

An investigation into genetic evaluation using an animal model when paternity is uncertain

F.W.C. Neser*

Glen Agricultural Development Institute, Private Bag X01, Glen, 9360 Republic of South Africa

K.V. Konstantinov and G.J. Erasmus

Department of Animal Science, University of the Orange Free State, Bloemfontein, 9300 Republic of South Africa

Received 19 February 1993; accepted 27 August 1993

A study was conducted to examine the effect of using Henderson's (1988) 'average numerator relationship matrix' (ANRM) on the prediction error variance (PEV) of breeding values obtained using an animal model with maternal effects. Genetic trends with and without using ANRM were also estimated and compared. The results suggest that, in situations where multiple-sire joining is practised, the use of ANRM in mixed model equations will result in lower PEVs than when an ordinary numerator relationship matrix, treating sires as unknown, is used. Very little bias in genetic trend estimates was observed when ANRM was used.

'n Studie is onderneem om die effek van die gebruik van Henderson (1988) se 'average numerator relationship matrix' (ANRM) op die voorspellings foutvariansie (PEV) van teelwaardes, wat met behulp van 'n diereemodel met maternale-effekte verkry is, te ondersoek. Genetiese tendense, met en sonder die gebruik van ANRM is ook beraam en vergelyk. Uit die resultate blyk dit dat, in gevalle waar groepparing toegepas word, die gebruik van ANRM in gemengde model vergelykings tot laer PEVs sal lei as wanneer 'n gewone verwantskapsmatriks waar vaars as onbekend beskou word, gebruik word. Baie min sydigheid in beramings van genetiese tendense is waargeneem wanneer ANRM gebruik is.

Keywords: Animal model, genetic evaluation, multiple-sire joining.

* To whom correspondence should be addressed.

Introduction

In animal breeding operations, there are practical situations where uncertainty with respect to the assignment of progeny to sires occurs. For example, multiple-sire groups are commonly used in pastoral mating systems. The genetic evaluation of animals in such instances is complicated by this uncertainty. The Koopmansfontein Dorper sheep control flock, on which this study was conducted, represents such a situation, since single-sire mating was only practised in the last eight of the 30 years of its existence.

Poivey & Elsen (1984) were among the first to study this problem within the framework of selection index theory and its restrictive assumptions. Foulley *et al.* (1987) presented more general methods which permit non-linear models and unknown probability of parentage. Recently, Im (1989) showed how the regular mixed model equations can be used to obtain the best linear unbiased predictors when paternity is uncertain.

A different approach was used by Henderson (1988) who derived an 'average numerator relationship matrix' (ANRM) with uncertain paternity, along with its inverse. This method is easy to apply and permits evaluation of all animals using an animal model. Moreover, if maternal effects are important it is relatively straightforward to incorporate these effects into the model. Also, as pointed out by Famula (1992), the variances attributable to Mendelian sampling are dependent on the probabilities of parentage even in populations with inbreeding.

The models of Foulley *et al.* (1987) fail to account for this fact in the structure of their residual covariance matrix.

The objectives of this study were: (i) to examine the effect of using Henderson's (1988) ANRM with uncertain paternity on

prediction error variances (PEV) of breeding values obtained using an animal model with maternal effects, and (ii) to estimate and compare genetic trends with and without using ANRM.

Material and Methods

The data were obtained from the Dorper control flock kept at the Koopmansfontein Research Station. In total, 1086 weaning weight records, the progeny of 127 sires, collected over a period of 8 years from 1985 to 1992 were used for the analyses. A detailed description of the data, environment and management practices is given in Nesar (1993).

