Original article

Multibreed analysis by splitting the breeding values

Luis Alberto García-Cortés[a,b*], Miguel Ángel Toro[a]

[a] Departamento de Mejora Genética Animal, SGIT-INIA, Carretera de la Coruña, Km. 7, 28040-Madrid, Spain
[b] Departamento de Producción Animal, ETSI Agrónomos, Universidad Politécnica de Madrid, Spain

Abstract – An equivalent model for multibreed variance covariance estimation is presented. It considers the additive case including or not the segregation variances. The model is based on splitting the additive genetic values in several independent parts depending on their genetic origin. For each part, it expresses the covariance between relatives as a partial numerator relationship matrix times the corresponding variance component. Estimation of fixed effects, random effects or variance components provided by the model are as simple as any model including several random factors. We present a small example describing the mixed model equations for genetic evaluations and two simulated examples to illustrate the Bayesian variance component estimation.

segregation variance / multibreed population / genetic evaluation / mixed model equations / Gibbs sampler

1. INTRODUCTION

The additive genetic covariance between relatives in multibreed cases was described by Elzo [6]. Nevertheless, this definition ignored the segregation terms, which explain the difference in the additive variance of the F1 and F2 groups, as was pointed out by Lo *et al.* [17]. These authors describe the rules to obtain **G**, the covariance matrix of the additive values, including both the pure breed contributions and the segregation deviations

$$\mathbf{g}_{ii} = \sum_{p=1}^{P} f_p^i \sigma_p^2 + 2 \sum_{p=1}^{P} \sum_{p'>p}^{P} \left(f_p^S f_{p'}^S + f_p^D f_{p'}^D \right) \sigma_{pp'}^2 + \frac{1}{2} \mathbf{g}_{SD} \tag{1}$$

---

* Corresponding author: garcia@inia.es

and

$$\mathbf{g}_{ij} = \frac{1}{2} \left(\mathbf{g}_{iS'} + \mathbf{g}_{iD'} \right), \tag{2}$$

where P is the number of breeds involved in the founder generation, f_p^i is the proportion of genes of the animal i coming from breed p, S and D are the sire and the dam of animal i, S' and D' are the sire and the dam of animal j, σ_p^2 is the additive variance component of breed p and $\sigma_{pp'}^2$ is the segregation variance between breeds p and p'. These rules are closely related to the tabular method to obtain inbreeding coefficients [9]. Lo *et al.* [17] also showed details about the calculation of $\mathbf{G}^{-1}$, closely derived from the conventional rules to obtain the inverse of the numerator relationship matrix [12, 19].

After this particular definition of $\mathbf{G}$, the analysis of multibreed populations is fully standard. The model $\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e}$ is defined as

$$E\left(\mathbf{y}\right) = \mathbf{Xb} \tag{3}$$
$$Var\left(\mathbf{y}\right) = \mathbf{ZGZ'} + \mathbf{I}\sigma_e^2, \tag{4}$$

where $\mathbf{b}$ are the fixed effects including the breed effects and other systematic effects, $\mathbf{X}$ is the incidence matrix relating fixed effects and records, $\mathbf{Z}$ is the incidence matrix for random effects and σ_e^2 is the residual variance. It must be noticed that the columns of $\mathbf{X}$ corresponding to breed effects consist of the genetic contributions of each breed to each phenotype. The missing data of the model, that is, the additive genetic values, are normally distributed as $\mathbf{a} \sim N\left(\mathbf{0}, \mathbf{G}\right)$ after discarding the breed effects. Genetic evaluations for this model are also standard, that is, fixed and random effects can be predicted with the mixed model equations [13] as described in [3].

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \mathbf{G}^{-1}\sigma_e^2 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \end{bmatrix} \tag{5}$$

where the additive covariance matrix $\mathbf{G}$ can not be expressed as the numerator relationship matrix times the additive variance components as in the conventional animal model because both the pure breed and the segregation components are mixtured into $\mathbf{G}$.

For that reason, genetic variance components for the model defined in formulas (3) and (4) are difficult to estimate. For instance, restricted maximum likelihood estimates for multibreed models became more complicated than in a conventional animal model; the details can be found in Elzo [7] and Birchmeier *et al.* [2]. On the contrary, Bayesian inferences applied on this model resulted in non standard full conditionals for the variance components,

making a Metropolis-Hastings implementation involving the repeated evalua-
tion of the determinants of **G** necessary [4].

In this paper we will describe an equivalent model to (3) and (4) where **G**
will be splitted into several parts allowing a simple analysis of the variance
components. As an example, the Bayesian inference will be described for this
case in a simpler way than that described in the literature.

