldsymbol{\varepsilon}^* = \mathbf{y} - \mathbf{X}\beta - \mathbf{Z}\mathbf{u}$ as a function of $\mathbf{r}$,

$$L(\boldsymbol{\gamma}; \mathbf{x}) = L(\mathbf{r}; \boldsymbol{\varepsilon}^*) + L(\mathbf{g}; \mathbf{u}) + const., \tag{8}$$

where $\mathbf{X}_{(N \times p)} = (\mathbf{X}_1', \mathbf{X}_2', \ldots, \mathbf{X}_i', \ldots, \mathbf{X}_I')'$
and $\mathbf{Z}_{(N \times KI)} = (\mathbf{Z}_1', \mathbf{Z}_2', \ldots, \mathbf{Z}_i', \ldots, \mathbf{Z}_I')'$.

Under normality assumptions, the two log-likelihoods in (8) can be expressed as:

$$L(\mathbf{g}; \mathbf{u}) = -1/2 \left[KI\ln 2\pi + I\ln|\mathbf{G}| + \sum_{i=1}^{I} \mathbf{u}_i' \mathbf{G}^{-1} \mathbf{u}_i \right], \tag{9}$$

$$L(\mathbf{r}; \boldsymbol{\varepsilon}^*) = -1/2 \left[N\ln 2\pi + \sum_{i=1}^{I} \ln|\mathbf{R}_i| + \sum_{i=1}^{I} \boldsymbol{\varepsilon}_i^{*\prime} \mathbf{R}_i^{-1} \boldsymbol{\varepsilon}_i^* \right]. \tag{10}$$

The E-step consists of evaluating the conditional expectation of the complete data log-likelihood $L(\boldsymbol{\gamma}; \mathbf{x}) = \ln p(\mathbf{x}|\boldsymbol{\gamma})$ given the observed data $\mathbf{y}$ with $\boldsymbol{\gamma}$ set at its current value $\boldsymbol{\gamma}^{[t]}$ i.e., evaluating the function

$$Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{[t]}) = E[L(\boldsymbol{\gamma}; \mathbf{x})|\mathbf{y}, \boldsymbol{\gamma} = \boldsymbol{\gamma}^{[t]}], \tag{11}$$

while the M-step updates $\boldsymbol{\gamma}$ by maximizing (11) with respect to $\boldsymbol{\gamma}$ i.e.,

$$\boldsymbol{\gamma}^{[t+1]} = \arg \max_{\boldsymbol{\gamma}} Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{[t]}). \tag{12}$$

The formula in (8) allows the separation of $Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{[t]})$ into two components, the first $Q_u(\mathbf{g}|\boldsymbol{\gamma}^{[t]})$ corresponding to $\mathbf{g}$, and the second $Q_\varepsilon(\mathbf{r}|\boldsymbol{\gamma}^{[t]})$ corresponding to $\mathbf{r}$, i.e.,

$$Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{[t]}) = Q_u(\mathbf{g}|\boldsymbol{\gamma}^{[t]}) + Q_\varepsilon(\mathbf{r}|\boldsymbol{\gamma}^{[t]}). \tag{13}$$

We will not consider the maximization of $Q_u(\mathbf{g}|\boldsymbol{\gamma}^{[t]})$ with respect to $\mathbf{g}$ in detail; this is a classical result: see e.g., Henderson [11], Foulley et al. [6] and Quaas [23]. The (k, l) element of $\mathbf{G}$ can be expressed as

$$(\mathbf{G}^{[t+1]})_{kl} = E \left(\sum_{i=1}^{I} u_{ik} u_{il} | \mathbf{y}, \boldsymbol{\gamma}^{[t]} \right). \tag{14}$$

If individuals are not independent (as happens in genetical studies), one has to replace $\sum_{i=1}^{I} u_{ik} u_{il}$ by $\mathbf{u}_k' \mathbf{A}^{-1} \mathbf{u}_l$ where $\mathbf{u}_k = \{u_{ik}\}$ for $i = 1, 2, \ldots, I$ and $\mathbf{A}$ is a $(I \times I)$ symmetric, positive definite matrix of known coefficients.

