

Genetic evaluation with uncertain parentage: a comparison of methods

M. Perez-Enciso* and R. L. Fernando

Department of Animal Sciences, 1207 W. Gregory Drive, University of Illinois, Urbana IL 61801, USA

Received July 1, 1991; Accepted October 28, 1991

Communicated by L. D. van Vleck

Summary. The most common method for genetic evaluation when parents are unknown is best linear unbiased prediction with genetic groups (BLUP-G). With this method unknown parents are assumed to be unrelated to any other animals in the population. This assumption is unrealistic in most situations. If a finite number of potential parents can be identified and the probabilities of being the true parent can be assigned to these, genetic evaluation can be obtained given the uncertainty of parentage without introducing genetic groups into the model. The correct numerator relationship matrix with uncertain parentage (\bar{A}) is derived. Rules are given to efficiently compute \bar{A} and \bar{A}^{-1} . Computer simulation was used to compare BLUP-G with BLUP using \bar{A} . The simulated population consisted of ten sires and 200 dams per breeding season. The dams were always known; the sires were unknown for 10% or 30% of the males and 30% of the females. The number of potential sires was three (BLUP- \bar{A}_1) or ten (BLUP- \bar{A}_2), including the true sire in both cases. Equal probabilities were assigned to each potential sire. The increase in response with BLUP- \bar{A}_1 and BLUP- \bar{A}_2 relative to BLUP-G ranged from 4% to 8% in the fifth breeding season. Selection with BLUP- \bar{A}_1 or BLUP- \bar{A}_2 resulted in higher inbreeding, 17% and 12%, respectively, than with BLUP-G. Estimates of response to selection were unbiased with BLUP- \bar{A}_1 and BLUP- \bar{A}_2 , but not unbiased with BLUP-G. Mean square error of estimated genetic means and mean prediction error variance were higher with BLUP-G than with BLUP- \bar{A}_1 or BLUP- \bar{A}_2 .

Key words: Genetic evaluation – Unknown parentage

* Correspondence to: M. Perez-Enciso

Present address: Department of Meat and Animal Science, University of Wisconsin, 1675 Observatory Drive, Madison, WI 53706-1284, USA

Introduction

Animals with unknown parents are common in most breeding programs. Unknown parentage can affect genetic progress in two ways. First, it can reduce selection intensity if animals with unknown parents are not considered for selection. Second, with uncertain parentage, the accuracy of evaluation will be lower. Additionally, mis-identification of parents can bias downwards estimates of heritability (Van Vleck 1970).

The most common method for genetic evaluation when the parents of animals are unknown is based on assigning 'phantom parents' to these animals (Westell et al. 1988). These phantom parents are classified into genetic groups, usually according to path of selection and date of birth (Wiggans et al. 1988). Evaluations are then obtained by best linear unbiased prediction (BLUP), assuming a mixed linear model for phenotypes and with genetic groups as fixed effects in addition to the usual fixed and random effects. An implicit assumption in this model is that the phantom parents are unrelated and noninbred (Quaas 1988). In most situations, however, it would be unrealistic to assume that unknown parents are unrelated to any other individuals in the population. Thus, in most situations, evaluations obtained as above, hereafter referred to as BLUP-G, would not have the properties of BLUP.

Genetic evaluations, however, can be obtained without making the assumption that unknown parents are unrelated to any individuals in the population (Poivey and Elsen 1984; Foulley et al. 1987; Henderson 1988; Dempfle 1990). In these methods, probabilities of parentage are assigned to a finite number of potential candidates. Evaluations are then obtained given this uncertainty of parentage, without introducing genetic groups into the model.

The relationship between these two methods of genetic evaluation when the parents of some individuals are unknown, however, has not been examined previously. The objectives of the work presented here are: (1) to generalize and formalize some previous results on evaluation with uncertain parentage (Poivey and Elsen 1984; Henderson 1988; Dempfle 1990), and (2) to compare response to selection using BLUP-G with those obtained using BLUP with uncertain parentage (Henderson 1988).

Theory

Model

Consider the animal model

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

where \mathbf{y} is a vector of observations, $\boldsymbol{\beta}$ is a vector of fixed effects, \mathbf{X} and \mathbf{Z} are known incidence matrices, and \mathbf{u} and \mathbf{e} are random vectors of additive breeding values and residuals, respectively.

