*Génét. Sél. Evol.*, 1987, **19** (1), 49-68

# Restricted Maximum Likelihood to estimate variance components for mixed models with two random factors

Karin MEYER

*Institute of Animal Genetics, University of Edinburgh West Mains Road, Edinburgh EH9 3JN, Scotland U.K.*
*and*
*Genetic Improvement of Livestock, Department of Animal and Poultry Science, University of Guelph, Guelph, Ontario N1G 2W1, Canada*

## Summary

A Restricted Maximum Likelihood procedure is described to estimate variance components for a univariate mixed model with two random factors. An EM-type algorithm is presented with a reparameterisation to speed up the rate of convergence. Computing strategies are outlined for models common to the analysis of animal breeding data, allowing for both a nested and a cross-classified design of the 2 random factors. Two special cases are considered : firstly, the total number of levels of fixed effects is small compared to the number of levels of both random factors ; secondly, one fixed effect with a large number of levels is to be fitted in addition to other fixed effects with few levels. A small numerical example is given to illustrate details.

*Key words : Restricted Maximum Likelihood, variance component estimation, nested design, full sib family structure.*

## Résumé

*Estimation des composantes de la variance par le Maximum de Vraisemblance Restreint dans un modèle mixte à deux facteurs aléatoires*

Une méthode d'estimation des composantes de la variance par le Maximum de Vraisemblance Restreint est décrite dans le cas d'un modèle mixte à une seule variable avec 2 facteurs aléatoires. Un algorithme de calcul du type E.M. est présenté avec une reparamétrisation pour accélérer la vitesse de convergence. Des stratégies de calcul sont abordées pour les modèles d'analyse génétique les plus courants avec 2 facteurs aléatoires hiérarchiques ou croisés. Deux cas particuliers sont décrits : premièrement, le nombre total de niveaux des effets fixés est faible comparativement à celui des facteurs aléatoires ; deuxièmement, un effet fixé avec un grand nombre de niveaux est ajouté aux précédents. Un petit exemple numérique illustre les détails.

*Mots clés : Maximum de Vraisemblance Restreint, estimation des composantes de la variance, modèle hiérarchique, familles de pleins frères.*

## I. Introduction

Recently Maximum Likelihood (ML) and related procedures to estimate variance components for unbalanced data have become popular. Restricted Maximum Likelihood (REML), developed by PATTERSON & THOMPSON (1971), which in contrast to ML accounts for the loss in degrees of freedom due to fitting fixed effects, has become accepted as the preferred method to estimate variance components for animal breeding data.

HENDERSON (1973) described an EM-type ML algorithm for several uncorrelated random effects, based on the Mixed Model Equations (MME) for Best Linear Unbiased Prediction (BLUP). Its REML analogue (e.g. HARVILLE, 1977 ; HENDERSON, 1984) is widely used although it is slower to converge than an algorithm using Fisher's Method of Scoring (THOMPSON, 1982). However, it is guaranteed to yield non-negative estimates (HARVILLE, 1977). THOMPSON (1976) outlined an ML procedure to estimate direct and maternal variances. Using small examples HENDERSON (1984) illustrated REML algorithms for a variety of more complex cases, including models accommodating additive and dominance, direct and maternal effects and a three-way classification where variance component estimates for one random factor and all random interactions were required. His algorithm permits a general form of the matrix of residual errors. In a different context, LAIRD & WARE (1982) discussed ML and REML estimation for longitudinal data, invoking a two-stage model which accommodated both growth and repeated measurement models.

In spite of well documented theory, most applications of REML in animal breeding have been restricted to models which include only a single random factor apart from the random residual error. This paper describes a univariate REML procedure for models where three variance components are to be estimated. This encompasses cases with 2 uncorrelated random effects and situations where the variance components for one random factor and its random interaction with a fixed effect are of interest. With an appropriate coding for the interaction, the latter is a special case of the 2 random factor model. For animal breeding data, these are commonly sires and dams. Frequently, there are considerably more dams than sires, in particular with artificial insemination, and sires are used across a wider range of fixed effects than dams. The algorithm has been developed with such a data structure in mind and will be presented in terms pertaining to the animal breeding situation.

## II. The model

Let **y**, of length N, denote the data vector and **b**, of length NF, denote the vector of fixed effects including any regression coefficients for covariables to be fitted. Similarly let **s**, of length NS, and **d**, of length ND, stand for the vectors of the first (e.g. sires) and second (e.g. dams) random effect and **e**, of length N, stand for the random vector of residuals. **X**, **Z** and **W** are the corresponding design matrices for **b**, **s** and **d** of order N × NF, N × NS and N × ND, respectively. The model of analysis can then be written as :

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zs} + \mathbf{Wd} + \mathbf{e} \tag{1}$$

with $E(\mathbf{y}) = \mathbf{Xb}$, $E(\mathbf{s}) = \mathbf{0}$, $E(\mathbf{d}) = \mathbf{0}$ and $E(\mathbf{e}) = \mathbf{0}$ and variances and covariances $V(\mathbf{s}) = \mathbf{G}_S$, $V(\mathbf{d}) = \mathbf{G}_D$, $V(\mathbf{e}) = \mathbf{R}$, $Cov(\mathbf{s},\mathbf{d}') = \mathbf{0}$, $Cov(\mathbf{s},\mathbf{e}') = \mathbf{0}$ and $Cov(\mathbf{d},\mathbf{e}') = \mathbf{0}$

Then $V(\mathbf{y}) = \mathbf{V} = \mathbf{ZG}_S\mathbf{Z}' + \mathbf{WG}_D\mathbf{W}' + \mathbf{R}$. Assuming errors to be uncorrelated and variances to be homogeneous for each random factor, this simplifies to :

$$\mathbf{V} = \mathbf{ZA}_S\mathbf{Z}'\sigma_S^2 + \mathbf{WA}_D\mathbf{W}'\sigma_D^2 + \mathbf{I}_N\sigma_W^2 \tag{2}$$

where $\sigma_S^2 = V(s_j)$, $\sigma_D^2 = V(d_k)$ and $\sigma_W^2 = V(e_m)$ for $j = 1, \ldots, NS$, $k = 1, \ldots, ND$ and $m = 1, \ldots, N$. $\mathbf{A}_S$ and $\mathbf{A}_D$ describe the covariance structure among the levels of each of the 2 random effects. In animal breeding terms, assuming an additive genetic model, for sires and dams, these are the numerator relationship matrices.

The MME for (1) are then (HENDERSON, 1973) :

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} & \mathbf{X'W} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \lambda_S\mathbf{A}_S^{-1} & \mathbf{Z'W} \\ \mathbf{W'X} & \mathbf{W'Z} & \mathbf{W'W} + \lambda_D\mathbf{A}_D^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{s}} \\ \hat{\mathbf{d}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \\ \mathbf{W'y} \end{bmatrix} \tag{3}$$

with variance ratios $\lambda_S = \sigma_W^2/\sigma_S^2$ and $\lambda_D = \sigma_W^2/\sigma_D^2$ (assumed to be the known parameter values).

## III. REML algorithm

To account for the loss in degrees of freedom due to fitting of fixed effects, REML, in contrast to ML, maximizes only the part of the likelihood of the data vector **y** which is independent of the fixed effects. This is achieved by operating on a vector of so-called « error contrasts », **Sy**, with $\mathbf{SX} = \mathbf{0}$ and hence $E(\mathbf{Sy}) = \mathbf{0}$. A suitable matrix **S** arises when absorbing the fixed into the random effects in (3) (THOMPSON, 1973).

$$\begin{bmatrix} \mathbf{Z'SZ} + \lambda_S\mathbf{A}_S^{-1} & \mathbf{Z'SW} \\ \mathbf{W'SZ} & \mathbf{W'SW} + \lambda_D\mathbf{A}_D^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{s}} \\ \hat{\mathbf{d}} \end{bmatrix} = \begin{bmatrix} \mathbf{Z'Sy} \\ \mathbf{W'Sy} \end{bmatrix} \tag{4}$$

with :

$$\mathbf{S} = \mathbf{I}_N - \mathbf{X}(\mathbf{X'X})^{-}\mathbf{X'} \tag{5}$$

Differentiating the log likelihood of **Sy** with respect to the variance components to be estimated then gives the general REML equations :

$$\mathbf{y'P}\delta\mathbf{v}/\delta\theta_i\,\mathbf{Py} = tr(\mathbf{P}\delta\mathbf{v}/\delta\theta_i) \tag{6}$$

where $\theta_i$ stands in turn for $\sigma_S^2$, $\sigma_D^2$ and $\sigma_W^2$. **P** is a projection matrix :

$$\mathbf{P} = \mathbf{S} - \mathbf{S}\,[\mathbf{Z} \vdots \mathbf{W}]\,\mathbf{C} \begin{bmatrix} \mathbf{Z}' \\ \mathbf{W}' \end{bmatrix} \mathbf{S} \tag{7}$$

with :