Regarding $\mathbf{r}$, $Q_\varepsilon(\mathbf{r}|\boldsymbol{\gamma}^{[t]})$ can be made explicit from (10) as

$$Q_\varepsilon(\mathbf{r}|\boldsymbol{\gamma}^{[t]}) = -1/2 \left[\sum_{i=1}^{I} \ln|\mathbf{R}_i| + \sum_{i=1}^{I} \mathrm{tr}(\mathbf{R}_i^{-1} \boldsymbol{\Omega}_i) \right] + const., \tag{15}$$

where $\mathbf{\Omega}_{i(n_i \times n_i)} = E(\boldsymbol{\varepsilon}_i^* \boldsymbol{\varepsilon}_i^{*\prime}|\mathbf{y}, \boldsymbol{\gamma}^{[t]})$ which can be computed from the elements of Henderson's mixed model equations [10, 11].

Using the decomposition of $\mathbf{R}_i$ in (7), this expression reduces to (16)

$$Q_\varepsilon(\mathbf{r}|\boldsymbol{\gamma}^{[t]}) = -1/2 \Big[N\ln \sigma^2 + \sum_{i=1}^{I} \ln|\tilde{\mathbf{H}}_i(\rho, \lambda)|$$

$$+ \sigma^{-2} \sum_{i=1}^{I} \mathrm{tr}\{[\tilde{\mathbf{H}}_i(\rho, \lambda)]^{-1}\mathbf{\Omega}_i\} \Big] + const.$$

In order to maximize $Q_\varepsilon(\mathbf{r}|\boldsymbol{\gamma}^{[t]})$ in (16) with respect to $\mathbf{r}$, we suggest using the gradient-EM technique [12] i.e., solving the M-step by one iteration of a second order algorithm. Since here $E(\mathbf{\Omega}_i) = \sigma^2 \tilde{\mathbf{H}}_i$, calculations can be made easier using the Fisher information matrix as in [31]. Letting $\dot{\mathbf{Q}} = \partial\mathbf{Q}/\partial\mathbf{r}$, $\overline{\overline{\mathbf{Q}}} = E(\partial^2\mathbf{Q}/\partial\mathbf{r}\partial\mathbf{r}')$ the system to solve can be written

$$-\overline{\overline{\mathbf{Q}}}\mathbf{\Delta r} = \dot{\mathbf{Q}}, \tag{17}$$

where $\mathbf{\Delta r}$ is the increment in $\mathbf{r}$ from one iteration to the next.

Here, elements of $\dot{\mathbf{Q}}$ and $\overline{\overline{\mathbf{Q}}}$ can be expressed as:

$$\dot{q}_1 = N\sigma^{-2} - \sigma^{-4} \sum_{i=1}^{I} \mathrm{tr}(\tilde{\mathbf{H}}_i^{-1}\mathbf{\Omega}_i)$$

$$\dot{q}_2 = \sum_{i=1}^{I} \mathrm{tr}\left[\frac{\partial\mathbf{H}_i}{\partial\rho}(\tilde{\mathbf{H}}_i^{-1} - \sigma^{-2}\tilde{\mathbf{H}}_i^{-1}\mathbf{\Omega}_i\tilde{\mathbf{H}}_i^{-1}) \right]$$

$$\dot{q}_3 = \sum_{i=1}^{I} \mathrm{tr}(\tilde{\mathbf{H}}_i^{-1} - \sigma^{-2}\tilde{\mathbf{H}}_i^{-1}\mathbf{\Omega}_i\tilde{\mathbf{H}}_i^{-1})$$

and

$$\overline{\overline{q}}_{11} = N\sigma^{-4}; \quad \overline{\overline{q}}_{12} = \sigma^{-2} \sum_{i=1}^{I} \mathrm{tr}\left(\frac{\partial\mathbf{H}_i}{\partial\rho}\tilde{\mathbf{H}}_i^{-1} \right)$$

$$\overline{\overline{q}}_{13} = \sigma^{-2} \sum_{i=1}^{I} \mathrm{tr}(\tilde{\mathbf{H}}_i^{-1}); \quad \overline{\overline{q}}_{22} = \sum_{i=1}^{I} \mathrm{tr}\left(\frac{\partial\mathbf{H}_i}{\partial\rho}\tilde{\mathbf{H}}_i^{-1}\frac{\partial\mathbf{H}_i}{\partial\rho}\tilde{\mathbf{H}}_i^{-1} \right)$$

