

# ON THE PRECISION OF REML ESTIMATES OF ADDITIVE AND DOMINANCE VARIANCE COMPONENTS IN MAIZE USING RELATIONSHIP MATRICES

R.T. Boca<sup>1,2</sup>, R.J.C. Cantet<sup>2,3,\*</sup>

<sup>1</sup> Instituto Nacional de Tecnología Agropecuaria, CNIA Castelar

<sup>2</sup> Universidad de Buenos Aires, Facultad de Agronomía

<sup>3</sup> Consejo Nacional de Investigaciones Científicas y Técnicas, (CONICET), Argentina

Received June 9, 2004

**ABSTRACT** - Best Linear Unbiased Prediction (BLUP) of combining abilities increases the efficiency of hybrid breeding programmes in maize, by combining the information contained in the data and in the pedigree. Calculation of BLUP requires that the variance components are known. However, as the dispersion parameters are unknown, they have to be estimated from the same data for which BLUP are obtained. This can be done by Restricted Maximum Likelihood (REML). In the present research, we estimated variance components for additive and dominance effects using REML. In addition, the asymptotic covariance matrix of REML estimates of the dispersion parameters was computed as the inverse of the observed information matrix. In doing so, formulae to calculate the latter matrix in terms of matrices related to the 'mixed model equations' for BLUP are presented, which make perform in the calculations with large data sets feasible. The model employed considers general combining abilities of related genotypes through the additive relationship matrix **A**, as well as specific combining abilities through the dominance relationship matrix **D**. Expressions are also given to include a variance component due to a fertility trend and dispersion parameters for genotype by environment interaction, into the asymptotic covariance matrix. Data from an Argentinean maize breeding program is used to illustrate the developments.

KEY WORDS: BLUP; REML; Mixed model equations; Asymptotic covariance matrix; Maize.

## INTRODUCTION

Mixed linear models have come to play a role in the genetic evaluation of single-cross maize hybrids. This is due to the introduction of Best Linear Unbiased Prediction (BLUP; HENDERSON, 1984) of combining abilities, and of Restricted Maximum Likelihood

estimation (REML; PATTERSON and THOMPSON, 1971) of the genetic variance components by BERNARDO (1994, 1996). Predictions by BLUP make use of the information contained in the data and in the pedigree. In practice, BLUP of genetic merit is calculated by means of Henderson's Mixed Model Equations (MME; HENDERSON, 1984) with the true values of the genetic and error variances replaced by their REML estimates, as these parameters are usually unknown. This two-stage procedure to calculate BLUP yields unbiased estimators of linear functions of the fixed effects and predictors of functions of random effects (KACKAR and HARVILLE, 1981). Nonetheless, the variability of BLUP increases due to the replacement of the variance components by their REML estimates (HARVILLE, 1985). On one hand, the use of genetic relationships allows for the predictions of combining abilities. However, it also introduces correlations among the joint REML estimates of the additive and dominance variance components which, in turn, decrease the precision of those estimates. As there are no analytical expressions for the sampling (co)variance matrix of likelihood based estimators, the joint variability of REML estimates can be assessed by the observed information matrix (HARVILLE, 1977). This method has not been used for estimating the joint variability of genetic dispersion parameters in plant breeding. Thus, the goal of this paper is to show how to calculate the asymptotic covariance matrix of REML estimates by means of the inverse of the REML information matrix (HARVILLE, 1977), written in terms of matrices related to the 'mixed model equations' (MME) of HENDERSON (1984), which provide BLUP of the combining abilities. The expressions presented here are particularly useful for large data sets, as they reduce the computational complexity from a function of the number of records to a function of the usually smaller number of inbred lines plus the number of trials or block effects.

\* For correspondence (fax: +54 11 4524 8737; e-mail: rcantet@agro.uba.ar)

**MIXED MODEL TO PREDICT COMBINING  
ABILITIES NAD TO ESTIMATE  
THE VARIANCE COMPONENTS IN CORN**

In the presence of additive and dominance inheritance, a vector of records from hybrid genotypes (*i.e.* maize) can be explained by means of the following individual (genotypic) mixed linear model (HENDERSON, 1984):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{a} + \mathbf{Z}_2\mathbf{d} + \mathbf{e} \quad [1]$$

In [1],  $\mathbf{y}$  is an  $n \times 1$  vector of records, and  $\mathbf{b}$  is a  $p \times 1$  vector of fixed effects (year-experiment-block) related to  $\mathbf{y}$  by the incidence matrix  $\mathbf{X}$  ( $n \times p$ ). The vector  $\boldsymbol{\beta}$  is parameterized such that the rank  $[\mathbf{X}]$  is full, *i.e.* rank  $[\mathbf{X}] = p$ . Random additive genetic effects are the general combining abilities in  $\mathbf{a}$  ( $q \times 1$ ) and the specific combining abilities due to dominance effects in  $\mathbf{d}$  ( $q \times 1$ ), and  $q$  is the number of genotypes. The incidence matrices  $\mathbf{Z}_1$  ( $n \times q$ ) and  $\mathbf{Z}_2$  ( $n \times q$ ) relate  $\mathbf{y}$  to the elements of  $\mathbf{a}$  and  $\mathbf{d}$ , respectively. Finally, the  $n \times 1$  vector  $\mathbf{e}$  contains the error terms. The joint distribution of all random variables in [1] is the following:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{a} \\ \mathbf{d} \\ \mathbf{e} \end{bmatrix} \sim N \left[ \begin{bmatrix} \mathbf{X}\boldsymbol{\beta} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{V} & \mathbf{Z}_1\mathbf{A}\sigma_a^2 & \mathbf{Z}_2\mathbf{D}\sigma_d^2 & \mathbf{I}\sigma_e^2 \\ \mathbf{AZ}_1'\sigma_a^2 & \mathbf{A}\sigma_a^2 & \mathbf{0} & \mathbf{0} \\ \mathbf{DZ}_2'\sigma_d^2 & \mathbf{0} & \mathbf{D}\sigma_d^2 & \mathbf{0} \\ \mathbf{I}\sigma_e^2 & \mathbf{0} & \mathbf{0} & \mathbf{I}\sigma_e^2 \end{bmatrix} \right] \quad [2]$$

where  $\mathbf{V} = \mathbf{Z}_1\mathbf{A}\mathbf{Z}_1'\sigma_a^2 + \mathbf{Z}_2\mathbf{D}\mathbf{Z}_2'\sigma_d^2 + \mathbf{I}\sigma_e^2$ . The matrix  $\mathbf{A}$  contains the additive relationships among genotypes (HENDERSON, 1984; BERNARDO, 1994), and represent the covariance among general combining abilities (after multiplying by the additive variance  $\sigma_a^2$ ) in a quantitative genetics setting (COCKERHAM, 1954; KEMPTHORNE, 1954). In the same way,  $\mathbf{D}$  contains dominance relationships, *i.e.* probabilities that both genes in each of two hybrids are identical by descent. Prediction of genetic values is carried out by linear combinations of  $\mathbf{a}$  and  $\mathbf{d}$ , which are calculated by means of their corresponding BLUP predictors  $\hat{\mathbf{a}}$  and  $\hat{\mathbf{d}}$  using the following set of 'mixed model equations' (HENDERSON, 1984; BERNARDO, 1994):

