REDUCED ANIMAL MODEL FOR MARKER ASSISTED SELECTION

USING BEST LINEAR UNBIASED PREDICTION

R.J.C.Cantet[1] and C.Smith

Centre for Genetic Improvement of Livestock,

Department of Animal and Poultry Science,

University of Guelph, Guelph, Ontario, N1G 2W1, Canada

Running title: Reduced animal model for MAS

---

[1]. On leave from Departamento de Zootecnia, Facultad de Agronomía, Universidad de Buenos Aires, Argentina.

Summary

A reduced animal model (RAM) version of the animal model (AM) incorporating independent marked quantitative trait loci (MQTL's) of Fernando and Grossman (1989) is presented. Both AM and RAM permit obtaining Best Linear Unbiased Predictions of MQTL effects plus the remaining portion of the breeding value that is not accounted for by independent MQTL's. RAM reduces computational requirements by a reduction in the size of the system of equations. Non-parental MQTL effects are expressed as a linear function of parental MQTL effects using marker information and the recombination rate (r) between the marker locus and the MQTL. The resulting fraction of the MQTL variance that is explained by the regression on parental MQTL effects is $2[(1-r)^2 + r^2]/2$ when the individual is not inbred and both parents are known. Formulae are obtained to simplify the computations when backsolving for non-parental MQTL and breeding values in case all non-parents have one record. A small numerical example is also presented.

Key words: marker assisted selection, best linear unbiased prediction, reduced animal model, genetic marker.

Résumé

Un modèle animal réduit pour la sélection assistée de marqueurs avec BLUP

Une version du modèle animal réduit (RAM) basée sur le modèle animal (AM) de Fernando et Grossman (1989) avec loci indépendants de caractères quantitatifs avec marqueurs (MQTL) est présentée. Dans les deux cas, RAM et AM, on obtient les meilleures prédictions linéaires sans biais (BLUP) des effets des MQTL en plus de la portion restante de la valeur d'élevage inexpliquée par les MQTL indépendants. L'emploi de RAM diminue les exigences de calcul par une réduction de la taille du système d'équations. Les effets des MQTL non-parentaux sont exprimés sous la forme d'une fonction linéaire des effets des MQTL parentaux à l'aide de l'information provenant du marqueur et du taux de recombinaison (r) entre le locus de marqueur et le MQTL. La proportion résultante de la variance du MQTL expliquée par la régression des effets des MQTL parentaux est donnée par l'expression $2[(1-r)^2 + r^2]/2$ dans le cas d'un individu non-consanguin avec parents connus. Des formules sont dérivées pour simplifier les calculs lorque l'on résout pour les effets des MQTL et les valeurs d'élevage non-parentaux dans le cas où tous les individus non-parents possèdent une seule observation. Un exemple numérique est également donné.

Mots-clés: sélection assistée de marqueurs, BLUP, modèle animal réduit, marqueur génétique

Introduction

In a recent paper, Fernando and Grossman (1989) obtained best linear unbiased predictors (Henderson, 1984) of the additive effects for alleles at a marked quantitative trait locus (MQTL) and of the remaining portion of the breeding value. They used an animal model (AM; Henderson, 1984) under a purely additive mode of inheritance. Letting p be the number of fixed effects in the model, n the number of animals in the pedigree file and m the number of MQTL's, the number of equations in the system for this AM is $p + n(2m + 1)$. For large m, n or both solving such a system may not always be feasible. The reduced animal model (RAM; Quaas and Pollak, 1980) is an equivalent model in the sense of Henderson (1985), to the AM and provides the same results, but with a smaller number of equations to be solved. In this paper, the RAM version of the model of Fernando and Grossman (1989) is obtained. The resulting system of equations is of order $p + s(2m + 1)$, s being the number of parents. In general s is much smaller than n. Therefore, the advantage due to the reduction in the number of equations by using RAM is considerable. A numerical example is included to illustrate the application.

Theory

For simplicity, derivations are presented for a model with one MQTL. The extension to the case of 2 or more independent MQTL's is covered in section 4.

In the notation of Fernando and Grossman (1989), M_i^p and M_i^m are alleles at the marker locus that individual i inherited from its paternal (p) and its maternal (m) parents, and v_i^p and v_i^m are the additive effects of the paternal and maternal MQTL's, respectively. The recombination frequency between the marker allele and the MQTL is denoted as r. We will use the expression "breeding value" to refer to the additive effects of all genes that affect the trait excluding the MQTL(s).

1. Matrix expressions for the animal model with genetic marker information.

A matrix version of equation (3) in Fernando and Grossman (1989) is:

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{W}\boldsymbol{v} + \boldsymbol{e}$$

[1]

where y is an nx1 vector of records, X, Z and W are nxp, nxn and nx2n incidence matrices which relate data to the unknown vector of fixed effects β, the random vector of additive breeding values u and the random vector v of additive effects of the individual MQTL effects, respectively. The 2nx1 vector v is ordered within animal such that v_i^p always precedes v_i^m. The matrices Z and W will have zero rows for animals that do not have records on themselves but that are related to animals with records. Non-zero rows of Z and W have one and two elements equal to 1, respectively, with the remaining elements being zero. First and second moments of y are given by:

$$\mathrm{E}(\mathbf{y}) = \mathbf{X}\beta \qquad \mathrm{Var}(\mathbf{y}) = \mathbf{Z}\mathbf{A}\mathbf{Z}'\sigma_A^2 + \mathbf{W}\mathbf{G_v}\mathbf{W}'\sigma_v^2 + \mathbf{I}\sigma_e^2$$

where $A\sigma_A^2$ and $G_v\sigma_v^2$ are the variance-covariance matrices of u and v, respectively. The scalars σ_A^2, σ_v^2 and σ_e^2 are the variance components of the additive effects of breeding values, the MQTL additive effects and of the environmental effects.

