Original article

Prediction of breeding values with additive animal models for crosses from 2 populations

RJC Cantet [1], RL Fernando [2]

[1] *Universidad de Buenos Aires, Departamento de Zootecnia, Facultad de Agronomía, Av San Martín, 4453, (1417) Buenos Aires, Argentina;*
[2] *University of Illinois, Department of Animal Sciences, 1207, West Gregory Drive, Urbana, IL 61801, USA*

Summary – Recent developments in the theory of the covariance between relatives in crosses from 2 populations, under additive inheritance, are used to predict breeding values (BV) by best linear unbiased prediction (BLUP) using animal models. The consequences of incorrectly specifying the covariance matrix of BV is discussed. The theory of the covariance between relatives in crosses from 2 populations is extended for predicting BV in models with multiple traits. A numerical example illustrates the prediction procedures.

cross / heterogeneous additive variance / segregation variance / genetic group / BLUP-animal model

Résumé – Prédiction des valeurs génétiques avec des modèles individuels additifs pour des croisements à partir de 2 populations. *De récentes avancées de la théorie de la covariance entre apparentés dans les croisements entre 2 populations, en hérédité additive, sont utilisées pour prédire les valeurs génétiques (VG) par le BLUP-modèle animal. Les conséquences résultant d'une définition incorrecte de la matrice de covariance des VG sont discutées. La théorie de la covariance entre apparentés en croisement à partir de 2 populations est étendue à la prédiction de VG pour plusieurs caractères. Un exemple numérique illustre les procédures de prédiction.*

croisement / BLUP-modèle animal / variance génétique additive hétérogène / variance de ségrégation / groupe génétique

INTRODUCTION

A common method used to genetically improve a local population is by planned migration of genes from a superior one. For example, in developing countries, US Holstein sires are mated to local cows in order to genetically improve the local Holstein population. Genetic evaluation in such populations must take into consideration the genetic differences between the local and the superior populations.

Best linear unbiased prediction (BLUP) is widely used for genetic evaluation (Henderson, 1984). BLUP methodology requires modelling genotypic means and covariances. Genetic groups are used to model differences in genetic means between populations (Quaas, 1988). However, populations can also have different genetic variances. Under additive inheritance, Elzo (1990) provided a theory to incorporate heterogeneous genetic variances in genetic evaluation by BLUP. His procedure is based on computing the additive variance for a crossbred animal as a weighted mean of the additive variances of the parental populations plus one half the covariance between parents. Lo *et al* (1993) showed that Elzo's theory did not account for additive variation created by segregation of alleles between populations with different gene frequencies. For example, even though the additive variance for an F2 individual should be higher than for an F1, due to segregation (Lande, 1981), Elzo's formulation gives the same variance for both. Lo *et al* (1993) provided a theory to incorporate segregation variance in computing covariances between crossbred relatives, and to invert the genetic covariance matrix efficiently.

The objectives of this paper are: 1) to demonstrate how the theory of Lo *et al* (1993) can be used for genetic evaluation, *ie* to predict breeding values (BV), by BLUP; 2) to study the consequences of using an incorrect genetic covariance matrix on prediction of BV; and 3) to extend the theory of Lo *et al* (1993) to accommodate multiple traits. A numerical example is used to illustrate the principles introduced here.

MODEL

Even though the theory presented by Lo *et al* (1993) allowed for several breeds or strains within a breed, we focus on the case of 2. A typical situation in beef or dairy cattle is when a 'local' (L) strain or breed is crossed with an 'imported' (I) one. Usually the program starts by mating genetically superior L females with I males to produce F1 progeny. Then superior F1 females are mated to I sires to produce backcross progeny. The program is continued by repeatedly mating superior backcross females to I sires. It should be noted that L, F1 and backcross sires are also used to produce progeny. Thus, the crosses generated by such a program may include F1 = I × L, F2 = F1 × F1, BI = I × F1, BL = F1 × L, BII = BI × F1, 5/8I = BI × F1, 3/8I = BI × L, *etc*. It is shown below how genetic evaluations for such a mixture of crossbred animals can be obtained by BLUP using Henderson's (1984) mixed-model equations (MME).

