MULTIPLE TRAIT LINKAGE ACROSS FLOCKS

A.E. Huisman, B. Tier, and D.J. Brown
Animal Genetics and Breeding Unit*, University of New England, Armidale, NSW, 2351, Australia

SUMMARY
To estimate contrasts between groups of animals or herds or flocks, or to compare groups of animals or herds or flocks, there has to be genetic linkage between those groups. A common problem for all extensively farmed livestock is the absence of such linkage. The main objective of assessing linkage is to determine the accuracy of comparisons between EBVs estimated in different herds or flocks. Several methods have been proposed to evaluate connectedness. However if the main objective of a linkage statistic is to identify flocks where EBVs are poorly contrasted in comparison to EBVs from other flocks, then a method which assesses the accuracy of such comparisons is most appropriate. This paper describes and gives an example from a multiple trait linkage analysis tool for data extracted from the Sheep Genetics Australia database.
Keywords: animal models, animal breeding methods, groups

INTRODUCTION

Estimated breeding values (EBVs) or combinations of EBVs (indexes) are estimates of an animal's genetic merit for a trait or group of traits. When comparing EBVs we are not only interested in their values, but also in the amount of trust we can place in the EBVs. This is described by the accuracy of the EBV; the accuracy describes the correlation between true and estimated breeding value, and is a function of the prediction error variance (PEV) of the EBV, which can be derived from the diagonal of the inverse of the coefficient matrix in the mixed model equations (MME), and the assumed additive genetic variance. In order to calculate the prediction error variance of the difference (PEVD) between two EBVs, we also need their prediction error covariance (PEC), which is the off-diagonal element of the inverse of the coefficient matrix in the MME.

To compare groups of animals or herds or flocks at the genetic level in the same analysis, the groups must be contrasted at the phenotypic level. This could be done directly or indirectly through comparison with other animals or their relatives. The amount or level of such comparisons is described as connectedness or linkage; the absence of such comparisons is referred to as disconnectedness or lack of linkage. A common problem for extensively farmed livestock is the absence of such linkage because of limited use of sires across flocks. The main objective of assessing 'connectedness' or linkage is to determine whether EBVs can be compared across group and with what level of accuracy those across group comparisons are done. Several methods have been proposed to evaluate connectedness (eg Fernando *et al.* 1983; Foulley *et al.* 1992; Kennedy and Trus 1993; Laloë *et al.* 1996). Definition of connectedness is not always the same across methods, and most methods are demanding computationally, which makes routine application difficult. The objective of this paper is to describe a simple method to assess linkage across groups of animals as implemented in the Sheep Genetics Australia (SGA) genetic evaluation.

* AGBU is a joint venture of NSW Department of Primary Industries and the University of New England

MATERIAL AND METHODS

We consider the sire model, $y_{ijk} = cg_i + s_j + e_{ijk}$, where y_{ijk} is an observation on an offspring of the j^{th} sire (s_j) made in the i^{th} contemporary group (cg_i), and e_{ijk} is the random residual term associated with the ijk^{th} observation. Henderson's (1973) mixed model equations for this model are:

$$\begin{bmatrix} \mathbf{X'R^{-1}X} & \mathbf{X'R^{-1}Z} \\ \mathbf{Z'R^{-1}X} & \mathbf{Z'R^{-1}Z+G^{-1}} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{s}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'R^{-1}y} \\ \mathbf{Z'R^{-1}y} \end{bmatrix}$$ Where E(**s**)=E(**e**)=0; E(**y**)= **Xb** ; Var(**s**)= **G** ; Var(**e**) = **R** and Var(**y**) = **ZGZ'**+**R** ; $\mathbf{G} = \mathbf{A}\sigma_s^2$; **A** is the numerator relationship matrix for sires; and $\sigma_s^2 = \frac{1}{4}\sigma_a^2$, where σ_a^2 is the additive genetic variance. $\mathbf{R} = \mathbf{I}\sigma_\varepsilon^2$; **I** is an identity matrix, and σ_ε^2 includes $\frac{3}{4}\sigma_a^2$. The inverse of Henderson's mixed model equations provides estimates of the standard errors of the fixed and random effects. The inverse of the coefficient matrix is

