Proc. Assoc. Advmt. Anim. Breed. Genet. ***17****: 399-402*

WHAT GENETIC GROUP STRUCTURE TO FIT? A BAYESIAN APPROACH APPLIED TO YEARLING WORM EGG COUNT DATA IN MERINO SHEEP

A.E. Huisman and D.J. Brown

Animal Genetics and Breeding Unit*, University of New England, Armidale, NSW, 2351, Australia

SUMMARY

Genetic groups need to be fitted in a genetic evaluation model to accommodate animals with unknown parents that come from a wide variety of sources. Different genetic grouping strategies were investigated for worm egg count data extracted from Sheep Genetics Australia's Merino database. A Bayesian approach was implemented that tested whether the genetic group variance was significantly different from zero. Four genetic grouping strategies were compared, 1) grouping on average fibre diameter, 2) groups by flock, 3) groups by flock-period, and 4) groups by flock-year. Two additional strategies were grouping strategies 3 and 4 with an assumed autocorrelation structure between genetic groups within flock.

INTRODUCTION

A genetic evaluation model needs to account for genetic groups when the animals in the data come from a wide variety of sources. Genetic groups are fitted to make comparison between animals of different genetic merit with unknown parentage possible (Westell *et al.* 1988). One of the features of the Sheep Genetics Australia's (SGA) Merino database is a high level of incomplete or missing pedigree, with many flocks recording only the sire for a high proportion of progeny. In the current Merino analyses conducted by SGA two analyses are performed; one for the main production traits and a separate analysis for worm egg count. In the worm egg count analysis genetic groups are formed on flock of origin of the animals with unknown parentage (Brown *et al.* 2006).

However despite much scientific investigation, until now the different applied genetic grouping strategies were based on data structure and expert opinion, and not on statistical model comparison. The aim of this work is to identify the 'best' genetic grouping strategy for worm egg count, using a Bayesian approach in the form of Bayes Factors (Varona *et al.* 2001).

MATERIAL AND METHODS

Data. Data were extracted from the SGA Merino database in September 2006. The trait of interest was yearling worm egg count. Contemporary groups were based on the combination of breed, flock, year of birth, date of measurement and user defined management group. The final dataset comprised 34,566 animals with data, and 49,547 animals in the pedigree. Worm egg count was expressed on the cube root scale; the average value for the animals with data on the cube root scale was 8.25, which is equivalent to approximately 560 eggs per gram of faeces.

Genetic Grouping Strategies. The following genetic grouping strategies were considered:

μ) Average phenotypic fibre diameter observations. The average phenotypic value for each flock was used to allocate flocks into distinct whole micron categories.

F) Each flock formed their own genetic group if there is enough data available (at least 100

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

animals with data). This is the genetic grouping strategy currently used in the SGA Merino analysis of worm egg count (Brown *et al.*, 2006).

FP) Each flock genetic group was divided in 4 periods (up to 1990; 1991-1995; 1996-2000; 2001 to current) if enough data (at least 100 animals with data) were available; if the resulting groups were too small, they were merged within flock.

FY) Each flock genetic group was divided in years, small groups were merged as above.

For the two methods with multiple genetic groups per flock (FP and FY), an additional analysis was performed where genetic groups within flock were assumed to have a correlation of 0.75 between adjacent groups and declining if groups were further apart. The correlation of 0.75 was chosen because roughly 25% of animals are replaced on a yearly basis. These two grouping strategies were denoted $FP_{0.75}$ and $FY_{0.75}$. This correlation between groups within flocks was fitted as an autoregressive covariance structure so that as the distance between periods or years became larger the correlation decreased.

Model. The standard model with genetic groups as defined by Westell *et al.* (1988) was:

$$y = X\beta + Zu + ZQg + e \quad [1]$$

where y, β, u, Qg, and e are vectors containing the trait observations, contemporary group effects, random animal genetic effects, genetic group effects, and random residual effects, respectively, **X** and **Z** are incidence matrices relating **y** to the respective effects. The q_{ij} elements of **Q** are a fraction relating the contribution of the *j*th genetic group to the total genetic value of the *i*th individual, and **g** is the vector containing the effects of genetic groups (Westell *et al.* 1988). The effects were assumed to have the following normal distributions;

$$\mathbf{u} \sim N(\mathbf{0}, \mathbf{A}\sigma_u^2);\ \mathbf{g} \sim N(\mathbf{0}, \mathbf{I}_g\sigma_g^2);\ \mathbf{e} \sim N(\mathbf{0}, \mathbf{I}_e\sigma_e^2)$$

A is the numerator relationship matrix, and **I**'s are identity matrices associated with different effects and σ^2 are the variances components associated with the different effects. Following Varona *et al.* (2001) and Casellas *et al.* (2006) the model in equation [1] can be reparameterised as;