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z}_1 & \mathbf{X}'\mathbf{Z}_2 \\ \mathbf{Z}_1'\mathbf{X} & \mathbf{Z}_1'\mathbf{Z}_1 + \mathbf{A}^{-1}\alpha & \mathbf{Z}_1'\mathbf{Z}_2 \\ \mathbf{Z}_2'\mathbf{X} & \mathbf{Z}_2'\mathbf{Z}_1 & \mathbf{Z}_2'\mathbf{Z}_2 + \mathbf{D}^{-1}\delta \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{a}} \\ \hat{\mathbf{d}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}_1'\mathbf{y} \\ \mathbf{Z}_2'\mathbf{y} \end{bmatrix} \quad [3]$$

with  $\alpha = \frac{\sigma_e^2}{\sigma_a^2}$  and  $\delta = \frac{\sigma_e^2}{\sigma_d^2}$ . It is useful to denote  $\mathbf{C}$  as the inverse of the coefficient matrix of the MME in [3]:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z}_1 & \mathbf{X}'\mathbf{Z}_2 \\ \mathbf{Z}_1'\mathbf{X} & \mathbf{Z}_1'\mathbf{Z}_1 + \mathbf{A}^{-1}\alpha & \mathbf{Z}_1'\mathbf{Z}_2 \\ \mathbf{Z}_2'\mathbf{X} & \mathbf{Z}_2'\mathbf{Z}_1 & \mathbf{Z}_2'\mathbf{Z}_2 + \mathbf{D}^{-1}\delta \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{C}^{\beta\beta} & \mathbf{C}^{\beta\alpha} & \mathbf{C}^{\beta d} \\ \mathbf{C}^{\alpha\beta} & \mathbf{C}^{\alpha\alpha} & \mathbf{C}^{\alpha d} \\ \mathbf{C}^{d\beta} & \mathbf{C}^{d\alpha} & \mathbf{C}^{dd} \end{bmatrix} \sigma_e^2$$

To estimate the variance components  $\sigma_a^2$ ,  $\sigma_d^2$  and  $\sigma_e^2$ , we use a modification of the EM (Expectation-Maximization; DEMPSTER *et al.*, 1977) algorithm, as implemented by CANTET *et al.* (1993) to improve the speed of convergence. The estimating equations in iteration  $[p + 1]$  for the variance components are:

$$\sigma_a^{2[p+1]} = \frac{[\hat{\mathbf{a}}' \mathbf{A}^{-1} \hat{\mathbf{a}}]^{[p+1]} - \phi \sigma_a^{2[p]} [q - \text{tr}(\mathbf{A}^{-1} \mathbf{C}^{\alpha\alpha})^{[p+1]}] \sigma_e^{2[p]}}{[q - \text{tr}(\mathbf{A}^{-1} \mathbf{C}^{\alpha\alpha})^{[p+1]}] (1 - \phi)}$$

$$\sigma_d^{2[p+1]} = \frac{[\hat{\mathbf{d}}' \mathbf{D}^{-1} \hat{\mathbf{d}}]^{[p+1]} - \phi \sigma_d^{2[p]} [q - \text{tr}(\mathbf{D}^{-1} \mathbf{C}^{dd})^{[p+1]}] \sigma_e^{2[p]}}{[q - \text{tr}(\mathbf{D}^{-1} \mathbf{C}^{dd})^{[p+1]}] (1 - \phi)} \quad [4]$$

$$\sigma_e^{2[p+1]} = \frac{[\hat{\mathbf{e}}' \hat{\mathbf{e}}]^{[p+1]} + [p + 2q - (\text{tr}(\mathbf{A}^{-1} \mathbf{C}^{\alpha\alpha})^{[p+1]} + \text{tr}(\mathbf{D}^{-1} \mathbf{C}^{dd})^{[p+1]})] \sigma_e^{2[p]}}{n}$$

where  $\mathbf{C}^{\alpha\alpha}$  is the partition of  $\mathbf{C}$  associated with the general combining abilities in  $\hat{\mathbf{a}}$ , and  $\mathbf{C}^{dd}$  is the corresponding partition associated with the specific combining abilities in  $\hat{\mathbf{d}}$ . Also,  $0 < \phi < 1$  is a scalar that controls the speed of convergence. In the current research  $\phi$  was taken to be equal to 0.6, as suggested by CANTET *et al.* (1993).

**ASYMPTOTIC VARIANCE OF REML ESTIMATES  
OF VARIANCE COMPONENTS**

The variability of REML estimates is usually not reported because the sampling distribution of the vector of variance components can not be written in closed form (LEHMAN, 1983). This is due to the REML estimates being obtained by iteration, which in turn precludes a feasible calculation of the variances of those estimates. Let  $\boldsymbol{\theta}$  be the vector of variance components in the mixed linear model [1]-[2]. HARVILLE (1977) derived formulae to calculate the information matrix  $\mathbf{I}(\boldsymbol{\theta})$  for REML estimates of  $q$ . Under asymptotic normality of REML estimates of variance components, the inverse of  $\mathbf{I}(\boldsymbol{\theta})$  is the variance of  $\boldsymbol{\theta}$  (CRESSIE and LAHIRI, 1993; JIANG, 1996). The expression given by HARVILLE (1977) for element  $i, j$  of  $\mathbf{I}(q)$  is equal to:

$$[\mathbf{I}(\boldsymbol{\theta})]_{i,j} = \frac{1}{2} \text{tr} \left( \mathbf{P} \frac{\partial \mathbf{V}}{\partial \theta_i} \mathbf{P} \frac{\partial \mathbf{V}}{\partial \theta_j} \right) \quad [5]$$

where  $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1}$ . In the model [1]-[2] the vector  $\boldsymbol{\theta}$  is:

$$\boldsymbol{\theta} = \begin{bmatrix} \sigma_a^2 \\ \sigma_d^2 \\ \sigma_e^2 \end{bmatrix}$$

so that  $\mathbf{I}(\mathbf{q})$  is equal to:

$$\mathbf{I}_{REML} \begin{bmatrix} \sigma_a^2 \\ \sigma_d^2 \\ \sigma_e^2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} \text{tr} \left( \mathbf{P} \frac{\partial \mathbf{V}}{\partial \sigma_a^2} \mathbf{P} \frac{\partial \mathbf{V}}{\partial \sigma_a^2} \right) & \text{tr} \left( \mathbf{P} \frac{\partial \mathbf{V}}{\partial \sigma_a^2} \mathbf{P} \frac{\partial \mathbf{V}}{\partial \sigma_d^2} \right) & \text{tr} \left( \mathbf{P} \frac{\partial \mathbf{V}}{\partial \sigma_a^2} \mathbf{P} \frac{\partial \mathbf{V}}{\partial \sigma_e^2} \right) \\ \text{tr} \left( \mathbf{P} \frac{\partial \mathbf{V}}{\partial \sigma_d^2} \mathbf{P} \frac{\partial \mathbf{V}}{\partial \sigma_a^2} \right) & \text{tr} \left( \mathbf{P} \frac{\partial \mathbf{V}}{\partial \sigma_d^2} \mathbf{P} \frac{\partial \mathbf{V}}{\partial \sigma_d^2} \right) & \text{tr} \left( \mathbf{P} \frac{\partial \mathbf{V}}{\partial \sigma_d^2} \mathbf{P} \frac{\partial \mathbf{V}}{\partial \sigma_e^2} \right) \\ \text{tr} \left( \mathbf{P} \frac{\partial \mathbf{V}}{\partial \sigma_e^2} \mathbf{P} \frac{\partial \mathbf{V}}{\partial \sigma_a^2} \right) & \text{tr} \left( \mathbf{P} \frac{\partial \mathbf{V}}{\partial \sigma_e^2} \mathbf{P} \frac{\partial \mathbf{V}}{\partial \sigma_d^2} \right) & \text{tr} \left( \mathbf{P} \frac{\partial \mathbf{V}}{\partial \sigma_e^2} \mathbf{P} \frac{\partial \mathbf{V}}{\partial \sigma_e^2} \right) \end{bmatrix} [6]$$