Genetic evaluations are based on a vector of phenotypic records ($\mathbf{y}$), which can be modelled as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{e} \qquad [1]$$

where $\boldsymbol{\beta}$ is a vector of non-genetic fixed effects, $\mathbf{a}$ is a vector of additive genetic values or BV and $\mathbf{e}$ is a vector of random residuals, independent of $\mathbf{a}$, with null mean and covariance matrix $\mathbf{R}$. Although $\mathbf{R}$ can be any general symmetric matrix, in general it is taken to be diagonal, and this simplifies computing solutions of $\boldsymbol{\beta}$ and predictions of $\mathbf{a}$. The incidence matrices $\mathbf{X}$ and $\mathbf{Z}$ relate $\boldsymbol{\beta}$ and $\mathbf{a}$, respectively, to $\mathbf{y}$. The mean and the covariance matrix of the vector of BV ($\mathbf{a}$) for crossbred individuals are modelled as:

$$\mathrm{E}(\mathbf{a}) = \mathbf{Qg} \qquad [2]$$

and

$$\mathrm{Var}(\mathbf{a}) = \mathbf{G} \qquad [3]$$

where $\mathbf{g}$ is a vector of genetic group effects for individuals in the I and L populations, $\mathbf{Q}$ is a matrix relating $\mathbf{a}$ with the genetic groups. If there is only 1 group on each breed, $\mathbf{Q}$ specifies the breed composition for each individual. The matrix $\mathbf{G}$ contains the variances and covariance among BV as defined by Lo *et al* (1993).

In modelling the mean of $\mathbf{a}$, genetic groups are only assigned to 'phantom' parents of known animals following the method proposed by Westell *et al* (1988). Quaas (1988) showed that $\mathbf{Q}$ can be expressed as:

$$\mathbf{Q} = (\mathbf{I} - \mathbf{P})^{-1} \, \mathbf{P}_b \mathbf{Q}_b \qquad [4]$$

where $\mathbf{P}$ relates progeny to parents, $\mathbf{P}_b$ progeny to phantom parents, and $\mathbf{Q}_b$ is an incidence matrix that relates phantom parents to genetic groups. Elements in each row of $[\mathbf{P}_b{:}\mathbf{P}]$ are all zero, except for two $1/2$'s in the columns pertaining to the parents of the animals in $\mathbf{a}$. It should be stressed that the above model for $\mathbf{a}$ assumes additive inheritance (Thompson, 1979; Quaas, 1988; Lo *et al*, 1993).

In the genetic grouping theory of Quaas (1988), all the groups are assumed to have the same additive variance. In this model, however, we allow the I and L populations to have different additive variances, and the variances and covariances of crossbred animals are computed following the theory of Lo *et al* (1993). They showed that the covariance between crossbred relatives can be computed using the tabular method for purebreds (Emik and Terril, 1949; Henderson, 1976), provided that the variance of a crossbred individual i is computed as:

$$\mathrm{Var}(a_i) = f_{i\mathrm{L}}\sigma_{\mathrm{AL}}^2 + f_{i\mathrm{I}}\sigma_{\mathrm{AI}}^2 + \frac{1}{2}\,\mathrm{cov}(a_j, a_k) + 2(f_{j\mathrm{L}}f_{j\mathrm{I}} + f_{k\mathrm{L}}f_{k\mathrm{I}})\sigma_{\mathrm{AIL}}^2 \qquad [5]$$

where j and k are the parents of i, and $f_{j\mathrm{I}}$, for example, is the breed I composition of dam j, σ_{AL}^2 is the additive variance of population L, σ_{AI}^2 is the additive variance for population I, and σ_{ALI}^2 is the segregation variance, which results from differences in gene frequencies between the L and I populations. The term segregation variance was used by Wright (1968) and Lande (1981) to refer to the additional genetic variance due to segregation in the F2 generation over that in the F1. Following Quaas (1988), Lo *et al* (1993) further showed that the inverse of the genetic covariance matrix ($\mathbf{G}$), required to setup Henderson's MME, can be constructed as:

$$\mathbf{G}^{-1} = (\mathbf{I} - \mathbf{P}') \, \mathbf{G}_{\varepsilon}^{-1}(\mathbf{I} - \mathbf{P}) \qquad [6]$$

where $\mathbf{G}_\varepsilon$ is a diagonal matrix with the ith diagonal element defined as:

$$\mathbf{G}_{\varepsilon_i} = \mathrm{Var}(a_i) - \frac{1}{4}\left[\mathrm{Var}(a_j) + \mathrm{Var}(a_k)\right] - \frac{1}{2}\mathrm{cov}(a_j, a_k) \qquad [7]$$

Note that these elements are linear functions of σ^2_{AL}, σ^2_{AI} and σ^2_{ALI}.

