Proc. Assoc. Advmt. Breed. Genet. 21: 469-472

A COMPARISON OF GENETIC CONNECTEDNESS MEASURES USING DATA FROM THE NZ SHEEP INDUSTRY

J.B. Holmes[1], B. Auvray[1], S.A. Newman[2], K.G. Dodds[2] and M.A. Lee[1]

[1] Department of Mathematics and Statistics, University of Otago, Dunedin, New Zealand
[2] AgResearch, Invermay Agricultural Centre, Mosgiel, New Zealand

SUMMARY

The New Zealand Sheep industry, via Sheep Improvement Limited (SIL), estimates genetic connectedness across flocks as a function of progeny counts. This estimate is derived separately from the model fitted to estimate breeding values. As it ignores sources of genetic linkages other than direct parent-progeny links, it may under-estimate the level of connectedness present in the flocks assessed. In this paper, we compared this estimate to another derived from the variance-covariance (relationship) matrix of additive effects when pedigree information was available and when genotype information was available on some of the animals assessed. For the example of a single trait model using weaning weight records, we found an increase in the level of connectedness estimated compared to the existing method, particularly when genotype information was incorporated in the relationship matrix.

INTRODUCTION

To optimise genetic gain in livestock programs, breeding values need to be predictable between flocks. In animal breeding literature this is referred to as connectedness. In the BLUP methods used to estimate breeding values, the most appropriate measure of connectedness is the prediction error variance-covariance matrix (PEV). However, this calculation is computationally demanding and many proxies have been proposed.

The standard error of differences in breeding value means between flocks can be estimated as a function of the number of progeny born to common parents across flocks. This approximation is often used in traditional evaluations, where only pedigree information is used, but it is problematic when genotype information is also incorporated. As genetic evaluations for New Zealand sheep are increasingly using genotype data, a measure of connectedness derived from the model is preferred so that we can quantify genetic connectedness that is due to including genotype data.

In this paper we compared the standard error of differences in breeding value means calculated from a model based proxy to PEV, (i.e. genetic drift variance (Kennedy and Trus 1993)) to the current measure. This was done for scenarios where only pedigree information was available, and when some animals had genotype information available.

MATERIALS AND METHODS

Data. The data was from 64,841 animals from 19 flocks born from 2011 to 2013 with weaning weight records. The pedigree file containing the recorded animals and parents without records consisted of 84,802 animals. Genotype information (50K Illumina SNP Chip) was available for 269 of these animals of which 21 were in the initial set with weaning weight records. There were 31,884 animals that were either genotyped or had a genotyped ancestor. Table 1 shows the distribution of animals with genotype records or a genotyped ancestor across flocks.

Table 1. Distribution of animals with weaning weight records and genotype records on either themselves or at least one parent across flocks

Flock	Number with records	Number with a genotyped ancestor	Percentage with genotypes.
1	641	0	0.00
2	2533	2065	81.52
3	21240	14404	67.82
4	1996	1314	65.83
5	2344	1513	64.55
6	1110	0	0.00
7	16761	8231	49.11
8	1984	1785	89.97
9	815	769	94.36
10	3535	0	0.00
11	787	0	0.00
12	953	0	0.00
13	1025	699	68.20
14	2412	293	12.15
15	1193	528	44.26
16	368	0	0.00
17	2226	222	9.97
18	984	0	0.00
19	1934	61	3.15

SIL measure. The measure of connectedness between two flocks used for genetic evaluations performed in SIL is proportional to the standard error of the weighted average of differences of breeding values (u) between flocks across parents namely

$$\sqrt{\sum_i \lambda_i^2 \left(\frac{1}{n_{A_i}} + \frac{1}{n_{B_i}}\right)} = \sqrt{1/\sum_j \left(\frac{1}{n_{A_j}} + \frac{1}{n_{B_j}}\right)^{-1}} \quad [1]$$

where n_{A_j} is the number of progeny of parent *j* in flock A, n_{B_j} is the number of progeny of parent *j* in flock B and $\lambda_i \propto \left(\frac{1}{n_{A_i}} + \frac{1}{n_{B_i}}\right)^{-1} \Big/ \sum_j \left(\frac{1}{n_{A_j}} + \frac{1}{n_{B_j}}\right)^{-1}$. The standard error of differences has a range of $(0, \sqrt{2}]$. If there are no progeny from common parents in flock A and flock B, the standard error of the difference was arbitrarily set to 2. Only progeny born in a set time period are considered when calculating this measure. This is usually taken to be the previous three years, and has also been applied in this paper.