In Section 2, the equivalent model for multiple breed cases will be proposed.
In Section 3 we will illustrate the implementation of the genetic evaluation
in a small example and we will describe both the Bayesian and frequentist
inference on variance components in Section 4.

2. THE MODEL

Formula (1) shows that the variability of a genetic value $\mathbf{g}_{ii}$ can be splitted
into several independent sources, some of them related with the pure breed
variabilities and others related with segregation terms. The matrix **G** can then
be partitioned into several pieces just considering only one source of variability
and assigning 0 to the others. Lacy *et al.* [15,16] used this procedure to analyze
the partial inbreeding of the animals due to each founder of the population. We
will use it here to split the source of each genetic covariance $\mathbf{g}_{ij}$, *i.e.*, each
element of **G**.

For instance, the partial **G** related with the breed p, $\mathbf{G}^{(p)}$, can be obtained
by considering every genetic component but σ_p^2 null. Then, the rules in equa-
tions (1) and (2) become

$$\mathbf{g}_{ii}^{(p)} = f_p^i \sigma_p^2 + \frac{1}{2} \mathbf{g}_{SD}^{(p)} \tag{6}$$

and

$$\mathbf{g}_{ij}^{(p)} = \frac{1}{2} \left(\mathbf{g}_{iS'}^{(p)} + \mathbf{g}_{iD'}^{(p)} \right). \tag{7}$$

The partial **G** due to the segregation between breeds p and p', $\mathbf{G}^{(pp')}$, can be
obtained after the same argument over $\sigma_{pp'}^2$.

$$\mathbf{g}_{ii}^{(pp')} = 2 \left(f_p^S f_{p'}^S + f_p^D f_{p'}^D \right) \sigma_{pp'}^2 + \frac{1}{2} \mathbf{g}_{SD}^{(pp')} \tag{8}$$

and

$$\mathbf{g}_{ij}^{(pp')} = \frac{1}{2} \left(\mathbf{g}_{iS'}^{(pp')} + \mathbf{g}_{iD'}^{(pp')} \right). \tag{9}$$

It can be noticed that $\mathbf{G}^{(p)}$ or $\mathbf{G}^{(pp')}$ depend each on a single variance compo-
nent and they can be expressed as $\mathbf{G}^{(p)} = \mathbf{A}_p \sigma_p^2$ or $\mathbf{G}^{(pp')} = \mathbf{A}_{pp'} \sigma_{pp'}^2$, being

both $\mathbf{A}_p$ and $\mathbf{A}_{pp'}$ the partial numerator relationship matrices due to pure breeds or segregation terms respectively. Although not explicitly described as partial numerator relationship matrices, splitting the inbreeding of the animals into several independent parts was described yet in [15, 16]. Following this idea, Rodrigañez *et al.* [20] calculated the partial inbreeding coefficients by founder origin by using a modification of the tabular method.

It is verifiable that

$$\mathbf{G} = \sum_p \mathbf{A}_p \sigma_p^2 + \sum_p \sum_{p'>p} \mathbf{A}_{pp'} \sigma_{pp'}^2. \tag{10}$$

Replacing (10) into (4) we have the variance of the model defined as

$$Var(\mathbf{y}) = \mathbf{Z} \left(\sum_p \mathbf{A}_p \sigma_p^2 + \sum_p \sum_{p'>p} \mathbf{A}_{pp'} \sigma_{pp'}^2 \right) \mathbf{Z}' + \mathbf{R} \tag{11}$$

which can be analyzed by including the corresponding missing data $\mathbf{a}_p$ and $\mathbf{a}_{pp}$ for each component as

$$\mathbf{y} = \mathbf{Xb} + \sum_p \mathbf{Za}_p + \sum_p \sum_{p'>p} \mathbf{Za}_{pp'} + \mathbf{e} \tag{12}$$

which is just a conventional animal model with several random factors. For instance, considering a case with two pure breeds, the model becomes $\mathbf{y} = \mathbf{Xb} + \mathbf{Za}_1 + \mathbf{Za}_2 + \mathbf{Za}_{12} + \mathbf{e}$, being $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_{12}$ the breeding values splitted by origin. Assuming the variance components to be known, the Henderson mixed model equations [13] will be

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} & \mathbf{X'Z} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \mathbf{A}_1^- \alpha_1 & \mathbf{Z'Z} & \mathbf{Z'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} & \mathbf{Z'Z} + \mathbf{A}_2^- \alpha_2 & \mathbf{Z'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} & \mathbf{Z'Z} & \mathbf{Z'Z} + \mathbf{A}_{12}^- \alpha_{12} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}}_1 \\ \hat{\mathbf{a}}_2 \\ \hat{\mathbf{a}}_{12} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \\ \mathbf{Z'y} \\ \mathbf{Z'y} \end{bmatrix} \tag{13}$$

where $\alpha_x = \sigma_e^2/\sigma_x^2$. Equations (5) and (13) provide equivalent results, being verifiable that $\mathbf{a} = \mathbf{a}_1 + \mathbf{a}_2 + \mathbf{a}_{12}$.