A less general expression is given by LYNCH and WALSH (1998, page 791). To calculate [6], the derivatives of  $\mathbf{V}$  with respect to the variance components are needed, and these are equal to:

$$\frac{\partial \mathbf{V}}{\partial \sigma_a^2} = \mathbf{Z}_1 \mathbf{A} \mathbf{Z}_1'; \quad \frac{\partial \mathbf{V}}{\partial \sigma_d^2} = \mathbf{Z}_2 \mathbf{D} \mathbf{Z}_2'; \quad \frac{\partial \mathbf{V}}{\partial \sigma_e^2} = \mathbf{I}_n \quad [7]$$

A problem with calculating [6] is its dependency on  $\mathbf{P}$ . To calculate  $\mathbf{P}$  requires inverting  $\mathbf{V}$  (of order  $n$ ) plus several other matrix multiplications of the same order. A more feasible algorithm can be obtained by expressing the elements of  $\mathbf{I}(\boldsymbol{\theta})$  as functions of matrices related to the MME. In doing so, advantage is taken of the relationship between terms such as  $\mathbf{Z}_i' \mathbf{P} \mathbf{Z}_j$  ( $i \neq j$ ,  $j = 1, 2$ ) and prediction error variance (HENDERSON, 1984) of  $\hat{\mathbf{a}}$  and  $\hat{\mathbf{d}}$ :  $\text{Var}[\hat{\mathbf{a}} - \mathbf{a}]$  and  $\text{Var}[\hat{\mathbf{d}} - \mathbf{d}]$ .

We now show how to calculate a diagonal element of  $\mathbf{I}(\boldsymbol{\theta})$  (element 1,1 or  $\mathbf{I}_{11}$ ). The derivation of an off-diagonal element is described in Appendix A, whereas Appendix B includes the derivation of  $\text{tr}(\mathbf{P})$ , which is needed for elements  $\mathbf{I}_{13}$  and  $\mathbf{I}_{23}$ , and  $\text{tr}(\mathbf{P}\mathbf{P})$  which is a multiple of  $\mathbf{I}_{33}$ . The proofs of the derivation of all remaining elements of  $\mathbf{I}(\boldsymbol{\theta})$  are available from request.

Using [5] and [7],  $\mathbf{I}_{11}$  is equal to  $1/2 \text{tr}(\mathbf{P}\mathbf{Z}_1\mathbf{A}\mathbf{Z}_1'\mathbf{P}\mathbf{Z}_1\mathbf{A}\mathbf{Z}_1')$ . By the use of formula (5.33) in page 46 of HENDERSON (1984) and a result proven in Appendix C, the following holds:

$$\mathbf{C}^{aa} \sigma_e^2 = \text{Var}[\hat{\mathbf{a}} - \mathbf{a}] = \sigma_a^2 \mathbf{A} - \mathbf{A}\mathbf{Z}_1'\mathbf{P}\mathbf{Z}_1\mathbf{A} (\sigma_a^2)^2 [8]$$

Algebraic manipulation of [8] produces:

$$\sigma_a^2 \mathbf{A} - \mathbf{C}^{aa} \sigma_e^2 = \mathbf{A}\mathbf{Z}_1'\mathbf{P}\mathbf{Z}_1\mathbf{A} (\sigma_a^2)^2$$

so that:

$$(\sigma_a^2)^2 \mathbf{A}^{-1} (\sigma_a^2 \mathbf{A} - \mathbf{C}^{aa} \sigma_e^2) \mathbf{A}^{-1} \mathbf{Z}_1' \mathbf{P} \mathbf{Z}_1 \quad [9]$$

Now, by substituting into  $\text{tr}(\mathbf{P}\mathbf{Z}_1\mathbf{A}\mathbf{Z}_1'\mathbf{P}\mathbf{Z}_1\mathbf{A}\mathbf{Z}_1')$  with [9] the following expression results

$$\frac{1}{2} \text{tr} \left[ (\sigma_a^2)^2 \mathbf{A}^{-1} (\mathbf{A} \sigma_a^2 - \mathbf{C}^{aa} \sigma_e^2) \mathbf{A}^{-1} (\sigma_a^2)^2 \mathbf{A} (\sigma_a^2)^2 \mathbf{A}^{-1} (\mathbf{A} \sigma_a^2 - \mathbf{C}^{aa} \sigma_e^2) \mathbf{A}^{-1} (\sigma_a^2)^2 \mathbf{A} \right] [10]$$

We can now write element  $\mathbf{I}_{11}$  as:

$$\frac{1}{2} \text{tr} \left[ (\mathbf{I} - \mathbf{A}^{-1} \mathbf{C}^{aa} \alpha) (\sigma_a^2)^2 (\mathbf{I} - \mathbf{A}^{-1} \mathbf{C}^{aa} \alpha) (\sigma_a^2)^2 \right] =$$

and, after expanding the trace operator, we finally obtain:

$$\mathbf{I}_{11} = \frac{\left[ q - 2 \text{tr}(\mathbf{A}^{-1} \mathbf{C}^{aa}) + \text{tr}(\mathbf{A}^{-1} \mathbf{C}^{aa} \mathbf{A}^{-1} \mathbf{C}^{aa} (\sigma_e^2 / \sigma_a^2)^2) \right]}{(2 \sigma_a^2)^2} [11]$$

Note that [11] does not require inverting  $\mathbf{V}$  which is of order equal to the number of observations. For the purpose of calculation, the REML estimates of the variance components replace the true values of the dispersion parameters.

## DATA

Data used in this study were originated from the national maize program (INTA, National Institute for Agricultural Technology, Pergamino Experimental Station, Buenos Aires province), and were collected from 1988 to 1990. The genetically broad-based maize populations used to start the program were BS13, a derivative of the Iowa Stiff Stalk Synthetic maize population, one maize composite line and one maize synthetic line from the germplasm collection of INTA: CII and S34. From 1989 to 1995, individual genotypes were propagated through successive generations by self-pollination (see Table 1), such as to obtain 32 inbred lines. Finally, each of these lines was crossed to either SB73, a dent tester derived from B73, or LP521, or LP611. The last two testers are flint lines developed by INTA Pergamino. Thus, the total number of individuals in the pedigree file was 169: 3 testers, 3 original lines, 5  $S_0$  genotypes, 22  $S_1$ , 23  $S_2$ , 24  $S_3$ , 25  $S_4$ , 32  $S_5$ , and 32 hybrids. Phenotypic data corresponded only to the hybrids, and they were tested in 1996 and 1997.

The experimental design used was an  $8 \times 8$  lattice with 3 replicates. The trial was repeated in four locations close to Pergamino, in northwestern Buenos Aires province. A plot consisted of two rows, 5.00 m long with 0.70 m between rows. All plots were over planted by machine and thinned to a uniform stand density of approximately 62124 plants  $\text{ha}^{-1}$  at the five-leaf stage. All yield trials were machine cultivated and/or hand weeded as neces-

TABLE 1 - Lines, testers and hybrids used during the program.

Initial lines	S <sub>0</sub> 1989/90	S <sub>1</sub> 1990/91	S <sub>2</sub> 1991/92	S <sub>3</sub> 1992/93	S <sub>4</sub> 1993/94	S <sub>5</sub> 1994/95	Testers	Trials 1996/97
BS13	BS13							
CII	CII-35 CII-47 CII-102	22 lines	23 lines	24 lines	25 lines	32 lines	LP521, SB73, LP611	32 hybrids
S34	S34-182 S34-228							

sary. Plots were hand harvested. The response variable was grain yield (in tn ha<sup>-1</sup>) analyzed on a 15% grain moisture basis.

