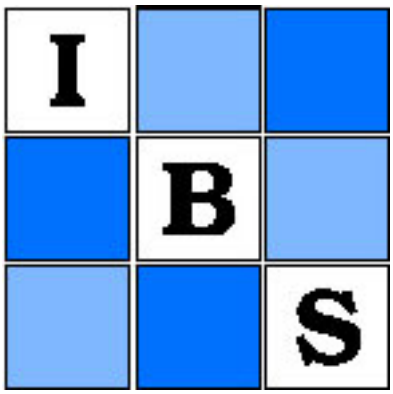



BIOMETRICS 51, 1440–1450
December 1995

# Average Information REML: An Efficient Algorithm for Variance Parameter Estimation in Linear Mixed Models

**Arthur R. Gilmour,[1] Robin Thompson,[2] and Brian R. Cullis[3]**

[1]Agricultural Research and Veterinary Centre, Orange, New South Wales 2800, Australia,
[2]IACR-Rothamsted, Harpenden, AL5 2JQ, United Kingdom,
[3]Agricultural Research Institute, Wagga Wagga, New South Wales 2650, Australia

SUMMARY

A strategy of using an average information matrix is shown to be computationally convenient and efficient for estimating variance components by restricted maximum likelihood (REML) in the mixed linear model. Three applications are described. The motivation for the algorithm was the estimation of variance components in the analysis of wheat variety means from 1,071 experiments representing 10 years and 60 locations in New South Wales. We also apply the algorithm to the analysis of designed experiments by incomplete block analysis and spatial analysis of field experiments.

## 1. Introduction

The estimation of variance components in mixed models assuming Gaussian random terms by restricted maximum likelihood (REML) has been widely adopted. The REML procedure maximises the joint likelihood of all error contrasts rather than of all contrasts as in ordinary maximum likelihood. See Searle, Casella, and McCulloch (1992) for details. In the original description of REML, Patterson and Thompson (1971) suggest that the score equations for the variance components may be solved iteratively using the Fisher scoring (FS) algorithm. For many applications, particularly in animal breeding, this strategy presents computational difficulties due to the large size of the matrices to be inverted and multiplied. Thompson (1977) presented an overview of the methodology with particular reference to animal breeding applications and showed how some of the computational burdens of the FS algorithm may be overcome. He suggested that in some situations use of the observed information rather than the expected information may be preferable.

Various other computational strategies have been presented to overcome these problems for large datasets. Graser, Smith, and Tier (1987) described a derivative free (DF) algorithm used by Meyer (1989) in her suite of animal breeding programs. Essentially, their procedure calculates the REML log likelihood from the mixed model equations. They obtain the REML estimates of the variance parameters by directly searching the likelihood surface using the Nelder–Mead (1965) simplex algorithm. This procedure performs adequately when there are only a few variance components. However, its convergence rate declines with an increase in variance components (Searle et al., 1992). Some authors (for example, Smith (1990)) suggest the EM algorithm (Dempster, Laird, and Rubin, 1977) as an efficient procedure for obtaining REML estimates of variance components. This utilises the score equations and requires about three times the computing needed for evaluating the REML likelihood per iteration. However, convergence can still be slow, although there have been several improvements to the original EM algorithm (see Searle et al. (1992), for a discussion).

Patterson and Nabugoomu (1992) reviewed the estimation of variance components in the genotype × environment analysis of variety means from a series of variety trials over years and locations. They recommend REML estimation and used the "REML" directive in GENSTAT (Payne, 1993). For larger problems they suggest various approximate analyses to overcome size limitations in the software. The algorithm we present was devised for a large analysis of this sort and is suitable for estimation in both small and large problems.

Gleeson and Cullis (1987), Cullis et al. (1989), and Cullis and Gleeson (1991) present an approach

*Key words:* EM algorithm; Spatial analysis; Variance components; REML.

to the spatial analysis of field experiments. They advocate estimation of variance and covariance parameters by REML and this performs adequately with low levels of bias (Cullis, Gleeson, and Thomson, 1992; Lill, Gleeson, and Cullis, 1988). Cullis and Gleeson (1989) demonstrated an increase in the efficiency of estimation of variety contrasts using spatial analysis relative to complete or incomplete block analysis. Gilmour (1992) implemented these models in the FORTRAN program TwoD, using a FS algorithm. Convergence of the algorithm is reliable; however, process time and memory requirements are often prohibitive for larger problems or for models with many covariance parameters. This seriously limits the adoption of the procedures, particularly in a PC environment.

In this article we present three applications of the average information (AI) algorithm. It is similar to the original FS algorithm proposed by Patterson and Thompson (1971) but avoids the evaluation of the traces of large matrices which appear in both the expected and observed (REML) information matrices. Johnson and Thompson (1995) applied the AI algorithm in an animal breeding context. They found it performed well when compared to the DF and EM algorithms.

The structure of the paper is as follows. In Section 2 we present the models and broadly describe the algorithm. In Section 3 we apply the algorithm to variance components estimation using two examples. In Section 4 we illustrate its use in REML estimation for mixed models with correlated errors.

## 2. Description of Models and AI Algorithm

We consider the model

$$\mathbf{y}|\mathbf{u},\ \mathbf{e} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}\mathbf{u} + \mathbf{e} \tag{1}$$

where $\mathbf{y}$ is an $n$ vector of data, $\mathbf{X}$ and $\mathbf{Z}$ are design matrices, $\boldsymbol{\tau}$ and $\mathbf{u}$ are $t$ and $b$ vectors of fixed and random effects and $\mathbf{e}$ is the residual error. We assume that $\mathbf{X}$ has full column rank and $\mathbf{Z} = [\mathbf{Z}_1, \ldots, \mathbf{Z}_q]$, where each $\mathbf{Z}_i$ represents the design matrix for the $i$th random factor and $\mathbf{u} = [\mathbf{u}_1', \ldots, \mathbf{u}_q']$, where $\mathbf{u}_i$ is a $b_i$ vector. Note $b = \Sigma\, b_i$. We assume further that

$$\begin{bmatrix}\mathbf{u}\\ \mathbf{e}\end{bmatrix} \sim N\left(\mathbf{0},\ \sigma^2\begin{bmatrix}\mathbf{G}(\boldsymbol{\gamma}) & \mathbf{0}\\ \mathbf{0} & \mathbf{R}(\boldsymbol{\phi})\end{bmatrix}\right)$$

where $\boldsymbol{\gamma}$ is the $q$ vector of variance parameters relating to $\mathbf{u}$ and $\boldsymbol{\phi}$ is the $s$ vector of variance and covariance parameters relating to $\mathbf{e}$ (through $\mathbf{R}$). The dimension of $\boldsymbol{\gamma}$ may exceed $q$ if the random terms are correlated. For many applications, it is convenient to assume the $q$ components of $\mathbf{u}$ are mutually independent with variance matrices $\sigma^2\gamma_i\mathbf{I}_{b_i}$, and so $\mathbf{G} = \oplus\{\gamma_i\mathbf{I}_{b_i}\}$ is a block diagonal matrix. Typically $s$ is much smaller than $n(n+1)/2$.

Particularly in field experiments, the data vec $\mathbf{y}$ may be the vector of a matrix representing an $r \times c$ rectangular grid of plots. In these applications $\mathbf{R}$ is sometimes assumed to be proportional to the variance matrix of a separable error process such that $\mathbf{R} = \mathbf{R}_c \otimes \mathbf{R}_r$ (see Section 4). We note that (1) can be easily extended to incorporate differencing (see Cullis and Gleeson (1991), for details).