$$\overline{\overline{q}}_{23} = \sum_{i=1}^{I} \mathrm{tr}\left(\tilde{\mathbf{H}}_i^{-1}\frac{\partial\mathbf{H}_i}{\partial\rho}\tilde{\mathbf{H}}_i^{-1} \right); \quad \overline{\overline{q}}_{33} = \sum_{i=1}^{I} \mathrm{tr}(\tilde{\mathbf{H}}_i^{-1}\tilde{\mathbf{H}}_i^{-1})$$

where 1, 2 and 3 refer to σ^2, ρ and λ respectively.

The expressions for $\dot{\mathbf{Q}}$ and $\overline{\overline{\mathbf{Q}}}$ are unchanged for models without measurement error; one just has to reduce the dimension by one and use $\mathbf{H}_i$ in place of $\tilde{\mathbf{H}}_i$.

The minimum of $-2L$ can be easily computed from the general formula given by Meyer [20] and Quaas [23]

$$-2L_{\mathrm{m}} = [N - r(\mathbf{X})]\ln 2\pi + \ln|\mathbf{G}^{\#}| + \ln|\mathbf{R}^{\#}| + \ln|\mathbf{M}^{\#}| + \mathbf{y}'\,\mathbf{R}^{\#-1}\mathbf{y} - \hat{\boldsymbol{\theta}}'\,\mathbf{R}^{\#-1}\mathbf{y}$$

where $\mathbf{G}^{\#} = \mathbf{A} \otimes \mathbf{G}$ ($\mathbf{A}$ is usually the identity matrix), $\mathbf{R}^{\#} = \oplus_{i=1}^{I}\mathbf{R}_i$, ($\otimes$ and $\oplus$ standing for the direct product and sum respectively) $\mathbf{M}^{\#} = \mathbf{M}/\sigma^2$ with $\mathbf{M}$ the coefficient matrix of Henderson's mixed model equations in $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}', \hat{\mathbf{u}}')'$ i.e., for $\mathbf{T}_i = (\mathbf{X}_i, 0, 0, \ldots, \mathbf{Z}_i, \ldots, 0)$ and $\boldsymbol{\Gamma}^{-} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{\#-1} \end{bmatrix}$,

$$\mathbf{M} = \sum_{i=1}^{I} \mathbf{T}_i' \tilde{\mathbf{H}}_i^{-1} \mathbf{T}_i + \sigma^2 \boldsymbol{\Gamma}^{-}.$$

Here $\mathbf{y}'\mathbf{R}^{\#-1}\mathbf{y} - \hat{\boldsymbol{\theta}}'\,\mathbf{R}^{\#-1}\mathbf{y} = [N - r(\mathbf{X})]\hat{\sigma}^2/\sigma^2$ which equals to $N - r(\mathbf{X})$ for σ^2 evaluated at its REML estimate, so that eventually

$$-2L_{\mathrm{m}} = [N - r(\mathbf{X})](1 + \ln 2\pi) + K\ln|\mathbf{A}| + I\ln|\mathbf{G}| + \ln|\mathbf{M}|$$

$$+ [N - \dim(\mathbf{M})]\ln \hat{\sigma}^2 + \sum_{i=1}^{I} \ln|\tilde{\mathbf{H}}_i|. \quad (18)$$

This formula is useful to compute likelihood ratio test statistics for comparing models, as advocated by Foulley and Quaas [5] and Foulley et al. [7,8].

4. NUMERICAL APPLICATION

The procedures presented here are illustrated with a small data set due to Pothoff and Roy [22]. These data shown in Table I contain facial growth measurements made on 11 girls and 16 boys at four ages (8, 10, 12 and 14 years) with the nine deleted values at age 10 defined in Little and Rubin [14].

The mean structure considered is the one selected by Verbeke and Molenberghs [32] in their detailed analysis of this example and involves an intercept and a linear trend within each sex such that

$$E(y_{ijk}) = \mu + \alpha_i + \beta_i t_j, \quad (19)$$

where μ is a general mean, α_i is the effect of sex ($i = 1, 2$ for female and male children respectively), and β_i is the slope within sex i of the linear increase with time t measured at age j ($t_j = 8, 10, 12$ and 14 years).