With uncertain parentage, if the sire of animal i is unknown, a number s_i of potential sires can be assigned with probabilities p_{i1}, \dots, p_{is_i} ; and similarly, if the dam is unknown, d_i potential dams can be assigned with probabilities q_{i1}, \dots, q_{id_i} . Each combination of potential parents for animals with unknown parents can be represented by a different pedigree, \mathbf{P}_k , $k=1, \dots, v$; where \mathbf{P} is a random variable with probability function $\pi(\mathbf{P})$. This function will be usually given by information external to the production records, such as management (e.g., multiple-sire joining) or molecular information (blood types, DNA markers). The joint distribution of \mathbf{y} and \mathbf{u} is

$$f(\mathbf{y}, \mathbf{u}) = \sum_{k=1}^v f(\mathbf{y}, \mathbf{u} | \mathbf{P} = \mathbf{P}_k) \pi(\mathbf{P}_k) \quad (1)$$

with

$$f(\mathbf{y}, \mathbf{u} | \mathbf{P} = \mathbf{P}_k) = N \left[\begin{pmatrix} \mathbf{X}\boldsymbol{\beta} \\ 0 \end{pmatrix}, \begin{pmatrix} \mathbf{Z}\mathbf{A}_k\mathbf{Z}'\sigma_a^2 + \mathbf{R}\sigma_e^2 & \mathbf{A}_k\sigma_a^2 \\ \mathbf{A}_k\sigma_a^2 & \mathbf{A}_k\sigma_a^2 \end{pmatrix} \right] \quad (2)$$

where \mathbf{A}_k is the numerator relationship matrix (NRM) obtained with $\mathbf{P} = \mathbf{P}_k$, σ_a^2 is the additive variance, \mathbf{R} is the residual covariance matrix, and σ_e^2 is the residual variance.

Best prediction

Selection based on the conditional mean or best predictor maximizes expected genetic progress when the proportion selected is fixed (Fernando and Gianola 1986). Note that given \mathbf{P} , \mathbf{y} and \mathbf{u} are normally distributed (2), but not marginally distributed (1). Thus, the conditional mean of \mathbf{u} given below

$$E(\mathbf{u} | \mathbf{y}) \propto \sum_{k=1}^v E(\mathbf{u} | \mathbf{y}, \mathbf{P} = \mathbf{P}_k) f(\mathbf{y} | \mathbf{P} = \mathbf{P}_k) \pi(\mathbf{P}_k)$$

is not a linear function of the data (Poivey and Elsen 1984). $E(\mathbf{u} | \mathbf{y}, \mathbf{P} = \mathbf{P}_k)$ is the best linear predictor of \mathbf{u} given $\mathbf{P} = \mathbf{P}_k$, and $f(\mathbf{y} | \mathbf{P} = \mathbf{P}_k)$ is the likelihood of the data given $\mathbf{P} = \mathbf{P}_k$. If the location parameters, $\boldsymbol{\beta}$, are not known, $E(\mathbf{u} | \mathbf{y})$ cannot be calculated. In this situation, prediction can be based on

$$E(\mathbf{u} | \mathbf{w}) \propto \sum_{k=1}^v E(\mathbf{u} | \mathbf{w}, \mathbf{P} = \mathbf{P}_k) f(\mathbf{w} | \mathbf{P} = \mathbf{P}_k) \pi(\mathbf{P}_k)$$

where $E(\mathbf{u} | \mathbf{w}, \mathbf{P} = \mathbf{P}_k)$ is BLUP of \mathbf{u} given $\mathbf{P} = \mathbf{P}_k$ with $\mathbf{w} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$, $\mathbf{X}\hat{\boldsymbol{\beta}}$ is any unbiased estimate of $\mathbf{X}\boldsymbol{\beta}$ (Fernando and Gianola 1986).

In the above formulas, summation is over all possible values of \mathbf{P} . Consider 20 animals with four possible parents each. Then \mathbf{P} may take $4^{20} = 1.1 \times 10^{12}$ different values. Thus, the calculation of the conditional mean will not be computationally feasible in most situations. Therefore, we now consider linear prediction.