The elements of the  $\mathbf{A}$  and  $\mathbf{D}$  matrices were calculated using standard formulae (for example, LYNCH and WALSH, 1998, page 763), whereas  $\mathbf{A}^{-1}$  was calculated using Henderson's simple rules (HENDERSON, 1976) and  $\mathbf{D}^{-1}$  was obtained by direct inversion of  $\mathbf{D}$ .

## RESULTS

### Variance component estimation

The REML estimates (in (tn/ha)<sup>2</sup>) were:  $\hat{\sigma}_e^2 = 0.631$ ;  $\hat{\sigma}_a^2 = 0.132$ ;  $\hat{\sigma}_d^2 = 0.176$ ; the ratio between dominance and estimated total genetic variance  $\hat{\sigma}_g^2 = \hat{\sigma}_a^2 + \hat{\sigma}_d^2$  was  $\hat{\sigma}_d^2/\hat{\sigma}_g^2 = 0.56$ . The estimated heritability ( $\hat{h}^2 = \hat{\sigma}_a^2/(\hat{\sigma}_a^2 + \hat{\sigma}_d^2 + \hat{\sigma}_e^2)$ ) for grain yield was equal to 0.14. BERNARDO (1994) estimated  $\hat{\sigma}_d^2/\hat{\sigma}_g^2 = 0.12$  and  $\hat{h}^2 = 0.81$  using REML with a similar model to the one presented here. Later (BERNARDO, 1996), and using 332 lines belonging to seven heterotic groups, he obtained the following estimates:  $\hat{\sigma}_a^2 = 0.512$  to 0.193, and  $\hat{\sigma}_d^2 = 0.107$  to 0.181. The ratio  $\hat{\sigma}_d^2/\hat{\sigma}_g^2$  varied from 0.4 to 0.75. Comparison with estimates obtained using other assumptions to the ones used here (inclusion of  $\mathbf{A}$  and  $\mathbf{D}$ ), such as those that involve quadratic type estimators (*i.e.* ANOVA), should be done with caution. For example, BOCA *et al.* (2002) found that estimates of  $\hat{h}^2$  were higher when using relationships among genotypes, than when those relationships were ignored. The mean of 99 estimates of  $\hat{h}^2$  reviewed by HALLAUER and MIRANDA (1988) was 0.18.

### Asymptotic variances of REML estimators of variance components

The asymptotic variances of REML estimates of  $\hat{\sigma}_a^2$ ,  $\hat{\sigma}_d^2$  and  $\hat{\sigma}_e^2$  were calculated using the formulae obtained in the previous section. The calculated values were: asymptotic variance of  $\hat{\sigma}_a^2 = 0.077$ ; asymptotic variance of  $\hat{\sigma}_d^2 = 0.025$ , asymptotic variance of  $\hat{\sigma}_e^2 = 0.0045$ , asymptotic covariance ( $\hat{\sigma}_a^2, \hat{\sigma}_d^2$ ) =  $-0.037$ , asymptotic covariance ( $\hat{\sigma}_a^2, \hat{\sigma}_e^2$ ) =  $-0.00055$ , and the asymptotic covariance ( $\hat{\sigma}_d^2, \hat{\sigma}_e^2$ ) =  $-0.00072$ . In terms of correlations ( $r$ ), the covariances led to  $r(\hat{\sigma}_a^2, \hat{\sigma}_d^2) = -0.82$ ,  $r(\hat{\sigma}_a^2, \hat{\sigma}_e^2) = -0.29$  and  $r(\hat{\sigma}_d^2, \hat{\sigma}_e^2) = -0.067$ . The magnitude of the correlations for the estimates of the genetic variance components ( $\hat{\sigma}_a^2$  and  $\hat{\sigma}_d^2$ ) with  $\hat{\sigma}_e^2$  are low. However, the high value of  $r(\hat{\sigma}_a^2, \hat{\sigma}_e^2)$  suggests that the observed data structure does not allow a precise joint estimation of  $\hat{\sigma}_a^2$  and  $\hat{\sigma}_d^2$ .

## DISCUSSION

The research discussed here complements the methodological developments presented by BERNARDO (1994, 1996), by providing formulae for the asymptotic variances of REML estimates of additive and dominance variance components in terms of matrices associated with HENDERSON'S (1984) 'mixed model equations'. Expressions for the information matrix such as [6] or the one in page 791 in LYNCH and WALSH (1998), require the inversion of the variance-covariance matrix of records  $\mathbf{V}$  (of order  $n$ ) plus several matrix multiplications, in order to calculate  $\mathbf{P}$  (which is also of order  $n$ ). The advantage of our formulation is the removal of the dependence on the number of records to a function of the

number of genotypes plus the number of trials or block effects:  $p + 2q$ , the order of the MME. As a consequence, the calculation becomes feasible for large data sets.

As in BERNARDO (1994, 1996), the mixed model presented here includes additive and dominance genetic effects, to predict general and specific combining abilities in maize by using additive (by means of matrix  $\mathbf{A}$ ) and dominance relationships (by means of matrix  $\mathbf{D}$ ). The difference between the model of the present research with the one used by BERNARDO (1994), is in his decomposition of the additive effects due to parental origin of the genes. In addition and in the second paper, BERNARDO (1996) included check hybrid effects. This differentiation is needed to take into account linkage disequilibrium of the parental populations (MELCHINGER, 1988), and it requires additional additive variances to estimate. As indicated by a reviewer, plant breeders rarely assume that variances are equal in different reference populations. In case more additive variances are included in the model, the diagonal element in  $\mathbf{I}(\boldsymbol{\theta})$  for the  $i^{th}$  additive variance can be seen to be equal to:

$$\frac{\left[ q_i - 2 \operatorname{tr} \left( \mathbf{A}_i^{-1} \mathbf{C}^{a_i} \frac{\sigma_a^2}{\sigma_a} \right) + \operatorname{tr} \left( \mathbf{A}_i^{-1} \mathbf{C}^{a_i} \mathbf{A}_i^{-1} \mathbf{C}^{a_i} \left( \frac{\sigma_a^2}{\sigma_a} \right)^2 \right) \right]}{(2\sigma_a^2)^2}$$

However, to obtain reliable parameter estimates of different genetic variance components requires more hybrids than the ones used in the current study. In this sense, having an estimate of the asymptotic variance of the dispersion parameters allows evaluating whether the REML estimates are reasonably accurate, as both the variance of the individual estimates and the covariance between them are taken into account. Either the models used by BERNARDO (1994, 1996), or the one employed here, take into account that genetic effects of related genotypes are not independent. In breeding programs of self-pollinated species to obtain pure lines for crossing, the additive relationships among selected materials at the end of the program are known. However, if linear models with fixed effects are employed to calculate combining abilities, additive and dominance relationships are not taken into account. As observed by BERNARDO (1994, 1996), mixed model methodology is a useful and precise tool to predict combining abilities in maize. By using REML under properly specified covariance matrices  $\mathbf{A}$  and  $\mathbf{D}$ , the method recovers the variance

components in the base population prior to selection (SORENSEN and KENNEDY, 1984).