RAM requires partitioning the data vector y into records of individuals with progeny (y_P; parents) and records of individuals without progeny (y_N; non-parents) so that $y' = [y_P', y_N']$. A conformable partition can be used in X, Z, W, u, v and e. Using this idea [1] can be written as:

$$\begin{bmatrix} \mathbf{y}_P \\ \mathbf{y}_N \end{bmatrix} = \begin{bmatrix} \mathbf{X}_P \\ \mathbf{X}_N \end{bmatrix} \beta + \begin{bmatrix} \mathbf{Z}_P & 0 \\ 0 & \mathbf{Z}_N \end{bmatrix} \begin{bmatrix} \mathbf{u}_P \\ \mathbf{u}_N \end{bmatrix} + \begin{bmatrix} \mathbf{W}_P & 0 \\ 0 & \mathbf{W}_N \end{bmatrix} \begin{bmatrix} \mathbf{v}_P \\ \mathbf{v}_N \end{bmatrix} + \begin{bmatrix} \mathbf{e}_P \\ \mathbf{e}_N \end{bmatrix}$$

[2]

To obtain RAM, u_N and v_N should be expressed as linear functions of u_P and v_P, respectively. Since an individual's breeding value can be described as the average of the breeding value of its parents plus an independently distributed Mendelian sampling residual (ϕ) (Quaas and Pollak, 1980), for u_N we can write:

$$\mathbf{u}_N = \mathbf{P}\mathbf{u}_P + \phi_N$$

[3]

where *P* is an (n-s)xs matrix relating non-parental to parental breeding values. Each row of *P* contains at most two 0.5 values in the columns pertaining to the BV's of the sire and of the dam. Now, $E(\phi) = 0$ and $Var(\phi) = D_A\sigma^2_A$, where D_A is a diagonal matrix with diagonal elements equal to:

$1 - \frac{1}{4}(a_{ss} + a_{dd})$, if both sire and dam of the non-parent are known

$1 - \frac{1}{4}a_{ss}$, if only the sire is known

$1 - \frac{1}{4}a_{dd}$, if only the dam is known

1, if both parents are unknown

with a_{ss} and a_{dd} being the diagonal elements of *A* corresponding to the sire and the dam, respectively.

A scalar version of the relationship between v_N and v_P can be obtained from equations (8a) and (8b) in Fernando and Grossman (1989) and these are:

$$v_o^p = b_1 v_s^p + b_2 v_s^m + e_o^p$$

$$v_o^m = b_3 v_d^p + b_4 v_d^m + e_o^m$$

The subscripts o, s and d denote the individual, its sire and its dam, respectively. The coefficients b_i's are either 1-r or r according to any of these four possible patterns of inheritance of the marker alleles:

Paternal marker	Maternal marker	b_1	b_2	b_3	b_4
M_s^p	M_d^p	1-r	r	1-r	r
M_s^p	M_d^m	1-r	r	r	1-r
M_s^m	M_d^p	r	1-r	1-r	r
M_s^m	M_d^m	r	1-r	r	1-r

The above developments lead us to the following relationship between v_N and v_P:

$$v_N = F v_P + \epsilon$$

[4]

The $2(n-s) \times 2s$ matrix F relates the additive effects of the MQTL of non-parents to the additive effects of the MQTL of parents and ϵ is the vector with element i equal to the residual ϵ_o^p and element i+1 equal to the residual ϵ_o^m. Each row of F contains at most, two non-zero elements: the b_i's. Let i and k be the row indices for the MQTL marked by M_o^p and M_o^m, respectively. Let j and j+1 be the column indices corresponding to the additive effects of the MQTL for the sire that transmits i: j refers to the paternal grandsire and j+1 to the paternal granddam. Also, let l and l+1 be the column indices corresponding to the dam that transmits i+1: l corresponds to the maternal grandsire and l+1 to the maternal granddam. Then, $F_{i,j} = b_1$, $F_{i,j+1} = b_2$, $F_{k,l} = b_3$ and $F_{k,l+1} = b_4$. All remaining elements of F are 0. When marker information is unavailable, r is taken to be 0.5 (Fernando and Grossman, 1989) and

all b_i's are 0.5. To exemplify, consider individuals 1 (male), 2 (female) and 3 (progeny of 1 and 2). Animals 1 and 2 are unrelated and 3 has paternal and maternal marker alleles originating from the dams of 1 and 2, namely alleles M_1^d and M_2^d, respectively. Then, $v = [v_1^p, v_1^m, v_2^p, v_2^m, v_3^p, v_3^m]'$, with $v^P = [v_1^p, v_1^m, v_2^p, v_2^m]'$ and $v^M = [v_3^p, v_3^m]'$. The matrix W is:

$$
W = \begin{bmatrix} W_P & 0_{2\times2} \\ 0_{1\times4} & W_N \end{bmatrix} = \left[\begin{array}{cccc|cc} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{array} \right]
$$