Best linear unbiased prediction

The best linear unbiased predictor of \mathbf{u} , $\hat{\mathbf{u}}$, can be obtained by solving Henderson's (1973) mixed model equations (MME) given below

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \text{Var}(\mathbf{u})^{-1}\sigma_e^2 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \quad (3)$$

With uncertain parentage, the variance of (\mathbf{y}, \mathbf{u}) can be written as:

$$\text{Var}(\mathbf{y}, \mathbf{u}) = E_{\mathbf{P}}[\text{Var}(\mathbf{y}, \mathbf{u} | \mathbf{P} = \mathbf{P}_k)] + \text{Var}_{\mathbf{P}}[E(\mathbf{y}, \mathbf{u} | \mathbf{P} = \mathbf{P}_k)]$$

where the subscript \mathbf{P} means that expectation (E) and variance (Var) are taken with respect to all possible values \mathbf{P} . Note that

$$E(\mathbf{y}, \mathbf{u} | \mathbf{P} = \mathbf{P}_k) = (\mathbf{X}\boldsymbol{\beta}, 0)$$

is constant for all values of \mathbf{P} . Thus,

$$\begin{aligned} \text{Var}(\mathbf{y}, \mathbf{u}) &= E_{\mathbf{P}}[\text{Var}(\mathbf{y}, \mathbf{u} | \mathbf{P} = \mathbf{P}_k)] \\ &= \sum_{k=1}^v \text{Var}(\mathbf{y}, \mathbf{u} | \mathbf{P} = \mathbf{P}_k) \pi(\mathbf{P}_k) = \begin{pmatrix} \bar{\mathbf{V}} & \bar{\mathbf{A}}\sigma_a^2 \\ \bar{\mathbf{A}}\sigma_a^2 & \bar{\mathbf{A}}\sigma_a^2 \end{pmatrix} \end{aligned}$$

with

$$\text{Var}(\mathbf{u}) = \bar{\mathbf{A}}\sigma_a^2 = \sum_{k=1}^v \mathbf{A}_k \pi(\mathbf{P}_k) \sigma_a^2 \quad (4)$$

and

$$\text{Var}(\mathbf{y}) = \bar{\mathbf{V}} = \mathbf{Z}\bar{\mathbf{A}}\mathbf{Z}'\sigma_a^2 + \mathbf{R}\sigma_e^2$$

However, computing $\text{Var}(\mathbf{u})$ from (4) may not be computationally feasible because the summation is over all possible pedigrees. In the following section we give an alternative derivation for $\text{Var}(\mathbf{u})$ that leads to computationally feasible algorithms for obtaining $\bar{\mathbf{A}}$ and $\bar{\mathbf{A}}^{-1}$.

Construction of \bar{A} and of \bar{A}^{-1}

Suppose that the i -th individual has males $m_1, \dots, m_j, \dots, m_{s_i}$ as potential sires and females $f_1, \dots, f_k, \dots, f_{d_i}$ as potential dams. The covariance between the breeding values of the l -th and the i -th individual is

$$\text{Cov}(u_l, u_i) = E(u_l u_i) = a_{li} \sigma_a^2$$

because $E(\mathbf{u}) = \mathbf{0}$ and where $\bar{A} = \{a_{ij}\}$. This can be written as:

$$E(u_l u_i) = E_{j,k} [E(u_l u_i | m_j, f_k)] \quad (5)$$

where the condition is on m_j and f_k being the true parents of i . If l is not a direct descendent of i (Chang et al. 1991):

$$\begin{aligned} E(u_l u_i | m_j, f_k) &= 0.5 [E(u_l u_{m_j}) + E(u_l u_{f_k})] \\ &= 0.5 [a_{lm_j} + a_{lf_k}] \sigma_a^2 \end{aligned} \quad (6)$$

Substituting (6) in (5) and taking the expectation over j and k , we obtain

$$\text{Cov}(u_l, u_i) = a_{li} \sigma_a^2 = 0.5 \left[\sum_{j=1}^{s_i} p_{ij} a_{lm_j} + \sum_{k=1}^{d_i} q_{ik} a_{lf_k} \right] \sigma_a^2. \quad (7)$$

Similarly, the variance of the breeding value of individual i is

$$\begin{aligned} \text{Var}(u_i) &= a_{ii} \sigma_a^2 = E_{j,k} [E(u_i^2 | m_j, f_k)] \\ &= \left[1 + 0.5 \sum_{j=1}^{s_i} \sum_{k=1}^{d_i} p_{ij} q_{ik} a_{m_j f_k} \right] \sigma_a^2 = (1 + F_i) \sigma_a^2 \end{aligned} \quad (8)$$

where F_i is the inbreeding coefficient of i -th individual.

Rules (7) and (8) are a generalization of those given by Henderson (1988), allowing uncertainty of parentage in both sexes, and those from Dempfle (1990) by accommodating pedigrees that span more than two generations.