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{SS} & \mathbf{C}_{SD} \\ \mathbf{C}_{DS} & \mathbf{C}_{DD} \end{bmatrix} = \begin{bmatrix} \mathbf{Z'SZ} + \lambda_S \mathbf{A}_S^{-1} & \mathbf{Z'SW} \\ \mathbf{W'SZ} & \mathbf{W'SW} + \lambda_D \mathbf{A}_D \end{bmatrix}^{-1} \tag{8}$$

From (2), the derivatives of **V** required are :

$\delta \mathbf{v}/\delta\sigma_S^2 = \mathbf{ZA}_S\mathbf{Z}'$, $\delta \mathbf{v}/\delta\sigma_D^2 = \mathbf{WA}_D\mathbf{W}'$ and $\delta \mathbf{v}/\delta\sigma_W^2 = \mathbf{I}_N$

This gives the following estimating equations :

$$\begin{aligned} \hat{\sigma}_S^2 &= [\hat{\mathbf{s}}'\mathbf{A}_S^{-1}\hat{\mathbf{s}} + \sigma_W^2 \operatorname{tr}(\mathbf{A}_S^{-1}\mathbf{C}_{SS})]/\mathrm{NS} \text{ or} \\ \hat{\sigma}_S^2 &= \hat{\mathbf{s}}'\mathbf{A}_S^{-1}\hat{\mathbf{s}}/[\mathrm{NS} - \lambda_S \operatorname{tr}(\mathbf{A}_S^{-1}\mathbf{C}_{SS})] \end{aligned} \tag{9}$$

$$\begin{aligned} \hat{\sigma}_D^2 &= [\hat{\mathbf{d}}'\mathbf{A}_D^{-1}\hat{\mathbf{d}} + \sigma_W^2 \operatorname{tr}(\mathbf{A}_D^{-1}\mathbf{C}_{DD})]/\mathrm{ND} \text{ or} \\ \hat{\sigma}_D^2 &= \hat{\mathbf{d}}'\mathbf{A}_D^{-1}\hat{\mathbf{d}}/[\mathrm{ND} - \lambda_D \operatorname{tr}(\mathbf{A}_D^{-1}\mathbf{C}_{DD})] \end{aligned} \tag{10}$$

and :

$$\hat{\sigma}_W^2 = \hat{\mathbf{e}}'\hat{\mathbf{e}}/[\mathrm{NDFW} + \lambda_S \operatorname{tr}(\mathbf{A}_S^{-1}\mathbf{C}_{SS}) + \lambda_D \operatorname{tr}(\mathbf{A}_D^{-1}\mathbf{C}_{DD})] \tag{11}$$

where $\hat{\mathbf{e}} = \mathbf{y} - \mathbf{X}\hat{\mathbf{b}} - \mathbf{Z}\hat{\mathbf{s}} - \mathbf{W}\hat{\mathbf{d}} = \mathbf{S}(\mathbf{y} - \mathbf{Z}\hat{\mathbf{u}} - \mathbf{W}\hat{\mathbf{d}})$ ; and $\mathrm{NDFW} = \mathrm{N} - \mathrm{NS} - \mathrm{ND} - \operatorname{rank}(\mathbf{X})$ denotes the degrees of freedom for residual. Equivalent expressions to (9) to (11) have been given by HARVILLE (1977), SEARLE (1979) and HENDERSON (1984). Estimates are usually obtained employing an iterative solution scheme. Above and in the following, $\sigma_i^2$, and $\lambda_i$ (or $\alpha_i$) are then thought of as starting values while a superscript « ^ » denotes estimates for the current round of iteration. These equations, (9) to (11), utilize only first derivatives of the likelihood function, resulting in an EM algorithm (DEMPSTER *et al.*, 1977). Alternatively, the right hand side of (6) can be expanded to include second derivatives, resulting in an algorithm equivalent to Fisher's Method of Scoring. Details are given in the Appendix (A).

While the EM algorithm requires only the diagonal blocks ($\mathbf{C}_{SS}$ and $\mathbf{C}_{DD}$) of the inverse of the coefficient matrix for random effects and traces of their simple products with the corresponding inverse of the numerator relationship matrix, off-diagonal blocks and more complicated traces are required for the Method of Scoring algorithm (see (A3) in relation to (9) to (11)). Hence computational requirements per round of iteration for the latter are considerably higher. Though the EM algorithm can be slow to converge, in particular for ratios of variance components common to animal breeding data (THOMPSON, 1982) it is often preferred for its computational ease and the fact that it guarantees estimates in the parameter space.

## IV. Reparameterisation

THOMPSON & MEYER (1986) described a reparameterisation to speed up convergence of a REML algorithm based on first derivatives of the likelihood function. It was derived considering the expectations of mean squares, resulting from the orthogonal partitioning of sums of squares due to factors in the model, in a balanced design. For a model with one random factor, for instance, where the variance components within ($\sigma^2_W$) and between ($\sigma^2_B$) random groups are of interest, it was suggested to estimate parameters $\alpha_W = \sigma^2_W$ and $\alpha_B = \sigma^2_B + \sigma^2_W/K$. The latter is the variance of a group mean if K is the group size. For $K \to \infty$, $\alpha_B$ reduces to $\sigma^2_B$. For a balanced design with K equal to the group size, estimates of $\alpha_B$ and $\alpha_W$ were obtained in one round of iteration. For the unbalanced case a value of K equal to the average group size increased speed of convergence markedly over the EM algorithm on the original scale ($K = \infty$), especially if $\sigma^2_B$ was small compared to $\alpha^2_W$.

### A. *Nested design*

For a model with 2 random factors it is necessary to distinguish between a nested and a cross-classified design. If the second random factor, for instance dams (**d**), is nested within the first, for instance sires (**s**), expectations of mean squares in a balanced hierarchical analysis of variance suggest a reparameterisation to $\alpha_W = \sigma^2_W$, $\alpha_D = \sigma^2_D + \sigma^2_W/K_D$ and $\alpha_S = \sigma^2_S + \alpha_D/K_S = \sigma^2_S + \sigma^2_D/K_S + \sigma^2_W/K_SK_D$. THOMPSON & MEYER (1986) demonstrated for $K_D$ equal to the average dam group size and $K_S$ equal to the average number of dams per sire a considerable reduction in rounds of iteration required for convergence, as compared to values of $K_S = K_D = \infty$. Again, in the balanced case estimates were obtained in one round.

Differentiating the log likelihood of **Sy** with respect to the new parameters $\alpha_S$, $\alpha_D$ and $\alpha_W$ and equating the resulting expressions to zero, « improved » estimates for the three variance components can be derived. The first variance component, $\sigma^2_S$, is derived as before, i.e. according to (9), while (10) is replaced by :

$$\hat{\sigma}^2_D = [\hat{\mathbf{d}}'\mathbf{A}_D^{-1}\hat{\mathbf{d}} - (\sigma^2_D/\sigma^2_S)^2/K_S\hat{\mathbf{s}}'\mathbf{A}_S^{-1}\hat{\mathbf{s}}] \;/ \left(ND - \lambda_D \operatorname{tr}(\mathbf{A}_D^{-1}\mathbf{C}_{DD}) - (\sigma^2_D/\sigma^2_S)\;/\; K_S\,[NS - \lambda_S \operatorname{tr}(\mathbf{A}_S^{-1}\mathbf{C}_{SS})]\right) \tag{12}$$

The residual variance is then found as :

$$\hat{\sigma}^2_W = [\mathbf{y}'\mathbf{S}\mathbf{y} - \mathbf{y}'\mathbf{S}\mathbf{Z}\hat{\mathbf{s}} - \mathbf{y}'\mathbf{S}\mathbf{W}\hat{\mathbf{d}} - \lambda_S\hat{\mathbf{s}}'\mathbf{A}_S^{-1}\hat{\mathbf{s}} - \lambda_D\,(1 + \lambda_D/K_D)\hat{\mathbf{d}}'\mathbf{A}_D^{-1}\hat{\mathbf{d}}] \;/ \left(N - \operatorname{rank}(\mathbf{X}) - [NS - \lambda_S \operatorname{tr}(\mathbf{A}_S^{-1}\mathbf{C}_{SS})] - (1 + \lambda_D/K_D)\,[ND - \lambda_D \operatorname{tr}(\mathbf{A}_D^{-1}\mathbf{C}_{DD})]\right) \tag{13}$$

Clearly, (12) and (13) reduce to (10) and (11) respectively, if $K_S$ and $K_D$ are $\infty$.