The mixed model equations are

$$\begin{bmatrix}\mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}\end{bmatrix}\begin{bmatrix}\hat{\boldsymbol{\tau}}\\ \tilde{\mathbf{u}}\end{bmatrix} = \begin{bmatrix}\mathbf{X}'\mathbf{R}^{-1}\mathbf{y}\\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y}\end{bmatrix}. \tag{2}$$

The solution of (2) requires values for $\boldsymbol{\gamma}$ and $\boldsymbol{\phi}$. In practice we replace $\boldsymbol{\gamma}$ and $\boldsymbol{\phi}$ by their REML estimates $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\phi}}$ which maximise the likelihood of error contrasts.

The log REML likelihood can be written as

$$\begin{aligned}\ell &= -\tfrac{1}{2}(\log\det\mathbf{X}'\mathbf{H}^{-1}\mathbf{X} + \log\det\mathbf{H} + \nu\log\sigma^2 + \mathbf{y}'\mathbf{P}\mathbf{y}/\sigma^2)\\ &= -\tfrac{1}{2}(\log\det\mathbf{C} + \log\det\mathbf{R} + \log\det\mathbf{G} + \nu\log\sigma^2 + \mathbf{y}'\mathbf{P}\mathbf{y}/\sigma^2),\end{aligned} \tag{3}$$

where $\nu = n - t$, $\mathbf{H} = \mathbf{R} + \mathbf{Z}\mathbf{G}\mathbf{Z}'$, $\mathbf{C}$ is the coefficient matrix in (2),

$$\begin{aligned}\mathbf{P} &= \mathbf{H}^{-1} - \mathbf{H}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{H}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{H}^{-1}\\ &= \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{W}\mathbf{C}^{-1}\mathbf{W}'\mathbf{R}^{-1}\end{aligned}$$

and $\mathbf{W} = [\mathbf{X} \vdots \mathbf{Z}]$. Letting $\boldsymbol{\kappa} = (\boldsymbol{\gamma}, \boldsymbol{\phi})$, the REML estimates of $\sigma^2$ and $\kappa_i$ satisfy

$$\partial \ell/\partial \sigma^2 = -\tfrac{1}{2}(\nu/\sigma^2 - \mathbf{y}'\mathbf{P}\mathbf{y}/\sigma^4) = 0$$

$$\partial \ell/\partial \kappa_i = -\tfrac{1}{2}[\mathrm{tr}(\mathbf{P}\mathbf{H}_i) - \mathbf{y}'\mathbf{P}\mathbf{H}_i\mathbf{P}\mathbf{y}/\sigma^2] = 0, \tag{4}$$

where $\mathbf{H}_i = \partial \mathbf{H}/\partial \kappa_i$.

Thus, given $\boldsymbol{\kappa}$, $\tilde{\sigma}^2 = \mathbf{y}'\mathbf{P}\mathbf{y}/\nu$. However, in general the solution to (4) requires an iterative scheme. Given an initial estimate, an update of $\boldsymbol{\kappa}$, using the FS algorithm, is

$$\boldsymbol{\kappa}^{(0)} + \mathbf{B}^{(0)}\partial \ell/\partial \boldsymbol{\kappa}(\boldsymbol{\kappa} = \boldsymbol{\kappa}^{(0)}),$$

where $\mathbf{B}$ represents the partition, relevant to $\boldsymbol{\kappa}$, of the inverse of the expected information matrix of $(\sigma^2, \boldsymbol{\kappa})$.

The elements of the observed information matrix are

$$-\partial^2 \ell/\partial \sigma^4 = \mathbf{y}'\mathbf{P}\mathbf{y}/\sigma^6 - \tfrac{1}{2}\nu/\sigma^4$$

$$-\partial^2 \ell/\partial \kappa_i \partial \sigma^2 = \tfrac{1}{2}\mathbf{y}'\mathbf{P}\mathbf{H}_i\mathbf{P}\mathbf{y}/\sigma^4$$

$$-\partial^2 \ell/\partial \kappa_i \partial \kappa_j = \tfrac{1}{2}\mathrm{tr}(\mathbf{P}\mathbf{H}_{ij}) - \tfrac{1}{2}\mathrm{tr}(\mathbf{P}\mathbf{H}_i\mathbf{P}\mathbf{H}_j) + \mathbf{y}'\mathbf{P}\mathbf{H}_i\mathbf{P}\mathbf{H}_j\mathbf{P}\mathbf{y}/\sigma^2 - \tfrac{1}{2}\mathbf{y}'\mathbf{P}\mathbf{H}_{ij}\mathbf{P}\mathbf{y}/\sigma^2, \tag{5}$$

where $\mathbf{H}_{ij} = \partial^2 \mathbf{H}/\partial \kappa_i \partial \kappa_j$.

The elements of the expected information matrix are

$$\mathrm{E}(-\partial^2 \ell/\sigma^4) = \tfrac{1}{2}\nu/\sigma^4$$

$$\mathrm{E}(-\partial^2 \ell/\partial \kappa_i \partial \sigma^2) = \tfrac{1}{2}\mathrm{tr}(\mathbf{P}\mathbf{H}_i)/\sigma^2$$

$$\mathrm{E}(-\partial^2 \ell/\partial \kappa_i \partial \kappa_j) = \tfrac{1}{2}\mathrm{tr}(\mathbf{P}\mathbf{H}_i\mathbf{P}\mathbf{H}_j). \tag{6}$$

The evaluation of some traces in either (5) or (6) can be either not feasible or very computer intensive. We therefore consider the matrix denoted by $\mathcal{I}_A$ which is a simplified average of the terms in (5) and (6). The elements of $\mathcal{I}_A$ are

$$\mathcal{I}_A(\sigma^2, \sigma^2) = \tfrac{1}{2}\mathbf{y}'\mathbf{P}\mathbf{y}/\sigma^6$$

$$\mathcal{I}_A(\sigma^2, \kappa_i) = \tfrac{1}{2}\mathbf{y}'\mathbf{P}\mathbf{H}_i\mathbf{P}\mathbf{y}/\sigma^4$$

$$\mathcal{I}_A(\kappa_i, \kappa_j) = \tfrac{1}{2}\mathbf{y}'\mathbf{P}\mathbf{H}_i\mathbf{P}\mathbf{H}_j\mathbf{P}\mathbf{y}/\sigma^2. \tag{7}$$

The first element ($\mathcal{I}_A(\sigma^2, \sigma^2)$) is the average of the terms in (5) and (6). The second term is obtained by considering only the data part (divided by $\sigma^2$) of the score for $\kappa_i$, or equivalently, approximating $\mathrm{tr}(PH_i)$ by $\mathbf{y}'\mathbf{P}\mathbf{H}_i\mathbf{P}\mathbf{y}/\sigma^2$ since these terms are equal at the solution to (4). This approximation ensures that $\mathcal{I}_A$ will be positive semidefinite. $\mathcal{I}_A(\kappa_i, \kappa_j)$ is obtained by averaging (5) and (6) and approximating $\mathbf{y}'\mathbf{P}\mathbf{H}_{ij}\mathbf{P}\mathbf{y}/\sigma^2$ by its expectation, $\mathrm{tr}(\mathbf{P}\mathbf{H}_{ij})$ in those cases when $\mathbf{H}_{ij} \neq 0$. For variance components models (i.e., those linear in $\mathbf{H}$), the terms in $\mathcal{I}_A$ except $\mathcal{I}_A(\sigma^2, \kappa_i)$ are exact averages of those in (5) and (6). $\mathcal{I}_A$ is the average for these models, if parameterized in terms of components rather than ratios (Johnson and Thompson, 1995). We call this matrix the AI matrix and use it in place of the expected information matrix to update $\boldsymbol{\kappa}$.