As an example, consider the pedigree given in Fig. 1, where dotted lines indicate uncertain parentage. Assigning equal probabilities for 1 or 2 being the sire of 4, and 3 or 4 being the dam of 5, the NRM in Fig. 2 was obtained using (7) and (8). It is easily verified that this NRM is equal to the average of the four NRMs shown in Fig. 3 corresponding to the four possible pedigrees.

Now we derive the rules to obtain the inverse of \bar{A} . Consider the following linear model for u_i :

$$u_i = 0.5 \sum_{j=1}^{s_i} p_{ij} u_{m_j} + 0.5 \sum_{k=1}^{d_i} q_{ik} u_{f_k} + \varepsilon_i \quad (9)$$

where ε_i is a genetic residual term. The variance of (9) is:

$$\begin{aligned} a_{ii} \sigma_a^2 &= \left\{ 0.25 \sum_{j=1}^{s_i} \sum_{j'=1}^{s_i} p_{ij} p_{ij'} a_{m_j m_{j'}} \right. \\ &\quad + 0.25 \sum_{k=1}^{d_i} \sum_{k'=1}^{d_i} q_{ik} q_{ik'} a_{f_k f_{k'}} \\ &\quad \left. + 0.5 \sum_{j=1}^{s_i} \sum_{k=1}^{d_i} p_{ij} q_{ik} a_{m_j f_k} \right\} \sigma_a^2 + \text{Var}(\varepsilon_i) \end{aligned}$$

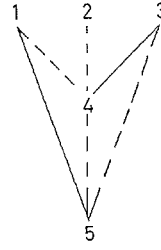


Fig. 1. Pedigree used as example. Continuous line known parent, dashed line uncertain parent

	1	2	3	4	5
1	1	0	0	0.25	0.5625
2	0	1	0	0.25	0.0625
3	0	0	1	0.5	0.375
4	0.25	0.25	0.5	1	0.5
5	0.5625	0.0625	0.375	0.5	1.0625

Fig. 2. Numerator relationship matrix for the pedigree of Fig. 1 computed using the rules given in Eqs. (7) and (8)

	1	2	3	4	5
1	1	0	0	0.5	0.5
2	0	1	0	0	0
3	0	0	1	0.5	0.5
4	0.5	0	0.5	1	0.5
5	0.5	0	0.5	0.5	1

a

	1	2	3	4	5
1	1	0	0	0	0.5
2	0	1	0	0.5	0
3	0	0	1	0.5	0.5
4	0	0.5	0.5	1	0.25
5	0.5	0	0.5	0.25	1

b

	1	2	3	4	5
1	1	0	0	0	0.5
2	0	1	0	0.5	0.25
3	0	0	1	0.5	0.25
4	0	0.5	0.5	1	0.75
5	0.75	0	0.25	0.75	1.25

c

	1	2	3	4	5
1	1	0	0	0	0.5
2	0	1	0	0.5	0.25
3	0	0	1	0.5	0.25
4	0	0.5	0.5	1	0.5
5	0.5	0.25	0.25	0.5	1

d

Fig. 3a-d. The four numerator relationship matrices corresponding to the four pedigrees in Fig. 1. a 1 is sire of 4, 3 is dam of 5; b 1 is sire of 4, 4 is dam of 5; c 2 is sire of 4, 3 is dam of 5; d 2 is sire of 4, 4 is dam of 5

since the covariances involving u and ε are zero (see Appendix). From (8)

$$\begin{aligned} \text{Var}(\varepsilon_i) &= \left\{ 1 - 0.25 \sum_{j=1}^{s_i} \sum_{j'=1}^{s_i} p_{ij} p_{ij'} a_{m_j m_{j'}} \right. \\ &\quad \left. - 0.25 \sum_{k=1}^{d_i} \sum_{k'=1}^{d_i} q_{ik} q_{ik'} a_{f_k f_{k'}} \right\} \sigma_a^2 = \delta_i \sigma_a^2. \end{aligned} \quad (10)$$

Note that when $s_i = d_i = 1$ (10) reduces to the usual formula when both parents are known:

$$\text{Var}(\varepsilon_i) = \{1 - 0.25 a_{mm} - 0.25 a_{ff}\} \sigma_a^2$$

If, on the other hand, each p_{ij} and/or q_{ik} are very small, usually s_i and/or d_i are very large, (10) also reduces to the usual expression when one or both parents are unknown (Henderson 1976). It is customary to interpret $\text{Var}(\varepsilon_i)$ as an error term produced by Mendelian segregation (Sorensen and Kennedy 1984). In the case of uncertain

parentage, however, it is shown in (10) that there is an additional source of residual genetic variance due to variability within potential sires and within potential dams (Dempfle 1990).