Variance-covariance matrix measure. The standard error of differences in average breeding values between flock A and B was calculated to proportionality from the elements of $\boldsymbol{V} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$ (Kennedy and Trus 1993) corresponding to flock A and B.

$$S.E.(\bar{u}_A - \bar{u}_B) \propto \sqrt{\boldsymbol{V_{AA}} + \boldsymbol{V_{BB}} - 2\boldsymbol{V_{AB}}} \quad [2]$$

where $\boldsymbol{G}$ is an additive relationship matrix, $\boldsymbol{Z}$ is the incidence matrix of animals with records and $\boldsymbol{X}$ is the flock incidence matrix. Two formulations for $\boldsymbol{G}$ were used. When only pedigree information was available, which we refer to as the pedigree measure, $\boldsymbol{G} = \boldsymbol{A}$, the pedigree additive relationship matrix. When some animals had genotype information available, which we refer to as the single step measure $\boldsymbol{G} = \boldsymbol{H}$. To calculate $\boldsymbol{H}$ both $\boldsymbol{A}$ and a genomic relationship matrix $\boldsymbol{G_1}$ was

required. $\boldsymbol{G_1}$ was calculated for the genotyped animals using the first method of VanRaden (2008). The $\boldsymbol{H}$ matrix was constructed using the method in Aguilar et al. (2010), where $\boldsymbol{A_{11}}$ and $\boldsymbol{A_{22}}$ are the additive relationship matrices for, and $\boldsymbol{A_{12}}$ is the matrix of additive relationship covariances between the un-genotyped and genotyped animals respectively.

$$\boldsymbol{H} = \boldsymbol{A} + \begin{bmatrix} \boldsymbol{A_{12}A_{22}^{-1}(G_1 - A_{22})A_{22}^{-1}A_{21}} & \boldsymbol{A_{12}A_{22}^{-1}(G_1 - A_{22})} \\ \boldsymbol{(G_1 - A_{22})A_{22}^{-1}A_{21}} & \boldsymbol{G_1 - A_{22}} \end{bmatrix} \quad [3]$$

When $\boldsymbol{V_{AB}} = 0$, the standard error of the difference in average breeding value is set to 2, analogously to the situation of no progeny from common parents in the SIL measure.

RESULTS AND DISCUSSION

Clusters of connected flocks. The connectedness estimated from the different methods is given in Table 2 where clusters of flocks estimated to be connected are shown. The criterion to be connected is a standard error of difference less than 2. The measure currently used in the NZ genetic evaluation was the most conservative in estimating connectedness across flocks and the single step measure was the least conservative. Changes in the clustering between the three measures were due to the admission of previously isolated flocks into clusters, or cluster merging rather the shifting of flocks from one cluster to another. This made intuitive sense since any linkage coming from shared parents is also contained in the pedigree along with linkage from more distant ancestors, such as grandparents. In turn in the genomic relationship matrix, almost all off-diagonals are non-zero, even for animals thought to be unrelated.

Table 2. Clusters of linked flocks (identified by flock code) according to the three measures of connectedness used

SIL measure									
Cluster 1	2	8	13						
Cluster 2	3	7	17						
Isolated Flocks	1	4	5	6	9	10	11	12	14
	15	16	18	19					
Pedigree measure									
Cluster 1	2	4	8	13					
Cluster 2	3	7	14	17					
Cluster 3	1	12	19						
Cluster 4	5	16							
Cluster 5	6	10							
Cluster 6	11	18							
Isolated Flocks	9	15							
Single step measure									
Cluster 1	1	2	3	4	5	7	8	9	12
	13	14	15	16	17	19			
Cluster 2	6	10							
Cluster 3	11	18							

Comparison of standard error of differences. Figure 1 plots the standard error of differences for the three measures considered. A reduction in the number of standard errors being arbitrarily set to 2 was found moving from the SIL measure to the pedigree and single step measures. This corresponded to the reduction in isolated flocks found in the cluster analysis. For flock pairs where connections were found using the pedigree measure, the rank correlation of the standard error of

differences between the pedigree and single step measure was 0.9902.