The model was applied using a full rank parameterisation of the fixed effects defined as $\boldsymbol{\beta}' = (\mu + \alpha_1, \alpha_2 - \alpha_1, \beta_1, \beta_2 - \beta_1)$. Given this mean structure, six models were fitted with different covariance structures. These models are symbolized as follows with their number of parameters indicated within brackets:

{1}	intercept + error	(2)
{2}	POW	(2)
{3}	POW + measurement error	(3)
{4}	intercept + POW	(3)
{5}	intercept + linear trend + error	(4)
{6}	unspecified	(10)

Table I. Growth measurements in 11 girls and 16 boys (from Pothoff and Roy [22] and Little and Rubin [14]).

	Age (years)					Age (years)			
Girl	8	10	12	14	Boy	8	10	12	14
1	210	200	215	230	1	260	250	290	310
2	210	215	240	255	2	215		230	265
3	205		245	260	3	230	225	240	275
4	235	245	250	265	4	255	275	265	270
5	215	230	225	235	5	200		225	260
6	200		210	225	6	245	255	270	285
7	215	225	230	250	7	220	220	245	265
8	230	230	235	240	8	240	215	245	255
9	200		220	215	9	230	205	310	260
10	165		190	195	10	275	280	310	315
11	245	250	280	280	11	230	230	235	250
					12	215		240	280
					13	170		260	295
					14	225	255	255	260
					15	230	245	260	300
					16	220		235	250

Distance from the centre of the pituary to the pteryomaxillary fissure (unit 10^{-4} m).

Table II. Covariance structures associated with the models considered.

Models[a]	$\mathbf{Z}_i$	$\mathbf{G}$	$\mathbf{R}_i$
{1}	$\mathbf{1}_{n_i}$	g_{00}	$\sigma_e^2 \mathbf{I}_{n_i}$
{2}	$\mathbf{0}_{n_i}$	—	$\sigma^2 \mathbf{H}_i$
{3}	$\mathbf{0}_{n_i}$	—	$\sigma^2 \mathbf{H}_i + \sigma_e^2 \mathbf{I}_{n_i}$
{4}	$\mathbf{1}_{n_i}$	g_{00}	$\sigma^2 \mathbf{H}_i$
{5}	$(\mathbf{1}_{n_i}, \mathbf{t}_i)$	$\begin{pmatrix} g_{00} & g_{01} \\ g_{01} & g_{11} \end{pmatrix}$	$\sigma_e^2 \mathbf{I}_{n_i}$
{6}	$\mathbf{0}_{n_i}$	—	$\{\sigma_{e_{[t_i, t_i']}}\}$

[a] {1} = intercept + error; {2} = POW; {3} = POW + measurement error; {4} = intercept + POW; {5} = intercept + linear trend + error; {6} = unspecified where POW is defined as $\sigma^2 \mathbf{H}_i$ with $\mathbf{H}_i = \{h_{i,tt'} = \rho^{|t_i - t_i'|}\}$; $\mathbf{t}_i$ is the $n_i \times 1$ vector of ages at wich measurements are made on individual i.

Variance covariance structures associated with each of these six models are shown in Table II. Due to the data structure, the power function $f(\rho, d) = \rho^d$ (in short POW) reduces here to an autoregressive first order process (AR1) having as correlation parameter ρ^2.

EM-REML estimates of the parameters of those models were computed via the techniques presented previously. Iterations were stopped when the norm

$$\sqrt{\left(\sum_i \Delta\gamma_i^2\right) \bigg/ \left(\sum_i \gamma_i^2\right)}$$ of both $\mathbf{g}$ and $\mathbf{r}$, was smaller than 10^{-6}. Estimates of

$\mathbf{g}$ and $\mathbf{r}$, $-2L$ values and the corresponding elements of the covariance structure for each model are shown in Tables III and IV.