Alternatively, an estimator of the general form :

$$\hat{\theta}_i = \theta_i + [\theta_i\,(\delta L/\delta\theta_i)\,\theta_i]/M \tag{14}$$

can be used to determine $\theta_i = \alpha_S$, $\alpha_D$ and $\alpha_W$, where $\delta L/\theta_i$ denotes the partial derivative of the log likelihood of **Sy** with respect to $\theta_i$. M stands for the number of levels or degrees of freedom pertaining to the respective random factor (see THOMPSON & MEYER (1986) for a reasoning for the latter). Estimates of the variance components are then found as $\hat{\sigma}_W^2 = \hat{\alpha}_W$, $\hat{\sigma}_D^2 = \hat{\alpha}_D - \hat{\alpha}_W/k_D$ and $\hat{\sigma}_S^2 = \hat{\alpha}_S - \hat{\alpha}_D/K_S$.

This implies that, in contrast to the scheme above (i.e. (12) and (13)), estimates of $\sigma_W^2$ and $\sigma_D^2$ rather than the starting values are used in back transforming from the reparameterised to the original scale. This appears to be advantageous. For $\theta_i = \alpha_S$, $\alpha_D$ and $\alpha_W$ in turn, this gives (from 14) :

$$\hat{\alpha}_S = \alpha_S + (\alpha_S/\sigma_S^2)^2 \, [\hat{\mathbf{s}}'\mathbf{A}_S^{-1}\hat{\mathbf{s}} - NS\sigma_S^2 + \sigma_W^2 \, tr(\mathbf{A}_S^{-1}\mathbf{C}_{SS})]/(NS - 1) \tag{15}$$

$$\hat{\alpha}_D = \alpha_D + (\alpha_D/\sigma_D^2)^2 \Big( [\hat{\mathbf{d}}'\mathbf{A}_D^{-1}\hat{\mathbf{d}} - ND\sigma_D^2 + \sigma_W^2 \, tr(\mathbf{A}_D^{-1}\mathbf{C}_{DD})] - (\sigma_D^2/\sigma_S^2)^2/K_S$$
$$[(\hat{\mathbf{s}}'\mathbf{A}_S^{-1}\hat{\mathbf{s}} - NS\sigma_S^2 + \sigma_W^2 \, tr(\mathbf{A}_S^{-1}\mathbf{C}_{SS})] \Big)/(ND - NS) \tag{16}$$

$$\hat{\alpha}_W = \alpha_W + \Big( [\hat{\mathbf{e}}'\hat{\mathbf{e}} - (NDFW + \lambda_S \, tr(\mathbf{A}_S^{-1}\mathbf{C}_{SS}) + \lambda_D \, tr(\mathbf{A}_D^{-1}\mathbf{C}_{DD}))\sigma_W^2]$$
$$- \lambda_D^2/K_D \, [\hat{\mathbf{d}}'\mathbf{A}_D^{-1}\hat{\mathbf{d}} - ND\sigma_D^2 + \sigma_W^2 \, tr(\mathbf{A}_D^{-1}\mathbf{C}_{DD})] \Big)/NDFW \tag{17}$$

Obviously, with $\alpha_W = \sigma_W^2$ rearranging (17) yields (13).

### B. *Crossclassified design*

Reparameterised variables for the crossclassified design are $\alpha_W = \sigma_W^2$, $\alpha_D = \sigma_D^2 + \sigma_W^2/K_D$ and $\alpha_S = \sigma_S^2 + \sigma_W^2/K_S$ where suitable values for $K_D$ and $K_S$ may be the average number of records per dam and sire, respectively. From (14),

$$\hat{\alpha}_D = \alpha_D + (\alpha_D/\sigma_D^2)^2 \, [\hat{\mathbf{d}}'\mathbf{A}_D^{-1}\hat{\mathbf{d}} - ND\sigma_D^2 + \sigma_W^2 \, tr(\mathbf{A}_D^{-1}\mathbf{C}_{DD})]/(ND - 1) \tag{18}$$

and

$$\hat{\alpha}_W = \alpha_W + \Big( [\hat{\mathbf{e}}'\hat{\mathbf{e}} - (NDFW + \lambda_S \, tr(\mathbf{A}_S^{-1}\mathbf{C}_{SS}) + \lambda_D \, tr(\mathbf{A}_D^{-1}\mathbf{C}_{DD}))\sigma_W^2]$$
$$- \lambda_S^2/K_S \, [\hat{\mathbf{s}}'\mathbf{A}_S^{-1}\hat{\mathbf{s}} - NS\sigma_S^2 + \sigma_W^2 \, tr(\mathbf{A}_S^{-1}\mathbf{C}_{SS})]$$
$$- \lambda_D^2/K_D \, [\hat{\mathbf{d}}'\mathbf{A}_D^{-1}\hat{\mathbf{d}} - ND\sigma_D^2 + \sigma_W^2 \, tr(\mathbf{A}_D^{-1}\mathbf{C}_{DD})] \Big)/NDFW \tag{19}$$

for $\theta_i = \alpha_D$ and $\alpha_W$, respectively, and (15) for $\theta_i = \alpha_S$. Estimates of $\sigma_W^2$ and $\sigma_D^2$ are then determined as for the nested design and $\sigma_S^2 = \hat{\alpha}_S - \hat{\alpha}_W/K_S$.

## V. Computing strategy

The REML algorithm as described so far centres around the matrix **S** which is of order equal to the number of observations. For most applications, **S** cannot be calculated directly but often special features of the data structure can be exploited to obtain the required terms indirectly.

### A. *Few fixed effects*

Consider a model where the total number of levels of fixed effects, including any regression coefficients for covariables, is small compared to the number of levels of the first random effects.

Assume further that :

i) there are more levels for the second than for the first random effect

ii) $\mathbf{A}_D = \mathbf{I}_{ND}$

iii) $\mathbf{A}_S = \mathbf{I}_{NS}$

The steps are then :

1) Absorb **d** into **s** and **b**. This gives MME

$$\begin{bmatrix} \mathbf{X'KX} & \mathbf{X'KZ} \\ \mathbf{Z'KX} & \mathbf{Z'KZ} + \lambda_S \mathbf{A}_S^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{s}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'Ky} \\ \mathbf{Z'Ky} \end{bmatrix} \tag{20}$$

with $\mathbf{K} = \mathbf{I}_N - \mathbf{W}(\mathbf{W'W} + \lambda_D \mathbf{A}_D^{-1})^{-1}\mathbf{W}'$

If $\mathbf{A}_D = \mathbf{I}_{ND}$, $(\mathbf{W'W} + \lambda_D \mathbf{A}_D^{-1})$ is diagonal and **d** can be absorbed one level at a time.

2) Absorb **s** into **b** giving

$$\mathbf{X'MX}\,\hat{\mathbf{b}} = \mathbf{X'My} \tag{21}$$

with $\mathbf{M} = \mathbf{K} - \mathbf{KZ}(\mathbf{Z'KZ} + \lambda_S \mathbf{A}_S^{-1})^{-1}\mathbf{Z'K}$

If **d** is nested within **s**, **Z'KZ** is diagonal and, for $\mathbf{A}_S = \mathbf{I}_{NS}$, $(\mathbf{Z'KZ} + \lambda_S \mathbf{A}_S^{-1})$ is easily inverted.

3) Obtain solutions for the fixed effects as :

$$\hat{\mathbf{b}} = (\mathbf{X'MX})^{-}\mathbf{X'My} \tag{22}$$

and backsolutions for the random effects

$$\hat{\mathbf{s}} = (\mathbf{Z'KZ} + \lambda_S \mathbf{A}_S^{-1})^{-1}(\mathbf{Z'Ky} - \mathbf{Z'KX}\hat{\mathbf{b}}) \tag{23}$$

and

$$\hat{\mathbf{d}} = (\mathbf{W'W} + \lambda_D \mathbf{A}_D^{-1})^{-1}(\mathbf{W'y} - \mathbf{W'Z}\hat{\mathbf{s}} - \mathbf{W'X}\hat{\mathbf{b}}) \tag{24}$$

4) The REML algorithm requires traces involving the diagonal blocks, $\mathbf{C}_{SS}$ and $\mathbf{C}_{DD}$, of the inverse of the coefficient matrix. These can be derived using partitioned matrix results, utilising inverses and matrix products arising during the absorption steps.