The $\mathcal{I}_A$ matrix is the (scaled) residual sums of squares and products matrix of $\mathbf{Y} = [\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_{q+s}]$ where $\mathbf{y}_i$, $i > 0$ is the "working" variate for $\kappa_i$ and is given by

$$\mathbf{y}_i = \mathbf{H}_i\mathbf{P}\mathbf{y} = \mathbf{H}_i\mathbf{R}^{-1}\tilde{\mathbf{e}},$$

where $\tilde{\mathbf{e}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\tau}} - \mathbf{Z}\tilde{\mathbf{u}}$, $\hat{\boldsymbol{\tau}}$ and $\tilde{\mathbf{u}}$ are solutions to (2) and $\mathbf{y}_0 = \mathbf{y}$, the data vector.

## 3. Variance Components Models

In this section we consider the common and widely applicable model in which the matrix $\mathbf{R}$, the variance structure for the residual term $\mathbf{e}$, is considered known and typically diagonal. In this case $\boldsymbol{\kappa}$ reduces to $\boldsymbol{\gamma}$ and for $\mathbf{H} = \mathbf{R} + \Sigma\, \gamma_i \mathbf{Z}_i \mathbf{Z}_i'$, the score for $\gamma_i$ is

$$\partial \ell / \partial \gamma_i = -\tfrac{1}{2}[b_i/\gamma_i - \mathrm{tr}(\mathbf{C}^{ii})/\gamma_i^2 - \tilde{\mathbf{u}}_i' \tilde{\mathbf{u}}_i / (\sigma^2 \gamma_i^2)], \tag{8}$$

where $\mathbf{C}$ is partitioned conformably with $\mathbf{W}$ in which the partition for $\mathbf{X}$ is denoted by $\mathbf{C}^{00}$. Further details are in Appendix A.

As an initialisation we form the matrix

$$\mathbf{M} = \begin{bmatrix} \mathbf{y}'\mathbf{R}^{-1}\mathbf{y} & \mathbf{y}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{y}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} \end{bmatrix}. \tag{9}$$

Since $\mathbf{M}$ is typically sparse, especially in $\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}$, it should be stored compactly and ordered so that computational efficiency is maximised. Algorithms for determining the optimal ordering are readily available (See Duff, Erisman, and Reid (1989)).

An outline of the steps in the iteration, given an initial value of $\boldsymbol{\gamma} = \boldsymbol{\gamma}^{(0)}$ follows. Further computational detail is presented in Appendix B.

1. Perform absorption (Gaussian Elimination, Golub and Van Loan (1983)) of $\mathbf{W}$ onto $\mathbf{y}$, after adding $\mathbf{G}^{-1}$ to $\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}$ in (9). The intermediate terms formed need to be held for use in steps 2, 4, and 5. The log likelihood is calculated.
2. Backsubstitute to obtain $\hat{\boldsymbol{\tau}}$ and $\tilde{\mathbf{u}}$, using the intermediate terms formed in step 1.
3. Calculate terms involving $\mathbf{y}_i$, namely $\mathbf{X}'\mathbf{R}^{-1}\mathbf{y}_i$, $\mathbf{Z}'\mathbf{R}^{-1}\mathbf{y}_i$, $\mathbf{y}'\mathbf{R}^{-1}\mathbf{y}_i = \mathbf{y}_0'\mathbf{R}^{-1}\mathbf{y}_i$ and $\mathbf{y}_j'\mathbf{R}^{-1}\mathbf{y}_i$. All these terms can be calculated directly from the elements of $\mathbf{M}$, since, for example,
$$\mathbf{Z}'\mathbf{R}^{-1}\mathbf{y}_i = \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}_i\tilde{\mathbf{u}}_i/\gamma_i$$
4. Perform absorption of $\mathbf{W}$ onto the multivariate working variate $\mathbf{Y} = [\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_q]$ to produce the terms essential for the matrix $\mathcal{I}_A$.
5. Evaluate sufficiently $\mathbf{C}^{-1}$ to obtain $\mathrm{tr}(\mathbf{C}^{ii})$. All elements of $\mathbf{C}^{-1}$, or those pertaining to $\mathbf{X}$, required for conducting significance tests on effects, may be calculated in the final iteration.
6. Calculate the update $\boldsymbol{\gamma}^{(1)} = \boldsymbol{\gamma}^{(0)} + \mathcal{I}_A^{22}\partial\ell/\partial\boldsymbol{\gamma}$, where $\mathcal{I}_A^{22}$ is the partition of the inverse of $\mathcal{I}_A$ pertaining to $\boldsymbol{\gamma}$. Exit or return to step 1, based on convergence criteria.

We note that the DF algorithm requires only step 1 for each iteration, whilst an iteration of the EM or enhanced EM algorithm requires steps 1, 2, and 5. Steps 1 and 5 are the time-consuming steps.

### 3.1 *Incomplete Block Analysis with Recovery of Interblock Information*

In this section we illustrate the convergence performance of the AI algorithm relative to the FS algorithm as implemented in GENSTAT (Payne, 1993) for the analysis of a spring wheat variety trial conducted at Slate Hall Farm, UK in 1976. The experiment was designed to compare the yields of 25 varieties, and these varieties were laid out in six replicates according to a balanced lattice square design. The field layout and plot yields are presented in Tables 1 and 2.

**Table 1**
*Field layout of the Slate Hall wheat variety trial*

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 4 | 3 | 5 | 19 | 23 | 2 | 6 | 15 | 18 | 25 | 9 | 11 | 2 |
| 6 | 7 | 9 | 8 | 10 | 8 | 12 | 16 | 25 | 4 | 5 | 7 | 16 | 23 | 14 |
| 21 | 22 | 24 | 23 | 25 | 11 | 20 | 24 | 3 | 7 | 6 | 13 | 22 | 4 | 20 |
| 11 | 12 | 14 | 13 | 15 | 22 | 1 | 10 | 14 | 18 | 24 | 1 | 15 | 17 | 8 |
| 16 | 17 | 19 | 18 | 20 | 5 | 9 | 13 | 17 | 21 | 12 | 19 | 3 | 10 | 21 |
| 3 | 18 | 8 | 13 | 23 | 16 | 24 | 10 | 13 | 2 | 10 | 4 | 17 | 11 | 23 |
| 1 | 16 | 6 | 11 | 21 | 12 | 20 | 1 | 9 | 23 | 12 | 6 | 24 | 18 | 5 |
| 5 | 20 | 10 | 15 | 25 | 4 | 7 | 18 | 21 | 15 | 19 | 13 | 1 | 25 | 7 |
| 2 | 17 | 7 | 12 | 22 | 25 | 3 | 14 | 17 | 6 | 21 | 20 | 8 | 2 | 14 |
| 4 | 19 | 9 | 14 | 24 | 8 | 11 | 22 | 5 | 19 | 3 | 22 | 15 | 9 | 16 |