PREDICTION OF BREEDING VALUES

Following Quaas (1988), MME for a model with genetic groups can be written as:

$$[\mathbf{H} + \boldsymbol{\Sigma}]\begin{bmatrix}\widehat{\boldsymbol{\beta}} \\ \widehat{\mathbf{g}} \\ \widehat{\mathbf{a}}\end{bmatrix} = \begin{bmatrix}\mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{0} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y}\end{bmatrix} \qquad [8]$$

where

$$\mathbf{H} = \begin{bmatrix}\mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{0} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{0} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}\end{bmatrix} \qquad [9]$$

and

$$\boldsymbol{\Sigma} = \begin{bmatrix}\mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}'\mathbf{G}^{-1}\mathbf{Q} & -\mathbf{Q}'\mathbf{G}^{-1} \\ \mathbf{0} & -\mathbf{G}^{-1}\mathbf{Q} & \mathbf{G}^{-1}\end{bmatrix} \qquad 10]$$

The matrix $\mathbf{H}$ can be constructed efficiently using algorithms already available (*eg*, Groeneveld and Kovac, 1990). Quaas (1988) gave rules to construct $\boldsymbol{\Sigma}$ efficiently for a model with homogeneous additive variances across genetic groups. To construct $\boldsymbol{\Sigma}$ efficiently for a model with heterogeneous additive variances, replace $x(= 4/[\text{number of unknown parents} + 2])$ in the rules of Quaas (1988) with $1/\mathbf{G}_{\varepsilon i}$.

CONSEQUENCES OF USING AN INCORRECT G

Henderson (1975a) showed that using an incorrect $\mathbf{G}$ leads to predictions that are unbiased but do not have minimum variance. His results are employed here to examine the consequences of using the same additive variance ($\sigma^2_{\mathrm{A}*}$) for L, I and crossbred animals.

Let $\mathbf{C}^{aa}$ be the submatrix of a g-inverse of the right-hand-side of the MME corresponding to $\mathbf{a}$, but calculated with $\mathbf{G}^* = \mathbf{A}\sigma^2_{\mathrm{A}*}$. Then, as in Henderson (1975a) and Van Vleck (1993), the prediction error variance (PEV) of $\mathbf{a}$ is not equal to $\mathbf{C}^{aa}$, but is:

$$\mathrm{Var}(\widehat{\mathbf{a}} - \mathbf{a}) = \mathbf{C}^{aa} + \mathbf{C}^{aa}\mathbf{G}^{*-1}(\mathbf{G} - \mathbf{G}^*)\mathbf{G}^{*-1}\mathbf{C}^{aa} \qquad [11]$$

where $\mathbf{G}$ is the correct covariance matrix of $\mathbf{a}$. Now, let $\mathbf{D}$ be a diagonal matrix with the ith diagonal element being equal to $0.5[1 - 0.5(F_{\mathrm{S}i} + F_{\mathrm{D}i})]$, if the father (S$i$) and the mother (D$i$) of i are known, and $F_{\mathrm{S}i}$ is the inbreeding coefficient of Si.

Also, $\mathbf{D}_{ii} = 0.25(3 - F_{\mathrm{S}i})$, if only the sire of i is known, and $\mathbf{D}_{ii} = 0.25(3 - F_{\mathrm{D}i})$ if only the dam of i is known. Finally, if both parents of i are unknown $\mathbf{D}_{ii} = 1$. With this definition of $\mathbf{D}$ and after some algebra, [11] becomes equal to:

$$\mathrm{Var}(\widehat{\mathbf{a}} - \mathbf{a}) = \mathbf{C}^{aa} + \mathbf{C}^{aa}(\mathbf{I} - \mathbf{P}')\mathbf{D}^{-1}(\mathbf{G}_\varepsilon - \mathbf{D})\mathbf{D}^{-1}(\mathbf{I} - \mathbf{P})\mathbf{C}^{aa} \qquad [12]$$

Therefore, PEV of $\mathbf{a}$ obtained from $\mathbf{C}^{aa}$ will be incorrectly estimated by the second term on the right of [11] or [12]. As this term depends on the structure of $\mathbf{G}_\varepsilon$ and $\mathbf{D}$, no general result can be given. However, if $\mathbf{C}^{aa}(\mathbf{I} - \mathbf{P}')\mathbf{D}^{-1}(\mathbf{G}_\varepsilon - \mathbf{D})\mathbf{D}^{-1}(\mathbf{I} - \mathbf{P})\mathbf{C}^{aa}$ is positive definite, it adds up to $\mathbf{C}^{aa}$ and true PEV is underestimated. This happens if both $(\mathbf{G}_\varepsilon - \mathbf{D})$ and $\mathbf{C}^{aa}(\mathbf{I} - \mathbf{P}')\mathbf{D}^{-1}$ are positive definite (see, for example, theorem A.9 in page 183 of Toutenburg, 1982). Now, the fixed effects are reparameterized so that $[\mathbf{X}{:}\mathbf{ZQ}]$ is a full rank matrix. Then, $[\mathbf{C}^{aa}(\mathbf{I} - \mathbf{P}')\mathbf{D}^{-1}]^{-1} = \mathbf{D}(\mathbf{I} - \mathbf{P}')^{-1}(\mathbf{C}^{aa})^{-1}$ and $\mathbf{C}^{aa}(\mathbf{I} - \mathbf{P}')\mathbf{D}^{-1}$ is positive definite. Finally, if $(\mathbf{G}_\varepsilon - \mathbf{D})$ is positive definite its diagonal elements are positive (Seber, 1977, page 388), which in turn happens when the diagonal elements of $\mathbf{G}_\varepsilon$ are strictly greater than corresponding elements of $\mathbf{D}$. For example, this may happen whenever σ^2_{ALI} contributes to the variance of crossbred individuals (such as F2 or 5/8I), and this variance parameter is ignored. Under these conditions PEV will be underestimated, and the amount of underestimation will depend on the magnitude of σ^2_{ALI}.