Let :

$\mathbf{H}_F = (\mathbf{X'MX})^{-}$

$\mathbf{H}_S = (\mathbf{Z'KZ} + \lambda_S \mathbf{A}_S^{-1})^{-1}$

$\mathbf{H}_D = (\mathbf{W'W} + \lambda_D \mathbf{A}_D^{-1})^{-1}$

$\mathbf{L}_{XS} = \mathbf{X'KZH}_S$

$\mathbf{L}_{XD} = \mathbf{X'WH}_D$

and :

$$\mathbf{L}_{SD} = \mathbf{Z'WH}_D$$

Then :

$$\mathbf{C}_{SS} = \mathbf{H}_S + \mathbf{H}_S\mathbf{Z'KXH}_F\mathbf{X'KZH}_S \tag{25}$$

and :

$$\mathbf{C}_{DD} = \mathbf{H}_D + \mathbf{H}_D\,[\mathbf{W'X} \,\vdots\, \mathbf{W'Z}] \begin{bmatrix} \mathbf{H}_F & \mathbf{H}_F\mathbf{X'KZH}_S \\ \mathbf{H}_S\mathbf{Z'KXH}_F & \mathbf{C}_{SS} \end{bmatrix} \begin{bmatrix} \mathbf{X'W} \\ \mathbf{Z'W} \end{bmatrix} \mathbf{H}_D \tag{26}$$

The traces are then :

$$\mathrm{tr}(\mathbf{A}_S^{-1}\mathbf{C}_{SS}) = \mathrm{tr}(\mathbf{A}_S^{-1}\mathbf{H}_S) + \mathrm{tr}(\mathbf{H}_F\mathbf{L}_{XS}\mathbf{A}_S^{-1}\mathbf{L'}_{XS}) \tag{27}$$

and :

$$\mathrm{tr}(\mathbf{A}_S^{-1}\mathbf{C}_{DD}) = \mathrm{tr}(\mathbf{A}_D^{-1}\mathbf{H}_D) + \mathrm{tr}(\mathbf{H}_S\mathbf{L}_{SD}\mathbf{A}_D^{-1}\mathbf{L'}_{SD}) + \mathrm{tr}(\mathbf{H}_F\mathbf{T}) \tag{28}$$

with :

$$\begin{aligned} \mathbf{T} &= [\mathbf{X'W} \,\vdots\, -\mathbf{X'KZH}_S\mathbf{Z'W}]\,\mathbf{H}_D\mathbf{A}_D^{-1}\mathbf{H}_D \begin{bmatrix} \mathbf{W'X} \\ -\mathbf{W'ZH}_S\mathbf{Z'KX} \end{bmatrix} \\ &= [\mathbf{L}_{XD} \,\vdots\, -\mathbf{L}_{XS}\mathbf{L}_{SD}]\,\mathbf{A}_D^{-1} \begin{bmatrix} \mathbf{L'}_{XD} \\ -\mathbf{L'}_{SD}\mathbf{L'}_{XS} \end{bmatrix} \end{aligned} \tag{29}$$

Hence, 3 additional symmetric matrices have to be determined to calculate the required traces indirectly : $\mathbf{L}_{SD}\mathbf{A}_D^{-1}\mathbf{L'}_{SD}$ of order equal to the number of levels of **s**, and $\mathbf{L}_{XS}\mathbf{A}_S^{-1}\mathbf{L'}_{XS}$ and **T**, both of order equal to the total number of levels of fixed effects including any regression coefficients. These can efficiently be calculated when absorbing the random effects.

The quadratics in the vector of random effects, $\hat{\mathbf{s}}'\mathbf{A}_S^{-1}\hat{\mathbf{s}}$ and $\hat{\mathbf{d}}'\mathbf{A}_D^{-1}\hat{\mathbf{d}}$, can be calculated directly. The corresponding term for residuals is then determined as :

$$\hat{\mathbf{e}}'\hat{\mathbf{e}} = \mathbf{y'y} - \mathbf{y'Z}\hat{\mathbf{s}} - \mathbf{y'W}\hat{\mathbf{d}} - \mathbf{y'X}\hat{\mathbf{b}} - \lambda_S\hat{\mathbf{s}}'\mathbf{A}_S^{-1}\hat{\mathbf{s}} - \lambda_D\hat{\mathbf{d}}'\mathbf{A}_D^{-1}\hat{\mathbf{d}} \tag{30}$$

### B. *One fixed effect with many levels*

Often the model of analysis includes one fixed effect with many levels, too many to pursue the approach described above. Usually, however, there are still considerably more levels of **d** so that it appears appropriate, first to absorb **d** and then to absorb the major fixed effect into **s** and any additional fixed effects or covariables to be fitted. This strategy requires that the levels of **d** are nested within the levels of the major

TABLE 1

*Data for numerical example*

| Sire | Dam | L size | Treatment 1 | | Treatment 2 | | Treatment 3 | |
|---|---|---|---|---|---|---|---|---|
| | | | Male | Female | Male | Female | Male | Female |
| | | | | | Time period I | | | |
| 1 | 1 | 4 | | | 99 | 130 | 124 | 119 |
| | 2 | 15 | 118 | 106, 109, 125 | 124, 97, 113, 115, 87 | 137, 112 | 93, 95 | 130, 97 |
| | 3 | 8 | 108 | 115 | 106, 116, 136, 103 | | 104 | 125 |
| | 4 | 13 | 109, 98 | 113, 111, 137 | 101, 116 | 122, 104 | 109, 111, 99, 107 | |
| 2 | 6 | 7 | | 138, 123 | 115 | 117, 106 | 119, 104 | |
| | 7 | 7 | 88 | | 116, 116 | 140, 114 | 100 | 109 |
| 3 | 11 | 8 | 123, 93, 107 | 107 | | 122, 114 | | 114, 127 |
| | 14 | 11 | 95 | 101 | 117, 103, 107, 76 | 118, 118, 96 | 117, 87 | |
| | 15 | 6 | | 91, 123, 124 | 98 | | 86, 107 | |
| 4 | 17 | 4 | | | 111, 116 | | 119, 116 | |
| | 18 | 4 | | | 111, 106 | 100 | 102 | |
| | 19 | 13 | 108, 116, 103 | 136 | 107, 116, 101, 107, 109 | | 112, 124 | 117, 103 |
| | 21 | 5 | | | 92 | 120, 107 | 123, 105 | |
| | 22 | 11 | 108, 122 | 117 | 119, 90, 107 | 128, 105, 111 | 83, 97 | |
| 5 | 24 | 12 | 104, 83, 103 | | 101 | 112, 108, 109, 109 | 81, 98 | 118, 120 |
| | 26 | 7 | 102 | 129 | 106, 108 | 123 | 99 | 105 |
| | 30 | 13 | 102, 116 | 106, 117 | 106, 96, 103, 127 | 129 | 116 | 134, 116, 126 |
| | | | | | Time period II | | | |
| 1 | 5 | 15 | | 108 | 90, 105, 92, 109, 110, 104, 118, 101, 108 | 106, 128, 120, 106 | | 130 |
| 2 | 8 | 16 | 94 | 104, 95, 112, 104 | 89, 101, 97 | 97, 92, 125 | 81 | 115, 99, 105, 108 |
| | 9 | 14 | 95 | 85, 117, 114 | 95, 91, 93 | 116, 102, 115, 115, 117, 110 | 91 | |
| | 10 | 13 | 108, 93 | 132, 109 | 106, 100, 100, 90, 89, 108 | 118 | 81, 121 | |
| 3 | 12 | 5 | 103 | 122 | 104 | | | 115, 119 |
| | 13 | 15 | 111, 92, 95 | 102 | 115, 99, 95 | 114, 96, 100 | 85, 108 | 114, 118, 104 |
| | 16 | 14 | 92 | 83 | 94, 111, 91 | 110, 108, 110, 96 | 83, 100, 110, 108 | 108 |
| 4 | 20 | 9 | 124 | | 114 | 132 | 119, 91, 97, 116 | 113, 121 |
| | 23 | 8 | 102 | 114 | 129, 96 | 122, 131 | 114 | 115 |
| 5 | 25 | 9 | 118 | 138 | 93, 114 | 140 | 118 | 125, 129, 114 |
| | 27 | 9 | 98 | | 105, 96, 90, 100 | 111, 110, 103 | | 102 |
| | 28 | 14 | | 117 | 102, 96 | 119, 128, 115, 101, 129, 108, 111, 95, 129 | 102, 94 | |
| | 29 | 5 | 102 | 122 | 112, 126 | 102 | | |

fixed effect or at least within a sufficiently small group thereof. Only then can the inverse required to absorb the fixed effect be calculated. A typical example is the analysis of dairy data where a large number of herd-year-season (HYS) effects has to be taken into account. Assuming cows do not change herds, repeated records for a cow, for instance for milking speed or calving ease, are nested within herds. Details for this case are outlined in the Appendix (B).