For r $=$ 0.2, the matrix F is 2x4 and equal to:

$$
F = \begin{bmatrix} 0.2 & 0.8 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.2 & 0.8 \end{bmatrix}
$$

The residuals ϵ have $E(\epsilon) = 0$ and $Var(\epsilon) = G_\epsilon \sigma^2_v$. Fernando and Grossman (1989) showed that $G_\epsilon \sigma^2_v$ is diagonal with non-zero elements equal to $Var(\epsilon_o^p) = 2r(1-r)(1-f_s)\sigma^2_v$ and $Var(\epsilon_o^m) = 2r(1-r)(1-f_d)\sigma^2_v$, where f_s and f_d are the inbreeding coefficients at the MQTL of the sire and of the dam, respectively. They express the probability that the paternal and maternal alleles of an individual for a given MQTL are the same. These f's are the off-diagonal elements in the 2x2 diagonal blocks of the matrix G_v (Fernando and Grossman, 1989).

Using [3] and [4] in [2] gives:

$$\begin{bmatrix} \boldsymbol{Y}_P \\ \boldsymbol{Y}_N \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_P \\ \boldsymbol{X}_N \end{bmatrix} \beta + \begin{bmatrix} \boldsymbol{Z}_P & 0 \\ \boldsymbol{Z}_N P & \boldsymbol{Z}_N \end{bmatrix} \begin{bmatrix} \boldsymbol{u}_P \\ \phi \end{bmatrix} + \begin{bmatrix} \boldsymbol{W}_P & 0 \\ \boldsymbol{W}_N F & \boldsymbol{W}_N \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_P \\ \epsilon \end{bmatrix} + \begin{bmatrix} \boldsymbol{e}_P \\ \boldsymbol{e}_N \end{bmatrix}$$

[5]

or

$$\begin{bmatrix} \boldsymbol{Y}_P \\ \boldsymbol{Y}_N \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_P \\ \boldsymbol{X}_N \end{bmatrix} \beta + \begin{bmatrix} \boldsymbol{Z}_P \\ \boldsymbol{Z}_N P \end{bmatrix} \boldsymbol{u}_P + \begin{bmatrix} \boldsymbol{W}_P \\ \boldsymbol{W}_N F \end{bmatrix} \boldsymbol{v}_P + \begin{bmatrix} \boldsymbol{e}_P \\ \boldsymbol{e}_N + \boldsymbol{Z}_N \phi + \boldsymbol{W}_N \epsilon \end{bmatrix}$$

[6]

On letting $e^* = e_N + Z_N \phi + W_N \epsilon$, we have that:

$$\mathrm{Var} \begin{bmatrix} \boldsymbol{e}_P \\ \boldsymbol{e}^* \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q} \end{bmatrix} \sigma_e^2$$

where $Q = I_{(n-s)} + Z_N D_A Z_N' \alpha_A + W_N G_\epsilon W_N' \alpha_v$, $\alpha_A = \sigma_A^2/\sigma_e^2$ and $\alpha_v = \sigma_v^2/\sigma_e^2$. Mixed model equations for [6] are:

$$\begin{bmatrix} X'_P X_P + X'_N Q^{-1} X_N & X'_P Z_P + X'_N Q^{-1} Z_N P & X'_P W_P + X'_N Q^{-1} W_N F \\ Z'_P X_P + P' Z'_N Q^{-1} X_N & Z'_P Z_P + P' Z'_N Q^{-1} Z_N P + A_P^{-1} \alpha_A & Z'_P W_P + P' Z'_N Q^{-1} W_N F \\ W'_P X_P + F' W'_N Q^{-1} X_N & W'_P Z_P + F' W'_N Q^{-1} Z_N P & W'_P W_P + F' W'_N Q^{-1} W_N F + G_{vP}^{-1} \alpha_v \end{bmatrix}$$

$$\begin{bmatrix} \hat{\beta} \\ \hat{u}_P \\ \hat{v}_P \end{bmatrix} = \begin{bmatrix} X'_P Y_P + X'_N Q^{-1} Y_N \\ Z'_P Y_P + P' Z'_N Q^{-1} Y_N \\ W'_P Y_P + F' W'_N Q^{-1} Y_N \end{bmatrix}$$

[7]

The matrices A_P and G_{vP} are the corresponding submatrices of A and G_v that belong to parents. Equations [7] give the solutions for RAM with genetic markers. Of practical importance is the case where all non-parents have only one record so that $Z_N = I$. Then, $W_N G_\epsilon W_N'$ and Q^{-1} are diagonal (see Appendix A). The diagonal elements of $W_N G_\epsilon W_N'$ are derived in Appendix A and they are equal to:

$2r(1-r)(2-f_s-f_d)$, when both the sire and the dam of the non-parent are known

$2r(1-r)(1-f_s) + 1$, when only the sire is known

$2r(1-r)(1-f_d) + 1$, when only the dam is known

2, if both the sire and the dam of the non-parent are unknown

If there is zero probability that the paternal and maternal alleles at the MQTL of parent p are the same (i.e. $f_p = 0$), the contribution to the diagonal element of

$W_N G_\epsilon W_N'$ is 2r(1-r) (if marker information is available) or ½ (if marker information

is unavailable). This occurs because, in the absence of marker information, there is

equal probability of receiving the MQTL from the grandsire and from the granddam,

and r $=$ 0.5 (Fernando and Grossman, 1989).