It has been shown that if all data employed to make selection decisions are available, then the BLUP of $\mathbf{a}$ can be computed ignoring selection (Henderson, 1975b; Goffinet, 1983; Fernando and Gianola, 1990). This result only holds when the correct covariance matrix of $\mathbf{a}$ is used to compute BLUP. Thus, in the improvement of a local breed by mating superior I sires to selected L females, the use of the same additive variance for the I and L populations will give biased results if the 2 populations are known to have different additive variances and the process of selection and mating to superior males is repeated.

MULTIPLE TRAITS

The theory presented by Lo *et al* (1993) can be extended to obtain BLUP with multiple traits. Consider the extension for 2 traits: X and Y. To obtain BLUP for X and Y the additive covariance matrices for traits X and Y, and between traits X and Y are needed. The covariance matrix for traits X and Y can be computed as described by Lo *et al* (1993). It is shown below how to compute the additive covariance matrix between traits X and Y.

Following the reasoning employed by Lo *et al* (1993) to derive the additive variance for a crossbred individual, it can be shown that the additive covariance between traits X and Y for a crossbred individual is:

$$\mathrm{cov}(a_{\mathrm{X}i}, a_{\mathrm{Y}i}) = f_{i\mathrm{I}}\sigma_{\mathrm{A(XY)I}} + f_{i\mathrm{L}}\sigma_{\mathrm{A(XY)L}}$$
$$+ \frac{1}{2}\mathrm{cov}(a_{\mathrm{X}j}, a_{\mathrm{Y}k}) + 2(f_{j\mathrm{I}}f_{j\mathrm{L}} + f_{k\mathrm{I}}f_{k\mathrm{L}})\sigma_{\mathrm{A(XY)LI}} \quad [13]$$

where $\sigma_{A(XY)L}$ and $\sigma_{A(XY)I}$ are the additive covariances for traits X and Y in populations L and I, respectively, and $\sigma_{A(XY)LI}$ is the additive segregation covariance for populations L and I.

Provided that the covariance between traits X and Y for a crossbred individual is computed using equation [9], the covariance between X and Y between crossbred individuals i and i' can be computed as:

$$\operatorname{cov}(a_{Xi}, a_{Yi'}) = \frac{1}{2}\left[\operatorname{cov}(a_{Xj}, a_{Yi'}) + \operatorname{cov}(a_{Xk}, a_{Yi'})\right] \qquad [14]$$

provided that i' is not a direct descendant of i.

Now let $\mathbf{a}(2q \times 1)$ be the vector containing the BV of q animals for the 2 traits, ordered by trait within animal. Then, $\mathbf{G}$ is the $2q \times 2q$ covariance matrix between traits and individuals. Following Elzo (1990) the inverse of $\mathbf{G}$, required to setup Henderson's MME, can be written as:

$$\mathbf{G}^{-1} = [(\mathbf{I} - \mathbf{P}') \otimes \mathbf{I}_2]\left(\sum_i \oplus \mathbf{G}_{\varepsilon_i}^{-1}\right)[(\mathbf{I} - \mathbf{P}) \otimes \mathbf{I}_2] \qquad [15]$$

where for individual l, the diagonal elements of $\mathbf{G}_{\varepsilon l}$ can be calculated by equation [7], and the off-diagonals elements can be computed as:

$$\mathbf{G}_{\varepsilon i XY} = \operatorname{cov}(a_{Xi}, a_{Yi}) - \frac{1}{4}\left[\operatorname{cov}(a_{Xj}, a_{Yj}) + \operatorname{cov}(a_{Xk}, a_{Yk})\right] - \frac{1}{2}\operatorname{cov}(a_{Xj}, a_{Yk}) \quad [16]$$

All covariances in the above equation are functions of $\sigma_{A(XY)L}$, $\sigma_{A(XYL)I}$ and $\sigma_{A(XY)LI}$ and can be computed using equations [13] and [14].