The inverses of the partial numerator relationship matrices involved in (13) can be easily obtained from $\mathbf{A}_x^- = (\mathbf{I} - \mathbf{P})' \mathbf{D}_x^- (\mathbf{I} - \mathbf{P})$, where $\mathbf{P}$ is a matrix relating progeny to parents and $\mathbf{D}_x$ are the pivots recursively obtained by the conventional algorithm [19]. Only $\mathbf{D}^-$ is specific for pure breeds or segregations and they will be described in the next section.

Note that formula (13) uses generalized inverses to take into account that some animals have null contributions from some breeds or segregations.

In these cases, the inverse of the whole $\mathbf{A}_x^{-1}$ does not exist because of the existence of null rows. From here on, we will use $\mathbf{A}_x^-$ as the inverse of the corresponding nonzero part of $\mathbf{A}$ but keeping the pattern of null rows as in $\mathbf{A}$. We will use $\mathbf{D}_x^-$ and $\mathbf{a}_x$ accordingly. We could not demonstrate the equivalence of both models in cases including null contributions, but their equivalence will be illustrated by using a small numerical example in Section 3.

In fact, the conventional strategies to set up the mixed model equations can be applied as if all contributions were present, but the equations corresponding to null contributions have to be discarded in order to solve the linear system in equation (13).

Calculation of the diagonal elements of the partial numerator relationship matrices

First, considering that in the new model $\mathbf{G}$ can be factored as in $\mathbf{A}\sigma^2$ for each partial covariance matrix, recursive formulas to compute the tabular method (6), (7), (8) and (9) can be written respectively as

$$\mathbf{A}_{ii}^{(p)} = f_p^i + \frac{1}{2}\mathbf{A}_{SD}^{(p)}$$

$$\mathbf{A}_{ij}^{(p)} = \frac{1}{2}\left(\mathbf{A}_{iS'}^{(p)} + \mathbf{A}_{iD'}^{(p)}\right)$$

$$\mathbf{A}_{ii}^{(pp')} = 2\left(f_p^S f_{p'}^S + f_p^D f_{p'}^D\right) + \frac{1}{2}\mathbf{A}_{SD}^{(pp')}$$

$$\mathbf{A}_{ij}^{(pp')} = \frac{1}{2}\left(\mathbf{A}_{iS'}^{(pp')} + \mathbf{A}_{iD'}^{(pp')}\right)$$

where $\mathbf{A}_{ij}$ is the element at the ith row and the jth column of the corresponding partial $\mathbf{A}$. Second, by defining $c_i = f_p^i$ for the pure breed contributions and $c_i = 2\left(f_p^S f_{p'}^S + f_p^D f_{p'}^D\right)$ for segregation effects, the tabular method can be applied in both cases by using the same algorithm

$$\mathbf{A}_{ii} = c_i + \frac{1}{2}\mathbf{A}_{SD}$$

$$\mathbf{A}_{ij} = \frac{1}{2}\left(\mathbf{A}_{iS'} + \mathbf{A}_{iD'}\right).$$

The Quaas procedure [19] can be easily reformulated to include the contributions of each genetic origin. In fact, following his procedure (Sect. 3.1.a of the

referred paper) for the complete numerator relationship matrix

$$
v_i = l_{ii} = \begin{cases} \sqrt{1 - .25\,(u_S + u_D)} & if \quad 0 < S, D \\ \sqrt{1 - .25 u_S} & if \quad D = 0 < S \\ \sqrt{1 - .25 u_D} & if \quad S = 0 < D \\ 1 & if \quad 0 = S = D \end{cases} \tag{14}
$$

where $\mathbf{L}$ is the Choleski factor of $\mathbf{A}$ and $\mathbf{v}$ and $\mathbf{u}$ are working vectors (details can be found in [19]), the partial relationship matrices $\mathbf{A}^-$ for pure breeds or segregation effects can be obtained only by setting a slight modification

$$
v_i = l_{ii} = \begin{cases} \sqrt{c_i - .25\,(u_S + u_D)} & if \quad 0 < S, D \\ \sqrt{c_i - .25 u_S} & if \quad D = 0 < S \\ \sqrt{c_i - .25 u_D} & if \quad S = 0 < D \\ \sqrt{c_i} & if \quad 0 = S = D. \end{cases} \tag{15}
$$

Other parts of the conventional algorithm to compute the diagonal elements of $\mathbf{A}^-$ do not suffer any modification, but skipping the animals with null contributions for some particular partial numerator relationship matrices. It is verifiable that (14) is a particular case of (15) for single breed cases with all $c_i = 1$.