A further simplification to [7] occurs when parents do not have records so that

Z_P and W_P are zero and the model becomes a sire-dam model. A program for RAM,

such as the one presented by Schaeffer and Wilton (1987) and modified to include

marker information can be employed to solve equations [7].

4. More than one MQTL

Multiple MQTL (k, say) can be dealt with assuming independence by the

following modification of model [1]:

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{Z}\boldsymbol{u} + (\boldsymbol{j}'_k \otimes \boldsymbol{W}) \begin{bmatrix} \boldsymbol{v}_1 \\ \cdots \\ \boldsymbol{v}_k \end{bmatrix} + \boldsymbol{e}$$

[8]

where j_k is a kx1 vector with all elements equal to 1. We will assume that Var(v_i) $=$

$G_{vi}\sigma^2_{vi}$ and Cov$(v_i, v_{i'})$ $=$ 0. For k $=$ 2 and letting $Q_* = I_{(n-s)} + Z_N D_A Z_N' \alpha_A + W_N(G_{\epsilon 1}$

$\alpha_{v1} + G_{\epsilon 2} \alpha_{v2}) W_N'$, RAM equations for [8] are:

$$
\begin{bmatrix}
\mathbf{X}'_p\mathbf{X}_p + \mathbf{X}'_N\mathbf{O}_*\mathbf{X}_N & \mathbf{X}'_p\mathbf{Z}_p + \mathbf{X}'_N\mathbf{O}_*\mathbf{Z}_p & \mathbf{X}'_p\mathbf{W}_1 + \mathbf{X}'_N\mathbf{O}_*\mathbf{W}_1 & \mathbf{X}'_p\mathbf{W}_2 + \mathbf{X}'_N\mathbf{O}_*\mathbf{W}_2 \\
\mathbf{Z}'_p\mathbf{X}_p + \mathbf{P}\,\mathbf{Z}'_N\mathbf{O}_*\mathbf{X}_N & \mathbf{Z}'_p\mathbf{Z}_p + \mathbf{P}\,\mathbf{Z}'_N\mathbf{O}_*\mathbf{Z}_p + \mathbf{A}\,\alpha_p^{-1} & \mathbf{Z}'_p\mathbf{W}_1 + \mathbf{P}\,\mathbf{Z}'_N\mathbf{O}_*\mathbf{W}_1 & \mathbf{Z}'_p\mathbf{W}_2 + \mathbf{P}\,\mathbf{Z}'_N\mathbf{O}_*\mathbf{W}_2 \\
\mathbf{W}'_1\mathbf{X}_p + \mathbf{W}'_1\mathbf{O}_*\mathbf{X}_N & \mathbf{W}'_1\mathbf{Z}_p + \mathbf{W}'_1\mathbf{O}_*\mathbf{Z}_p & \mathbf{W}'_1\mathbf{W}_1 + \mathbf{W}'_1\mathbf{O}_*\mathbf{W}_1 + \mathbf{G}\,\alpha_{v_1^p}^{-1} & \mathbf{W}'_1\mathbf{W}_2 + \mathbf{W}'_1\mathbf{O}_*\mathbf{W}_2 \\
\mathbf{W}'_2\mathbf{X}_p + \mathbf{W}'_2\mathbf{O}_*\mathbf{X}_N & \mathbf{W}'_2\mathbf{Z}_p + \mathbf{W}'_2\mathbf{O}_*\mathbf{Z}_p & \mathbf{W}'_2\mathbf{W}_1 + \mathbf{W}'_2\mathbf{O}_*\mathbf{W}_1 & \mathbf{W}'_2\mathbf{W}_2 + \mathbf{W}'_2\mathbf{O}_*\mathbf{W}_2 + \mathbf{G}\,\alpha_{v_2^p}^{-1}
\end{bmatrix}
$$

$$
\begin{bmatrix}
\beta \\
\beta_{u_p} \\
\beta_{v_1^p} \\
\beta_{v_2^p}
\end{bmatrix}
=
\begin{bmatrix}
\mathbf{X}'_p\mathbf{Y}_p + \mathbf{X}'_N\mathbf{O}_*\mathbf{Y}_N \\
\mathbf{Z}'_p\mathbf{Y}_p + \mathbf{P}\,\mathbf{Z}'_N\mathbf{O}_*\mathbf{Y}_N \\
\mathbf{W}'_1\mathbf{Y}_p + \mathbf{W}'_1\mathbf{O}_*\mathbf{Y}_N \\
\mathbf{W}'_2\mathbf{Y}_p + \mathbf{W}'_2\mathbf{O}_*\mathbf{Y}_N
\end{bmatrix}
\qquad [9]
$$