Equation [15] gives rise to simple rules to setup MME for 2 or more traits. By appropriately redefining all vector and matrices to include 2 or more traits, equations [8], [9] and [10] are valid for the multiple trait situation. Again, matrix $\mathbf{H}$ can be constructed efficiently by commonly used algorithms (Groeneveld and Kovac, 1990). Cantet $et\ al$ (1992) gave rules to construct $\boldsymbol{\Sigma}$ efficiently for a model with homogeneous additive (co)variances across genetic groups. Their algorithm can be modified to construct $\boldsymbol{\Sigma}$ efficiently for a model with heterogeneous additive (co)variances as follows. Let:

$i_r =$ equation number of individual i for the rth trait.
$j_r =$ equation number of the sire of i or its sire group (if base sire) for trait r.
$k_r =$ equation number of the dam of i or its dam group (if base dam) for trait r.

Now let t be the number of traits, for $m = 1$ to t and $n = 1$ to t, and add to $\boldsymbol{\Sigma}$ the following 9 contributions:

Contribution	To elements
$G_{\hat{\varepsilon}imn}^{-1}$	(i_m, i_n)
$-0.5\ G_{\hat{\varepsilon}imn}^{-1}$	$(i_m, j_n);\ (j_m, i_n);\ (i_m, k_n);(k_m, i_n)$
$0.25\ G_{\hat{\varepsilon}imn}^{-1}$	$(j_m, j_n);\ (j_m, k_n);\ (k_m, j_n);\ (k_m, k_n)$

where $G_{\hat{\varepsilon}imn}^{-1}$ is element (m, n) of the inverse of the $t \times t$ matrix $\mathbf{G}_{\varepsilon i}$, which is associated to individual i. If $\boldsymbol{\Sigma}$ is full-stored, every animal makes $9t^2$ contributions.

For example, if 2 traits are considered ($t = 2$), there are $9(2^2) = 36$ contributions. If Σ is half-stored, there are $[9t(t-1)]/2 + 6t$ contributions. For $t = 2$, each individual makes 21 contributions to the upper triangular part of half-stored matrix Σ.

To obtain BLUP under a maternal effects model (Willham, 1963), the additive covariance matrices for the direct effect, the maternal effects, and between the direct and maternal effects are needed. These matrices can be computed using the theory used to compute the covariance matrices for traits X and Y as described above.

NUMERICAL EXAMPLE

Consider a single trait situation where a sire from breed I (animal 1) serves 2 dams: animal 2, from breed I, and animal 3 from breed L. Individuals 1 and 2 are the parents of 4 (purebred I), and 1 and 3 are parents of 5 (an F1 male). Finally, the F2 animal 7 is the offspring of 5 and 6, the latter being an F1 dam with unknown parents. Individuals 1, 4 and 5 are males and the rest are females. Age at measure and observed data for animals 2–7 are 100 (age), 100 (data); 110, 103; 95, 160; 98, 175; 106, 105; and 100, 114; respectively. There are 2 genetic groups for breed I and 1 for L. The model of evaluation includes fixed effects of age (as a covariate), sex and genetic groups (A1, A2 and B), and random BV for animals 1 through 7. In order for $[\mathbf{X}{:}\mathbf{ZQ}]$ to have full rank we imposed the restriction: sex 1 + sex $2 = 0$, or sex $1 = -$sex 2. Hence, $\boldsymbol{\beta}$ contains only 2 parameters: 1) the age covariate; and 2) the sex 1 effect (or –sex 2). Matrices $\mathbf{y}$, $\mathbf{X}$ and $\mathbf{Q}$ are then equal to:

$$
\mathbf{y} = \begin{bmatrix} 100 \\ 103 \\ 160 \\ 175 \\ 105 \\ 114 \end{bmatrix}
\quad
\mathbf{X} = \begin{bmatrix} 100 & 1 \\ 110 & 1 \\ 95 & -1 \\ 98 & -1 \\ 106 & 1 \\ 100 & 1 \end{bmatrix}
\quad
\mathbf{Q} = \begin{bmatrix} 1.00 & 0.00 & 0.00 \\ 0.50 & 0.50 & 0.00 \\ 0.00 & 0.00 & 1.00 \\ 0.75 & 0.25 & 0.00 \\ 0.50 & 0.00 & 0.50 \\ 0.00 & 0.50 & 0.50 \\ 0.25 & 0.25 & 0.50 \end{bmatrix}
$$