Equation (9) can be expressed in matrix form as

$$\mathbf{u} = \mathbf{T}\mathbf{u} + \boldsymbol{\varepsilon}$$

where \mathbf{T} is a lower triangular transition matrix containing in the row of the i -th individual s_i terms equal to $0.5 p_{ij}$ ($j=1, \dots, s_i$) and d_i terms equal to $0.5 q_{ik}$ ($k=1, \dots, d_i$), and $\boldsymbol{\varepsilon}$ is a vector with elements ε_i . Then, following Quaas' (1988) approach,

$$\mathbf{u} = (\mathbf{I} - \mathbf{T})^{-1} \boldsymbol{\varepsilon}$$

and

$$\text{Var}(\mathbf{u}) = \bar{\mathbf{A}} \sigma_a^2 = (\mathbf{I} - \mathbf{T})^{-1} \mathbf{D} (\mathbf{I} - \mathbf{T}')^{-1} \sigma_a^2$$

where \mathbf{D} is a diagonal matrix (see Appendix) containing the elements δ_i given by (10). Consequently

$$\bar{\mathbf{A}}^{-1} = (\mathbf{I} - \mathbf{T}') \mathbf{D}^{-1} (\mathbf{I} - \mathbf{T}).$$

The rules to invert $\bar{\mathbf{A}}$ follow:

- 1) Set $\bar{\mathbf{A}}^{-1}$ to zero
- 2) Compute \mathbf{D} from (10)
- 3) For each individual i in the pedigree, make the following contributions to $\bar{\mathbf{A}}^{-1}$

position	contribution
(i, i)	δ_i^{-1}
$(i, m_j), (m_j, i)$	$-0.5 p_{ij} \delta_i^{-1} \quad j=1, \dots, s_i$
$(i, f_k), (f_k, i)$	$-0.5 q_{ik} \delta_i^{-1} \quad k=1, \dots, d_i$
$(m_j, m_{j'})$	$0.25 p_{ij} p_{ij'} \delta_i^{-1} \quad j=1, \dots, s_i; j'=1, \dots, s_i$
$(f_k, f_{k'})$	$0.25 q_{ik} q_{ik'} \delta_i^{-1} \quad k=1, \dots, d_i; k'=1, \dots, d_i$
$(m_j, f_k), (f_k, m_j)$	$0.25 p_{ij} q_{ik} \delta_i^{-1} \quad j=1, \dots, s_i; k=1, \dots, d_i$

This algorithm is a generalization of those given by Henderson (1988) and Dempfle (1990). A FORTRAN program to compute $\bar{\mathbf{A}}$, $\bar{\mathbf{A}}^{-1}$ and BLUP is available on request from the first author.

Simulation

A computer simulation was used to compare genetic evaluation by BLUP using genetic (BLUP-G) and that by BLUP using $\bar{\mathbf{A}}$ (BLUP- $\bar{\mathbf{A}}$). The simulated population consisted of ten sires and 200 dams per breeding season. Each breeding season each sire was mated to 20 dams, producing 10 male and 10 female offspring. Each dam produced one offspring per breeding season. Only females had records, which were assigned at birth, one record per female. The sire was unknown for 10% or 30% of the male offspring and 30% of the female offspring. The dam was always known. The 10 best males, according to the corresponding evaluation method, from among the

100 male offspring and from among the ten sires were selected. Similarly, the 200 best females from among the 100 female offspring and from among the 200 dams were selected. The unselected animals were culled. Five breeding seasons were simulated. Heritability was 0.25 and the phenotypic variance was 100. The number of replicates was 500 for each method of evaluation. Three methods of evaluation, described below, were compared.

- 1) BLUP-G: one genetic group was defined per breeding season;
- 2) BLUP- $\bar{\mathbf{A}}_1$: for each animal with unknown sire, the true sire and two random sires were assigned the same probability (1/3) of being the true sire;
- 3) BLUP- $\bar{\mathbf{A}}_2$: for each animal with unknown sire, all the ten sires of the corresponding breeding season were assigned the same probability (1/10) of being the true sire.