The data were analysed using the following mixed linear model:

$$y = X\beta + Z_1a + Z_2m + e \quad (1)$$

where y is an $n \times 1$ vector of records; X is an $n \times p$ incidence matrix that relates data to the unknown vector of location parameters β ; β included year-season, sex, birth status as fixed effects and two linear regressors, age of dam and age at weaning. The incidence matrices Z_1 and Z_2 relate the unknown random vectors of direct breeding value (a), and maternal breeding value (m), respectively, to y . The unknown vector e contains the random residuals due to environmental effects peculiar to individual records. The random vectors a and m are such that:

$$E \begin{bmatrix} a \\ m \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{Var} \begin{bmatrix} a \\ m \end{bmatrix} = \begin{bmatrix} \sigma_a^2 & \sigma_{am} \\ \sigma_{am} & \sigma_m^2 \end{bmatrix} * \mathbf{A}$$

The positive definite matrix \mathbf{A} contains the additive relationships among animals. Dispersion parameters σ_a^2 , σ_m^2 and σ_{am} are the

variances of additive direct and additive maternal effects and their covariance, respectively, and * is a direct product operator. The vector e is assumed to have zero expectation and to be independent of the vector of breeding values. Under these assumptions, the expected value of y in (1) is $E(y) = X\beta$ and the variance of y is

$$\text{Var}(y) = Z_1AZ_1'\sigma_a^2 + Z_2AZ_2'\sigma_m^2 + (Z_1AZ_2' + Z_2AZ_1')\sigma_{am} + I\sigma_e^2$$

where I is an identity matrix.

Breeding values for direct additive and maternal effects were estimated by solving Henderson's (1984) equations:

$$\begin{bmatrix} X'X \\ Z_1'X \\ Z_2'X \end{bmatrix} \begin{bmatrix} X'Z_1 & X'Z_2 \\ Z_1'Z_1 & Z_1'Z_2 \\ Z_2'Z_1 & Z_2'Z_2 \end{bmatrix} + \sigma_e^2 G^{-1} * A^{-1} \begin{bmatrix} \beta \\ \hat{a} \\ \hat{m} \end{bmatrix} = \begin{bmatrix} X'y \\ Z_1'y \\ Z_2'y \end{bmatrix} \quad (2)$$

$$\text{where } G^{-1} = \begin{bmatrix} \sigma_a^2 & \sigma_{am} \\ \sigma_{am} & \sigma_m^2 \end{bmatrix}^{-1}$$

Since the elements of the matrix G were not known and no literature estimates were available, they were estimated from the data using Meyer's (1991) DFREML program under model (1) and are shown in Table 1.

Table 1 Estimates of (co)variance components and heritabilities for weaning weight

σ_a^2	0.6952
σ_m^2	2.4517
σ_{am}	0.5331
σ_e^2	8.8605
σ_p^2	12.5406
h_a^2	0.0554
SE	0.0458
h_m^2	0.1955
SE	0.0409

The system (2) was solved directly, i.e. a generalized inverse of the coefficient matrix was obtained and multiplied by the right-hand side vector. This was necessary to obtain the prediction error variances.

Three different runs were performed. In the first run all sires were known and the relationship matrix was constructed according to the rules of Henderson (1976) and its inverse was obtained following the algorithm of Quaas (1976). In the second run all sires were deleted from the data file and the procedure was repeated. In the third run the inverse of the ANRM was obtained using the algorithm of Famula (1992). This algorithm was chosen because it is much faster and more flexible than Henderson's (1988) original one. A special pedigree file was created as follows:

Within each year 'dummy' mating groups were created. The inclusion of all sires, used in a specific year, would have been the most correct method. However, this would have increased the computing time considerably. It was therefore decided to assign only two more sires to each progeny with a known sire, chosen randomly from the total number of sires used in a specific year.

To each of these trio of sires, a probability was assigned. These probabilities were obtained using a random number generator, which generates an integer number uniformly distributed within an interval from 1 to 100. If, for example, the probability for the first sire was randomly generated as 60, the probability for the second sire was generated from interval 1 to (100-60), and so on.

The efficiency of using ANRM was estimated by comparing average PEVs from the three analyses. For known single sires PEVs are simply the diagonal elements of a generalized inverse of the coefficient matrix in (2). However, as pointed out by Henderson (1988), when using ANRM, the G-inverse is not the representation of PEV. To obtain the PEV, the results of Henderson (1975) were extended to accommodate the incorporation of maternal effects into an animal model.