Variance components are $\sigma^2_{AL} = 80$, $\sigma^2_{AI} = 120$ and $\sigma^2_{ALI} = 50$. Using [7], the diagonal elements of matrix $\mathbf{G}_\varepsilon$ (the variances of Mendelian residuals) are 80, 80, 120, 40, 50, 100 and 100 for animals 1–7 respectively. Residual variance is $\mathbf{R} = \mathbf{I}_6(400)$. Matrix $\mathbf{G}$ is:

$$
\mathbf{G} = \begin{bmatrix}
80 & 0 & 0 & 40 & 40 & 0 & 20 \\
0 & 80 & 0 & 40 & 0 & 0 & 0 \\
0 & 0 & 120 & 0 & 60 & 0 & 30 \\
40 & 40 & 0 & 80 & 20 & 0 & 10 \\
40 & 0 & 60 & 20 & 100 & 0 & 50 \\
0 & 0 & 0 & 0 & 0 & 100 & 50 \\
20 & 0 & 30 & 10 & 50 & 50 & 150
\end{bmatrix}
$$

MME are equal to the following matrix:

$$
\begin{bmatrix}
154.9 & 0.558 & 0 & 0 & 0 & 0 & 0.25 & 0.275 & 0.238 & 0.245 & 0.265 & 0.25 \\
0.558 & 0.015 & 0 & 0 & 0 & 0 & 0.0025 & 0.0025 & -0.0025 & -0.025 & 0.0025 & 0.0025 \\
0 & 0 & 0.015625 & 0.0031 & 0 & -0.0125 & -0.0063 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.0031 & 0.0056 & 0.0025 & 0 & 0 & 0 & 0 & 0 & -0.005 & 0 \\
0 & 0 & 0 & 0.0025 & 0.0108 & 0 & 0 & -0.0083 & 0 & 0 & -0.005 & 0 \\
0 & 0 & -0.0125 & 0 & 0 & 0.0238 & 0.0063 & 0.005 & -0.0125 & -0.01 & 0 & 0 \\
0.25 & 0.0025 & -0.006 & -0.0063 & 0 & 0.0063 & 0.02125 & 0 & -0.0125 & 0 & 0 & 0 \\
0.275 & 0.0025 & 0 & 0 & -0.0083 & 0.005 & 0 & 0.0158 & 0 & -0.01 & 0 & 0 \\
0.238 & -0.0025 & 0 & 0 & 0 & -0.0125 & -0.0125 & 0 & 0.0275 & 0 & 0 & 0 \\
0.245 & -0.025 & 0 & 0 & 0 & -0.01 & 0 & -0.01 & 0 & 0.025 & 0.0025 & -0.005 \\
0.265 & 0.0025 & 0 & -0.005 & -0.005 & 0 & 0 & 0 & 0 & 0.0025 & 0.015 & -0.005 \\
0.25 & 0.0025 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.005 & -0.005 & 0.0125
\end{bmatrix}
\begin{bmatrix}
b(\text{age}) \\ s1 \\ gA1 \\ gA2 \\ gB \\ a1 \\ a2 \\ a3 \\ a4 \\ a5 \\ a6 \\ a7
\end{bmatrix}
=
\begin{bmatrix}
190.525 \\ 0.2175 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0.25 \\ 0.2575 \\ 0.4 \\ 0.4375 \\ 0.2625 \\ 0.285
\end{bmatrix}
$$

Solutions are -2.903201, -28.95692, 402.73785, 431.59906, 452.07844, 402.68086, 417.28243, 452.16393, 409.6967, 427.70734, 441.69628 and 434.41686. The large absolute values of the solutions are due to multicollinearity associated with genetic groups in the model for the small example worked. This is evidenced by a small eigenvalue (4.36×10^{-7}) in the coefficient matrix. Van Vleck (1990) obtained a similar result in an example with genetic groups for direct and maternal effects. If groups are left out of the model, solutions are 1.3540226 for the age effect, -36.19053 for the sex 1 effect, and 0.602659, -0.247433, -1.030572, -0.276959, 1.1075823, 0.6457529, 3.6589865 for the BV of animals 1–7, respectively. The smallest eigenvalue is 0.0046413, almost 10 000 times larger than the situation where genetic groups are in the model.

The consequences of assuming an incorrect $\mathbf{G}$ can be seen by taking $\mathbf{G}^* = \mathbf{A}(100)$. The value of 100 for σ^2_{A*} is chosen because it is the average between $\sigma^2_{AL} = 80$ and $\sigma^2_{AI} = 120$. To alleviate the problem associated with multicollinearity, the system is solved using regular MME (Henderson, 1984) rather than the QP-transformed system [8]. Therefore, PEV are estimated for BV deviated from their means. Incorrect PEV for animals 1–7 are 99.91, 99.66, 99.91, 98.30, 98.66, 99.66 and 99.91, respectively. Whereas true PEV for the same animals computed by means of [11] or [12] (or direct inversion of the MME) are: 79.94, 79.79, 119.88, 78.98, 98.52, 99.66 and 149.59 for animals 1–7, respectively.