**Table 2**
*Yields (grams/m$^2$) for the Slate Hall wheat variety trial*

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1,003 | 1,356 | 1,412 | 1,239 | 1,508 | 1,967 | 1,572 | 1,969 | 1,747 | 1,598 | 1,630 | 1,633 | 1,255 | 1,277 | 1,572 |
| 1,531 | 1,540 | 1,250 | 1,658 | 1,185 | 1,605 | 1,550 | 1,500 | 1,642 | 1,504 | 1,680 | 1,526 | 1,452 | 1,480 | 1,482 |
| 1,126 | 1,400 | 1,329 | 1,287 | 1,555 | 1,395 | 1,696 | 1,570 | 1,404 | 1,285 | 1,473 | 1,761 | 1,695 | 1,364 | 1,790 |
| 1,261 | 1,423 | 1,110 | 1,735 | 1,617 | 1,820 | 1,351 | 1,297 | 1,412 | 1,506 | 1,512 | 1,355 | 1,524 | 1,478 | 1,371 |
| 1,458 | 2,036 | 2,119 | 1,912 | 1,893 | 1,748 | 1,450 | 1,740 | 1,450 | 1,523 | 1,364 | 1,690 | 1,334 | 1,239 | 1,557 |
| 1,623 | 1,862 | 1,645 | 1,888 | 1,527 | 1,606 | 1,842 | 1,186 | 1,462 | 1,242 | 1,082 | 1,304 | 1,267 | 1,266 | 1,200 |
| 1,331 | 1,417 | 1,611 | 1,454 | 1,790 | 1,767 | 1,917 | 1,264 | 1,060 | 951 | 1,130 | 1,266 | 1,289 | 1,260 | 1,174 |
| 1,211 | 1,411 | 1,183 | 1,550 | 1,660 | 1,526 | 1,681 | 1,545 | 1,290 | 976 | 1,240 | 1,181 | 917 | 1,287 | 975 |
| 1,388 | 1,453 | 1,384 | 1,669 | 1,738 | 1,845 | 1,700 | 1,528 | 1,373 | 1,240 | 1,252 | 1,591 | 1,428 | 1,509 | 1,273 |
| 1,443 | 1,667 | 1,549 | 1,459 | 1,722 | 1,583 | 1,490 | 1,607 | 1,315 | 1,174 | 1,443 | 1,649 | 1,407 | 1,315 | 1,318 |

In the notation of Wilkinson and Rogers (1973), the blocking structure for this design is replicate/(row + column), where row and column are both factors with five levels, nested within replicates. Thus $\boldsymbol{\gamma}' = (\gamma_1, \gamma_2, \gamma_3)$, and $\sigma^2\boldsymbol{\gamma}$ represents the vector of variance components for replicate, replicate.row, and replicate.column. The REML estimates of the variance components are 4,262, 15,595, and 14,812, respectively. The residual variance is 8,062.

Table 3 presents the convergence sequence for the FS and AI algorithms. The FS algorithm first step is closer to the final solution, although convergence is attained in three iterations for both algorithms. No timing comparisons were made. The means and average standard error of difference (SED) from this analysis are presented in Table 4.

**Table 3**
*Convergence sequences for the interblock analysis of the Slate Hall wheat variety trial using the FS and AI algorithms, starting values all set to* 1

| | FS | | | | AI | | | |
|---|---|---|---|---|---|---|---|---|
| Iteration | $\gamma_1$ | $\gamma_2$ | $\gamma_3$ | $\ell$ | $\gamma_1$ | $\gamma_2$ | $\gamma_3$ | $\ell$ |
| 1 | .525 | 1.972 | 1.875 | −648.507 | .578 | 1.683 | 1.642 | −648.597 |
| 2 | .529 | 1.933 | 1.836 | −648.505 | .535 | 1.917 | 1.829 | −648.505 |
| 3 | .529 | 1.934 | 1.837 | −648.505 | .529 | 1.934 | 1.837 | −648.505 |

**Table 4**
*Variety means and standard error of difference (grams/m$^2$) from the interblock analysis of the Slate Hall wheat variety trial. Standard error of difference is* 62.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1,284 | 2 | 1,549 | 3 | 1,421 | 4 | 1,452 | 5 | 1,533 |
| 6 | 1,527 | 7 | 1,401 | 8 | 1,457 | 9 | 1,299 | 10 | 1,193 |
| 11 | 1,327 | 12 | 1,484 | 13 | 1,619 | 14 | 1,327 | 15 | 1,498 |
| 16 | 1,346 | 17 | 1,498 | 18 | 1,592 | 19 | 1,670 | 20 | 1,640 |
| 21 | 1,493 | 22 | 1,644 | 23 | 1,329 | 24 | 1,546 | 25 | 1,631 |

### 3.2 *Variance Component Estimation in Nonorthogonal Genotype × Environment Data*

To illustrate the performance of the AI algorithm for large problems we present a combined analysis of the NSW wheat variety trials from 1982 to 1991 inclusive. Briefly the data consists of genotype (G) mean yields from 1,071 experiments (E). The experiments were classified by year (Y) and location (L, with 60 levels). The 107 genotypes in the data were each tested for at least 2 years. Since experiments were not sown at each location in each year and often more than one experiment was sown at a location in a year, the G × Y × L table is incomplete and unbalanced.

The aim of this analysis is to estimate the variance components for G, G.Y, G.L, and G.Y.L. We consider a complex model which also includes random interaction terms for the interaction of experiments with genotype covariates Acid (A), Winter (W), and Maturity (M). These are sources of genotype × environment variation which can be taken into account by the breeders. Acid indicates the tolerance of the genotype to soil acidity. The Winter factor indicates whether the variety is a winter or spring type, since both types are sown in the same trials in NSW. Maturity is a measure of genotype maturity. There are six variance components (Table 5). Experiment (E), fitted as fixed, subsumes Year, Location, Year.Location, and sowing date of the trial.

**Table 5**
*Summary of the performance of three algorithms for the REML estimation of six variance parameters using the NSW wheat variety database*

| $\gamma$ | DF | EM | AI |
|---|---|---|---|
| Genotype | .3510 | .3489 | .3510 |
| Geno.Year | .1183 | .1119 | .1113 |
| Geno.Locn | .0911 | .0908 | .0920 |
| Expt.Acid | .4496 | .3894 | .4284 |
| Expt.Winter | .6780 | .6318 | .6348 |
| Expt.Maturity | 1.3478 | 1.3550 | 1.4368 |
| $\ell$ | 9035.205 | 9035.074 | 9035.886 |
| CPU time (mm:ss) | 4236:46 | 298:04 | 111:42 |
| Iterations | 927 | 23 | 7 |
| Time/iteration (seconds) | 274 | 769 | 957 |

This example provided the stimulus for the development of the AI algorithm. The size of the problem precluded using the FS algorithm, and initial analyses were attempted with the DF algorithm. These were unsatisfactory as convergence was extremely slow because of the large number of variance components in the model. To implement the DF algorithm, the equations were ordered A, W, M, G, G.Y, [G.L, E, A.E, W.E, M.E] with the equations in [ ] nested within experiment within location. By processing the equations by experiment within location, the working size of **C** remained below 720 equations, although the real size was 8,765 equations (see Appendix C for details).

The implementation of the AI algorithm used the same ordering of equations as the DF implementation so that the number of operations involved in steps 1 and 5 was minimised. For comparison, we also used the EM algorithm as described by Searle et al. (1992), in which the equations (4) are solved directly for $\gamma_i$. The stopping rule for both AI and EM was based on a percentage change in the log likelihood, which caused the EM algorithm to stop prematurely, relative to the AI algorithm. The stopping rule for the DF algorithm was that the sample variance of the REML log likelihood at the current set of vertices of the simplex not exceed .01 (Nelder and Mead, 1965). Starting values for the variance parameters were (.1, .1, .1, .1, .1, .1).

Table 5 presents the summary of the performances of the three algorithms. The EM algorithm took 23 iterations to reach a log likelihood of 9,035.074 compared to the AI algorithm which took seven iterations to reach a log likelihood of 9,035.886. The DF algorithm took 927 iterations to reach a log likelihood of 9,035.205. The CPU times given are for a SPARC 10+ and should only be taken as a guide, however the CPU time per iteration from these runs are in line with the approximate timings presented in Appendix B. The AI algorithm took 3.5 longer per iteration than the DF algorithm, while the EM algorithm took 2.8 longer per iteration compared to the DF algorithm. The AI times are inflated, because the program did not reuse the results from step 1 in step 4 but repeated the whole absorption. The correct factor is about 3.