## VI. Numerical example

Consider records on progeny of 5 sires and 30 dams, subject to 3 treatments in 2 time periods, as summarized in table 1. Dams are nested within sires and within time periods. Let the model of analysis include the 6 time × treatment subclasses ($h_h$) and two sexes ($b_i$) as fixed effects, litter size ($X_{hijkl}$) as linear covariable and sires ($s_j$) and dams ($d_k$) as random factors,

$$Y_{hijkl} = h_h + b_i + s_j + d_k + b_3 (X_{hijkl} - \overline{X}) + e_{hijkl}$$

where $b_3$ denotes the regression on litter size and $e_{hijkl}$ the residual error associated with $Y_{hijkl}$, the record for the l-th progeny of dam k and sire j and sex i in treatment × time class h. Assume both sires and dams are unrelated, i.e. $\mathbf{A}_S = \mathbf{I}_{NS}$ and $\mathbf{A}_D = \mathbf{I}_{ND}$

### A. *Absorption strategy for few fixed effects*

For $\sigma_S^2 = 10, \sigma_D^2 = 12$ and $\sigma_W^2 = 120$, submatrices for time × treatment classes in period I are :

$$\mathbf{B}_I'\mathbf{K}_I\mathbf{B}_I = \begin{bmatrix} 32.651 & & \text{sym.} \\ -\ 7.952 & 50.999 & \\ -\ 5.552 & -\ 8.856 & 36.855 \end{bmatrix} \quad \mathbf{B}_I'\mathbf{K}_I\mathbf{y}_I = \begin{bmatrix} 2\,140.66 \\ 3\,813.11 \\ 2\,423.77 \end{bmatrix}.$$

and :

$$(\mathbf{L}_{BD}\mathbf{A}_D^{-1}\mathbf{L}_{BD}')_I = \begin{bmatrix} 0.3095 & & \text{sym.} \\ 0.3782 & 0.7576 & \\ 0.2664 & 0.4478 & 0.3131 \end{bmatrix}$$

$$\mathbf{B}_I'\mathbf{K}_I\mathbf{X}_A = \begin{bmatrix} 8.585 & 10.561 & 198.54 \\ 20.845 & 13.345 & 318.09 \\ 15.807 & 6.639 & 205.53 \end{bmatrix}$$

and :

$$(\mathbf{L}_{BD}\mathbf{A}_D^{-1}\mathbf{L}_{XD}')_l = \begin{bmatrix} 0.5434 & 0.4107 & 10.3124 \\ 0.9586 & 0.6250 & 15.9732 \\ 0.6116 & 0.4167 & 10.2803 \end{bmatrix}$$

$$\mathbf{B}_l'\mathbf{K}_l\mathbf{Z} = \begin{bmatrix} 4.885 & 1.765 & 5.050 & 3.168 & 4.279 \\ 8.190 & 4.118 & 5.069 & 10.602 & 6.211 \\ 5.879 & 2.353 & 3.313 & 6.168 & 4.734 \end{bmatrix}$$

and :

$$(\mathbf{L}_{BD}\mathbf{A}_D^{-1}\mathbf{L}_{SD}')_l = \begin{bmatrix} 0.2683 & 0.0727 & 0.2190 & 0.1731 & 0.2211 \\ 0.4059 & 0.1696 & 0.2474 & 0.4412 & 0.3195 \\ 0.2845 & 0.0969 & 0.1461 & 0.2539 & 0.2459 \end{bmatrix}$$

Absorbing all dams,

$$\mathbf{X}_A'\mathbf{K}\mathbf{X}_A = \begin{bmatrix} 113.438 & & \text{sym.} \\ -\ 34.596 & 98.165 & \\ 821.58 & 694.31 & 18061. \end{bmatrix}, \quad \mathbf{X}_A'\mathbf{K}\mathbf{y} = \begin{bmatrix} 7\,850.17 \\ 7\,707.22 \\ 164{,}276. \end{bmatrix}$$

$$\mathbf{L}_{XD}\mathbf{A}_D^{-1}\mathbf{L}_{XD}' = \begin{bmatrix} 2.2663 & & \text{sym.} \\ 1.6165 & 1.5977 & \\ 43.330 & 37.289 & 999.91 \end{bmatrix}$$

With dams nested within sires, the coefficient matrix for sires absorbing dams is diagonal.

$\mathbf{Z}'\mathbf{K}\mathbf{Z}$ = Diag. {24.954  25.875  28.599  29.119  33.865},
$(\mathbf{Z}'\mathbf{K}\mathbf{y})'$ = (2 786.4  2 762.2  3 017.0  3 246.8  3 745.0) and
$\mathbf{L}_{SD}\mathbf{A}_D^{-1}\mathbf{L}_{SD}'$ = Diag. {1.3186  1.3776  1.4239  1.2901  1.6867}

The first term required to calculate $\mathrm{tr}(\mathbf{C}_{DD})$ is $\mathrm{tr}(\mathbf{A}_D^{-1}\mathbf{H}_D) = 1.57588$.

Absorbing sires, (sub)matrices corresponding to $\mathbf{X}_A'\mathbf{K}\mathbf{X}_A$ are :

$$\mathbf{X}_A'\mathbf{M}\mathbf{X}_A = \begin{bmatrix} 82.114 & & \text{sym.} \\ -\ 52.742 & 77.484 & \\ 238.41 & 212.17 & 6\,586.8 \end{bmatrix}$$

$$\begin{bmatrix} 0.7738 & & \text{sym.} \\ 0.5933 & 0.5051 & \\ 14.447 & 11.874 & 285.01 \end{bmatrix} \text{ of } \mathbf{L}_{XS}\mathbf{A}_S{}^{\mathbf{I}}\mathbf{L}_{XS}$$

and :

$$\begin{bmatrix} 0.2882 & & \text{sym.} \\ 0.0524 & 0.2341 & \\ 3.9472 & 4.1987 & 179.953 \end{bmatrix} \text{ of } \mathbf{T}$$

The first term in (27) is then $\mathrm{tr}(\mathbf{A}_S{}^{\mathbf{I}}\mathbf{H}_S) = 0.1752778$, and the second term in (28) is $\mathrm{tr}(\mathbf{H}_S\mathbf{L}_{SD}\mathbf{A}_D{}^{\mathbf{I}}\mathbf{L}_{SD}') = 0.1242176$.

With more than one fixed effect fitted, the coefficient matrix is not of full rank. Hence the row and column of $\mathbf{X'MX}$ pertaining to the first level of each additional, i.e. other than the first, fixed effect are set to zero. Obtaining a generalized inverse gives $\mathrm{tr}(\mathbf{H}_F\mathbf{L}_{XS}\mathbf{A}_S^{-1}\mathbf{L}_{XS}') = 0.0634841$, $\mathrm{tr}(\mathbf{H}_F\mathbf{T}) = 0.1160263$, $\mathrm{tr}(\mathbf{A}_S^{-1}\mathbf{C}_{SS}) = 0.1877017$ and $\mathrm{tr}(\mathbf{A}_D^{-1}\mathbf{C}_{DD}) = 1.867190$.

Corresponding results pursuing a computing strategy suitable for a model with one fixed effect with many levels are given in the Appendix (C).

## B. *Solutions*

For both computing strategies, solutions (or backsolutions) for the fixed effects are $\hat{\mathbf{h}}' = [112.672\ 112.862\ 111.485\ 110.480\ 111.532\ 111.116]$ and $\hat{\mathbf{b}}_A' = [0\ 11.349\ -0.71834]$, while sire and dam effects are predicted as $\hat{\mathbf{s}}' = [2.4608\ -1.3884\ -2.8995\ 1.4868\ 0.3403]$ and $\hat{\mathbf{d}} = [0.1614\ 0.6646\ 0.930\ \ldots\ 0.1335\ 3.5630]$. This gives products of solutions and right hand sides $\hat{\mathbf{b}}_A'\mathbf{X}_A\mathbf{y} = -85{,}022.4$, $\hat{\mathbf{h}}'\mathbf{By} = 3{,}576{,}705.2$, $\hat{\mathbf{s}}'\mathbf{Zy} = 285.5$ and $\hat{\mathbf{d}}'\mathbf{Wy} = 2\,636.4$. With a total sum of squares (SS) of 3,526,153, the residual SS is 31,548.2. The quadratics required in the estimation equations are then $\hat{\mathbf{s}}'\mathbf{A}_S^{-1}\hat{\mathbf{s}} = 18.716404$, $\hat{\mathbf{d}}'\mathbf{A}_D^{-1}\hat{\mathbf{d}} = 119.472337$ and $\hat{\mathbf{e}}'\hat{\mathbf{e}} = 30{,}128.9$.