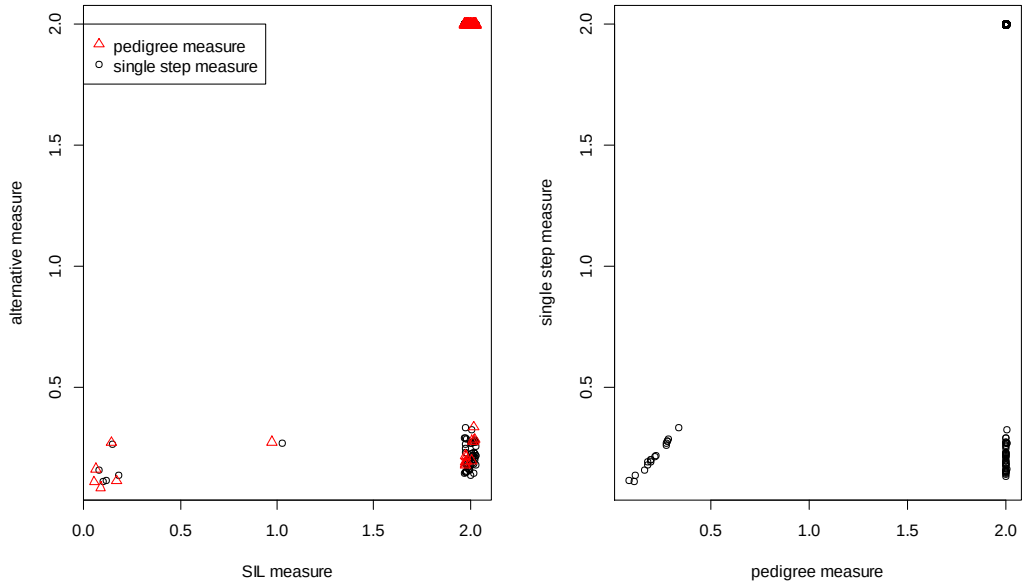

Figure 1. Comparison of standard error of differences using the three measures.

The loss of connections through removed data. In this paper, we used three years of data in the calculation of the connectedness measure. In routine genetic evaluations, there are many more years of records and pedigree data available. It may be inappropriate to develop a flock based connectedness measure from the full relationship matrix from a routine genetic evaluation, since connections from old animals would be given equal weighting to younger animals. Kennedy and Trus (1993) discussed changing the incidence matrix $\boldsymbol{X}$ in genetic drift variance from flock to flock by year. This method would utilise the connections lost through data removal while removing bias in measured connectedness through equal weighting of older and younger animals.

Single step method results. The single step measure assigned all flocks with genotype information and any flock related to such a flock through pedigree into a single cluster of related flocks. This means single step BLUP would lead to an increase in the number of animals with comparable breeding values compared to traditional BLUP. It is unclear how reasonable this result is, but this warrants further investigation. However the genomic relationship matrix used uses IBS to estimate relatedness. As a result the accuracy and comparability of estimated breeding values may be inflated. The degree of this inflation will be dependent on the group of markers used in the calculation of the genomic relationship matrix and the ancestry of the animals. To overcome this, developing an unbiased estimation of a genomic relationship matrix may be of interest.

REFERENCES

Kennedy B.W., and Trus D. (1993) *J. Dairy Sci.* **71**: 2341.
VanRaden P.M. (2008) *J. Dairy Sci.* **91**: 4414.
Aguilar I., Misztal I., Johnson D.L., Legarra A., Tsurata S. and Lawlor T.J. (2010) *J. Dairy Sci.* **93**: 743