## 4. Mixed Models with Unknown, Nondiagonal R

In this section we present an implementation of the AI algorithm for the analysis of Gaussian mixed linear models with correlated errors. This includes a wide class of problems including spatial analysis of field experiments (Cullis and Gleeson, 1991) and repeated measures analysis (Diggle, 1988; Laird and Ware, 1982; McGilchrist and Cullis, 1991; Verbyla and Cullis, 1992). In general,

$$\mathbf{H} = \mathbf{R}(\boldsymbol{\phi}) + \sum \gamma_i \mathbf{Z}_i \mathbf{Z}_i',$$

where $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_s)$ is the vector of covariance parameters representing spatial and/or temporal variance and covariance parameters. Thus, $\partial \mathbf{H}/\partial \phi_i = \mathbf{R}_i$, $\mathbf{R}_i = \partial \mathbf{R}/\partial \phi_i$, and so the score for $\phi_i$ is

$$\begin{aligned} \partial \ell/\partial \phi_i &= -\tfrac{1}{2}[\mathrm{tr}(\mathbf{P}\mathbf{R}_i) - \mathbf{y}'\mathbf{P}\mathbf{R}_i\mathbf{P}\mathbf{y}/\sigma^2] \\ &= -\tfrac{1}{2}[\mathrm{tr}(\mathbf{R}^{-1}\mathbf{R}_i) - \mathrm{tr}(\mathbf{C}^{-1}\mathbf{C}_i) - \mathbf{y}'_{q+i}\mathbf{P}\mathbf{y}/\sigma^2], \end{aligned} \tag{10}$$

where $\mathbf{C}_i = \mathbf{W}'\mathbf{R}^{-1}\mathbf{R}_i\mathbf{R}^{-1}\mathbf{W}$ and $\mathbf{y}_{q+i} = \mathbf{R}_i\mathbf{P}y = \mathbf{R}_i\mathbf{R}^{-1}\tilde{\mathbf{e}}$. The score for $\gamma_i$ is the same as before (see (8)). In (10), $\mathbf{y}_{q+i}$ is the working variate for $\phi_i$, and $\mathbf{C}_i$ is the weighted coefficient matrix formed using weights $\mathbf{R}^{-1}\mathbf{R}_i\mathbf{R}^{-1}$ instead of $\mathbf{R}^{-1}$. It follows that the elements of $\mathcal{I}_A$ for $\boldsymbol{\kappa} = (\boldsymbol{\gamma}, \boldsymbol{\phi})$ are

$$\tfrac{1}{2}\mathbf{y}_i'\mathbf{P}\mathbf{y}_j/\sigma^2,$$

where $\mathbf{y}_i|\,\gamma_i = \mathbf{Z}_i\tilde{\mathbf{u}}_i/\gamma_i$ and $\mathbf{y}_{q+i}|\,\phi_i = \mathbf{R}_i\mathbf{R}^{-1}\tilde{\mathbf{e}}$.

The steps in the iterative solution to the score equations are similar to those described in Section 3. There are additional terms required in the score for $\boldsymbol{\phi}$ and an additional $s$ working variates. Given an initial value for $\boldsymbol{\kappa} = \boldsymbol{\kappa}^{(0)}$ we proceed as follows:

1. Perform formation of $\mathbf{R}$, $\mathbf{R}^{-1}$, and $\mathbf{R}_i$ and calculate $\mathbf{M}$,
2. Perform absorption of $\mathbf{W}$ onto $\mathbf{y}$ using $\mathbf{M}$ after addition of $\mathbf{G}^{-1}$ to $\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}$ in $\mathbf{M}$, saving intermediate terms required in steps 3 and 5,
3. Backsubstitute to evaluate $\hat{\boldsymbol{\tau}}$ and $\tilde{\mathbf{u}}$ and calculate $\mathbf{C}^{-1}$,
4. Calculate $\tilde{\mathbf{e}}$, the working variates $\mathbf{y}_i$, and $\mathbf{C}_i$,
5. Absorb $\mathbf{W}$ onto the multivariate-dependent variate $[\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_{q+s}]$, and
6. Calculate the update for $\boldsymbol{\kappa}$. Return to 1 if not converged.

We note that the inversion of $\mathbf{R}$ and subsequent calculation of matrix products involving $\mathbf{R}^{-1}$ can be computer-intensive for moderate and large $n$. Assuming separability of the error process alleviates much of this burden using results concerning Kronecker products and vec( ) results (see Cullis and Gleeson (1991) and Martin (1990), for computational details).

### 4.1 *Spatial Analysis of Slate Hall Spring Wheat Trial*

In this section we present a comparison of the AI and FS algorithms for the estimation of some commonly used spatial models for field experiments using the data described in Section 3.1. We assume throughout that the trend process is a realisation of a separable process. In the notation of Cullis and Gleeson (1991) we model trend + error = $\boldsymbol{\xi} + \boldsymbol{\eta}$ by the following:

a. ARIMA(0,1,1) × ARIMA(0,1,1): the standard first difference model (Besag and Kempton, 1986) in two dimensions proposed by Cullis and Gleeson (1991);
b. ARIMA(0,1,0) × ARIMA(0,1,0) + white noise: an interesting alternate two-dimensional extension of the Besag and Kempton (1986) model;
c. ARIMA(1,0,0) × ARIMA(1,0,0): the autogressive model proposed by Zimmerman and Harville (1991); and
d. ARIMA(1,0,0) × ARIMA(1,0,0) + white noise: an extension to (c) fitting a nugget effect.

Variety effects are taken as fixed and no additional terms are included in $\mathbf{X}$. Table 6 presents a summary of the convergence sequences for the FS and AI algorithms for the four models. The convergence sequences for the two algorithms are very similar for all models, although the AI algorithm converged in one less iteration. Additionally there is a significant reduction in CPU and memory using the AI implementation of TwoD.

The loglikelihoods are directly comparable for models (a) and (b), and for the interblock analysis (Section 3.1) with models (c) and (d). We conclude that the variance model (b) has higher parsimony than (a) and that (d) is better than (c) which is better than the interblock analysis.

Table 7 provides a comparison of the variety effects among the five models. Models (a), (b), and (d) give very similar results, followed by (c) and then the interblock analysis. By ignoring the nugget effect, model (c) appears to underestimate the standard error of difference. However, the nugget effect is often not significant.

## 5. Discussion

In this paper we have presented an improved derivative-based procedure for the estimation of variance and covariance parameters by REML. The algorithm appears to have similar convergence properties to the FS algorithm, yet avoids the computing burdens of that approach. As for the FS algorithm, it may be used to estimate negative variance components by removing the restriction that the components be positive. We have tested its performance on a much wider class of models and datasets and have implemented the algorithm in the TwoD software (Gilmour, 1992), an S-plus routine (which is available form the third author on request) and a FORTRAN 77 program for variance component estimation in large genotype × environment problems.