The EM algorithm on the original scale gives estimates $\sigma_S^2 = 8.2481$ (first line of (9)) or $\sigma_S^2 = 6.8120$ (second line of (9)), $\sigma_D^2 = 11.4512$ (first line of (10)) or $\sigma_D^2 = 10.5465$ (second line of (10)) and $\sigma_W^2 = 110.7988$ (eq. (11)). The average number of progeny per dam is $k_D = 294/30 = 9.8$ and the average number of dams per sire $k_S = 30/5 = 6.0$. This gives $\alpha_D = 24.2449$ and $\alpha_S = 14.0408$. Using estimators of form (14) then gives $\hat{\alpha}_S = 9.72366$, $\hat{\alpha}_D = 21.89974$ and $\hat{\alpha}_W = \hat{\sigma}_W^2 = 110.70115$ (from (15), (16) and (17)) with estimates of the original components of $\hat{\sigma}_D^2 = 10.6037$ and $\hat{\sigma}_S^2 = 6.0737$. Estimates for subsequent rounds of iteration are given in table 2 for both the reparameterisation (using (15), (16) and (17)) and the « better » version of the EM algorithm on the original scale (using (11) and the second lines of (9) and (10)).

*Received November 12, 1985.*
*Accepted September 5, 1986.*

TABLE 2

*Estimates over rounds of iteration for the numerical example.*

| Round | $\hat{\sigma}_S^2$ | $\hat{\sigma}_D^2$ | $\hat{\sigma}_W^2$ |
|---|---|---|---|
| | Starting values | | |
| 0 | 10.000000000 | 12.00000000 | 120.0000000 |
| | Reparameterised scale | | |
| 1 | 6.073697343 | 10.60368758 | 110.7011524 |
| 2 | 5.793292231 | 10.38246830 | 111.0148001 |
| 3 | 5.774875145 | 10.36519071 | 111.0016223 |
| 4 | 5.773818666 | 10.36309267 | 111.0020272 |
| 5 | 5.773855456 | 10.36278748 | 111.0020245 |
| 6 | 5.773887264 | 10.36272901 | 111.0020299 |
| 7 | 5.773896807 | 10.36271619 | 111.0020312 |
| 8 | 5.773899257 | 10.36271321 | 111.0020315 |
| 9 | 5.773899859 | 10.36271249 | 111.0020316 |
| 10 | 5.773900005 | 10.36271232 | 111.0020316 |
| 11 | 5.773900041 | 10.36271228 | 111.0020316 |
| 12 | 5.773900049 | 10.36271227 | 111.0020316 |
| 13 | 5.773900051 | 10.36271227 | 111.0020316 |
| 14 | 5.773900052 | 10.36271227 | 111.0020316 |
| | Original scale | | |
| 1 | 6.811960832 | 10.54653761 | 110.7988338 |
| 2 | 6.162192464 | 10.38779261 | 110.9331919 |
| 3 | 5.934679073 | 10.34431931 | 110.9844251 |
| 4 | 5.846001718 | 10.33784182 | 111.0011447 |
| 5 | 5.808836612 | 10.34174703 | 111.0054555 |
| 6 | 5.792134787 | 10.34732035 | 111.0057549 |
| 7 | 5.784054722 | 10.35207264 | 111.0050067 |
| 8 | 5.779848630 | 10.35557730 | 111.0041680 |
| 9 | 5.777512317 | 10.35800506 | 111.0034923 |
| 10 | 5.776146465 | 10.35963494 | 111.0030054 |
| 11 | 5.775318220 | 10.36071089 | 111.0026719 |
| 12 | 5.774803618 | 10.36141452 | 111.0024494 |
| 13 | 5.774478933 | 10.36187223 | 111.0023031 |
| 14 | 5.774272135 | 10.36216905 | 111.0022075 |
| 16 | 5.774054544 | 10.36248546 | 111.0021052 |
| 20 | 5.773926863 | 10.36267281 | 111.0020444 |
| 25 | 5.773903062 | 10.36270784 | 111.0020331 |
| 30 | 5.773900390 | 10.36271177 | 111.0020318 |
| 35 | 5.773900090 | 10.36271221 | 111.0020316 |
| 40 | 5.773900056 | 10.36271226 | 111.0020316 |
| 44 | 5.773900052 | 10.36271227 | 111.0020316 |

## Acknowledgements

Financial support has been provided by the Agricultural and Food Research Council (A.F.R.C.), U.K., and the Canadian Association of Animal Breeders. I am grateful to R. THOMPSON for helpful comments and L.R. SCHAEFFER for comments on the manuscript.


## References


DEMPSTER A.P., LAIRD N.M., RUBIN D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc.*, Series B, **39**, 1-22.

HARVILLE D.A., 1977. Maximum likelihood approaches to variance component estimation and to related problems. *J. Am. Stat. Assoc.*, **72**, 320-340.

HENDERSON C.R., 1973. Sire evaluation and genetic trends. *Proc. Anim. Breed. Genet. Symp. in Honor of Dr J.L. Lush, Blacksburg, Virginia, July 29, 1972.* 10-41, ASAS, Champaign, IL.

HENDERSON C.R., 1984. *Applications of Linear Models in Animal Breeding.* 462 pp., University of Guelph, Guelph, Ontario.

LAIRD N.M., WARE J.H., 1982. Random-effects models for longitudinal data. *Biometrics*, **38**, 963-974.

PATTERSON H.D., THOMPSON R., 1971. Recovery of interblock information when block sizes are unequal. *Biometrika*, **58**, 545-554.

SEARLE S.R., 1966. *Matrix Algebra for the Biological Sciences.* 296 pp., Wiley, New York.

SEARLE S.R., 1979. *Notes on variance component estimation : A detailed account of maximum likelihood and kindred methodology.* Paper BU-673 M, Biometrics Unit, Cornell University, Ithaca, N.Y.

THOMPSON R., 1973. The estimation of variance and covariance components with an application when records are subject to culling. *Biometrics*, **22**, 527-550.

THOMPSON R., 1976. The estimation of maternal genetic variances. *Biometrics*, **32**, 903-917.

THOMPSON R., 1982. Methods of estimation of genetic parameters. *Proc. Second Int. Congr. Genet. Applied Livest. Prod., Madrid, Spain,* vol. V, 95-103, Edit. Garsi, Madrid.

THOMPSON R., MEYER K., 1986. Estimation of variance components : What is missing in the EM algorithm ? *J. Stat. Comput. Simulation* (in press).


## Appendix

### A. *Method of scoring*

Utilizing that $\mathbf{PVP} = \mathbf{P}$ and that $\mathbf{V}$ is linear in the parameters to be estimated (see (2)), (6) can be rewritten as :

$$\mathbf{y}'\mathbf{P}\delta\mathbf{V}/\delta\theta_i\mathbf{P}\mathbf{y} = \sum_j \operatorname{tr}(\mathbf{P}\delta\mathbf{V}/\delta\theta_i\ \mathbf{P}\delta\mathbf{V}/\delta\theta_j)\ \theta_j \tag{A1}$$