The following parameters were studied: response to selection, bias of estimated genetic mean in the last breeding season (BEGM), mean squared error of estimated genetic mean, averaged over breeding seasons (MSEGM), and mean prediction error variance (MPEV). BEGM was estimated as

$$\sum_i^n (\hat{u}_{ti} - \bar{u}_{ti}) / n$$

where n is the number of replicates; \hat{u}_{ti} , the mean of the predicted breeding values (\hat{u}_{ti}) in the last breeding season (t); and \bar{u}_{ti} , the mean of the true breeding values (u_{ti}). MSEGM was estimated as

$$\sum_t \sum_i^n (\hat{u}_{ti} - \bar{u}_{ti})^2 / (n-1) / t$$

MPEV was estimated as

$$\sum_i^n \sum_j^N (\hat{u}_{ij} - u_{ij})^2 / (N-1) / n$$

where N is the total number of animals per replicate. Predicted breeding values to obtain BEGM, MSEGM, and MPEV were computed with all information at the end of each replicate.

Simulation results are in Table 1. This table also gives the results when all of the parents were identified (BLUP-A), where response to selection is expected to be maximum in this situation. Response to selection using BLUP- $\bar{\mathbf{A}}_1$ and BLUP- $\bar{\mathbf{A}}_2$ was significantly higher than with BLUP-G. When sires of 10% of the male offspring and 30% of the female offspring were unknown, and with three potential sires (BLUP- $\bar{\mathbf{A}}_1$), the loss in response with respect to BLUP-A was less than 2%. The increase in response with BLUP- $\bar{\mathbf{A}}$ relative to BLUP-G ranged from 4% to 8%, the advantage increasing as both the number of candidates and the percentage of animals with unknown sires decreased. Selection with BLUP- $\bar{\mathbf{A}}_1$ or BLUP- $\bar{\mathbf{A}}_2$ resulted in higher levels of inbreeding than selection with

Table 1. Simulation results using \bar{A} with three potential sires (BLUP- \bar{A}_1), \bar{A} with ten potential sires (BLUP- \bar{A}_2), and genetic groups (BLUP-G). The method with no unknown parents (BLUP-A) is shown for comparison. The percentage of females with unknown sire was 30%. Inbreeding and all relationships were taken into account to obtain the inverse of the additive relationship matrix. Five breeding seasons and 500 replicates per method were run

Method	% of males with unknown sire	Response ± 0.08	$F^a \times 100$	BEGM ^b	MSEGM ^c	MPEV ^d
BLUP-A	0.00	7.88	7.71	-0.02	0.38	16.99
BLUP- \bar{A}_1	0.10	7.75	7.00	-0.02	0.41	17.52
BLUP- \bar{A}_1	0.30	7.48	6.70	0.02	0.40	17.58
BLUP- \bar{A}_2	0.10	7.54	6.78	-0.01	0.41	17.55
BLUP- \bar{A}_2	0.30	7.27	6.38	0.04	0.38	17.62
BLUP-G	0.10	7.17	6.17	0.30	0.49	24.30
BLUP-G	0.30	6.97	5.57	0.29	0.63	23.97

^a Mean inbreeding coefficient in the last breeding season computed using the real pedigree

^b Bias of the estimated genetic mean in the last breeding season

^c Mean squared error of the genetic mean

^d Mean prediction error variance

Table 2. Central processing unit time used to compute the numerator relationship matrix (NRM) and to solve the mixed model equations (MME) in a SUN 4/260 workstation. Thirty percent of male and female offspring with unknown sire except with BLUP-A

Method	Number of potential sires	Inbreeding included	NRM time ^a	MME time ^b
BLUP-A	1	yes	8.96	8.99
BLUP- \bar{A}_1	3	yes	28.16	15.72
BLUP- \bar{A}_1	3	no	0.00	15.26
BLUP- \bar{A}_2	10	yes	39.42	28.83
BLUP- \bar{A}_2	10	no	0.00	28.98
BLUP-G	-	yes	8.70	10.39
BLUP-G	-	no	0.00	10.55

^a Mean time per replicate in the last breeding season

^b Mean time per replicate per breeding season

BLUP-G. This is due to the fact that level of inbreeding increases as both accuracy of selection and amount of family information increases (Robertson 1961).

Estimation of response to selection seemed to be unbiased with BLUP- \bar{A}_1 and BLUP- \bar{A}_2 , but was biased with BLUP-G (Table 1). Both MSEGM and MPEV were smaller with BLUP- \bar{A}_1 and BLUP- \bar{A}_2 than with BLUP-G (Table 1).