3. NUMERICAL EXAMPLE TO ILLUSTRATE THE GENETIC EVALUATION FOR THE MULTIPLE BREED MODEL

In Table I, an example is presented including a herd effect with two levels ($\mathbf{h}$) and 11 animals. The four base animals were classified into two breeds ($\mathbf{g}$) and the true variance components are $\sigma_e^2 = 4$, $\sigma_1^2 = 1$, $\sigma_2^2 = 2$ and $\sigma_{12}^2 = 0.25$.

We implemented the Cantet and Fernando [3] method for the animal model $\mathbf{y} = \mathbf{X}_1\mathbf{h} + \mathbf{X}_2\mathbf{g} + \mathbf{a} + \mathbf{e}$ as presented in equation (5), where $\mathbf{b}' = (\mathbf{h}'\ \mathbf{g}')$. In this case, the number of equations is 15. The incidence matrix for the breed effect, $\mathbf{X}_2$, includes the contribution of genes of each breed for each phenotype. We discarded the equation for the first herd to allow the estimability of the system. The relevant part of this formulation is the matrix $\mathbf{G}$, which has been calculated following the rules of Lo et al. [17] at equations (1) and (2).

Table I. Pedigree and data file for the small example.

Animal	Sire	Dam	Breed*	Herd	Record
1	-	-	1	2	11.
2	-	-	1	2	12.
3	-	-	2	2	13.
4	-	-	2	1	14.
5	1	2	-	1	15.
6	3	2	-	2	16.
7	3	4	-	2	17.
8	5	6	-	2	18.
9	7	6	-	1	19.
10	9	8	-	1	20.
11	5	8	-	1	21.

* Only base animals.

$$\mathbf{G} = \begin{bmatrix} 1.00 & 0.00 & 0.00 & 0.00 & 0.50 & 0.00 & 0.00 & 0.25 & 0.00 & 0.12 & 0.37 \\ & 1.00 & 0.00 & 0.00 & 0.50 & 0.50 & 0.00 & 0.50 & 0.25 & 0.37 & 0.50 \\ & & 2.00 & 0.00 & 0.00 & 1.00 & 1.00 & 0.50 & 1.00 & 0.75 & 0.25 \\ & & & 2.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.50 & 0.25 & 0.00 \\ & & & & 1.00 & 0.25 & 0.00 & 0.62 & 0.12 & 0.37 & 0.81 \\ & & & & & 1.50 & 0.50 & 0.87 & 1.00 & 0.93 & 0.56 \\ & & & & & & 2.00 & 0.25 & 1.25 & 0.75 & 0.12 \\ & & & & & & & 1.62 & 0.56 & 1.09 & 1.12 \\ & & Sym. & & & & & & 2.25 & 1.40 & 0.34 \\ & & & & & & & & & 2.15 & 0.73 \\ & & & & & & & & & & 1.62 \end{bmatrix}$$

The results for the fixed and random effects are presented in column 1 of Table II.

We also implemented the mixed model equations as presented in formula (13), which consisted of 37 equations. The relevant parts are $\mathbf{A}_1$, $\mathbf{A}_2$ and $\mathbf{A}_{12}$. In order to allow an easy verification of formula (10), that is, $\mathbf{G} = \mathbf{A}_1\sigma_1^2 + \mathbf{A}_2\sigma_2^2 + \mathbf{A}_{12}\sigma_{12}^2$, we present them including the corresponding null rows and multiplied by its variance component.

Table II. Mixed model solutions for Cantet and Fernando [3] (Model 1) and the reformulation presented here (Model 2) for the small data set.