The equivalent sampling expressions for the variance of the REML estimates of the variance components do not exist, as the estimating equations are iterative non-linear functions of the estimators. The REML estimates are asymptotically normal for breeding applications involving relationship matrices, as indicated by CRESSIE and LAHIRI (1993, remark 4.1., page 224). In this case, twice the inverse of the REML information matrix (HARVILLE, 1977) is the asymptotic covariance matrix of the REML estimators (JIANG, 1996). CRESSIE and LAHIRI (1993) stated that REML estimators of a variance components model are asymptotically normal if: 1) the information matrix exists; and 2) the eigenvalues of the matrices related to each variance component (matrices  $\mathbf{Z}_1 \mathbf{A} \mathbf{Z}_1'$  and of  $\mathbf{Z}_2 \mathbf{A} \mathbf{Z}_2'$  in model [1]-[2]) are all positive real numbers. The first is an existence condition, whereas the second one is proved for model [1]-[2] in Appendix D. The information matrix has been previously used in a tree breeding context by McCUTCHAN *et al.* (1985), to compare the efficiency of different designs to estimate heritability. However, their expressions do not allow for relationships among families, *i.e.* they assumed  $\mathbf{A} = \mathbf{I}$  and  $\mathbf{D} = \mathbf{I}$ . Besides obtaining the asymptotic variances of REML estimates, the analysis produces asymptotic covariances which discriminate whether the analysis allows for estimating pairs of dispersion parameters in a precise manner. As an example, the values obtained with the Argentinean maize data suggest that the estimates of  $\sigma_a^2$  and of  $\sigma_d^2$  were highly correlated, although both estimates were not correlated to the estimated error variance  $\sigma_e^2$ . Therefore, the design was not optimal to separate the information for estimating  $\sigma_a^2$  and  $\sigma_d^2$  simultaneously.

The model discussed in the present research focused on the covariance between genetic effects. However, it can include random environmental effects. For example, fertility trend as discussed by DURBAN *et al.* (2001). Let the covariance matrix of this random fertility effect to be equal to  $\sigma_s^2 \mathbf{G}$  (as described in their expression (11)). In a similar manner as it was derived  $\mathbf{I}_{11}$ , the diagonal element of  $\mathbf{I}(\boldsymbol{\theta})$  corresponding to  $\sigma_s^2$  can be seen to be equal to:

$$\frac{\left[ n - 2 - 2 \operatorname{tr} \left( \mathbf{G}^{-1} \mathbf{C}^{SS} \frac{\sigma_s^2}{\sigma_s} \right) + 2 \operatorname{tr} \left( \mathbf{G}^{-1} \mathbf{C}^{SS} \mathbf{G}^{-1} \mathbf{C}^{SS} \left( \frac{\sigma_s^2}{\sigma_s} \right)^2 \right) \right]}{(2\sigma_s^2)^2}$$

where  $\mathbf{C}^{SS}$  is that portion of the inverse of the coef-



cient matrix of the MME associated with the random fertility effect. By a similar reasoning, genotype by environment interactions can be included in the model. PIEPHO *et al.* (2003) observed that “ANOVA-type mixed models cannot always satisfactorily model interactions”, and then reviewed several proposed mixed models with individual genetic effects as in the framework of the current research. For example, FRENSHAM *et al.* (1998) considered the following model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\xi} + \mathbf{e}$$

The vectors  $\mathbf{y}$ ,  $\boldsymbol{\beta}$  and  $\mathbf{e}$  and the matrices  $\mathbf{X}$  and  $\mathbf{Z}$  are as in [1], while  $\mathbf{u}$  are random genotypic effects and  $\boldsymbol{\xi}$  are random  $G \times E$  effects. The specification is completed with  $E(\boldsymbol{\xi}) = \mathbf{0}$  and  $\text{Var}(\boldsymbol{\xi}) = \boldsymbol{\Omega}$ . Using [5] the element of  $\mathbf{I}(\boldsymbol{\theta})$  for the dispersion parameter  $\boldsymbol{\Omega}_{ij}$  is equal to

$$[\mathbf{I}(\boldsymbol{\theta})]_{ij} = \frac{1}{2} \text{tr} \left( \mathbf{P} \frac{\partial \boldsymbol{\Omega}}{\partial \boldsymbol{\Omega}_i} \mathbf{P} \frac{\partial \boldsymbol{\Omega}}{\partial \boldsymbol{\Omega}_j} \right)$$

Depending on the structure of  $\boldsymbol{\Omega}$ , the above formula can be expressed in terms of the inverse of the coefficient matrix of the mixed model equations, in a similar fashion as it was done for model [1]-[2].

## APPENDIX A

### Expressions for elements of $\mathbf{I}(\boldsymbol{\theta})$ in terms of matrices which are functions of the MME

Consider  $\mathbf{I}_{12} = 1/2 \text{tr}(\mathbf{PZ}_1\mathbf{AZ}_1'\mathbf{PZ}_2\mathbf{DZ}_2')$  and the following equalities for prediction error variance of dominance effects:

$$\mathbf{C}^{dd} \sigma_e^2 = \text{Var}[\hat{\mathbf{d}} - \mathbf{d}] = \mathbf{D} \sigma_d^2 - \sigma_d^2 \mathbf{D} \mathbf{Z}_2' \mathbf{P} \mathbf{Z}_2 \mathbf{D} \sigma_d^2 \quad [\text{A.1}]$$

Expression [A.1] allows us to write

$$\mathbf{D} \sigma_d^2 - \mathbf{C}^{dd} \sigma_e^2 = \mathbf{D} \sigma_d^2 \mathbf{Z}_2' \mathbf{P} \mathbf{Z}_2 \mathbf{D} \sigma_d^2$$

so that

$$\mathbf{D}^{-1} \frac{1}{\sigma_d^2} (\mathbf{D} \sigma_d^2 - \mathbf{C}^{dd} \sigma_e^2) \mathbf{D}^{-1} \frac{1}{\sigma_d^2} = \mathbf{Z}_2' \mathbf{P} \mathbf{Z}_2 \sigma_d^2 \quad [\text{A.2}]$$

To obtain an expression for  $\mathbf{I}_{12}$  in terms of  $\mathbf{C}^{aa}$  and  $\mathbf{C}^{dd}$ , replace with [9] and [A.2] in [A.1]:

$$\frac{1}{2} \text{tr} \left[ (\mathbf{A}^{-1} \sigma_a^2 (\mathbf{A} \sigma_a^2 - \mathbf{C}^{aa} \sigma_e^2) \mathbf{A}^{-1} \sigma_a^2) \mathbf{A} (\mathbf{D}^{-1} \sigma_d^2 (\mathbf{D} \sigma_d^2 - \mathbf{C}^{dd} \sigma_e^2) \mathbf{D}^{-1} \sigma_d^2) \mathbf{D} \right]$$

which, after taking traces results in:

$$\begin{aligned} & \frac{1}{2} \text{tr} \left[ \left( (\mathbf{I}_q - \mathbf{A}^{-1} \mathbf{C}^{aa} \frac{\sigma_e^2}{\sigma_a^2}) \mathbf{A}^{-1} \sigma_a^2 \right) \mathbf{A} \left( (\mathbf{I}_q - \mathbf{D}^{-1} \mathbf{C}^{dd} \frac{\sigma_e^2}{\sigma_d^2}) \mathbf{D}^{-1} \sigma_d^2 \right) \mathbf{D} \right] = \\ & \frac{1}{2} \text{tr} \left[ (\mathbf{I}_q - \mathbf{A}^{-1} \mathbf{C}^{aa} \alpha) \sigma_a^2 (\mathbf{I}_q - \mathbf{D}^{-1} \mathbf{C}^{dd} \delta) \sigma_d^2 \right] \end{aligned}$$

or

$$\mathbf{I}_{12} = \frac{1}{2 \sigma_a^2 \sigma_d^2} \left[ q - \text{tr}(\mathbf{A}^{-1} \mathbf{C}^{aa}) \alpha - \text{tr}(\mathbf{D}^{-1} \mathbf{C}^{dd}) \delta + \text{tr}(\mathbf{A}^{-1} \mathbf{C}^{aa} \mathbf{D}^{-1} \mathbf{C}^{dd}) \alpha \delta \right]$$