For large problems typically encountered in animal and plant breeding problems, the AI algorithm is a viable alternative to both the DF or EM algorithms. Both of these have great difficulty when

**Table 6**

*Summary of convergence sequences for four spatial models fitted to the Slate Hall wheat variety trial*

(a) ARIMA(0,1,1) × ARIMA(0,1,1) Starting values: $\phi_1 = \phi_2 = .5$, $\ell = -534.4$

| | AI | | | FS | | |
|---|---|---|---|---|---|---|
| Iteration | $\phi_1$ | $\phi_2$ | $\ell$ | $\phi_1$ | $\phi_2$ | $\ell$ |
| 1 | .386 | .608 | −533.4 | .392 | .554 | −533.5 |
| 2 | .383 | .607 | −533.4 | .390 | .592 | −533.4 |
| 3 | | | | .385 | .603 | −533.4 |

(b) ARIMA(0,1,0) × ARIMA(0,1,0) + white noise Starting values: $\phi_1 = 1.0$, $\ell = -533.0$

| | AI | | FS | |
|---|---|---|---|---|
| Iteration | $\phi_1$ | $\ell$ | $\phi_1$ | $\ell$ |
| 1 | 1.420 | −532.6 | 1.355 | −532.7 |
| 2 | 1.563 | −532.6 | 1.524 | −532.6 |
| 3 | | | 1.575 | −532.6 |

(c) ARIMA(1,0,0) × ARIMA(1,0,0) Starting values: $\phi_1 = \phi_2 = .5$, $\ell = -644.5$

| | AI | | | FS | | |
|---|---|---|---|---|---|---|
| Iteration | $\phi_1$ | $\phi_2$ | $\ell$ | $\phi_1$ | $\phi_2$ | $\ell$ |
| 1 | .679 | .463 | −641.0 | .725 | .508 | −641.4 |
| 2 | .684 | .459 | −641.0 | .684 | .459 | −641.0 |
| 3 | | | | .684 | .459 | −641.0 |

(d) ARIMA(1,0,0) × ARIMA(1,0,0) + white noise Starting values: $\phi_1 = .684$, $\phi_2 = .459$, $\phi_3 = .1$, $\ell = -640.0$

| | AI | | | | FS | | | |
|---|---|---|---|---|---|---|---|---|
| Iteration | $\phi_1$ | $\phi_2$ | $\phi_3$ | $\ell$ | $\phi_1$ | $\phi_2$ | $\phi_3$ | $\ell$ |
| 1 | .871 | .658 | .542 | −638.0 | .880 | .680 | .567 | −638.0 |
| 2 | .844 | .681 | .679 | −637.5 | .862 | .694 | .740 | −637.6 |
| 3 | .844 | .682 | .690 | −637.5 | .849 | .689 | .707 | −637.5 |
| 4 | | | | | .845 | .685 | .696 | −637.5 |

**Table 7**

*Correlations among solutions for variety effects and average standard errors of difference (grams/m$^2$) from the interblock and four spatial analyses of the Slate Hall wheat variety trial*

| Analysis | Interblock | (a) | (b) | (c) | (d) |
|---|---|---|---|---|---|
| (a) ARIMA(0,1,1) × ARIMA(0,1,1) | .986 | 1. | | | |
| (b) ARIMA(0,1,0) × ARIMA(0,1,0) + white noise | .982 | .999 | 1. | | |
| (c) ARIMA(1,0,0) × ARIMA(1,0,0) | .971 | .987 | .987 | 1. | |
| (d) ARIMA(1,0,0) × ARIMA(1,0,0) + white noise | .983 | .997 | .996 | .993 | 1. |
| average standard error of difference | 62.0 | 61.1 | 61.3 | 59.0 | 60.5 |

there are many correlated components. The computational requirement for AI is similar to EM and three times that of the DF algorithm (see Appendix B), and this was verified by the results in Section 3.2. Given the reduction in the number of iterations there should be a significant advantage to using the AI algorithm over both DF and EM.


## Résumé

Nous présentons une stratégie, utilisant une matrice moyenne d'information, qui est à la fois aisément calculable et efficace pour l'estimation des composantes de variance par la méthode du maximum de vraisemblance restreint dans le modèle linéaire mixte. Trois applications sont décrites. Le but de cet algorithme était l'estimation des composantes de variance dans une comparaison de variétés de blé issue de 1,071 essais qui couvraient 10 ans et 60 lieux d'expérimentation en Nouvelle

Galles du Sud. Nous appliquons l'algorithme pour un plan d'expérience en blocs incomplets et pour une analyse spatiale d'un champ d'essais.

## References

Besag, J. and Kempton, R. A. (1986). Statistical analysis of field experiments using neighbouring plots. *Biometrics* **42,** 231–251.

Cullis, B. R. and Gleeson, A. C. (1989). The efficiency of neighbour analysis for replicated variety trials in Australia. *Journal of Agricultural Science, Cambridge* **113,** 233–239.

Cullis, B. R. and Gleeson, A. C. (1991). Spatial analysis of field experiments—An extension to two dimensions. *Biometrics* **47,** 1449–1460.

Cullis, B. R., Gleeson, A. C., Lill, W. J., Fisher, J. A., and Read, B. J. (1989). A new procedure for the analysis of early generation variety trials. *Applied Statistics* **38,** 361–375.

Cullis, B. R., Gleeson, A. C., and Thomson, F. M. (1992). The response to selection of different procedures for the analysis of early generation variety trials. *Journal of Agricultural Science, Cambridge* **118,** 141–148.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM Algorithm. *Journal of the Royal Statistical Society, Series B* **39,** 1–38.

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

Duff, I. S., Erisman, A. M., and Reid, J. K. (1989). *Direct Methods for Sparse Matrices*. Oxford: Clarendon Press.

Erisman, A. M. and Tinney, W. F. (1975). On computing certain elements of the inverse of a sparse matrix. *Communications ACM* **18,** 177–179.

Gilmour, A. R. (1992). *TwoD. A Program to Fit a Mixed Linear Model with Two Dimensional Spatial Adjustment for Local Trend*. NSW Agriculture Biometric Bulletin No. 1, 73 pp.

Gleeson, A. C. and Cullis, B. R. (1987). Residual maximum likelihood (REML) estimation of a neighbour model for field experiments. *Biometrics* **43,** 277–287.

Golub, H. G. and Van Loan, C. F. (1983). *Matrix Computations*. Oxford: North Oxford Academic.

Graser, H.-U., Smith, S. P., and Tier, B. (1987). A derivative-free approach for estimating variance components in animal models by restricted maximum likelihood. *Journal of Animal Science* **64,** 1362–1372.

Johnson, D. L. and Thompson, R. (1995). Restricted maximum likelihood estimation of variance components for univariate animal models using sparse matrix techniques and a quasi Newton procedure. *Journal of Dairy Science* **78,** 449–456.

Laird, N. and Ware, J. (1982). Random effects models for longitudinal data. *Biometrics* **38,** 963–975.

Lill, W. J., Gleeson, A. C., and Cullis, B. R. (1988). Relative accuracy of a neighbour method for field experiments. *Journal of Agricultural Science, Cambridge* **111,** 339–346.

Martin, R. J. (1990). The use of time series models and methods in the analysis of agricultural field trials. *Communications in Statistics—Theory and Methods* **19,** 55–81.

McGilchrist, C. A. and Cullis, B. R. (1991). REML estimation for repeated measures analysis. *Journal of Statistical Computations and Simulation* **38,** 151–163.

Meyer, K. (1989). Restricted maximum likelihood to estimate variance components of animal models with several random effects using a derivative free algorithm. *Genetic Selection and Evolution* **21,** 318–340.

Misztal, I. C. (1990). Restricted maximum likelihood estimation of variance components in animal model using sparse matrix inversion and a supercomputer. *Journal of Dairy Science* **73,** 163–172.

Misztal, I. C. and Perez-Enciso, M. (1993). Sparse matrix inversion for Restricted Maximum Likelihood estimation of variance components by expectation maximization. *Journal of Dairy Science* **76,** 1479–1483.