Model 1	Model 2		
$\hat{h}_1 = 0$	$\hat{h}_1 = 0$		
$\hat{h}_2 = -3.198032$	$\hat{h}_2 = -3.198032$		
$\hat{g}_1 = 16.613064$	$\hat{g}_1 = 16.613064$		
$\hat{g}_2 = 17.408480$	$\hat{g}_2 = 17.408480$		
$\hat{a}_1 = -0.293401$	$\hat{a}_{1,1} = -0.293401$	$\hat{a}_{2,1} = 0$	$\hat{a}_{12,1} = 0$
$\hat{a}_2 = 0.293401$	$\hat{a}_{1,2} = 0.293401$	$\hat{a}_{2,2} = 0$	$\hat{a}_{12,2} = 0$
$\hat{a}_3 = 0.688673$	$\hat{a}_{1,3} = 0$	$\hat{a}_{2,3} = 0.688673$	$\hat{a}_{12,3} = 0$
$\hat{a}_4 = -0.688673$	$\hat{a}_{1,4} = 0$	$\hat{a}_{2,4} = -0.688673$	$\hat{a}_{12,4} = 0$
$\hat{a}_5 = 0.237007$	$\hat{a}_{1,5} = 0.237007$	$\hat{a}_{2,5} = 0$	$\hat{a}_{12,5} = 0$
$\hat{a}_6 = 1.216290$	$\hat{a}_{1,6} = 0.388451$	$\hat{a}_{2,6} = 0.827838$	$\hat{a}_{12,6} = 0$
$\hat{a}_7 = 0.671231$	$\hat{a}_{1,7} = 0$	$\hat{a}_{2,7} = 0.671231$	$\hat{a}_{12,7} = 0$
$\hat{a}_8 = 1.746217$	$\hat{a}_{1,8} = 0.749687$	$\hat{a}_{2,8} = 0.705225$	$\hat{a}_{12,8} = 0.291305$
$\hat{a}_9 = 1.262473$	$\hat{a}_{1,9} = 0.229638$	$\hat{a}_{2,9} = 0.962010$	$\hat{a}_{12,9} = 0.070825$
$\hat{a}_{10} = 1.778623$	$\hat{a}_{1,10} = 0.555868$	$\hat{a}_{2,10} = 0.966027$	$\hat{a}_{12,10} = 0.256728$
$\hat{a}_{11} = 1.456135$	$\hat{a}_{1,11} = 0.780909$	$\hat{a}_{2,11} = 0.441093$	$\hat{a}_{12,11} = 0.234133$

$$\mathbf{A}_1\sigma_1^2 = \begin{bmatrix} 1.00 & 0.00 & 0.00 & 0.00 & 0.50 & 0.00 & 0.00 & 0.25 & 0.00 & 0.12 & 0.37 \\ & 1.00 & 0.00 & 0.00 & 0.50 & 0.50 & 0.00 & 0.50 & 0.25 & 0.37 & 0.50 \\ & & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ & & & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ & & & & 1.00 & 0.25 & 0.00 & 0.62 & 0.12 & 0.37 & 0.81 \\ & & & & & 0.50 & 0.00 & 0.37 & 0.25 & 0.31 & 0.31 \\ & & & & & & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ & & & & & & & 0.87 & 0.18 & 0.53 & 0.75 \\ & & & Sym. & & & & & 0.25 & 0.21 & 0.15 \\ & & & & & & & & & 0.59 & 0.45 \\ & & & & & & & & & & 1.18 \end{bmatrix}$$

$$\mathbf{A}_2\sigma_2^2 = \begin{bmatrix} 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ & & 2.00 & 0.00 & 0.00 & 1.00 & 1.00 & 0.50 & 1.00 & 0.75 & 0.25 \\ & & & 2.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.50 & 0.25 & 0.00 \\ & & & & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ & & & & & 1.00 & 0.50 & 0.50 & 0.75 & 0.62 & 0.25 \\ & & & & & & 2.00 & 0.25 & 1.25 & 0.75 & 0.12 \\ & & & & & & & 0.50 & 0.37 & 0.43 & 0.25 \\ & & & Sym. & & & & & 1.75 & 1.06 & 0.18 \\ & & & & & & & & & 1.18 & 0.21 \\ & & & & & & & & & & 0.25 \end{bmatrix}$$

$$\mathbf{A}_{12}\sigma_{12}^2 = \begin{bmatrix} 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ & & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ & & & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ & & & & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ & & & & & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ & & & & & & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ & & & & & & & 0.25 & 0.00 & 0.12 & 0.12 \\ & & S\,ym. & & & & & & 0.25 & 0.12 & 0.00 \\ & & & & & & & & & 0.37 & 0.06 \\ & & & & & & & & & & 0.18 \end{bmatrix}$$

The $\mathbf{A}_x^-$ required to set up the mixed model equations are really the inverses formed by the nonzero rows of each partial numerator relationship matrix. The results of this model are presented in Table II in columns 2 to 4. It is easily verifiable that the sum of the partial predictors for each animal equals the complete predictor provided by the Cantet and Fernando's [3] formulation at column 1. The null partial results for some animals correspond to the discarded equations because of their null contributions.