Random coefficient models such as {5} are especially demanding in terms of computing efforts. Models involving time processes and measurement errors require a backtracking procedure [2] at the beginning of the iterative process i.e., one has to compute $\mathbf{r}^{[k+1]}$ as the previous value $\mathbf{r}^{[k]}$ plus a fraction $\overline{\omega}^{[k+1]}$ of the Fisher scoring increment $\Delta\mathbf{r}^{[k+1]}$ where $\mathbf{r}^{[k]}$ is the parameter vector defined as previously at iteration k. For instance, we used $\overline{\omega} = 0.10$ up to $k = 3$ in the case of model 3.

Model comparisons are worthwhile at this stage to discriminate between all the possibilities offered. However, within the likelihood framework, one has to check first whether models conform to nested hypotheses for the likelihood test procedure to be valid.

E.g. model 3 (POW + m-error) can be compared to model 2 (POW), as model 2 is a special case of model 3 for $\sigma_e^2 = 0$, and also to model 1 (intercept) which corresponds to $\rho = 1$. The same reasoning applies to the 3-parameter model 4 (intercept + POW) which can be contrasted to model 1 (equivalent to model 4 for $\rho = 0$) and also to model 2 (equivalent to model 4 for $g_{00} = 0$).

In these two examples, the null hypothesis (H_0) can be described as a point hypothesis with parameter values on the boundary of the parameter space which implies some change in the asymptotic distribution of the likelihood ratio statistic under H_0 [29, 30]. Actually, in these two cases, the asymptotic null distribution is a mixture $1/2X_0^2 + 1/2X_1^2$ of the usual chi-square with one degree of freedom X_1^2 and of a Dirac (probability mass of one) at zero (usually noted X_0^2) with equal weights. This results in a P-value which is half the standard one i.e., $P - \text{value} = 1/2\Pr[X_1^2 > \Delta(-2L)_{\text{obs}}]$; see also Robert-Granié et al. [26], page 556, for a similar application.

In all comparisons, model 2 (POW) is rejected while model 1 (intercept) is accepted. This is not surprising as model 2 emphasizes the effect of time separation on the correlation structure too much as compared to the values observed in the unspecified structure (Tab. IV). Although not significantly different from model 1, models 3 (POW + measurement error) and 5 (intercept + linear trend) might also be good choices with a preference to the first one due to the lower number of parameters.

As a matter of fact, as shown in Table III, one can construct several models with the same number of parameters which cannot be compared. There are two models with two parameters (models 1 and 2) and also two with three parameters (models 3 and 4). The same occurs with four parameters although only the random coefficient model was displayed because fitting the alternative model (intercept + POW + measurement error) reduces here to fitting the sub-model 3 (POW + measurement error) due to $\hat{g}_{00}$ becoming very small. Incidentally, running SAS Proc MIXED on this alternative model leads to $\hat{g}_{00} = 331.4071$, $\hat{\rho} = 0.2395$ and $\hat{\sigma}_e^2 = 1.0268$ i.e. to fitting model 4 (intercept +

Table III. Likelihood statistics (ML, REML, $-2L$) of several models for the analysis of facial growth in 11 girls and 16 boys (Pothoff and Roy [22]; Little and Rubbin [14]).

Method	Model	Random Effect[a]			Time process[b]		Error[c]	#iter[d]	#par	$-2L$	Likelihood			
		g_{00}	g_{01}	g_{11}	σ^2	ρ	σ_e^2				comp	$\Delta[-2L]$	Distr[e]	P-value
REML	{1}	337.27					207.48	16	2	843.6408				
	{2}				545.40	0.802		9	2	850.7416				
	{3}				380.96	0.966	164.99	13	3	842.8263	{3}-{2}	7.9153	0:1	0.0024
											{3}-{1}	0.8145	0:1	0.1833
	{4}	331.42			213.60	0.239		43	3	843.5586	{4}-{2}	7.1830	0:1	0.0037
											{4}-{1}	0.0822	0:1	0.3872
	{5}	835.50	−46.53	4.42			176.66	163	4	842.3559	{5}-{1}	1.2849	1:2	0.3914
	{6}[f]								10	835.3176	{6}-{1}	8.3222	8	0.4025
ML	{1}	309.53					201.74	15	2	857.2247				
	{2}				510.95	0.792		7	2	865.4353				
	{3}				342.73	0.971	168.69	13	3	856.7004	{3}-{2}	8.7349	0:1	0.0016
											{3}-{1}	0.5243	0:1	0.2345
	{4}	307.36			203.92	0.151		39	3	857.2106	{4}-{2}	8.2247	0:1	0.0021
											{4}-{1}	0.0141	0:1	0.4527
	{5}	678.63	−34.99	3.37			177.00	209	4	856.3640	{5}-{1}	0.8607	1:2	0.5019
	{6}[f]								10	849.1997	{6}-{1}	8.0250	8	0.4310