Table 2 shows the computer time consumed to obtain the NRM, which is necessary to compute the elements of \mathbf{D} , and the time to solve the MME by iteration on data (Misztal and Gianola 1987). It took 2.4 to over 4 times longer to calculate the NRM with uncertain parentage than with no unknown parents or with genetic groups. With BLUP- \bar{A} , the cost to solve the MME increased 1.7–3.2 times. The additional cost with BLUP-G relative

to BLUP-A is due to a slower convergence and the additional equations for genetic groups.

The computational costs of obtaining \bar{A} can be a handicap for the application of BLUP- \bar{A} with large pedigrees. Often \bar{A}^{-1} is approximated without computing \mathbf{A} by disregarding inbreeding of the parents. Similarly, \bar{A}^{-1} can be approximated by disregarding covariances between potential sires and between potential dams, in addition to their inbreeding coefficients. Then δ_i in (10) can be approximated by

$$\delta_i \cong 1 - 0.25 \sum_{j=1}^{s_i} p_{ij}^2 - 0.25 \sum_{k=1}^{d_i} q_{ik}^2 \quad (11)$$

and \bar{A}^{-1} can be calculated without needing \bar{A} . Furthermore, if the probabilities for each potential parent are the same

$$\delta_i \cong 1 - 0.25/s_i - 0.25/d_i.$$

Using (11) the statement by Henderson (1988) that “unfortunately, in the multiple sire problem \mathbf{D} must be computed (from \bar{A}) even for non-inbred populations” can be relaxed.

The rate of convergence to solve the MME using (10) or (11) was almost the same (Table 2). The results with (11) are shown in Table 3. In the case of genetic groups, \mathbf{A} was computed neglecting inbreeding. The use of the approximation for both \bar{A} or \mathbf{A} did not affect either response to selection or inbreeding attained. MSEGM and MPEV were only slightly greater than with the exact formulas. Estimates of response were unbiased with BLUP- \bar{A}_1 and BLUP- \bar{A}_2 . BEGM with BLUP-G, however, seemed to be greater. This may be of concern when the objective is to estimate genetic trends and not only to rank animals using BLUP-G. The effect of inbreeding in this case merits further research.

Table 3. Simulation results using \bar{A} with three potential sires (BLUP- \bar{A}_1), \bar{A} with ten potential sires (BLUP- \bar{A}_2), and genetic groups (BLUP-G). The percentage of female offspring with unknown sire was 30%. The approximation in Eq. (11) was used to obtain the inverse of the additive relationship matrix. Five breeding seasons and 500 replicates per method were run

Method	% of males with unknown sire	Response ± 0.08	F ^a $\times 100$	BEGM ^b	MSEGM ^c	MPEV ^d
BLUP- \bar{A}_1	0.10	7.53	6.80	0.01	0.39	17.41
BLUP- \bar{A}_1	0.30	7.55	6.72	0.03	0.39	17.62
BLUP- \bar{A}_2	0.10	7.48	6.63	0.04	0.40	17.62
BLUP- \bar{A}_2	0.30	7.20	6.49	0.05	0.44	17.74
BLUP-G	0.10	7.17	6.14	0.57	0.53	24.16
BLUP-G	0.30	6.94	5.68	0.36	0.63	23.98

^a Mean inbreeding coefficient in the last breeding season computed using the real pedigree

^b Bias of the estimated genetic mean in the last breeding season

^c Mean squared error of the genetic mean

^d Mean prediction error variance

Conclusions

We have shown that $\bar{A} \sigma_a^2$, as defined by Henderson (1988), is the correct covariance matrix of additive effects with uncertain parentage. Thus, using \bar{A} in the MME, genetic evaluation can be obtained by BLUP without introducing genetic groups into the model, together with the associated assumption that unknown parents are unrelated, as in BLUP-G. However, it should be borne in mind that with uncertain parentage, BLUP is not equal to the conditional mean (Poivey and Elsen 1984). Nonetheless, Poivey and Elsen (1984) found that response to selection was similar using the linear predictor and the conditional mean.

The simulation results showed that greater response to selection can be obtained using BLUP- \bar{A} than using BLUP-G. The advantage of BLUP- \bar{A} over BLUP-G increased as the percentage of animals with unknown parents and the number of potential parents were smaller. Estimates of response obtained by BLUP-G were biased, while those obtained by BLUP- \bar{A} were not.

To obtain genetic evaluations by BLUP- \bar{A} , potential parents must be identified for animals with unknown parents, and probabilities of parentage have to be assigned. Thus, more information is taken into account with BLUP- \bar{A} than with BLUP-G. In some situations, however, identification of potential parents may be difficult for some animals. In this situation, genetic evaluation can be obtained by using a combination of \bar{A} and genetic groups in the MME.