4. ESTIMATION OF VARIANCE COMPONENTS

Model (12) can be understood as a conventional animal model with several random factors with different covariance matrices. Variance component estimations on such kinds of models have been described in the literature under both the restricted maximum likelihood or the Bayesian inference. The main difference between the model presented here and other models involving, for instance, permanent environmental and litter effects, is the particular nature of the partial numerator relationship matrices $\mathbf{A}_1$, $\mathbf{A}_2$ and $\mathbf{A}_{12}$.

In this section we will use two simulated data sets and we will estimate the variance components *via* the Gibbs sampler. We will also outline the estimation of the variance components *via* the expectation maximization algorithm.

4.1. Data simulation

Two crossed populations were simulated in order to analyze the variance component estimation. In both cases we simulated 4000 base animals in

Table III. Marginal posterior moments for the variance components and the estimable difference between genetic means.

	Case I			Case II		
	True	Post. expect.	Post. var.	True	Post. expect.	Post. var.
$\sigma_{a_1}^2$	200	191.85	133.47	72	71.89	73.07
$\sigma_{a_2}^2$	168	169.94	144.30	128	140.38	118.41
σ_s^2	8	11.72	38.07	98	93.41	127.59
σ_e^2	400	401.52	46.47	400	403.16	47.39
$\mu_{a_2} - \mu_{a_1}$	80	80.29	0.34	280	280.78	0.26

two pure breeds and four non-overlapping generations of random mating. The total number of animals was 20 000 in both cases. Additive genetic values were obtained after genetic simulation with 100 loci and $AA = 2$, $Aa = 0$ and $aa = -2$, we simulated a fixed effect with 20 levels assigned at random and the continuous residual deviates were obtained from a normal distribution with $\sigma_e^2 = 400$. All animals were recorded and the pedigree was considered to be completely known.

In case I, all genetic frequencies were 0.5 and 0.7 for breeds 1 and 2 respectively, resulting in a segregation variance of $\sigma_{12}^2 = 8$. In case II, all genetic frequencies were 0.1 and 0.8, respectively, resulting in a segregation variance of $\sigma_{12}^2 = 98$. True values of both the variance components and the estimable differences between genetic means are presented in Table III.

Both data sets were simulated with the same structure. They have the same fixed levels and residual deviates and they only differ in the additive genetic values. The number of animals with nonzero contributions was 16 826 for breed 1, 16 716 for breed 2 and 10 958 for the segregation.

4.2. Bayesian analysis

We analyzed both data sets by using the Bayesian inference on model (12) which consists of a conventional mixed model with several random effects. Considering flat unbounded priors for both the variance components and the

fixed levels to simplify the implementation, the joint posterior distribution is

$$p\left(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_{12}, \mathbf{b}, \sigma_1^2, \sigma_2^2, \sigma_{12}^2, \sigma_e^2 | \mathbf{y}\right) \propto \left(\sigma_e^2\right)^{-\frac{n}{2}} \exp\left\{-\frac{\mathbf{e}'\mathbf{e}}{2\sigma_e^2}\right\}$$

$$\left(\sigma_1^2\right)^{-\frac{q_1}{2}} \exp\left\{-\frac{\mathbf{a}_1'\mathbf{A}_1^-\mathbf{a}_1}{2\sigma_1^2}\right\}$$

$$\left(\sigma_2^2\right)^{-\frac{q_2}{2}} \exp\left\{-\frac{\mathbf{a}_2'\mathbf{A}_2^-\mathbf{a}_2}{2\sigma_2^2}\right\}$$

$$\left(\sigma_{12}^2\right)^{-\frac{q_{12}}{2}} \exp\left\{-\frac{\mathbf{a}_{12}'\mathbf{A}_{12}^-\mathbf{a}_{12}}{2\sigma_{12}^2}\right\} \quad (16)$$

where $\mathbf{e} = \mathbf{y} - \mathbf{Xb} - \mathbf{Za}_1 - \mathbf{Za}_2 - \mathbf{Za}_{12}$, n is the number of records and q_1, q_2 and q_{12} are the number of animals with nonzero contributions of breed 1, 2 and segregation, respectively. Full conditional distributions from 16 are the well known normal distributions for the levels of both $\mathbf{b}$ and $\mathbf{a}$ and inverted chi-square distributions for variance components. Details of these conditionals can be found in [21] and are not presented in this paper because they are fully standard.