DISCUSSION

A model has been presented to predict BV of different crosses between 2 populations under an additive type of inheritance. It allows for different additive means and variances. Computations are as simple as when there is only 1 σ^2_A and, as usual, $\mathbf{R}$ is a diagonal matrix. A practical application is the analysis of data from crosses between 'foreign' and 'local' strains of a breed, as in dairy or beef breeding. Also, records from registered *vs* grade animals, or 'selected *vs* unselected', *etc*, can be analyzed in this fashion. Although the developments presented were in terms of 2 populations, inclusion of more than 2 can be done as indicated by Lo *et al* (1993). With p being the number of populations, the number of parameters in $\mathbf{G}$ is $[p(p+1)]/2$, so that for $p = 4$ there are 10 variances to consider. Some of these estimates may be highly correlated depending on the type and distribution of the crosses involved.

The approach taken in the present paper differs from Elzo (1990) in the inclusion of the segregation variance (σ^2_{ALI}). The magnitude of this parameter depends on differences in gene frequencies between the 2 populations (Lo *et al*, 1993). The change in gene frequency due to selection is inversely related to the number of loci because change in gene frequency at a locus due to selection is proportional to the magnitude of the average effect of gene substitution at that locus (Pirchner, 1969, page 145), and the magnitude of average effects across loci tend to be inversely related to the number of loci. Thus, σ^2_{ALI} due to different selection criteria in 2 populations is expected to be inversely related to the number of loci. The change in gene frequency due to other forces (mutation, migration and random drift) is not related to the magnitude of average effects. Thus, σ^2_{ALI} due to differences in gene frequency between populations brought about by these forces is not related to the

number of loci. Now, the greater the value of σ_{ALI}^2, the larger the difference between the predictors calculated following the approach of Elzo (1990) and the one used in the present work. This is due to σ_{ALI}^2 not only entering into the diagonal elements of $\mathbf{G}$, but also into off-diagonals which are functions of the diagonal elements (Lo *et al*, 1993). For example, consider the additive covariance between paternal half sibs (cov(PHS)) i and i', from common sire s and unrelated dams. By repeated use of expression [10] in Lo *et al* (1993), cov(PHS) is equal to:

$$\text{cov}(a_i, a_{i'}) = \frac{1}{2}\,\text{cov}(a_i, a_s) = \frac{1}{4}\,\text{cov}(a_s, a_s) = \frac{1}{4}\,\text{Var}(a_s) \qquad [17]$$

Expression [17] shows that cov(PHS) is a function of the additive variance of the sire, and is not a function of the additive variance present in progeny genotypes. For I, F1, BI or F2 sires, cov(PHS) are equal to:

Sire	cov(PHS)
I	$0.2500\ \sigma_{\text{AI}}^2$
F1	$0.1250\ \sigma_{\text{AI}}^2 + 0.1250\ \sigma_{\text{AL}}^2$
BI	$0.1875\ \sigma_{\text{AI}}^2 + 0.0625\ \sigma_{\text{AL}}^2 + 0.1250\ \sigma_{\text{ALI}}^2$
F2	$0.1250\ \sigma_{\text{AI}}^2 + 0.1250\ \sigma_{\text{AL}}^2 + 0.2500\ \sigma_{\text{ALI}}^2$

Note that σ_{ALI}^2 enters into the covariance of individuals whose sire belongs to later generations than the F1 (*eg*, BI or F2). It must be pointed out that although predictions are still unbiased, ignoring σ_{ALI}^2 would result in a larger shrinkage of predictions in [8] than otherwise. As there are no estimates of σ_{ALI}^2 so far, nothing can be said about the magnitude of the parameter on genetic variation of economic traits in livestock.

If dominance is not null, the model should be properly modified to take into account this non-additive genetic effect. Proper specification of the variance–covariance matrix for additive and dominance effects in crosses of 2 populations can involve as many as 25 parameters for a single trait (Lo *et al*, 1995). Therefore, predictors of BV and dominance deviations may be difficult to compute for a general situation, involving animals from several crossbred genotypes. Lo (1993) presented an efficient algorithm for computing BLUP in the case of 2- and 3-breed terminal crossbreeding systems under additive and dominance inheritance.