This yields a system of linear equations to be solved simultaneously :

$$\mathbf{B}\ \boldsymbol{\theta} = \mathbf{q} \tag{A2}$$

with $\boldsymbol{\theta} = \{\theta_i\}$ the vector of parameters to be estimated, $\mathbf{q} = \{q_i\} = \{\mathbf{y}'\mathbf{P}\delta\mathbf{V}/\delta\theta_i\mathbf{P}\mathbf{y}\}$ a vector of quadratics and $\mathbf{B} = \{b_{ij}\} = \{\mathrm{tr}(\mathbf{P}\delta\mathbf{V}/\delta\theta_i\ \mathbf{P}\delta\mathbf{V}/\delta\theta_j\}$ a symmetric matrix of coefficients. Apart from a factor of 1/2, $\mathbf{B}$ is equal to the information matrix for $\boldsymbol{\theta}$. The elements of $\mathbf{B}$ for the model considered here are :

$$
\begin{aligned}
b_{11} &= \lambda_S^2\,[NS - 2\lambda_S\,\mathrm{tr}(\mathbf{A}_S^{-1}\mathbf{C}_{SS}) + \lambda_S^2\,\mathrm{tr}(\mathbf{A}_S^{-1}\mathbf{C}_{SS})^2] \\
b_{12} &= \lambda_S^2\lambda_D^2\,\mathrm{tr}(\mathbf{A}_S^{-1}\mathbf{C}_{SD}\mathbf{A}_D^{-1}\mathbf{C}_{DS}) \\
b_{13} &= \lambda_S^2\,[\mathrm{tr}(\mathbf{A}_S^{-1}\mathbf{C}_{SS}) - \lambda_S\,\mathrm{tr}(\mathbf{A}_S^{-1}\mathbf{C}_{SS})^2 - \lambda_D\,\mathrm{tr}(\mathbf{A}_S^{-1}\mathbf{C}_{SD}\mathbf{A}_D^{-1}\mathbf{C}_{DS})] \qquad \text{(A3)} \\
b_{22} &= \lambda_D^2\,[ND - 2\lambda_D\,\mathrm{tr}(\mathbf{A}_D^{-1}\mathbf{C}_{DD}) + \lambda_D^2\,\mathrm{tr}(\mathbf{A}_D^{-1}\mathbf{C}_{DD})^2] \\
b_{23} &= \lambda_D^2\,[\mathrm{tr}(\mathbf{A}_D^{-1}\mathbf{C}_{DD}) - \lambda_D\,\mathrm{tr}(\mathbf{A}_D^{-1}\mathbf{C}_{DD})^2 - \lambda_S\,\mathrm{tr}\,(\mathbf{A}_S^{-1}\mathbf{C}_{SD}\mathbf{A}_D^{-1}\mathbf{C}_{DS})] \\
b_{33} &= NDFW + \lambda_S^2\,\mathrm{tr}(\mathbf{A}_S^{-1}\mathbf{C}_{SS})^2 + \lambda_D^2\,\mathrm{tr}(\mathbf{A}_D^{-1}\mathbf{C}_{DD})^2 \\
&\quad + 2\lambda_S\lambda_D\,\mathrm{tr}(\mathbf{A}_S^{-1}\mathbf{C}_{SD}\mathbf{A}_D^{-1}\mathbf{C}_{DS})
\end{aligned}
$$

The quadratics required are equel to those in the EM algorithm :

$$
\begin{aligned}
q_1 &= \hat{\mathbf{s}}'\mathbf{A}_S^{-1}\hat{\mathbf{s}}, \quad q_2 = \hat{\mathbf{d}}'\mathbf{A}_D^{-1}\hat{\mathbf{d}} \text{ and} \\
q_3 &= \hat{\mathbf{e}}'\hat{\mathbf{e}} = \mathbf{y}'\mathbf{S}\mathbf{y} - \mathbf{y}'\mathbf{S}\mathbf{Z}\hat{\mathbf{s}} - \mathbf{y}'\mathbf{S}\mathbf{W}\hat{\mathbf{d}} - \lambda_S\hat{\mathbf{s}}'\mathbf{A}_S^{-1}\hat{\mathbf{s}} - \lambda_D\hat{\mathbf{d}}'\mathbf{A}_D^{-1}\hat{\mathbf{d}} \qquad \text{(A4)}
\end{aligned}
$$

B. *Computing strategy for a model including a fixed effect with many levels*

Partition the vector of fixed effects and the design matrix in (1), according to the « major » fixed effect h with many levels and any additional fixed effects and covariables.

$$
\mathbf{X} = [\mathbf{B} : \mathbf{X}_A] \text{ and } \mathbf{b} = \begin{bmatrix} \mathbf{h} \\ b_A \end{bmatrix}
$$

Let the subscript h denote the submatrix or vector for the hth group of levels of $\mathbf{h}$. The MME absorbing $\mathbf{d}$, (20), can then be rewritten as :

$$
\begin{bmatrix}
\mathbf{B}'\mathbf{K}\mathbf{B} & \mathbf{B}'\mathbf{K}\mathbf{X}_A & \mathbf{B}'\mathbf{K}\mathbf{Z} \\
\mathbf{X}_A'\mathbf{K}\mathbf{B} & \mathbf{X}_A'\mathbf{K}\mathbf{X}_A & \mathbf{X}_A'\mathbf{K}\mathbf{Z} \\
\mathbf{Z}'\mathbf{K}\mathbf{B} & \mathbf{Z}'\mathbf{K}\mathbf{X}_A & \mathbf{Z}'\mathbf{K}\mathbf{Z} + \lambda_S\mathbf{A}_S^{-1}
\end{bmatrix}
\begin{bmatrix} \hat{\mathbf{h}} \\ \hat{\mathbf{b}}_A \\ \hat{\mathbf{s}} \end{bmatrix}
=
\begin{bmatrix} \mathbf{B}'\mathbf{K}\mathbf{y} \\ \mathbf{X}'_A\mathbf{K}\mathbf{y} \\ \mathbf{Z}'\mathbf{K}\mathbf{y} \end{bmatrix}
\qquad \text{(A5)}
$$

with $\mathbf{B}'\mathbf{K}\mathbf{B} = \sum_{h=1}^{NH}{}^{+}\ \mathbf{B}'_h\mathbf{K}_h\mathbf{B}_h$, where " $\Sigma^+$ " denotes the direct matrix sum (SEARLE, 1966) and NH the number of groups of the major fixed effect. This holds only if $\mathbf{A}_D$ has a corresponding block structure, i.e. if all covariances between levels of $\mathbf{d}$ in different groups are zero.

Absorbing $\mathbf{h}$ then gives the MME for sires and additional fixed effects as :

$$
\begin{bmatrix}
\mathbf{X}_A'\mathbf{N}\mathbf{X}_A & \mathbf{X}_A'\mathbf{N}\mathbf{Z} \\
\mathbf{Z}'\mathbf{N}\mathbf{X}_A & \mathbf{Z}'\mathbf{N}\mathbf{Z} + \lambda_S\mathbf{A}_S^{-1}
\end{bmatrix}
\begin{bmatrix} \hat{\mathbf{b}}_A \\ \hat{\mathbf{s}} \end{bmatrix}
=
\begin{bmatrix} \mathbf{X}_A'\mathbf{N}\mathbf{y} \\ \mathbf{Z}'\mathbf{N}\mathbf{y} \end{bmatrix}
\qquad \text{(A6)}
$$

with $\mathbf{N} = \mathbf{K} - \mathbf{KB}(\mathbf{B}'\mathbf{KB})^{-}\mathbf{B}'\mathbf{K}$. From (A5) it follows that $\mathbf{N}$ is block diagonal, i.e. $\mathbf{N} = \sum_{h=1} \mathbf{N}_h$ with :

$$\mathbf{N}_h = \mathbf{K}_h - \mathbf{K}_h\mathbf{B}_h\,(\mathbf{B}'_h\mathbf{K}_h\mathbf{B}_h)^{-}\mathbf{B}'_h\mathbf{K}_h \tag{A7}$$

Absorbing any additional fixed effects then leaves :

$$(\mathbf{Z}'\mathbf{FZ} + \lambda_S\mathbf{A}_S^{-1})\hat{\mathbf{s}} = \mathbf{Z}'\mathbf{Fy} \tag{A8}$$

with $\mathbf{F} = \mathbf{N} - \mathbf{NX}_A\,(\mathbf{X}_A'\mathbf{NX}_A)^{-}\mathbf{X}_A'\mathbf{N}$. Hence a direct inverse of order NS, equal to the number of levels of $\mathbf{s}$, is required,

$$\mathbf{C}_{SS} = (\mathbf{Z}'\mathbf{FZ} + \lambda_S\mathbf{A}_S^{-1}) \tag{A9}$$

to obtain solutions :

$$\hat{\mathbf{s}} = \mathbf{C}_{SS}\mathbf{Z}'\mathbf{Fy} \tag{A10}$$

After backsolving for any additional fixed effects or covariables,

$$\hat{\mathbf{b}}_A = (\mathbf{X}_A'\mathbf{NX}_A)^{-}\,(\mathbf{X}_A'\mathbf{Ny} - \mathbf{X}_A'\mathbf{NZ}\hat{\mathbf{s}}) \tag{A11}$$

backsolutions for $\mathbf{h}$ and $\mathbf{d}$ can be obtained group by group.