$$\begin{bmatrix} \mathbf{X'R^{-1}X} & \mathbf{X'R^{-1}Z} \\ \mathbf{Z'R^{-1}X} & \mathbf{Z'R^{-1}Z+G^{-1}} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{C}^{11} & \mathbf{C}^{12} \\ \mathbf{C'}^{12} & \mathbf{C}^{22} \end{bmatrix}$$ where $\mathbf{C}^{11}$, $\mathbf{C}^{12}$, and $\mathbf{C}^{22}$ are sub-matrices. The sampling covariance for the best linear unbiased prediction (BLUP) of $\hat{\mathbf{s}} - \mathbf{s}$ is given by $\sigma(\hat{\mathbf{s}} - \mathbf{s}) = \mathbf{C}^{22}$, which are the prediction error variances and prediction error covariances for the sire effects. The standardised prediction error variance of the difference (sPEVD) between the average sire breeding value of flock *i* and flock *j* is then equal to

$$sPEVD(\bar{x}_i - \bar{x}_j) = \frac{(\mathbf{w'}_i \mathbf{C}^{22} \mathbf{w}_i) + (\mathbf{w'}_j \mathbf{C}^{22} \mathbf{w}_j) - 2(\mathbf{w'}_i \mathbf{C}^{22} \mathbf{w}_j)}{(\mathbf{w'}_i \mathbf{G} \mathbf{w}_i) + (\mathbf{w'}_j \mathbf{G} \mathbf{w}_j) - 2(\mathbf{w'}_i \mathbf{G} \mathbf{w}_j)}$$ where $\mathbf{w}_i$ and $\mathbf{w}_j$ are vectors of size *number of sires* and contain ones and zeros relating sires to flocks *i* and *j*. The accuracy of the comparison between flocks *i* and *j* is then equal to $\sqrt{1 - sPEVD(\bar{x}_i - \bar{x}_j)}$. It is then possible to set threshold at which flocks *i* and *j* are determined to be linked. If $\mathbf{w}_i$ and $\mathbf{w}_j$ both have only one 1, then two sires are compared. When we extend this to a multivariate model the matrix $\mathbf{R} = \mathbf{E} \otimes \mathbf{I}$; and the matrix $\mathbf{G} = \mathbf{S} \otimes \mathbf{A}$ where **E** is the covariance matrix of residuals, **S** is the covariance matrix of additive sire effects, **A** and **I** are as before, and $\otimes$ denotes the Kronecker product.

Amalgamation. Let **P** be a matrix of size *number flocks* × *number flocks*, which is equal to $\mathbf{W'C^{22}W} / \mathbf{W'GW}$; where **W** is a matrix of all vectors **w**; and $\mathbf{C}^{22}$ and **G** are as before. This matrix **P** contains the standardised PEC between flock means on the off-diagonals, and the standardised PEV of flock means on the diagonal. Flocks that have a PEVD that meets the set threshold are then amalgamated into a linked set, or 'super-flock'. This amalgamation is done by adding the vectors **w** of both flocks into one new vector **w**, thus creating a 'new' flock. There could be several sets of linked flocks in the analysis, which may or may not be linked together. We distinguish five ways to amalgamate flocks into linked sets.

M1: The two flocks that have the lowest PEVD and have a PEVD below a threshold are amalgamated into a linked group, reducing the matrix **P** by one row and one column.

M2: Determine the flock that has the most PEVD below a threshold, all these flocks are

amalgamated at once into one linked group.

M3: Starting at the top left of the **P** matrix, the first two flocks that have a PEVD below the threshold are combined into a 'linked' flock, reducing the matrix **P** by one row and one column.

M4: As method 3, but starting at bottom right of the matrix **P**.

M5: The two flocks that have a PEVD closest to and below the threshold are amalgamated into a linked group, reducing the matrix **P** by one row and one column.

All five methods use iteration to amalgamate flocks; the process is repeated until there are no more flocks that can be added to the 'super-flock'. The EBVs from animals from flocks not part of the 'super-flock' cannot be compared to any animals outside that flock.