[a] Random effects model intercept (0) and/or slope (1): fixed part with sex, linear regression on age varying according to sex.
[b] POW process: $Cov_{jk} = \rho^{d_{jk}}\sigma^2$.
[c] Residual in random regression models and measurement error σ_e^2 ("nugget") for models involving $\mathbf{R}_i = \sigma^2 \mathbf{H}_i + \sigma_e^2 \mathbf{I}_{n_i}$.
[d] Stopping rule: norm set to power −6.
[e] Asymptotic distribution of the likelihood ratio under the null hypothesis: Chi-square or mixture of chi-squares.
[f] Model with an unspecified covariance structure (10 parameters).

Table IV. Variance and correlation matrix among measures within individuals generated by several models for the analysis of facial growth in 11 girls and 16 boys.

a) REML

Model		Variances				Correlations					
	σ_{11}	σ_{22}	σ_{33}	σ_{44}	r_{12}	r_{23}	r_{34}	r_{13}	r_{24}	r_{14}
{1} inter	544.75	544.75	544.75	544.75	0.6191	0.6191	0.6191	0.6191	0.6191	0.6191
{2} POW	545.40	545.40	545.40	545.40	0.6426	0.6426	0.6426	0.4129	0.4129	0.2654
{3} POW+ m-error	545.95	545.95	545.95	545.95	0.6511	0.6511	0.6511	0.6075	0.6075	0.5669
{4} inter + POW	545.02	545.02	545.02	545.02	0.6304	0.6304	0.6304	0.6094	0.6094	0.6082
{5} inter + slope	550.31	523.14	531.30	574.77	0.6546	0.6482	0.6651	0.6081	0.6145	0.5448
{6} unspecified	542.29	486.59	626.82	498.94	0.6334	0.4963	0.7381	0.6626	0.6164	0.5225

b) ML

Model		Variances				Correlations					
	σ_{11}	σ_{22}	σ_{33}	σ_{44}	r_{12}	r_{23}	r_{34}	r_{13}	r_{24}	r_{14}
{1} inter	511.26	511.26	511.26	511.26	0.6054	0.6054	0.6054	0.6054	0.6054	0.6054
{2} POW	510.95	510.95	510.95	510.95	0.6265	0.6265	0.6265	0.3925	0.3925	0.2459
{3} POW+ m-error	511.42	511.42	511.42	511.42	0.6323	0.6323	0.6323	0.5967	0.5967	0.5630
{4} inter + POW	511.28	511.28	511.28	511.28	0.6103	0.6103	0.6103	0.6014	0.6014	0.6012
{5} inter + slope	511.41	492.74	501.02	536.25	0.6341	0.6302	0.6461	0.5971	0.6041	0.5465
{6} unspecified	505.12	455.79	598.03	462.32	0.6054	0.4732	0.7266	0.6570	0.6108	0.5226

POW). However, since the value of $-2L_{\mathrm{m}}$ for model 4 is slightly higher than that for model 3, it is the EM procedure which gives the right answer.

5. DISCUSSION-CONCLUSION

This study clearly shows that the EM algorithm is a powerful tool for calculating maximum likelihood estimates of dispersion parameters even when the covariance matrix $\mathbf{V}$ is not linear in the parameters as postulated in linear mixed models.

The EM algorithm allows separation of the calculations involved in the $\mathbf{R}$ matrix parameters (time processes + errors) and those arising in the $\mathbf{G}$ matrix parameters (random coefficients), thus making its application to a large class of models very flexible and attractive.