5. Backsolving for non-parents

After solving for fixed effects, parental breeding values and parental effects of the MQTL, the breeding values and additive MQTL effects of non-parents can be calculated. This is accomplished by writing the equations for $\hat{\phi}$ and $\hat{\epsilon}$ from the mixed model equations of [5]. This gives:

$$\boldsymbol{Z}'_N \boldsymbol{X}_N \hat{\beta} + \boldsymbol{Z}'_N \boldsymbol{Z}_N \boldsymbol{P}\hat{\boldsymbol{u}}_P + (\boldsymbol{Z}'_N \boldsymbol{Z}_N + \boldsymbol{D}_A^{-1} \alpha_A) \hat{\phi} + \boldsymbol{Z}'_N \boldsymbol{W}_N \boldsymbol{F}\hat{\boldsymbol{v}}_P + \boldsymbol{Z}'_N \boldsymbol{W}_N \hat{\boldsymbol{\epsilon}} = \boldsymbol{Z}'_N \boldsymbol{Y}_N$$

$$\boldsymbol{W}'_N \boldsymbol{X}_N \hat{\beta} + \boldsymbol{W}'_N \boldsymbol{Z}_N \boldsymbol{P}\hat{\boldsymbol{u}}_P + \boldsymbol{W}'_N \boldsymbol{Z}_N \hat{\phi} + \boldsymbol{W}'_N \boldsymbol{W}_N \boldsymbol{F}\hat{\boldsymbol{v}}_P + (\boldsymbol{W}'_N \boldsymbol{W}_N + \boldsymbol{G}_e^{-1} \alpha_v) \hat{\boldsymbol{\epsilon}} = \boldsymbol{W}'_N \boldsymbol{Y}_N$$

and after a little algebra:

$$\begin{bmatrix} \boldsymbol{Z}'_N \boldsymbol{Z}_N + \boldsymbol{D}_A^{-1} \alpha_A & \boldsymbol{Z}'_N \boldsymbol{W}_N \\ \boldsymbol{W}'_N \boldsymbol{Z}_N & \boldsymbol{W}'_N \boldsymbol{W}_N + \boldsymbol{G}_e^{-1} \alpha_v \end{bmatrix} \begin{bmatrix} \hat{\phi} \\ \hat{\boldsymbol{\epsilon}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{Z}'_N \\ \boldsymbol{W}'_N \end{bmatrix} (\boldsymbol{Y}_N - \boldsymbol{X}_N \hat{\beta} - \boldsymbol{P}\hat{\boldsymbol{u}}_P - \boldsymbol{W}_N \boldsymbol{F}\hat{\boldsymbol{v}}_P)$$

[10]

Appendix B shows how to obtain solutions of equations [10], when all non-parents have one record, by solving (n-s) independent systems of order 2. Using the predictors obtained from [7] and [10] in [3] and [4], solutions for non-parents are:

$$\hat{u}_N = P\hat{u}_P + \hat{\phi}_N \qquad \hat{v}_N = F\hat{v}_P + \hat{e}$$

Example

We use the same data that Fernando and Grossman (1989) employed. There are 4 individuals, 3 of them are parents and 1 is a non-parent. The file is:

Individual	Sire	Dam	Marker inheritance:	
			Sire	Dam
.	.	.		
1	–	–	–	–
2	–	–	–	–
3	1	2	M_1^p	M_2^m
4	1	3	–	M_3^p

Notice that individual 4 is inbred. A fixed effect was included and the matrix resulting from adjoining the incidence matrix X and the vector of observations y, i.e. $[X/y]$ is:

$$
[X|y] = \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array}
\left[\begin{array}{cc|c}
1 & 0 & 235 \\
1 & 0 & 210 \\
0 & 1 & 250 \\
0 & 1 & 255
\end{array} \right]
$$

Variance components used were $\sigma^2_A = 100$, $\sigma^2_v = 10$ and $\sigma^2_e = 500$ and $r = 0.1$.

The matrices G_v and G^{-1}_v are presented in Fernando and Grossman (1989).

First, solutions for AM were obtained. The coefficient matrix for AM is:

$$
\begin{bmatrix}
2 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & \tfrac{1}{1} & 0 & 2 & 5 & \tfrac{1}{2} & 5 & \tfrac{1}{5} & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 2 & 5 & 8 & 5 & \tfrac{1}{5} & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & \tfrac{1}{2} & 5 & \tfrac{1}{5} & 0 & \tfrac{1}{3} & 5 & \tfrac{1}{5} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & \tfrac{1}{5} & 0 & 0 & 0 & \tfrac{1}{5} & 0 & \tfrac{1}{1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & \tfrac{1}{3} & 0 & \tfrac{1}{5} & 0 & 0 & 0 & 0 & 0 & \tfrac{23}{1} & 0 & 0 & 0 & 0 & 0 & \tfrac{1}{5} & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & \tfrac{1}{5} & 0 & 8 & 7 & 0 & 0 & 0 & 0 & \tfrac{1}{2} & 7 & 0 & 0 & \tfrac{1}{5} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 3 & 7 & \tfrac{26}{1} & 0 & 0 & 0 & \tfrac{21}{1} & 7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \tfrac{26}{1} & 0 & \tfrac{26}{1} & 0 & 0 & 0 & \tfrac{31}{1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & \tfrac{23}{1} & 0 & \tfrac{1}{2} & 7 & 0 & 0 & 0 & 0 & \tfrac{31}{1} & 7 & \tfrac{26}{1} & 0 & 0 & 0 & \tfrac{31}{1} & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & \tfrac{1}{2} & 7 & \tfrac{23}{1} & 0 & \tfrac{26}{1} & 0 & \tfrac{28}{1} & 0 & 5 & \tfrac{1}{1} & 0 & 0 & 7 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \tfrac{1}{5} & 0 & \tfrac{1}{5} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \tfrac{1}{1} & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \tfrac{23}{1} & 0 & \tfrac{1}{2} & 7 & 1 & 0 & 7 & \tfrac{22}{2} \\
\end{bmatrix}
$$