To evaluate the posterior density in (16) it suffices to compute for each quadratic form $\sum \phi_i^2 / d_i$, where ϕ represents the Mendelian sampling terms. Note that in this case, it only makes sense to consider the animals with nonzero contributions.

The Bayesian inference for both cases was implemented by using a Gibbs sampler. Just the special structure of the numerator relationship matrices and the use of different amounts of animals in each part of equation (16) differed in this analysis in relation to the conventional mixed models.

The burn-in lengths provided by the coupling argument [10] were 13 428 and 2426 in cases I and II, respectively, to reach a difference between chains of 10^{-4}. After burn-in, posterior distributions were obtained by averaging 50 000 cycles of a single chain. The effective chain lengths after burn-in were 274.56 and 131.08 for cases I and II, respectively.

Table III shows the true simulated values and the posterior expectations and variances for both cases. It must be noticed that the accuracy of the segregation variance estimate is worse than the additive genetic variances of the pure breeds.

4.3. Restricted maximum likelihood

Restricted maximum likelihood [18] estimates of the variance components for multibreed analysis is also a particular case of a mixed linear model with several random factors. Then, the expectation-maximization algorithm [5] can be implemented by using the iterative approach described in [11]. This algorithm requires for each iteration the fixed and random solutions of the linear system provided in formula (13). The genetic variance components can be calculated at each iteration with

$$\hat{\sigma}_i^2 = \frac{1}{q_i} \left[\mathbf{a}_i' \mathbf{A}_i^- \mathbf{a}_i + tr\left(\mathbf{A}_i^- \mathbf{C}^{ii} \right) \right],$$

where i represents pure breeds and segregations, $\hat{\sigma}_i^2$ is the variance component corresponding to pure breeds or segregation effects, q_i is the number of animals with nonzero contributions from this particular source of variability (number of nonzero c_i as described above) and $\mathbf{C}^{ii}$ is the corresponding random by random block of the generalized inverse of the coefficient matrix in formula (13). The animals with null contributions have to be discarded both in $\mathbf{a}_i' \mathbf{A}_i^- \mathbf{a}_i$ and $\mathbf{A}_i^- \mathbf{C}^{ii}$. The residual variance can be obtained at each round of iteration after the well known formula

$$\hat{\sigma}_e^2 = \frac{1}{n - r(\mathbf{X})} \left[\mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\mathbf{b} - \sum_i \mathbf{y}'\mathbf{Z}\mathbf{a}_i \right],$$

where $\hat{\sigma}_e^2$ is the residual variance estimate at the current round of iteration, n is the number of records, $r(\mathbf{X})$ is the column rank of $\mathbf{X}$ and i goes throughout the pure breeds and segregation effects. Note that we considered null the elements of $\mathbf{a}_i$ corresponding to animals with a null contribution from breed or segregation i.

We did not implement this algorithm because it requires a high computation demand. It requires the inversion of the coefficient matrix in (13) at each round of iteration, whose rank is greater than in a conventional single breed analysis.

5. DISCUSSION

The model proposed in this paper is equivalent to the current model of choice [2–4, 17] for genetic evaluations in crossbred populations, including the segregation components. When discarding the segregation terms, the proposed model can also be used as an equivalent model to that proposed by Elzo's group [6–8]. It will provide equivalent results because it assumes the same

distribution of records: the expectation is the same in both models and considering (10), equations (11) *vs.* (4) provide the same variance. They only differ in the definition of the missing data, that is, the complete additive values in the conventional model were replaced by the partial additive values. As far as these values will be integrated out, both models will provide equivalent results, as we illustrated in a small example assuming the variances to be known (Sect. 3).

Our proposal requires a linear mixed model very much larger than the classical one. For instance, examples including two founder breeds require a linear system around three times larger, depending on the number of equations discarded because of the null contributions. Furthermore, the resulting linear system in formula (13) is not expected to be more sparse than the linear system in formula (5), because the sparsity of both approaches depends mainly on the pedigree structure, that is, on $(\mathbf{I} - \mathbf{P})$.

Although its computational demand is expected to be higher, the proposed model is a more natural approach in the sense that it defines a set of missing data for each source of variability, as linear models in animal breeding usually do. For that reason, it provides very simple formulas for the variance component estimation. When using the Gibbs sampler, the variance components will have the well known scaled inverted chi-square full conditional distributions, avoiding the use of the non standard forms described in Cardoso and Tempelman [4]. Although not implemented in this paper, restricted maximum likelihood estimates of the variance components *via* the EM algorithm (details in [2]) will also be simpler when based on the quadratic forms of the partial additive values.