Let the G-inverse of the coefficient matrix in (2) be

$$\begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{12}' & C_{22} & C_{23} \\ C_{13}' & C_{23}' & C_{33} \end{bmatrix}$$

Then

$$\text{Var} \begin{bmatrix} \hat{a} - a \\ \hat{m} - m \end{bmatrix} = \begin{bmatrix} C_{22} & C_{23} \\ C_{23}' & C_{33} \end{bmatrix} + \begin{bmatrix} C_{22} & C_{23} \\ C_{23}' & C_{33} \end{bmatrix} T \begin{bmatrix} C_{22} & C_{23} \\ C_{23}' & C_{33} \end{bmatrix}$$

where $T = G^{-1} * A_0^{-1} (G * A - G * A_0) G^{-1} * A_0^{-1}$, A is the relationship matrix with all sires known and A_0 is either the relationship matrix without sires or the average relationship matrix.

Results and Discussion

The average PEVs, calculated for the three situations studied, are given in Table 2.

Table 2 Average prediction error variances for direct additive and maternal effects

Genetic effect	All sires		
	All sires known	deleted	ANRM
Direct additive	0.06215 σ_e^2	0.06693 σ_e^2	0.06400 σ_e^2
Maternal	0.18858 σ_e^2	0.20895 σ_e^2	0.19771

As expected, the average PEVs for direct additive genetic breeding values are in ascending order, i.e. lowest for known single sires, higher for multiple-sire joining using ANRM and highest when treating sires as unknown. The same results were obtained for the average PEVs for maternal breeding values.

The product moment correlation estimates between estimate breeding values are supplied in Table 3.

The correlations between BVs with single sires and multi-sire joining were higher than the correlations between single sires and treating sires as unknown. A possible reason for the high correlations between estimated breeding values with unknown sires and multi-sire joining could be the low number of progeny per sire. In the data set used, only 14% of the sires had 10 progeny, 24% had 2-5 progeny and 62% had 6-9 progeny. As individual performance records of all the sires were included, the small number of progeny per sire did not contribute much to their estimated breeding values.

Table 3 Product moment correlations coefficients between the three estimated breeding values*

	All sires known	All sires deleted	ANRM
All siresknown	1	0.849	0.947
All siresdeleted	0.899	1	0.922
ANRM	0.966	0.947	1

*Above diagonal — correlations between the direct additive estimates of breeding values

Below diagonal — correlations between estimates for maternal breeding values

Estimated direct additive and maternal genetic trends are presented in Figures 1 & 2. The results show that considerable underestimation of the genetic trends is observed when sires are treated as unknown. Similar results are reported by Scarth & Graser (1987). It does, however, appear that the use of ANRM results in very little bias in estimation of the genetic trends.

It should be pointed out that, although significant, the estimated genetic trends have little, if any, practical value. The reason for these low trend estimates is that no selection was practised for the trait under consideration. If no genetic change has been achieved, estimates from the analysis with unknown sires would be expected to be close to zero. Since genetic trends are underestimated, selection among sheep born in different years will be biased.

The results from this study suggest that in situations where multiple-sire joining is practised, the use of ANRM in mixed model equations will result in smaller PEVs than when an ordinary numerator relationship matrix, which treats sires as unknown, is used. Also, very little bias is observed in genetic trend estimates when ANRM is used. As pointed out by