and the right-hand side vector is [445, 505, 235, 210, 250, 255, 235, 235, 210, 210, 250, 250, 255, 255]'. The vector of solutions is [222.5, 251.764, 2.08109, -2.08109, -0.083214, 1.16537, 0.213435, 0.216098, -0.202783, -0.226749, 0.213102, -0.229745, 0.231409, 0.174809].

There are 11 equations in the system for RAM (as compared to 14 in AM) since there is only one non-parent (individual 4) and Q is a scalar: $1.1136 = 1 + (0.5/\alpha_A) + 2 [0.5 (0.5) + (0.9) (0.1)] / \alpha_v$. The vector of right-hand sides for equations [7] is [445, 478.987, 349.494, 210, 364.494, 349.494, 210, 210, 456.088, 272.899]' and the coefficient matrix is:

Solutions for RAM are 222.5, 251.764, 2.08109, -2.08109, -0.083214, 0.213435, 0.216098, -0.202783, -0.226749, 0.213102 and -0.229745. The next step is to backsolve for individual 4 (non-parent) using equation [B.2]. Since both parents of 4 are known, $d_{A44} = 0.5$ and $d_{o44} = 5/(5 + 0.5) = 10/11 = 0.90909...$ The diagonal elements of the 2x2 system in [B.2] are functions of r. However, as the information from the sire marker is unavailable, $r = 0.5$ for the first diagonal element. Also, $f_1 = f_3 = 0$ and $d_{o44} y_4^* = 0.90909 [255 - 251.764 - \frac{1}{2}(2.081090 + 0.083214 + 0.213435 + 0.216098) - 0.9(0.213102) - 0.1(-0.229745)] = 1.6848545$. For animal 4, we then have:

$$
\begin{bmatrix}
0.90909 + 100 & 0.90909 \\
0.90909 & 0.90909 + 277.7777
\end{bmatrix}
\begin{bmatrix}
\hat{e}_1^p \\
\hat{e}_3^m
\end{bmatrix}
=
\begin{bmatrix}
1.6848545 \\
1.6848545
\end{bmatrix}
$$

which has solutions $\hat{\epsilon}_1^p = 0.0166428$ and $\hat{\epsilon}_3^m = 0.00599141$. Putting these into [B.3] gives $\hat{\phi}_4 = 0.166428$. Therefore, BLUP(u_4) = $\frac{1}{2}$ BLUP(u_1) + $\frac{1}{2}$ BLUP(u_3) + BLUP(ϕ_4) = $\frac{1}{2}$ [2.08109 + (-0.083214)] + 0.166428 = 1.16537. Also, BLUP(v_4^p) = $\frac{1}{2}$ BLUP(v_1^p) + $\frac{1}{2}$ BLUP(v_1^m) + BLUP(ϵ_4^p) = $\frac{1}{2}$ [0.213435 + 0.216098] + 0.0166428 = 0.231409 and BLUP(v_4^m) = 0.9 BLUP(v_3^p) + 0.1 BLUP(v_3^m) + BLUP(ϵ_4^m) = 0.9 (0.213102) + 0.1 (-0.229745) + 0.0059141 = 0.174809. As expected, solutions obtained by both AM and RAM are the same.

Discussion

The advantage of RAM over AM increases as both the ratio between the number of non-parents to the number of parents and the number of independent MQTL increase. Goddard (1991) suggested the use of RAM to decrease the size of the resulting system of equations when working with information on flanking markers.

As shown in Appendix A and for a non-inbred individual, the fraction of the variance of the MQTL that is due to Mendelian segregation is $4r(1-r)/2$. Now, $1 = (r+1-r)^2 = r^2 + 2r(1-r) + (1-r)^2$, so that $2[1-2r(1-r)] = 2[r^2 + (1-r)^2]$. Therefore, the fraction of the variance of the MQTL that is explained by parental segregation is $2[r^2 + (1-r)^2]/2$. These proportions can also be worked out from equations (8a) and (8b) in Fernando and Grossman (1989) and they agree with formulae derived by Dekkers and Dentine (1991). A slight difference between their result and the one obtained here stems from the fact that they define the variance of the MQTL as one half the variance as defined by Fernando and Grossman (1989) ($\frac{1}{2}\sigma^2_v$).

Both AM and RAM rest on knowing the variance components as well as the recombination rate between the marker gene and the QTL. As the latter parameter enters into the variance-covariance matrix of QTL effects in a rather complex manner, its estimation by the classical methods employed in animal breeding seems to be difficult, as discussed by Fernando (1990).