## APPENDIX B

### Expressions for the traces of $\mathbf{P}$ and $\mathbf{PP}$ in terms of matrices related to the MME

SEARLE (1979) expressed  $\mathbf{P}$  as

$$\mathbf{P} = \left[ \mathbf{M} - \mathbf{M} \mathbf{Z} (\mathbf{Z}' \mathbf{M} \mathbf{Z} + \mathbf{G}_{AD}^{-1})^{-1} \mathbf{Z}' \mathbf{M} \right] \sigma_e^{-2} \quad [\text{B.1}]$$

where:

$$\mathbf{M} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}'; \quad \mathbf{G}_{AD}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} \alpha & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^{-1} \delta \end{bmatrix}; \quad \mathbf{Z} = [\mathbf{Z}_1 | \mathbf{Z}_2] \quad [\text{B.2}]$$

After taking traces [B.1] results in:

$$\text{tr}(\mathbf{P}) = \left[ \text{tr}(\mathbf{M}) - \text{tr}(\mathbf{Z}' \mathbf{M} \mathbf{Z} + \mathbf{G}_{AD}^{-1})^{-1} \mathbf{Z}' \mathbf{M} \mathbf{Z} \right] \sigma_e^{-2} \quad [\text{B.3}]$$

Using [B.2] in [B.3] and as  $\text{tr}(\mathbf{I}_n) = n$  and  $\text{tr}[\mathbf{X}(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}'] = \text{rank}[\mathbf{X}] = p$ , we have:

$$\begin{aligned} \text{tr}(\mathbf{P}) &= \left[ n - p - \text{tr} \left[ (\mathbf{Z}' \mathbf{M} \mathbf{Z} + \mathbf{G}_{AD}^{-1})^{-1} (\mathbf{Z}' \mathbf{M} \mathbf{Z} + \mathbf{G}_{AD}^{-1} - \mathbf{G}_{AD}^{-1}) \right] \right] \sigma_e^{-2} = \\ \text{tr}(\mathbf{P}) &= \left[ n - p - \text{tr}(\mathbf{I}_{2q}) + \text{tr} \begin{bmatrix} \mathbf{C}^{aa} & \mathbf{C}^{ad} \\ \mathbf{C}^{da} & \mathbf{C}^{dd} \end{bmatrix} \begin{bmatrix} \mathbf{A}^{-1} \alpha & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^{-1} \delta \end{bmatrix} \right] \sigma_e^{-2} = \\ \text{tr}(\mathbf{P}) &= \left[ n - p - 2q + \text{tr}(\mathbf{A}^{-1} \mathbf{C}^{aa}) \alpha + \text{tr}(\mathbf{D}^{-1} \mathbf{C}^{dd}) \delta \right] \sigma_e^{-2} \end{aligned}$$

Using [B.1],  $\text{tr}(\mathbf{PP})$  is equal to:

$$\begin{aligned} \text{tr}(\mathbf{PP}) &= \text{tr} \left[ \left( \mathbf{M} - \mathbf{M} \mathbf{Z} \begin{bmatrix} \mathbf{C}^{aa} & \mathbf{C}^{ad} \\ \mathbf{C}^{da} & \mathbf{C}^{dd} \end{bmatrix} \mathbf{Z}' \mathbf{M} \right) \right. \\ & \quad \left. \left( \mathbf{M} - \mathbf{M} \mathbf{Z} \begin{bmatrix} \mathbf{C}^{aa} & \mathbf{C}^{ad} \\ \mathbf{C}^{da} & \mathbf{C}^{dd} \end{bmatrix} \mathbf{Z}' \mathbf{M} \right) \right] (\sigma_e^2)^{-2}. \end{aligned} \quad [\text{C.1}]$$

[C.1] can be expressed as:

$$\left[ \begin{aligned} & \text{tr}[\mathbf{M}] - 2 \text{tr} \left[ \mathbf{Z}' \mathbf{M} \mathbf{Z} \begin{bmatrix} \mathbf{C}^{aa} & \mathbf{C}^{ad} \\ \mathbf{C}^{da} & \mathbf{C}^{dd} \end{bmatrix} \sigma_e^2 \right] + \\ & \text{tr} \left[ \mathbf{Z}' \mathbf{M} \mathbf{Z} \begin{bmatrix} \mathbf{C}^{aa} & \mathbf{C}^{ad} \\ \mathbf{C}^{da} & \mathbf{C}^{dd} \end{bmatrix} \mathbf{Z}' \mathbf{M} \mathbf{Z} \begin{bmatrix} \mathbf{C}^{aa} & \mathbf{C}^{ad} \\ \mathbf{C}^{da} & \mathbf{C}^{dd} \end{bmatrix} (\sigma_e^2)^2 \right] \end{aligned} \right] (\sigma_e^2)^{-2} =$$

The second terms in [C.2]: can be expressed as:

$$\text{tr} \left[ \left( \mathbf{Z}' \mathbf{M} \mathbf{Z} + \begin{bmatrix} \mathbf{A}^{-1} \sigma_a^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^{-1} \sigma_d^2 \end{bmatrix} - \begin{bmatrix} \mathbf{A}^{-1} \sigma_a^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^{-1} \sigma_d^2 \end{bmatrix} \right) \begin{bmatrix} \mathbf{C}^{aa} & \mathbf{C}^{ad} \\ \mathbf{C}^{da} & \mathbf{C}^{dd} \end{bmatrix} \right] (\sigma_e^2)^{-2} =$$

but

$$\mathbf{Z}'\mathbf{M}\mathbf{Z} + \begin{bmatrix} \mathbf{A}^{-1}\alpha & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^{-1}\delta \end{bmatrix} = \begin{bmatrix} \mathbf{C}^{aa} & \mathbf{C}^{ad} \\ \mathbf{C}^{da} & \mathbf{C}^{dd} \end{bmatrix}^{-1} \sigma_e^2$$

Then [C.3] is equal to

$$\begin{aligned} & \text{tr} \left[ \mathbf{I}_q - \begin{bmatrix} \mathbf{A}^{-1}\alpha & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^{-1}\delta \end{bmatrix} \begin{bmatrix} \mathbf{C}^{aa} & \mathbf{C}^{ad} \\ \mathbf{C}^{da} & \mathbf{C}^{dd} \end{bmatrix} \right] \sigma_e^2 = \\ & = [2q - \text{tr}(\mathbf{A}^{-1}\mathbf{C}^{aa})\alpha - \text{tr}(\mathbf{D}^{-1}\mathbf{C}^{dd})\delta] \sigma_e^2 \end{aligned}$$

The last term in [C.2] can be expressed as:

$$\text{tr} \left( \begin{bmatrix} \mathbf{Z}'\mathbf{M}\mathbf{Z} + \begin{bmatrix} \mathbf{A}^{-1}\alpha & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^{-1}\delta \end{bmatrix} - \begin{bmatrix} \mathbf{A}^{-1}\alpha & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^{-1}\delta \end{bmatrix} \begin{bmatrix} \mathbf{C}^{aa} & \mathbf{C}^{ad} \\ \mathbf{C}^{da} & \mathbf{C}^{dd} \end{bmatrix} \right) =$$