The model can also be used to analyze models including genetic groups with different genetic variances, as Alfonso and Estany [1] analyzed. Nevertheless, the required number of equations will grow dramatically even for models with a small number of genetic groups.

Even for single breed analysis, splitting the breeding values of the animals into several independent parts by genetic origin can also be useful. For instance, it can be used to evaluate the impact of some group of founders in the breeding values of their descendants, for instance, multiple ovulation embryo transfer nucleus or artificial insemination centers. Further research should be focused to develop models to evaluate the impact of some non founder animals in their descendants.

Although not described in this paper, the extension to multiple trait models is expected to have the same difficulties as the conventional models including several random factors.

6. IMPLICATIONS

Variance component estimates in multiple breed populations have been implemented by using particular and specific computer programs. General purpose packages to estimate variance components, such as VCE5 [14], can not be used currently for multiple breed analysis, but they can handle models with multiple random factors. These packages will only require an extra coding effort to calculate both the genetic contributions and the partial relationship matrices outlined in algorithm (15), which can be easily implemented from the code they actually include to calculate the diagonal elements of the numerator relationship matrix (algorithm (14)).

REFERENCES

[1] Alfonso L., Estany J., An expression of mixed animal model equations to account for different means and variances in the base population, Genet. Sel. Evol. 31 (1999) 105–113.

[2] Birchmeier A.N., Cantet R.J.C., Fernando R.L., Morris C.A., Holgado F., Jara A., Santos Cristobal M., Estimation of segregation variance for birth weight in beef cattle, Lives. Prod. Sci. 76 (2002) 27–35.

[3] Cantet R.J.C., Fernando R.L., Prediction of breeding values with additive animal models for crosses from 2 populations, Genet. Sel. Evol. 27 (1995) 323–334.

[4] Cardoso F.F., Tempelman R.J., Hierarchical bayes multi-breed inference with an application to genetic evaluation of a Nelore-Hereford population, J. Anim. Sci. 82 (2004) 1589–1601.

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

[6] Elzo M.A., Recursive procedures to compute the inverse of multiple trait additive genetic covariance matrix in inbreed and non inbreed multibreed populations, J. Anim. Sci. 68 (1990) 1215–1228.

[7] Elzo M.A., Restricted maximum likelihood estimation of additive and nonadditive genetic variances and covariances in multibreed populations, J. Anim. Sci. 72 (1994) 3055–3065.

[8] Elzo M.A., Wakeman D.L., Covariance components and prediction for additive and nonadditive preweaning growth genetic effects in an Angus-Brahman multibreed herd, J. Anim. Sci. 76 (1998) 1290–1302.

[9] Emik L.O., Terrill C.E., Systematic procedures for calculating inbreeding coefficients, J. Hered. 40 (1949) 51–55.

[10] García-Cortés L.A., Rico M., Groeneveld E., Coupling with the Gibbs sampler to assess convergence in animal models, J. Anim. Sci. 76 (1998) 441–447.

[11] Harville D.A., Maximum likelihood approaches to variance component estimation and to related problems, J. Amer. Stat. Assoc. 72 (1977) 320–340.

[12] Henderson C.R., A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values, Biometrics 32 (1976) 69–83.

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

[14] Kovac M., Groeneveld E., García-Cortés L.A., VCE-5, a package for the estimation of dispersion parameters, in: Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, 19–23 August 2002, Vol. 33, Montpellier, France, pp. 741–742.

[15] Lacy R.C., Errata, Evolution 51 (1997) 1025.

[16] Lacy R.C., Alaks G., Walsh A., Hierarchical analysis of inbreeding depression in *Peromyscus polionotus*, Evolution 50 (1996) 2187–2200.

[17] Lo L.L., Fernando R.L., Grossman M.G., Genotypic covariance between relatives in multibreed populations: additive model, Theor. Appl. Genet. 87 (1993) 423–430.

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

[19] Quaas R.L., Computing the diagonal elements and inverse of a large numerator relationship matrix, Biometrics 32 (1976) 949–953.

[20] Rodrigañez J., Toro M.A., Rodriguez C., Silió L., Effect of founder allele survival and inbreeding depression on litter size in a closed line of Large White pigs, Anim. Sci. 67 (1998) 573–582.

[21] Wang C.S., Rutledge J.J., Gianola D., Marginal inference about variance components in a mixed model using Gibbs sampling, Genet. Sel. Evol. 25 (1993) 41–62.