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *Computer Journal* **7,** 308–313.

Patterson, H. D. and Nabugoomu, F. (1992). REML and the analysis of a series of crop variety trials. In *Proceedings of the 25th International Biometric Conference*, 77–93. Hamilton, New Zealand: AgResearch.

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

Payne, R. W. (1993). *GENSTAT 5 Release 3 Reference Manual*. Oxford: Clarendon Press.

Searle, S. R., Casella, G., and McCulloch, C. E. (1992). *Variance Components*, J. W. Wiley: New York.

Smith, S. P. (1990). Estimation of genetic parameters in non linear models. *Advances in Statistical Methods for Genetic Improvement of Livestock*, D. Gianola and K. Hammond (eds), 190–206. New York: Springer-Verlag.

Takahashi, K., Fagan, J., and Chin, M. S. (1973). Formation of sparse bus impedance matrix and its application to short circuit theory. *Proceedings of the 8th International PICA Conference, Minneapolis*, 63–69.

Thompson, R. (1977). The estimation of heritability with unbalanced data. I. Observations available on parents and offspring. *Biometrics* **33,** 485–495.

Thompson, R., Wray, N. R., and Crump, R. E. (1994). Calculation of prediction error variances using sparse matrix methods. *Journal of Animal Breeding and Genetics* **111,** 102–109.

Verbyla, A. P. and Cullis, B. R. (1992). The analysis of multistratum and spatially correlated repeated measures data. *Biometrics* **48,** 1015–1032.

Wilkinson, G. N. and Rogers, C. E. (1973). Symbolic description of factorial models for analysis of variance. *Applied Statistics* **22,** 392–399.

Zimmerman, D. L. and Harville, D. A. (1991). A random field approach to the analysis of field-plot experiments and other spatial experiments. *Biometrics* **47,** 223–239.

*Received March* 1994; *revised September* 1994; *accepted January* 1995.

## Appendix A

In this appendix we derive the terms in (8). Let $\mathbf{Z}_i = \mathbf{W}\mathbf{S}_i$ and $\mathbf{C} = \mathbf{W}'\mathbf{R}^{-1}\mathbf{W} + \mathbf{G}^*$, where $\mathbf{G}^*$ is $\mathbf{G}^{-1}$ augmented with zeros corresponding to $\mathbf{X}$ in $\mathbf{W}$.

$$\begin{aligned}
\mathrm{tr}(\mathbf{H}_i\mathbf{P}) &= \mathrm{tr}[\mathbf{Z}_i\mathbf{Z}_i'(\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{W}\mathbf{C}^{-1}\mathbf{W}'\mathbf{R}^{-1})] \\
&= \mathrm{tr}[\mathbf{S}_i'\mathbf{W}'\mathbf{R}^{-1}\mathbf{W}\mathbf{C}^{-1}(\mathbf{C} - \mathbf{W}'\mathbf{R}^{-1}\mathbf{W})\mathbf{S}_i] \\
&= \mathrm{tr}[\mathbf{S}_i'(\mathbf{C} - \mathbf{G}^*)\mathbf{C}^{-1}\mathbf{G}^*\mathbf{S}_i] \\
&= \mathrm{tr}(\mathbf{S}_i'\mathbf{G}^*\mathbf{S}_i) - \mathrm{tr}(\mathbf{S}_i'\mathbf{G}^*\mathbf{C}^{-1}\mathbf{G}^*\mathbf{S}_i) \\
&= b_i/\gamma_i - \mathrm{tr}(\mathbf{C}^{ii})/\gamma_i^2.
\end{aligned}$$

The modified working variate for $\gamma_i$ follows by using the result $\mathbf{Z}'\mathbf{P}\mathbf{y} = \mathbf{G}^{-1}\tilde{\mathbf{u}}$.

## Appendix B

In this appendix we define the recursive procedures of absorption, backsubstitution, and partial inversion of $\mathbf{C}$ referred to in Sections 2, 3, and 4. We represent $\mathbf{M}$ (defined in (9)) with $\mathbf{G}^{-1}$ added to $\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}$ by the matrix

$$\begin{bmatrix} a^{(N)} & & \\ \mathbf{b}^{(N)} & \mathbf{C}^{(N)} & \\ d^{(N)} & \mathbf{e}^{(N)\prime} & f^{(N)} \end{bmatrix},$$

where $N$ is the order of the coefficient matrix $\mathbf{C}$, partitioned into $\mathbf{C}^{(N)}$, $\mathbf{e}^{(N)}$, and $f^{(N)}$, and $a^{(N)}$, $\mathbf{b}^{(N)}$, and $d^{(N)}$ represent the first column of $\mathbf{M}$, the terms involving $\mathbf{y}$.

The absorption recursion required in step 1 is defined for $i$ decremented from $N$ to 1 by

$$a^{(i-1)} = a^{(i)} - d^{(i)}d^{(i)}/f^{(i)}$$

$$\begin{bmatrix} \mathbf{b}^{(i-1)} \\ d^{(i-1)} \end{bmatrix} = \mathbf{b}^{(i)} - \mathbf{e}^{(i)}d^{(i)}/f^{(i)}$$

$$\begin{bmatrix} \mathbf{C}^{(i-1)} & \\ \mathbf{e}^{(i-1)\prime} & f^{(i-1)} \end{bmatrix} = \mathbf{C}^{(i)} - \mathbf{e}^{(i)}\mathbf{e}^{(i)\prime}/f^{(i)},$$

where $a^{(i)}$, $d^{(i)}$, and $f^{(i)}$ are scalars, $\mathbf{b}^{(i)}$ and $\mathbf{e}^{(i)}$ are vectors of length $i - 1$, and $\mathbf{C}^{(i)}$ is a symmetric matrix with $i - 1$ rows. The terms $a^{(i)}$, $\mathbf{b}^{(i)}$, $\mathbf{C}^{(i)}$, $d^{(i)}$, $\mathbf{e}^{(i)}$, and $f^{(i)}$ are sums of squares and cross products adjusted for rows $i + 1, \ldots, N$. The scalar $f^{(i)}$ is often referred to as the pivot and $\log\det\mathbf{C} = \Sigma_{i=1}^{N} \log(f^{(i)})$ and $\mathbf{y}'\mathbf{P}\mathbf{y} = a^{(0)}$.

Assuming we avoid multiplication by zero and that $\mathbf{e}^{(i)}$ has $n_i$ non-zero values, the number of multiplications is $(n_i + 1)(n_i + 2)/2$. Thus operations can be avoided by ordering the equations to minimise $n_i$ at each stage. In the recursion $a^{(i)}$, $\mathbf{b}^{(i)}$, and $\mathbf{C}^{(i)}$ are overwritten, whilst $d^{(i)}$, $\mathbf{e}^{(i)}$, and $f^{(i)}$ are the intermediate terms retained.

In step 2, we backsubstitute for $\boldsymbol{\beta}' = (\boldsymbol{\tau}', \mathbf{u}')$ starting with $\boldsymbol{\beta}^{(1)} = \beta_1 = d^{(1)}/f^{(1)}$, thence

$$\beta_i = (d^{(i)} - \mathbf{e}^{(i)\prime}\boldsymbol{\beta}^{(i-1)})/f^{(i)},$$

where $\boldsymbol{\beta}^{(i-1)}$ represents the first $i-1$ elements of $\boldsymbol{\beta}$, and $\beta_i$ is the $i$th element of $\boldsymbol{\beta}$. By avoiding unnecessary operations involving zero elements there are $2n_i$ operations at the $i$th recursion.