Another potential application of BLUP- \bar{A} is for the mixed model analysis of experiments with laboratory animals. In such experiments, mass selection (Falconer 1989) is usually practiced and the complete pedigree is rarely available, but identification of potential parents is generally possible. Thus, a mixed model analysis using \bar{A} is appropriate.

Acknowledgements. M. Perez-Enciso gratefully acknowledges support by a grant from INIA (Spain).

Appendix

Proof that D is diagonal

For any two individuals, one will not be a direct descendent of the other. Thus, first it will be shown that $\text{Cov}(u_i, \varepsilon_i) = 0$ for any i that is not a direct descendent of i . Consider the model given in (9):

$$u_i = 0.5 \sum_{j=1}^{s_i} p_{ij} u_{m_j} + 0.5 \sum_{k=1}^{d_i} q_{ik} u_{f_k} + \varepsilon_i.$$

From (7) it can be seen that

$$\text{Cov}(u_i, u_i) = \text{Cov}\left(u_i, 0.5 \sum_{j=1}^{s_i} p_{ij} u_{m_j} + 0.5 \sum_{k=1}^{d_i} q_{ik} u_{f_k}\right).$$

Thus, $\text{Cov}(u_i, \varepsilon_i) = 0$. Consider the linear model for the breeding value of animal i

$$u_i = 0.5 \sum_{j=1}^{s_i} p_{ij} u_{\mu_j} + 0.5 \sum_{k=1}^{d_i} q_{ik} u_{\phi_k} + \varepsilon_i$$

where μ_j and ϕ_k are a potential sire and dam of i , respectively. $\text{Cov}(u_{\mu_j}, \varepsilon_i) = \text{Cov}(u_{\phi_k}, \varepsilon_i) = 0$, for any j, k because μ_j and ϕ_k are not direct descendents of i . Then, since $\text{Cov}(u_i, \varepsilon_i) = 0$, $\text{Cov}(\varepsilon_i, \varepsilon_i)$ is zero and \mathbf{D} is diagonal.

References

- Chang, HL, Fernando RL, Grossman M (1991) On the principle underlying the tabular method to compute coancestry. *Theor Appl Genet* 81:233–238
- Dempfle L (1990) Problems in the use of the relationship matrix in animal breeding. In: Gianola D, Hammond K (eds) *Advances in statistical methods for genetic improvement of livestock*. Springer, Berlin Heidelberg New York, pp 454–473
- Falconer DS (1989) *Introduction to quantitative genetics*, 3rd edn. Longman, New York
- Fernando RL, Gianola D (1986) Optimal properties of the conditional mean as a selection criterion. *Theor Appl Genet* 72:822–825

- Foulley JL, Gianola D, Planchenault D (1987) Sire evaluation with uncertain paternity. *Genet Sel Evol* 19:83–102
- Henderson CR (1973) Sire evaluation and genetic trends. In: *Proc Anim Breed Genet Symp in Honor of Dr Jay L Lush*. Am Soc Anim Sci Am Dairy Sci Assoc, Champaign, Ill., pp 10–41
- Henderson CR (1976) A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* 32: 69–83
- Henderson CR (1988) Use of an average relationship for multiple-sire joining. *J Anim Sci* 66:1614–1621
- Misztal I, Gianola D (1987) Indirect solution of mixed model equations. *J Dairy Sci* 70:716–723
- Quaas RL (1988) Additive genetic model with groups and relationships. *J Dairy Sci* 71:1338–1345
- Poivey JP, Elsen JM (1984) Estimation de la valeur génétique des reproducteurs dans le cas d'incertitude sur les apparentements. I. Formulation des indices de sélection. *Genet Sel Evol* 16:445–454
- Robertson A (1961) Inbreeding in artificial selection programmes. *Genet Res* 2:189–194
- Sorensen D, Kennedy BW (1984) Estimation of response to selection using least-squares methods and mixed model methodology. *J Anim Sci* 58:1097–1106
- Van Vleck LD (1970) Misidentification in estimating the paternal sib correlation. *J Dairy Sci* 53:1469–1474
- Westell RA, Quaas RL, Van Vleck LD (1988) Genetic groups in an animal model. *J Dairy Sci* 71:1310–1318
- Wiggans GR, Misztal I, Van Vleck LD (1988) Animal model evaluation for Ayrshire milk yield with all lactations, herd-sire interaction, and groups based on unknown parents. *J Dairy Sci* 71:1319–1329