The procedure can also be easily adapted to get ML rather than REML estimates of parameters with very little change in the implementation, involving only an appropriate evaluation of the conditional expectation of $u_{ik}u_{il}$ and of $\varepsilon_i^* \varepsilon_i^{*\prime}$ along the same lines as given by Foulley et al. [8]. Corresponding results for the numerical example are shown in Tables III and IV suggesting as expected some downward bias for variances of random coefficient models and of time processes.

Several variants of the standard EM procedure are possible such as those based e.g., on conditional maximization [15, 18, 19] or parameter expansion [16]. In the case of models without "measurement errors", an especially simple ECME procedure consists of calculating $\rho^{[t+1]}$ for σ^2 fixed at $\sigma^{2[t]}$, with $\sigma^{2[t+1]}$ being updated by direct maximization of the residual likelihood (without recourse to missing data), i.e.,

$$\rho^{[t+1]} = \rho^{[t]} - \frac{\sum_{i=1}^{I} \mathrm{tr}\left[\frac{\partial \mathbf{H}_i}{\partial \rho}(\tilde{\mathbf{H}}_i^{-1} - \sigma^{-2}\tilde{\mathbf{H}}_i^{-1}\boldsymbol{\Omega}_i\tilde{\mathbf{H}}_i^{-1})\right]}{\sum_{i=1}^{I} \mathrm{tr}\left(\frac{\partial \mathbf{H}_i}{\partial \rho}\tilde{\mathbf{H}}_i^{-1}\frac{\partial \mathbf{H}_i}{\partial \rho}\tilde{\mathbf{H}}_i^{-1}\right)}, \tag{20}$$

$$\sigma^{2[t+1]} = \frac{\sum_{i=1}^{I}\left[\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}(\rho^{[t]}, \mathbf{D}^{[t]})\right]'[\mathbf{W}_i(\rho^{[t]}, \mathbf{D}^{[t]})]^{-1}[\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}(\rho^{[t]}, \mathbf{D}^{[\mathbf{t}]})]}{N - r(\mathbf{X})}$$

where $\mathbf{W}_i = \mathbf{Z}_i'\mathbf{D}\mathbf{Z}_i + \mathbf{H}_i$ with $\mathbf{D} = \mathbf{G}/\sigma^2$, and which can be evaluated using Henderson's mixed model equations by

$$\sigma^{2[t+1]} = \frac{\sum_{i=1}^{I} \mathbf{y}_i'\mathbf{H}_i^{-1}(\rho^{[t]})\mathbf{y}_i - \hat{\boldsymbol{\theta}}' \sum_{i=1}^{I} \mathbf{T}_i'\mathbf{H}_i^{-1}(\rho^{[t]})\mathbf{y}_i}{N - r(\mathbf{X})}. \tag{21}$$

Finally, random coefficient models can also be accommodated to include heterogeneity of variances both at the temporal and environmental levels [8, 24, 25, 27] which enlarges the range of potentially useful models for longitudinal data analysis.

ACKNOWLEDGEMENTS

REFERENCES

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

[2] Denis J.E., Schnabel R.B., Numerical methods for unconstrained optimization and non linear equations, Prentice-Hall Inc, Englewood Cliffs, New Jersey, 1983.

[3] Diggle P.J., An approach to the analysis of repeated measurements, Biometrics 44 (1988) 959–971.

[4] Diggle P.J., Liang K.Y., Zeger S.L., Analysis of longitudinal data, Oxford Science Publications, Clarendon Press, Oxford, 1994.

[5] Foulley J.L., Quaas R.L., Heterogeneous variances in Gaussian linear mixed models, Genet. Sel. Evol. 27 (1995) 211–228.

[6] Foulley J.L., Im S., Gianola D., Hoeschele I., Empirical Bayes estimation of parameters for n polygenic binary traits, Genet. Sel. Evol. 19 (1987) 197–224.

[7] Foulley J.L., Gianola D., San Cristobal M., Im S., A method for assessing extent and sources of heterogeneity of residual variances in mixed linear models, J. Dairy Sci. 73 (1990) 1612–1624.

[8] Foulley J.L., Quaas R.L., Thaon d'Arnoldi C., A link function approach to heterogeneous variance components, Genet. Sel. Evol. 30 (1998) 27–43.

[9] Harville D.A., Recursive estimation using mixed linear models with autoregressive random effects, in: Variance Components and Animal Breeding, in: Proceedings of a Conference in Honor of C.R. Henderson, Cornell University, Ithaca, NY, 1979.