When more than one MQTL is being considered, covariances between pairs

of MQTL effects are likely to be non-zero due to linkage disequilibrium caused by selection (Bulmer, 1985). Model [8] assumes that these covariances are zero. The extent of the error in predicting v (or functions of v) due to incorrectly assuming null covariances between MQTL effects will depend on the magnitude and sign of the covariance. If the covariances are mostly negative, which is likely to happen on a trait undergoing selection (Bulmer, 1985), MQTL effects may be overpredicted. Research is in progress to overcome this restriction of model [8].

Appendix A: Derivation of the diagonal elements of $W_N G_\epsilon W_N{}'$

when all non-parents have one record

First we shown that $W_N G_\epsilon W_N{}'$ is diagonal. Because G_ϵ is diagonal (Fernando and Grossman, 1989), we can write:

$$W_N G_\epsilon W_N{}' = \sum_{j=1}^{2(n-s)} w_j w_j{}' g_j$$

where w_j is the column j of W_N and g_j is diagonal element j of G_ϵ . Now, w_j has all its elements equal to zero except for a 1 in position j. Therefore, the matrix $w_j w_j{}' g_j$ has all elements equal to zero except for element j,j which is equal to g_j. The paternal and maternal MQTL additive effects of an animal are in consecutive columns of the matrix W (and W_N), w_j and w_{j+1} say, and these are equal. We then have:

$$\sum_{j=1}^{2(n-s)} \mathbf{w}_j \mathbf{w}'_j g_j = \sum_{j=1,3,5...}^{2(n-s)} \mathbf{w}_j \mathbf{w}'_j (g_j + g_{j+1})$$

and $W_N G_\epsilon W_N'$ is diagonal with non-zero elements equal to $g_j + g_{j+1}$.

Now, $(g_j + g_{j+1})\sigma^2_v = \text{Var}(\epsilon_o^p) + \text{Var}(\epsilon_o^m) = 2r(1-r)(1-f_s) + 2r(1-r)(1-f_d)$, where f_s and f_d are the inbreeding coefficients of sire and dam for the MQTL, respectively. The last equality follows from expressions (12a) and (12b) in Fernando and Grossman (1989). After some rearranging, the diagonal element of $W_N G_\epsilon W_N'$ is:

$$2r(1-r)(2-f_s-f_d)$$

when both parents of the individual are known. If the sire is unknown, $\epsilon_o^p = v_o^p$ and the diagonal element is $2r(1-r)(1-f_d) + 1$. If the dam is unknown, $\epsilon_o^m = v_o^m$ and the diagonal element is $2r(1-r)(1-f_s) + 1$. If both parents are unknown the diagonal element of $W_N G_\epsilon W_N'$ is 2.

Appendix B: Solutions of equations [10] when non-parents have one record

When non-parents have one record $(Z_N = I)$, equations [10] reduce to:

$$
\begin{bmatrix} \boldsymbol{I} + \boldsymbol{D}_A^{-1}\alpha_A & \boldsymbol{W}_N \\ \boldsymbol{W}_N' & \boldsymbol{W}_N'\boldsymbol{W}_N + \boldsymbol{G}_e^{-1}\alpha_v \end{bmatrix} \begin{bmatrix} \hat{\phi} \\ \hat{e} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{W}_N' \end{bmatrix} (\boldsymbol{y}_N - \boldsymbol{X}_N\hat{\beta} - \boldsymbol{P}\hat{u}_p - \boldsymbol{W}_N\boldsymbol{F}\hat{v}_p)
$$

[B.1]

On absorbing the equations for $\hat{\phi}$ the solution for $\hat{\epsilon}$ is:

$$
(\boldsymbol{W}_N'[\boldsymbol{I} - (\boldsymbol{I} + \boldsymbol{D}_A^{-1}\alpha_A)^{-1}]\boldsymbol{W}_N + \boldsymbol{G}_e^{-1}\alpha_v)\,\hat{e} = \boldsymbol{W}_N'[\boldsymbol{I} - (\boldsymbol{I} + \boldsymbol{D}_A^{-1}\alpha_A)^{-1}](\boldsymbol{y}_N - \boldsymbol{X}_N\hat{\beta} - \boldsymbol{P}\hat{u}_p - \boldsymbol{W}_N\boldsymbol{F}\hat{v}_p)
$$

The matrix $D_o = I - (I + D^{-1}_A\alpha_A)^{-1}$ is diagonal with element d_{oii} being equal to:

$$
1 - \left(1 + \frac{\alpha_A}{d_{Aii}}\right)^{-1} = \frac{\alpha_A}{\alpha_A + d_{Aii}}
$$

and d_{Aii} is diagonal element i of D_A. Since W (and W_N) has rows with two consecutive elements equal to 1 and the rest equal to 0, $W_N'D_oW_N$ is block diagonal, each block being of order 2x2 with all elements equal to d_{oii}. Adding $G^{-1}_\epsilon\alpha_v$ gives the coefficient matrix on the left-hand side of [B.1] and solutions for $\hat{\epsilon}$ can be obtained by solving (n-s) systems of order 2. The system for animal i is equal to:

$$\left[\begin{array}{cc} d_{oii} + \dfrac{\alpha_v}{2\,r\,(1-r)\,(1-f_s)} & d_{oii} \\[4mm] d_{oii} & d_{oii} + \dfrac{\alpha_v}{2\,r\,(1-r)\,(1-f_d)} \end{array}\right] \left[\begin{array}{c} \hat{e}_i^p \\[2mm] \hat{e}_i^m \end{array}\right] = \left[\begin{array}{c} 1 \\ 1 \end{array}\right] d_{oii}\,\boldsymbol{y}_i^*$$

[B.2]

and y_i^* is element i of the vector y_N - $X\hat{\beta}$ - $P\hat{u}_P$ - $W_N F \hat{v}_P$.