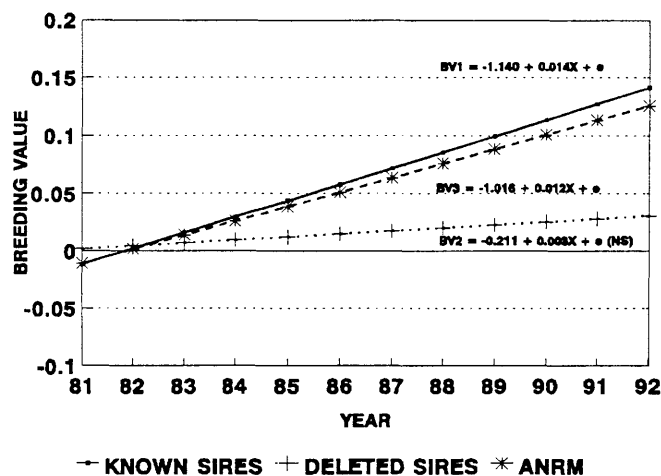


Figure 1 Estimated direct additive genetic trends for weaning weight.

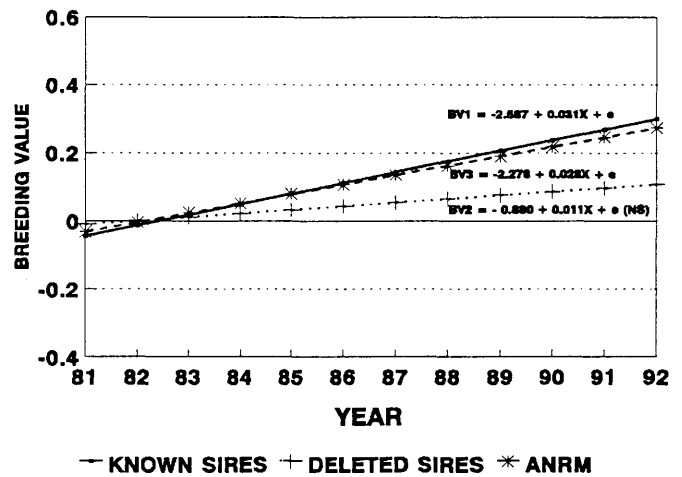


Figure 2 Estimated maternal genetic trends for weaning weight.

Henderson (1988), the goal in a breeding scheme should be single-sire mating groups. However, when this goal is not attainable, incorporating ANRM into mixed model equations, for the genetic evaluation of animals, should be preferred to eliminating useful, although imperfect information, by merely treating sires as unknown.

References

FAMULA, T.R., 1992. Simple and rapid inversion of additive relationship matrices incorporating parental uncertainty. *J. Anim. Sci.* 70, 1045.

FOULLEY, J.L., GIANOLA, D. & PLANCHENAULT, D., 1987. Sire evaluation with uncertain paternity. *Génet. Sél. Evol.* 19, 83.

HENDERSON, C.R., 1975. Use of relationships among sires to increase accuracy of sire evaluation. *J. Dairy Sci.* 58, 1731.

HENDERSON, C.R., 1976. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* 32, 69.

HENDERSON, C.R., 1984. Applications of linear models in animal breeding. University of Guelph.

HENDERSON, C.R., 1988. Use of an average numerator relationship matrix for multiple-sire joining. *J. Anim. Sci.* 66, 1614.

IM, S., 1989. A note on sire evaluation with uncertain paternity. *Biom. J.* 31(6), 749.

MEYER, K. 1991. Estimating variance and covariances for multivariate animal models by restricted maximum likelihood. *Génet. Sél. Evol.* 23, 67.

NESER, F.W.C. 1993. Estimation of genetic parameters when paternity is uncertain. M.Sc. Agric. treatise. Univ. of OFS, Bloemfontein.

POIVEY, J.P. & ELSER, J.M., 1984. Estimation de la valeur génétique des reproducteurs dans le cas d'incertitude sur les apparentements. I. Formulation des indices de sélection. *Génet. Sél. Evol.* 16, 445.

QUAAS, R.L., 1976. Computing the diagonal elements and inverse of a large numerator relationship matrix. *Biometrics* 32, 949.

SCARTH, R.D. & GRASER, H.U., 1987. Comparisons of genetic evaluation from animal models for single sire and multiple-sire joining. *Proc. 6th Conf. Aust. Assoc. Anim. Breed. Genet.*, Perth. 368.