[10] Henderson C.R., Sire evaluation and genetic trends, in: Proceedings of the animal breeding and genetics symposium in honor of Dr J Lush, American society of animal science-American dairy science association, Champaign, 1973, pp. 10–41.

[11] Henderson C.R., Applications of linear models in animal breeding, University of Guelph, Guelph, 1984.

[12] Lange K., A gradient algorithm locally equivalent to the EM algorithm, J. R. Stat. Soc B 57 (1995) 425–437.

[13] Laird N.M., Ware J.H., Random effects models for longitudinal data, Biometrics 38 (1982) 963–974.

[14] Little R.J.A., Rubin D.B., Statistical analysis with missing data. J. Wiley and Sons, New-York, 1977.

[15] Liu C., Rubin D.B., The ECME algorithm: a simple extension of EM and ECM with fast monotone convergence, Biometrika 81 (1994) 633–648.

[16] Liu C., Rubin D.B., Wu Y.N., Parameter expansion to accelerate EM: The PX-EM algorithm, Biometrika 85 (1998) 765–770.

[17] Longford N.T., Random coefficient models, Clarendon Press, Oxford, 1993.

[18] Meng X.L., Rubin D.B., Maximum likelihood estimation via the ECM algorithm: A general framework, Biometrika 80 (1993) 267–278.

[19] Meng X.L., van Dyk D.A., Fast EM-type implementations for mixed effects models, J. R. Stat. Soc. B 60 (1998) 559–578.

[20] Meyer K., Restricted maximum likelihood to estimate variance components for multivariate animal models with several random effects using a derivative-free algorithm, Genet. Sel. Evol. 21 (1989) 317–340.

[21] Patterson H.D., Thompson R., Recovery of interblock information when block sizes are unequal, Biometrika 58 (1971) 545–554.

[22] Pothoff R.F., Roy S.N., A generalized multivariate analysis of variance model useful especially for growth curve problems, Biometrika 51 (1964) 313–326.

[23] Quaas R.L., REML Note book, Mimeo, Cornell University, Ithaca, New York, 1992.

[24] Robert C., Foulley J.L., Ducrocq V., Inference on homogeneity of intra-class correlations among environments using heteroskedastic models, Genet. Sel. Evol. 27 (1995) 51–65.

[25] Robert C., Foulley J.L., Ducrocq V., Estimation and testing of constant genetic and intra-class correlation coefficients among environments, Genet. Sel. Evol. 27 (1995) 125–134.

[26] Robert-Granié C., Ducrocq V., Foulley J.L., Heterogeneity of variance for type traits in the Montbéliarde cattle breed, Genet. Sel. Evol. 29 (1997) 545–570.

[27] San Cristobal M., Foulley J.L., Manfredi E., Inference about multiplicative heteroskedastic components of variance in a mixed linear Gaussian model with an application to beef cattle breeding, Genet. Sel. Evol. 25 (1993) 3–30.

[28] Schaeffer L.R., Dekkers J.C.M., Random regressions in animal models for test-day production in dairy cattle, in: Proceedings 5th World Cong. Genet. Appl. Livest. Prod. 18 (1994) 443–446.

[29] Self S.G., Liang K.Y., Asymptotic properties of maximum likelihood estimation and likelihood ratio tests under nonstandard conditions, J. Am. Stat. Assoc. 82 (1987) 605–610.

[30] Stram D.O., Lee J.W., Variance components testing in the longitudinal mixed effects model, Biometrics 50 (1994) 1171–1177.

[31] Titterington D.M., Recursive parameter estimation using incomplete data. J.R. Stat. Soc. B 46 (1984) 257–267.

[32] Verbeke G., Molenberghs G., Linear mixed models in practice, Springer Verlag, New York, 1997.

[33] Wade K.M., Quaas R.L., Solutions to a system of equations involving a first-order autoregression process, J. Dairy Sci. 76 (1993) 3026–3032.

[34] Wade K.M., Quaas R.L., van Vleck L.D., Estimation of the parameters involved in a first-order autoregressive process for contemporary groups, J. Dairy Sci. 76 (1993) 3033–3040.