Replacing with the equality [C.4], and using [C.6] in the term [C.5] and after some algebra, we have:

$$\begin{aligned} & \text{tr} \left( \begin{bmatrix} \mathbf{I}_{2q} - \begin{bmatrix} \mathbf{A}^{-1}\alpha & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^{-1}\delta \end{bmatrix} \begin{bmatrix} \mathbf{C}^{aa} & \mathbf{C}^{ad} \\ \mathbf{C}^{da} & \mathbf{C}^{dd} \end{bmatrix} \right) \\ & \quad \left( \begin{bmatrix} \mathbf{I}_{2q} - \begin{bmatrix} \mathbf{A}^{-1}\alpha & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^{-1}\delta \end{bmatrix} \begin{bmatrix} \mathbf{C}^{aa} & \mathbf{C}^{ad} \\ \mathbf{C}^{da} & \mathbf{C}^{dd} \end{bmatrix} \right) = \\ & \text{tr}(\mathbf{I}_{2q}) - 2 \left[ \text{tr}(\mathbf{A}^{-1}\mathbf{C}^{aa})\alpha + \text{tr}(\mathbf{D}^{-1}\mathbf{C}^{dd})\delta \right] + \text{tr}[\mathbf{A}^{-1}\mathbf{C}^{aa}\mathbf{A}^{-1}\mathbf{C}^{aa}]\alpha^2 \\ & + 2 \text{tr}[\mathbf{D}^{-1}\mathbf{C}^{da}\mathbf{A}^{-1}\mathbf{C}^{ad}]\alpha\delta + \text{tr}[\mathbf{D}^{-1}\mathbf{C}^{dd}\mathbf{D}^{-1}\mathbf{C}^{dd}]\delta^2 = \\ & \frac{1}{(\sigma_e^2)^2} \left[ 2q - 2 \left[ \text{tr}(\mathbf{A}^{-1}\mathbf{C}^{aa})\alpha + \text{tr}(\mathbf{D}^{-1}\mathbf{C}^{dd})\delta \right] + \text{tr}[\mathbf{A}^{-1}\mathbf{C}^{aa}\mathbf{A}^{-1}\mathbf{C}^{aa}]\alpha^2 \right. \\ & \left. + 2 \text{tr}[\mathbf{D}^{-1}\mathbf{C}^{da}\mathbf{A}^{-1}\mathbf{C}^{ad}]\alpha\delta + \text{tr}[\mathbf{D}^{-1}\mathbf{C}^{dd}\mathbf{D}^{-1}\mathbf{C}^{dd}]\delta^2 \right] = \end{aligned}$$

The final expression for  $\text{tr}(\mathbf{PP})$  is equal to:

$$\begin{aligned} \text{tr}(\mathbf{PP}) &= \text{tr}[\mathbf{M}] - 2 \left[ q - \text{tr}(\mathbf{A}^{-1}\mathbf{C}^{aa})\alpha - \text{tr}(\mathbf{D}^{-1}\mathbf{C}^{dd})\delta \right] \left( \frac{1}{\sigma_e^2} \right)^2 + \\ & \frac{1}{(\sigma_e^2)^2} \left[ 2q - 2 \left[ \text{tr}(\mathbf{A}^{-1}\mathbf{C}^{aa})\alpha + \text{tr}(\mathbf{D}^{-1}\mathbf{C}^{dd})\delta \right] + \text{tr}[\mathbf{A}^{-1}\mathbf{C}^{aa}\mathbf{A}^{-1}\mathbf{C}^{aa}]\alpha^2 \right. \\ & \left. + 2 \text{tr}[\mathbf{D}^{-1}\mathbf{C}^{da}\mathbf{A}^{-1}\mathbf{C}^{ad}]\alpha\delta + \text{tr}[\mathbf{D}^{-1}\mathbf{C}^{dd}\mathbf{D}^{-1}\mathbf{C}^{dd}]\delta^2 \right] = \\ \text{tr}(\mathbf{PP}) &= \frac{1}{(\sigma_e^2)^2} \left[ n - p - 2q - \text{tr}[\mathbf{A}^{-1}\mathbf{C}^{aa}\mathbf{A}^{-1}\mathbf{C}^{aa}]\alpha^2 \right. \\ & \left. + 2 \text{tr}[\mathbf{D}^{-1}\mathbf{C}^{da}\mathbf{A}^{-1}\mathbf{C}^{ad}]\alpha\delta + \text{tr}[\mathbf{D}^{-1}\mathbf{C}^{dd}\mathbf{D}^{-1}\mathbf{C}^{dd}]\delta^2 \right] = \end{aligned}$$

## APPENDIX C

### Derivation of $\text{Var}[\hat{\mathbf{a}} - \mathbf{a}]$

Using expression (5.18e) in HENDERSON (1984, page 43) prediction error variance of  $\hat{\mathbf{a}}$  is  $\text{Var}[\hat{\mathbf{a}} - \mathbf{a}] = \text{Var}(\mathbf{a}) - \text{Var}\hat{\mathbf{a}}$  (also  $\hat{\mathbf{a}} = \text{cov}(\mathbf{a}, \mathbf{y})[\text{Var}(\mathbf{y})]^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$  (SEARLE *et al.*, 1993, equation (45) in page

273). Now,  $\text{Var}(\mathbf{a}) = \mathbf{A} \sigma_a^2$  and the BLUP of  $\mathbf{a}$  is equal to

$$\hat{\mathbf{a}} = \sigma_a^2 \mathbf{A} \mathbf{Z}_1' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \sigma_a^2 \mathbf{A} \mathbf{Z}_1' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}) = \sigma_a^2 \mathbf{A} \mathbf{Z}_1' \mathbf{V}^{-1} \mathbf{P} \mathbf{y}$$

Therefore

$$\text{Var}(\hat{\mathbf{a}}) = \text{Var}(\sigma_a^2 \mathbf{A} \mathbf{Z}_1' \mathbf{V}^{-1} \mathbf{P} \mathbf{y}) = \sigma_a^2 \mathbf{A} \mathbf{Z}_1' \mathbf{V}^{-1} \mathbf{P} \mathbf{V} \mathbf{P} \mathbf{V}^{-1} \mathbf{Z}_1 \mathbf{A} \sigma_a^2$$

As  $\mathbf{PVP} = \mathbf{P}$  (SEARLE *et al.*, 1993, page 274), we have that  $\text{Var}(\hat{\mathbf{a}}) = \mathbf{A} \mathbf{Z}_1' \mathbf{V}^{-1} \mathbf{PVPV}^{-1} \mathbf{Z}_1 \mathbf{A} (\sigma_a^2)^2$ .

Finally, to obtain (8) replace with the last expression into  $\text{Var}[\hat{\mathbf{a}} - \mathbf{a}]$ :

$$\text{Var}[\hat{\mathbf{a}} - \mathbf{a}] = \mathbf{A} \sigma_a^2 - \mathbf{A} \mathbf{Z}_1' \mathbf{V}^{-1} \mathbf{PVPV}^{-1} \mathbf{Z}_1 \mathbf{A} (\sigma_a^2)^2$$

## APPENDIX D

### Eigenvalues of $\mathbf{Z}_1 \mathbf{A} \mathbf{Z}_1'$

The second condition for the REML estimators of the variance components to be asymptotically normal in theorem 4.1 of CRESSIE and LAHIRI (1993), is that the smallest eigenvalues of  $\mathbf{Z}_1 \mathbf{A} \mathbf{Z}_1'$  and  $\mathbf{Z}_2 \mathbf{D} \mathbf{Z}_2'$  are greater than 0, *i.e.* positive. We now show that all eigenvalues of  $\mathbf{Z}_1 \mathbf{A} \mathbf{Z}_1'$  are positive. Similar results can be obtained for  $\mathbf{Z}_2 \mathbf{D} \mathbf{Z}_2'$ .