Up to this point the animal model has been employed. However, the 'reduced animal model' (Quaas and Pollak, 1980) can alternatively be used by properly writing matrix $\mathbf{Z}$ with 1/2's, whenever a BV of a 'non-parent' (an animal that has no progeny in the data set) is expressed as a function of its parental BV. Residual genetic variances are obtained by means of expression [7].

In order to solve equations [8] the parameters σ_{AL}^2, σ_{AI}^2, σ_{ALI}^2, and the residual components, have to be known. Usually variance components are unknown and should be estimated from the data. Elzo (1994) developed expressions for restricted maximum likelihood (REML) estimators of variance components (including σ_{ALI}^2) in multibreed populations, through the expectation–maximization algorithm.

CONCLUSIONS

For 1 or several traits governed by additive effects, predictions of BV from crosses between 2 populations can be obtained by means of animal models that allow for different additive means and heterogeneous additive genetic (co)variances. Calculations required are similar to those with homogeneous additive (co)variances.

ACKNOWLEDGMENT

REFERENCES

Bulmer MG (1985) *The Mathematical Theory of Quantitative Genetics*. Clarendon Press, Oxford, UK

Cantet RJC, Fernando RL, Gianola D, Misztal I (1992) Genetic grouping for direct and maternal effects with differential assignment of groups. *Genet Sel Evol* 24, 211-223

Emik LO, Terril CE (1949) Systematic procedures for calculating inbreeding coefficients. *J Hered* 40, 51-55

Elzo M (1990) Recursive procedures to compute the inverse of the multiple trait additive genetic covariance matrix in inbred and noninbred multibreed populations. *J Anim Sci* 68, 1215-1228

Elzo M (1994) Restricted maximum likelihood procedures for the estimation of additive and nonadditive genetic variances and covariances in multibreed populations. *J Anim Sci* 72, 3055-3065

Fernando RL, Gianola D (1990) Statistical inferences in populations undergoing selection or non-random mating. *In: Advances in Statistical Methods for Genetic Improvement of Livestock* (D Gianola, K Hammond, eds), Springer-Verlag, New York, USA

Goffinet B (1983) Selection on selected records. *Genet Sel Evol* 15, 91-98

Groeneveld E, Kovac M (1990) A generalized computing procedure for setting up and solving mixed linear models. *J Dairy Sci* 73, 513-531

Henderson CR (1975a) Comparison of alternative sire evaluation methods. *J Anim Sci* 41, 760-770

Henderson CR (1975b) Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31, 423-447

Henderson CR (1976) A simple method for computing the inverse of a numerator matrix used in the prediction of breeding values. *Biometrics* 32, 69-84

Henderson CR (1984) *Applications of Linear Models in Animal Breeding*. University of Guelph Press, 2nd edition, Guelph, ON, Canada

Lande R (1981) The minimum number of genes contributing to quantitative variation between and within populations. *Genetics* 99, 541-553

Lo LL (1993) Genetic evaluation and selection in multibreed population. PhD thesis, University of Illinois, Champaign-Urbana, IL, USA

Lo LL, Fernando RL, Grossman MG (1993) Genotypic covariance between relatives in multibreed populations: additive model. *Theor Appl Genet* 87, 423-430

Lo LL, Fernando RL, Cantet RJC, Grossman MG (1995) Theory for modelling means and covariances in a 2 breed population with dominance inheritance. *Theor Appl Genet* 90, 49-62

Pirchner F (1969) *Population Genetics in Animal Breeding.* Freeman, San Franciso, CA, USA

Quaas RL (1988) Additive models with groups and relationships. *J Dairy Sci* 71, 1338-1345

Quaas RL, Pollak J (1980) Mixed model methodology for farm and ranch beef cattle testing programs. *J Anim Sci* 51, 1277-1287

Seber GAF (1977) *Linear Regression Analysis.* J Wiley & Sons, New York, USA

Thompson R (1979) Sire evaluation. *Biometrics* 35, 339-353

Toutenburg H (1982) *Prior Information on Linear Models.* J Wiley & Sons, New York, USA

Van Vleck LD (1990) Breeding value prediction with maternal genetic groups. *J Anim Sci* 63, 3998-4013

Van Vleck LD (1993) Variance of prediction error with mixed-model equations when relationships are ignored. *Theor Appl Genet* 85, 545-549

Westell RA, Quaas RL, Van Vleck LD (1988) Genetic groups in an animal model. *J Dairy Sci* 71, 1310-1318

Willham RL (1963) The covariance between relatives for characters composed of components contributed by related individuals. *Biometrics* 19, 18-27

Wright S (1968) *Evolution and the Genetics of Populations. Vol 1. Genetics and Biometrical Foundations.* University of Chicago Press, Chicago, USA