In step 3, we use the solution for $\mathbf{u}$ with $\mathbf{M}$ to form the crossproducts of the working variables: $\mathbf{y}'\mathbf{R}^{-1}\mathbf{Z}_i\mathbf{u}_i/\gamma_i$, $\mathbf{u}_j'\mathbf{Z}_j'\mathbf{R}^{-1}\mathbf{Z}_i\mathbf{u}_i/(\gamma_i\gamma_j)$, $\mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}_i\mathbf{u}_i/\gamma_i$, and $\mathbf{Z}_j'\mathbf{R}^{-1}\mathbf{Z}_i\mathbf{u}_i/\gamma_i$.

In step 4, we repeat the absorption of $\mathbf{W}$ onto the multivariate-dependent vector $[\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_{q+s}]$. Define $\mathbf{A}^{(N)}$ as the crossproducts of the working variates namely $\mathbf{y}_i'\mathbf{R}^{-1}\mathbf{y}_j$, $i$ and $j \neq 0$, $\mathbf{B}^{(N)}$ as the crossproducts of the working variates analagous to $\mathbf{b}^{(N)}$ but with $\mathbf{y}_0'\mathbf{R}^{-1}\mathbf{y}_j$ in the first row and $\mathbf{D}^{(N)}$ as the crossproduct of the working variate analagous to $d^{(N)}$. The recursion is then

$$\mathbf{A}^{(i-1)} = \mathbf{A}^{(i)} - \mathbf{D}^{(i)}\mathbf{D}^{(i)\prime}/f^{(i)}$$

$$\begin{bmatrix} \mathbf{B}^{(i-1)} \\ \mathbf{D}^{(i-1)} \end{bmatrix} = \mathbf{B}^{(i)} - \mathbf{e}^{(i)}\mathbf{D}^{(i)\prime}/f^{(i)}.$$

Note that the $\mathbf{e}^{(i)}$ and $f^{(i)}$ were formed in step 1. The terms required for $\mathcal{J}_A$ are then $\mathbf{A}^{(0)}$, $\mathbf{B}^{(0)}$, and $\mathbf{a}^{(0)}$.

In step 5, we form the elements of $\mathbf{C}^{-1}$ required for the trace terms in the score equation for $\gamma_i$. Let $f^{*(1)} = 1/f^{(1)} = \mathbf{C}^{*(1)}$. For $i$ from 1 to $N-1$ then

$$\mathbf{C}^{*(i+1)} = \begin{bmatrix} \mathbf{C}^{*(i)} & \\ \mathbf{e}^{*(i)\prime} & f^{*(i)} \end{bmatrix},$$

where $\mathbf{e}^{*(i)} = -\mathbf{C}^{*(i)}\mathbf{e}^{(i)}/f^{(i)}$ and $f^{*(i)} = 1/f^{(i)} - \mathbf{e}^{(i)\prime}\mathbf{e}^{*(i)}/f^{(i)}$.

In the case considered in this paper, only $f^{*(i)}$ terms are involved in the traces and so we choose to evaluate only those terms in $\mathbf{e}^{*(i)}$ which are subsequently needed. Since $e_j^{*(i)} = C_{i+1,j}^{*(k)}$ for $k > i$ is multiplied by $e_j^{(k)}$, it is not needed if $e_j^{(k)}$ is zero for all $k > i$. This condition is satisfied if $C_{i+1,j}^{(k)}$ was always zero during the absorption phase. Using sparse storage techniques (Duff et al., 1989), we would not actually form this cell during absorption (Erisman and Tinney, 1975; Misztal, 1990; Misztal and Perez-Enciso, 1993; Takahashi, Fagan and Chin, 1973; Thompson, Wray and Crump, 1994) and so do not need to form it now (unless we actually want this portion of the whole inverse for some other purpose). Thus, the partial $\mathbf{C}^{-1}$ can be formed to overwrite the matrix existing after absorption. The elements $(C_{ij}^{*(N)})$ are proportional to the covariance between $\beta_i$ and $\beta_j$. The number of multiplications involved in forming $\mathbf{C}^{*(N)}$ at the $i$th stage is $n_i(n_i + 3)$.

## Appendix C

Utilising sparsity in our genotype × environment study with 60 locations and 1,071 experiments, we only hold one location and one experiment at a time. Thus we operate on a matrix of order about 720 rather than one of order 8,765, since $\mathbf{X}$ has 609 equations, $\mathbf{L}_i$ has up to 107, and $\mathbf{E}_i$ has 4 equations.

Consider the reduced coefficient matrix

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & & & & & & \\ \mathbf{L}_1'\mathbf{X} & \mathbf{L}_1'\mathbf{L}_1 & & & & & \\ \mathbf{E}_1'\mathbf{X} & \mathbf{E}_1'\mathbf{L}_1 & \mathbf{E}_1'\mathbf{E}_1 & & & & \\ \mathbf{E}_2'\mathbf{X} & \mathbf{E}_2'\mathbf{L}_1 & 0 & \mathbf{E}_2'\mathbf{E}_2 & & & \\ \mathbf{L}_2'\mathbf{X} & 0 & 0 & 0 & \mathbf{L}_2'\mathbf{L}_2 & & \\ \mathbf{E}_3'\mathbf{X} & 0 & 0 & 0 & \mathbf{E}_3'\mathbf{L}_2 & \mathbf{E}_3'\mathbf{E}_3 & \\ \mathbf{E}_4'\mathbf{X} & 0 & 0 & 0 & \mathbf{E}_4'\mathbf{L}_2 & 0 & \mathbf{E}_4'\mathbf{E}_4 \end{bmatrix},$$

where $\mathbf{X}$ relates to the equations not involving Location ($L$) or Experiment ($E$), $\mathbf{L}_i$ has up to 107 genotype.location equations and $\mathbf{E}_i$ has four experiment equations as described in Section 3.2.

We read in the data from experiment 1, accumulating into $\mathbf{X}'\mathbf{X}$, $\mathbf{L}_1'\mathbf{X}$, and $\mathbf{L}_1'\mathbf{L}_1$ and forming $\mathbf{E}_1'[\mathbf{X} \quad \mathbf{L}_1 \quad \mathbf{E}_1]$. Since the space spanned by the matrix $E_1$ is orthogonal to all other $\mathbf{E}_i$, we adjust $\mathbf{X}'\mathbf{X}$, $\mathbf{L}_1'\mathbf{X}$, and $\mathbf{L}_1'\mathbf{L}_1$ for $\mathbf{E}_1$ and write $\mathbf{E}_1'[\mathbf{X} \quad \mathbf{L}_1 \quad \mathbf{E}_1]$ to disk. We repeat this process for experiment 2 (in general we repeat this process for all experiments within location 1). Now, $\mathbf{X}'\mathbf{Q}_1\mathbf{X}$ is adjusted for $\mathbf{L}_1$ which are now of the form $\mathbf{L}_1'\mathbf{Q}_1\mathbf{X}$ and $\mathbf{L}_1'\mathbf{Q}_1\mathbf{L}_1$ and write them to disk. We repeat this process for location 2 (and so on in the general case).

To form the inverse elements required for the trace terms in the score, we reverse the process. That is, first invert $\mathbf{X}'\mathbf{Q}\mathbf{X}$ and then read back $\mathbf{L}_2$ and form its inverse. Read back $\mathbf{E}_4$ and form its inverse. Use and discard the $\mathbf{E}_4$ inverse, if required, before reading back $\mathbf{E}_3$ and forming its inverse. Then both the $\mathbf{L}_2$ and $\mathbf{E}_3$ inverses may be discarded and the process repeated for the next location (that is, location 1).