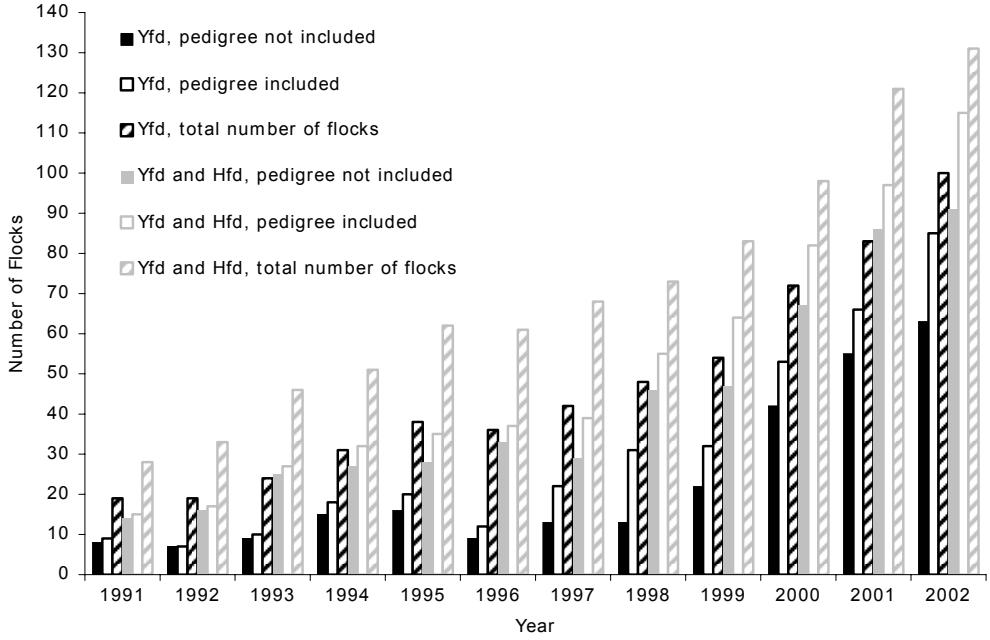

Figure 1. The number of flocks per year (vertical lines) and the number of linked flocks per year for yearling fibre diameter (Yfd) and yearling and hogget fibre diameter (Hfd) combined

Example data. The described methods were applied to fibre diameter data extracted from the SGA database. The traits of interest were fibre diameter measured at yearling and hogget age, these are repeated records of fibre diameter. The number of flocks with records for fibre diameter increased over time, from 19 in 1991 to 100 in 2002 for Yfd, from 19 to 85 in 2002 for Hfd, and from 29 in 1991 to 131 in 2002 for Yfd and Hfd combined (Figure 1).

RESULTS AND DISCUSSION

When multiple measures of fibre diameter were analysed more flocks were included in the analysis

and more flocks were determined to be linked together (Figure 1). The inclusion of pedigree increased the number of linked flocks (Figure 1), and more closely reflects true linkage. The use of related rams in different flocks, where individual rams are only used within flock, is accounted for by including pedigree in the analysis, which results in higher between flock comparison accuracies and potentially more flocks that are included in the linked set of flocks.

Table 1. Number of linked flocks for fibre diameter recorded in 2002 for the five amalgamation methods

		no pedigree included			
threshold	M1	M2	M3	M4	M5
50.0%	91	91	91	91	91
70.0%	25	38	26	37	31

The number of flocks that are in the linked subset for each amalgamation method and two different thresholds, 50% and 70%, are in Table 1. With the threshold at 50% we can estimate a difference between flock means of one genetic standard deviation at a 10% significance level. At a 70% threshold the corresponding significance level is 5%. Not all amalgamation methods give the same results, the size of the linked set for a threshold of 70% and no pedigree included ranges from 25 flocks, when M1 is used, to 38 flocks, when M2 is used (Table 1). When the threshold is set at 50% all five amalgamation methods give the same result. When pedigree was included, there was no difference between the five amalgamation methods.

CONCLUSION

The key question remains: *When are flocks considered linked?* According to Kennedy and Trus (1993) flocks are considered linked if they can be compared with certain accuracy. That certain accuracy still is and always will be a subjective matter. For the SGA genetic evaluation, amalgamation method 2 is applied because this is least demanding computationally, and one generation of pedigree is included to accommodate for the use of half sib sires across different flocks and the linkage threshold is set at 70% so that a difference of one genetic standard deviation between flock means can be estimated at a significance level of 5%. The described method also has relevance for other livestock species were animals are evaluated across flocks or herds.

ACKNOWLEDGEMENTS

This research was funded by Meat and Livestock Australia (MLA). Data were provided by Merino Genetic Services, Merino Benchmark and Australian Merino Sire Evaluation Association.

REFERENCES

Fernando, R.L., Gianola, D., and Grossman, M. (1983) *J. Dairy Sci.* **66**:1399.
Foulley, J.-L., Hanocq, E., and Boichard, D. (1992) *Genet. Sel. Evol.* **2 4**:315.
Henderson, C.R. 1973. In: Proc. Anim. Breeding and Genet. Symp. in Honor of Dr. Jay L. Lush. Am. Soc. Anim. Sci. and Am. Dairy Sci. Assoc., Champaign, IL.
Kennedy, B.W., and Trus, D. (1993) *J. Anim. Sci.* **71**:2341.
Laloe, D., Phocas, F., Menissier, F. (1996) *Genet. Sel. Evol.* **28**:359.