$$\hat{\mathbf{h}}_h = (\mathbf{B}_h'\mathbf{K}_h\mathbf{B}_h)^{-}\,(\mathbf{B}_h'\mathbf{K}_h\mathbf{y}_h - \mathbf{B}_h'\mathbf{K}_h\mathbf{X}_{Ah}\hat{\mathbf{b}}_A - \mathbf{B}_h'\mathbf{K}_h\mathbf{Z}_h\hat{\mathbf{s}}) \tag{A12}$$

$$\hat{\mathbf{d}}_h = (\mathbf{W}_h'\mathbf{W}_h + \lambda_D\mathbf{A}_{Dh}^{-1})^{-1}\,(\mathbf{W}_h'\mathbf{y}_h - \mathbf{W}_h'\mathbf{B}_h\hat{\mathbf{h}}_h$$
$$- \mathbf{W}_h'\mathbf{X}_{Ah}\hat{\mathbf{b}}_A - \mathbf{W}_h'\mathbf{Z}_h\hat{\mathbf{s}}) \tag{A13}$$

The quadratic forms and traces for REML are the same as before except :

(i) $\mathbf{y}'\mathbf{X}\hat{\mathbf{b}}$ (in (30)) expands to $\mathbf{y}'\mathbf{B}\hat{\mathbf{h}} + \mathbf{y}'\mathbf{X}_A\hat{\mathbf{b}}_A$,

(ii) $\mathrm{tr}(\mathbf{A}_S^{-1}\mathbf{C}_{SS})$ can be calculated directly, and

(iii) $\mathrm{tr}(\mathbf{A}_D^{-1}\mathbf{C}_{DD}) = \mathrm{tr}(\mathbf{A}_D^{-1}\mathbf{H}_D) + \mathrm{tr}(\mathbf{H}_B\mathbf{L}_{BD}\mathbf{A}_D^{-1}\mathbf{L}'_{BD})$

$$+ \mathrm{tr}(\mathbf{H}_X\mathbf{T}_{XX}) + \mathrm{tr}(\mathbf{C}_{SS}\mathbf{T}) \tag{A14}$$

with :

$$\mathbf{H}_D = (\mathbf{W}'\mathbf{W} + \lambda_D\mathbf{A}_D^{-1})^{-1}$$
$$\mathbf{H}_B = (\mathbf{B}'\mathbf{KB})^{-}$$
$$\mathbf{H}_X = (\mathbf{X}_A'\mathbf{NX}_A)^{-}$$
$$\mathbf{L}_{BD} = \mathbf{B}'\mathbf{WH}_D$$
$$\mathbf{L}_{AD} = \mathbf{X}_A'\mathbf{WH}_D$$
$$\mathbf{L}_{SD} = \mathbf{Z}'\mathbf{WH}_D$$
$$\mathbf{L}_{AB} = \mathbf{X}_A'\mathbf{KBH}_B$$
$$\mathbf{L}_{SB} = \mathbf{Z}'\mathbf{KBH}_B$$
$$\mathbf{L}_{SA} = \mathbf{Z}'\mathbf{NX}_A\mathbf{H}_X$$
$$\mathbf{T} = \mathbf{T}_{SS} - \mathbf{L}_{SX}\mathbf{T}'_{SX} - \mathbf{T}_{SX}\mathbf{L}'_{SX} + \mathbf{L}_{SX}\mathbf{T}_{XX}\mathbf{L}'_{SX} \tag{A15}$$

and :

$$\mathbf{T}_{XX} = [\mathbf{L}_{AD} \ \vdots \ - \mathbf{L}_{AB}\mathbf{L}_{BD}] \quad \mathbf{A}_D^{-1} \quad \begin{bmatrix} \mathbf{L'}_{AD} \\ - \mathbf{L'}_{BD}\mathbf{L'}_{AB} \end{bmatrix} \tag{A16}$$

$$\mathbf{T}_{SX} = [\mathbf{L}_{SD} \ \vdots \ - \mathbf{L}_{SB}\mathbf{L}_{BD}] \quad \mathbf{A}_D^{-1} \quad \begin{bmatrix} \mathbf{L'}_{AD} \\ - \mathbf{L'}_{BD}\mathbf{L'}_{AB} \end{bmatrix} \tag{A17}$$

and :

$$\mathbf{T}_{SS} = [\mathbf{L}_{SD} \ \vdots \ - \mathbf{L}_{SB}\mathbf{L}_{BD}] \quad \mathbf{A}_D^{-1} \quad \begin{bmatrix} \mathbf{L'}_{SD} \\ - \mathbf{L'}_{BD}\mathbf{L'}_{SB} \end{bmatrix} \tag{A18}$$

C. *Numerical example : absorbing a fixed effect with many levels*

Absorbing treatments for one time period after the other, intermediate results are as follows.

Processing data for period I gives :

$$\mathbf{X}_A{}'\mathbf{N}\mathbf{X}_A = \begin{bmatrix} 107.340 & & \text{sym.} \\ -\ 34.374 & 107.374 & \\ 893.281 & 926.719 & 25{,}131.3 \end{bmatrix}$$

$$\mathbf{T}_{XX} = \begin{bmatrix} 0.0489 & & \text{sym.} \\ -\ 0.0489 & 0.0489 & \\ 0.0060 & -\ 0.0060 & 40.376 \end{bmatrix}$$

and $\mathrm{tr}(\mathbf{H}_B\mathbf{L}_{BD}\mathbf{A}_D^{-1}\mathbf{L}_{BD}{}') = 0.0497559$. After absorbing all dams and treatments, $\mathrm{tr}(\mathbf{H}_B\mathbf{L}_{BD}\mathbf{A}_D^{-1}\mathbf{L}_{BD}{}') = 0.1089976$,

$$\mathbf{T}_{XX} = \begin{bmatrix} 0.1238 & & \text{sym.} \\ -\ 0.1238 & 0.1238 & \\ -\ 0.3848 & 0.3848 & 77.240 \end{bmatrix}$$

and :

$$\mathbf{X_A'NX_A} = \begin{bmatrix} 69.030 & & \\ -69.030 & 69.030 & \\ -\ 8.768 & 8.768 & 1\,708.54 \end{bmatrix}$$

$$(\mathbf{X_A'Ny})' = [-\ 767.511 \quad 767.511 \quad -\ 1\,107.39]$$

$$\mathbf{Z'NZ} = \begin{bmatrix} 19.547 & & & & \text{sym.} \\ -\ 3.633 & 20.185 & & & \\ -\ 4.603 & -\ 5.429 & 22.521 & & \\ -\ 5.700 & -\ 4.540 & -\ 5.649 & 22.317 & \\ -\ 5.611 & -\ 6.429 & -\ 6.839 & -\ 6.429 & 25.462 \end{bmatrix}$$

$$\mathbf{Z'Ny} = \begin{bmatrix} 44.704 \\ -\ 48.678 \\ -\ 103.484 \\ 54.079 \\ 53.379 \end{bmatrix}$$

$$\mathbf{T_{SS}} = \begin{bmatrix} 1.0118 & & & & \text{sym.} \\ -\ 0.1941 & 1.0291 & & & \\ -\ 0.2353 & -\ 0.2922 & 1.1119 & & \\ -\ 0.2665 & -\ 0.2019 & -\ 0.2522 & 1.0078 & \\ -\ 0.3158 & -\ 0.3407 & -\ 0.3321 & -\ 0.2870 & 1.2757 \end{bmatrix}$$

$$\mathbf{T_{SX}} = \begin{bmatrix} 0.0405 & -\ 0.0629 & -\ 0.0098 & 0.1470 & -\ 0.1147 \\ -\ 0.0405 & 0.0629 & 0.0098 & -\ 0.1470 & 0.1147 \\ 2.5173 & 0.9691 & -\ 0.2343 & -\ 2.1407 & -\ 1.1114 \end{bmatrix}$$

and $\text{tr}(\mathbf{H_B L_{BD} A_D^{-1} L_{BD}'}) = 0.1089976$.

Again, setting the first level of each additional effect to zero and obtaining a generalized inverse, yields $\text{tr}(\mathbf{H_X T_{XX}}) = 0.0469752$. Absorbing the additional fixed effects and covariables into sires,

$$\mathbf{T} = \begin{bmatrix} 0.9303 & & & & \text{sym.} \\ -0.2245 & 1.0150 & & & \\ -0.2271 & -0.2900 & 1.1109 & & \\ -0.2005 & -0.1703 & -0.2569 & 0.9363 & \\ -0.2782 & -0.3302 & -0.3369 & -0.3087 & 1.2546 \end{bmatrix}$$

and the fourth term of (A14) is $\mathrm{tr}(\mathbf{C}_{ss}\mathbf{T}) = 0.1353313$.