Consider first the  $n \times q$  matrix  $\mathbf{Z}_1$  and the square positive definite matrix  $\mathbf{A}$  of order  $q$ . Without loss of generality we can reorder the genotypes in  $\mathbf{a}$  such as  $(q - t)$  genotypes without phenotypic records precede those of the  $t$  genotypes with records (in our case, the hybrids). Then,  $\mathbf{Z}_1 = [\mathbf{0} \ \mathbf{D}]$ , with matrix  $\mathbf{0}$  having its  $n \times (q - t)$  elements equal to 0. Matrix  $\mathbf{D}$  is diagonal of order  $t$  and all diagonal elements equal to  $n_i$  (number of observations for the  $i^{th}$ -genotype,  $i = 1, \dots, t$ ). A similar partition can be practiced in  $\mathbf{A}$  so that

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{NN} & \mathbf{A}_{NO} \\ \mathbf{A}_{ON} & \mathbf{A}_{OO} \end{bmatrix}$$

where subindexes  $N$  and  $O$  stand for 'non-observed phenotype' and 'observed phenotype', respectively. Pre and postmultiplying  $\mathbf{A}$  by  $\mathbf{Z}_1$  and its transpose respectively, produces:

$$\mathbf{Z}_1 \mathbf{A} \mathbf{Z}_1' = [\mathbf{0} \ \mathbf{D}] \begin{bmatrix} \mathbf{A}_{NN} & \mathbf{A}_{NO} \\ \mathbf{A}_{ON} & \mathbf{A}_{OO} \end{bmatrix} \begin{bmatrix} \mathbf{0}' \\ \mathbf{D}' \end{bmatrix} = \mathbf{D} \mathbf{A}_{OO} \mathbf{D} \quad [1]$$

Matrix  $\mathbf{D}$  has an inverse as its diagonal elements are  $n_i \geq 1$ . Also,  $\text{Var}(\mathbf{a}_O) = \mathbf{A}_{OO} \sigma_a^2$  is the covariance matrix of the general combining abilities of the lines with phenotype recorded,  $\mathbf{A}_{OO}$  is positive definite.

Therefore, by theorem A.9 (page 185) in TOUTENBURG (1982),  $\mathbf{DA}_{OO}\mathbf{D}$  is positive definite and its eigenvalues are all positive.

ACKNOWLEDGEMENTS - Thanks to two anonymous reviewers whose comments greatly improved the original version. Any mistake is responsibility of the authors. This research was financed with grants from Secretaría de Ciencia y Técnica, UBA, UBACyT G035; and Consejo Nacional de Investigaciones Científicas y Técnicas PIP 0934-98. Thanks are also given to Dr. G. Eyherabide and INTA for allowing us to analyse their data.

## REFERENCES

- BERNARDO R., 1994 Prediction of maize single cross performance using information from related hybrids. *Crop Sci.* **34**: 20-25.
- BERNARDO R., 1996 Best linear prediction of maize single cross performance. *Crop Sci.* **36**: 50-56.
- BOCA R.T., R.J.C. CANTET, C.G. LOPEZ, 2002 Efectos de la incorporación de las relaciones de parentesco sobre la estimación REML de la heredabilidad para rendimiento en soja [*Glycine max* (L.) Merr.]. *Rev. Invest. Agr.* **17**: 67-76.
- CANTET R.J.C., D.GIANOLA, I. MISZTAL, R.L. FERNANDO, 1993 Estimates of dispersion parameters and genetic and environmental trends for weaning weight in Angus cattle using a maternal animal model with genetic grouping. *Livestock Prod. Sci.* **34**: 203-212.
- COCKERHAM C.C., 1954 An extension of the concept of partitioning hereditary variance for the analysis of covariance between relatives when epistasis is present. *Genetics* **39**: 859-882.
- CRESSIE N., S.N. LAHIRI, 1993 The asymptotic distribution of REML estimators. *J. Multivariate Anal.* **45**: 217-233.
- DEMPSTER A., N. LAIRD, D. RUBIN, 1977 Maximum likelihood estimation from incomplete data using the EM algorithm. *J. Royal Statistical Soc.* **39**: 1-38.
- DURBAN M., I.D. CURRIE, R.A. KEMPTON, 2001 Adjusting for fertility and competition in variety trials. *J. Agric.Sci. Cambridge* **136**: 129-140.
- FRENHAM A., B. CULLIS, A. VERBYLA, 1997 Genotype by environment variance heterogeneity in a two-stage analysis. *Biometrics* **53**: 1373-1383.
- HALLAUER A.R., J.B. MIRANDA, 1988 Quantitative genetics in maize breeding. Iowa State University Press. Ames, IA, USA.
- HARVILLE D.A., 1977 Maximum-Likelihood approaches to variance component estimation and related problems. *J. Am. Statistical Assoc.* **72**: 320-340.
- HARVILLE D.A., 1985 Decomposition of prediction error. *J. Am. Statistical Assoc.* **80**: 132-138.
- HENDERSON C.R., 1976 A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding value. *Biometrics* **32**: 69-83.
- HENDERSON C.R., 1984 Applications of linear models in animal breeding. Univ. of Guelph, Guelph, ON, Canada.
- JIANG J., 1996 REML Estimation asymptotic behavior and related topics. *Ann. Statistics* **24**: 225-286.
- KACKAR R.N., D.A. HARVILLE, 1981 Unbiasedness of two-stage estimation and prediction procedures for mixed linear models. *Communication in Statistics. Theory and Methods* **10**: 1249-1261.
- KEMPTHORNE O., 1954 The correlation between relatives in a random mating population. *Proc. Royal Soc. London B.* **143**: 103-113.
- LEHMAN E.L., 1983 Theory of point estimation. Wiley-Interscience, New York, NY, USA.
- LYNCH M., B. WALSH, 1998 Genetics and analysis of quantitative traits. Sinauer. Sunderland, MA, USA.
- MCCUTCHAN B.G., J.X. OU, G. NAMKOONG, 1985 A comparison of planned unbalanced designs for estimating heritability in perennial tree crops. *Theor. Appl. Genet.* **71**: 536-544.
- MELCHINGER A.E., 1988 Means, variances, and covariances between relatives in hybrid populations with disequilibrium in the parent populations. pp. 400-415. *In*: B.S. Weir (Ed.), Proceedings of the Second Intl. Conf. Quantitative Genetics. Sinauer.
- PATTERSON H.D., R. THOMPSON, 1971 Recovery of interblock information when block sizes are unequal. *Biometrika* **58**: 545-554.
- PIEPHO H.P., A. BÜCHSE, K. EMRICH, 2003 A hitchhikers's guide to mixed model for randomized experiments. *J. Agron. Crop Sci.* **189**: 310-322.
- SAS USER'S GUIDE, 1999 SAS Inst Cary, NC. SAS Institute.
- SEARLE S.R., 1979 Notes on variance components estimation a detailed account of Maximum Likelihood and kindred methodology. Technical Report BU-673-M, Biometrics Unit, Cornell University, Ithaca, New York.
- SORENSEN D., B.W. KENNEDY, 1984 Estimation of genetic variances from unselected and selected populations. *J. Anim. Sci.* **59**: 1213-1223.
- TOUTENBURG H., 1982 Prior information in linear models. J.Wiley & Sons, Bath, UK.