After solving for $\hat{\epsilon}$, the first equation in [B.1] can be solved for as follows:

$$(\boldsymbol{I} + \boldsymbol{D}_A^{-1}\alpha_A)\ \hat{\phi} = \boldsymbol{y}^* - \boldsymbol{W}_N\hat{e}$$

Since the coefficient matrix of this system is diagonal, BLUP(ϕ_i) is:

$$\hat{\phi}_i = \left(\dfrac{d_{Aii}}{d_{Aii} + \alpha_A}\right)\ (\boldsymbol{y}_i^* - \hat{e}_i^p - \hat{e}_i^m)$$

[B.3]

If there is more than one MQTL the matrix of system [10] becomes poorly conditioned. The reason is that all off-diagonal elements are equal to 1 and the diagonals are relatively large (may be in the order of the hundreds, depending on the α's). An exact solution can be obtained by writing the matrix of system [10] for each animal as $jj' + S$, where j is a $1 + 2m$ vector with all elements equal to one and S is a diagonal matrix. Using the inverse of the sum of matrices formula (Henderson and

Searle, 1981), we have that:

$$(\boldsymbol{jj}' + \boldsymbol{S})^{-1} = \boldsymbol{S}^{-1} - \boldsymbol{S}^{-1}\boldsymbol{j}(\boldsymbol{j}'\boldsymbol{S}^{-1}\boldsymbol{j}+1)^{-1}\boldsymbol{j}'\boldsymbol{S}^{-1}$$

[B.4]

Notice that the expression in parentesis on the right-hand side of [B.4] is a scalar. Using [B.4] the inverse of the matrix of system [10] when there are m MQTL's for a non-parent is:

$$(\boldsymbol{jj}' + \boldsymbol{S})^{-1} = \begin{bmatrix} S_1(1 - \dfrac{S_1}{g}) & \dfrac{-S_1 S_2}{g} & \dfrac{-S_1 S_2}{g} & \cdots \\[2ex] \dfrac{-S_2 S_1}{g} & S_2(1 - \dfrac{S_2}{g}) & \dfrac{-S_2 S_3}{g} & \cdots \\[2ex] \dfrac{-S_3 S_1}{g} & \dfrac{-S_3 S_2}{g} & S_3(1 - \dfrac{S_3}{g}) & \cdots \\[2ex] \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

[B.5]

and $g = 1 + S_1 + S_2 + \ldots + S_{2m}$. The S's are such that $S_1 = D_{Aii}\alpha_A^{-1}$, $S_2 = G_{\epsilon is}\alpha_{vI}^{-1}$ $= 2r_1(1-r_1)(1-f_{1s})\alpha_{vI}^{-1}$, $S_3 = G_{\epsilon id}\alpha_{vI}^{-1} = 2r_1(1-r_1)(1-f_{1d})\alpha_{vI}^{-1}$, and so on for MQTL's 2 to m. Expression [B.5] is easy to program, does not require iteration and, more importantly, it is not subject to the numerical problems that occur when solving such a system of equations.

Acknowledgement

References

Bulmer MG (1985) *The mathematical theory of quantitative genetics.* Clarendon Press, Oxford

Dekkers JCM, Dentine MR (1991) Quantitative genetic variance associated with chromosomal markers in segregating populations. Theor Appl Gen (in press)

Falconer DS (1990) *Introduction to quantitative genetics.* Longman, London

Fernando RL (1990) Statistical problems in marker assisted selection for QTL. *In: 4th. World Congress on Genetics Applied to Livestock Production, Edinburgh, 23-27 July 1990,* vol. 13 (WG Hill, R Thompson, JA Woolliams, eds), pp 433-436

Fernando RL, Grossman M (1989) Marker assisted selection using best linear unbiased prediction. Gen Sel Evol 21, 467-477

Goddard ME (1991) A mixed model for analyses of data on multiple genetic markers. Theor Appl Gen (submitted)

Henderson CR (1984) *Applications of linear models in animal breeding.*

University of Guelph, Guelph

Henderson CR (1985) Equivalent linear models to reduce computations. J Dairy Sci 68,2267-2277

Henderson HV, SR Searle (1981) On deriving the inverse of a sum of matrices. SIAM Review 23,53-59

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

Schaeffer LR, JW Wilton (1987) RAM computing strategies and multiple traits. *In: Prediction of Genetic Values for Beef Cattle: Proceedings of Workshop II, Kansas City, 10-11 March 1987.* Winrock International institute for Agricultural Development, Morrilton, Arkansas, USA, pp 25-52.