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Zu} + \boldsymbol{\varepsilon} \quad [2]$$

where $\boldsymbol{\varepsilon} = \mathbf{ZQg} + \mathbf{e}$ and consequently $\varepsilon = N(\mathbf{0}, \mathbf{V})$ and $\sigma^2_\varepsilon = \sigma^2_g + \sigma^2_e$, with;

$$\mathbf{V} = \mathbf{ZQQ}^T\mathbf{Z}^T\sigma_g^2 + \mathbf{I}_e\sigma_e^2 = \sigma_\varepsilon^2\left[\mathbf{ZQQ}^T\mathbf{Z}^T\rho_g^2 + \mathbf{I}_e\left(1-\rho_g^2\right)\right]$$

Ratios of Bayes Factor (RBF) which were used to compare models without and with genetic groups are defined as:

$$\text{RBF} = p\left(\rho_g^2 = 0\right) \Big/ p\left(\rho_g^2 = 0 \middle| \mathbf{y}\right) = 1 \Big/ p\left(\rho_g^2 = 0 \middle| \mathbf{y}\right) \quad [3]$$

because $p\left(\rho_g^2 = 0\right) = 1$. Thus $\rho^2_g = \sigma^2_g + \sigma^2_\varepsilon$. Assumptions about prior distributions were as described by Varona *et al.* (2001). Testing the null hypothesis of no genetic group variance only requires the ordinate at zero of the marginal posterior density of ρ_g^2 (García-Cortés *et al.* 2001). Values for the Bayes factors were classified according to the levels of evidence of Kass and Raftery (1995).

A Gibbs sampler was implemented to estimate fixed and random effects for the described model. The coupling method (Garcia-Cortes *et al.* 1998) was used to assess the convergence of the Gibbs sampler. Three different sets of starting values were used. Based on these assessments the Gibbs sampler was run for 50,000 iterations with a burn-in period of 10,000 iterations.

RESULTS AND DISCUSSION

Variance Components. Estimates for the additive genetic and the residual variance were very similar across the different grouping strategies (Table 1). The estimated genetic group variance varied across the different grouping strategies, and ranged from 0.014 (μ) to approximately 1.4 for FP and FY when a correlation structure between genetic groups was assumed.

Table 1. Estimated direct genetic (σ^2_a), genetic group (σ^2_g) and residual (σ^2_e) variance components, σ^2_g/σ^2_e (ρ^2_g) and the ratio of Bayes Factor (RBF) for yearling worm egg count (standard errors as subscripts) using each genetic grouping strategy

Grouping Method	No. Genetic Groups	σ^2_a	σ^2_g	σ^2_e	ρ^2_g	RBF
μ	6	$1.91_{0.15}$	$0.01_{0.33}$	$5.90_{0.13}$	$0.00_{0.04}$	0.105
F	110	$1.82_{0.15}$	$0.54_{0.28}$	$5.98_{0.13}$	$0.09_{0.04}$	2.088
FP	161	$1.84_{0.15}$	$0.34_{0.18}$	$5.95_{0.13}$	$0.05_{0.03}$	0.555
FY	311	$1.82_{0.15}$	$0.31_{0.11}$	$5.95_{0.13}$	$0.05_{0.02}$	3.440
$FP_{0.75}$	161	$1.85_{0.16}$	$1.44_{0.47}$	$5.96_{0.14}$	$0.20_{0.05}$	700.5
$FY_{0.75}$	311	$1.80_{0.15}$	$1.37_{0.34}$	$5.99_{0.13}$	$0.19_{0.04}$	150598

Genetic Grouping Strategies. As expected the number of genetic groups that were formed varied with each grouping strategy (Table 1), and ranged from 6 (μ) to 311 (FY).

The ratio of Bayes Factors for the different grouping strategies are presented in Table 1. Ratios of Bayes Factors of 3 and higher (Kass and Raftery 1995) indicate positive evidence against the null hypothesis of no genetic group variance. Without an auto correlation within flocks only the FY grouping strategy results in positive evidence against the null hypothesis. However, if an autocorrelation between the genetic groups is fitted within flock both the ratio of Bayes Factors for $FP_{0.75}$ and $FY_{0.75}$ indicate very strong evidence against the null hypothesis. From results presented in Table 1 we can conclude that grouping strategy μ is outperformed by all other strategies, including not fitting genetic groups at all. Grouping strategies F and FY are both better than grouping strategy FP. Fitting a within flock correlation between genetic groups has an enormous impact on the estimation of genetic group variance. This correlation also impacts on the genetic group solutions (Figure 1). When the within flock correlation between genetic groups is taken into account genetic group solutions are spread out more (Figure 1). Within flock correlations between estimated breeding values (with breeding values expressed as u + Qg), where high and close to unity.

Alternate assumptions about prior distributions and covariances between genetic groups are currently under further investigation.

CONCLUSION

Genetic groups formed on the combination of flock and year of birth appears to be the most accurate representation of animals with unknown genetic merit. Within the grouping strategies that do not fit a within flock correlation between genetic groups, grouping strategies F and FY appear to be slightly favourable than fitting no genetic groups. The grouping strategy FY with an autocorrelation between genetic groups within flocks is vastly superior when compared to the other grouping strategies.

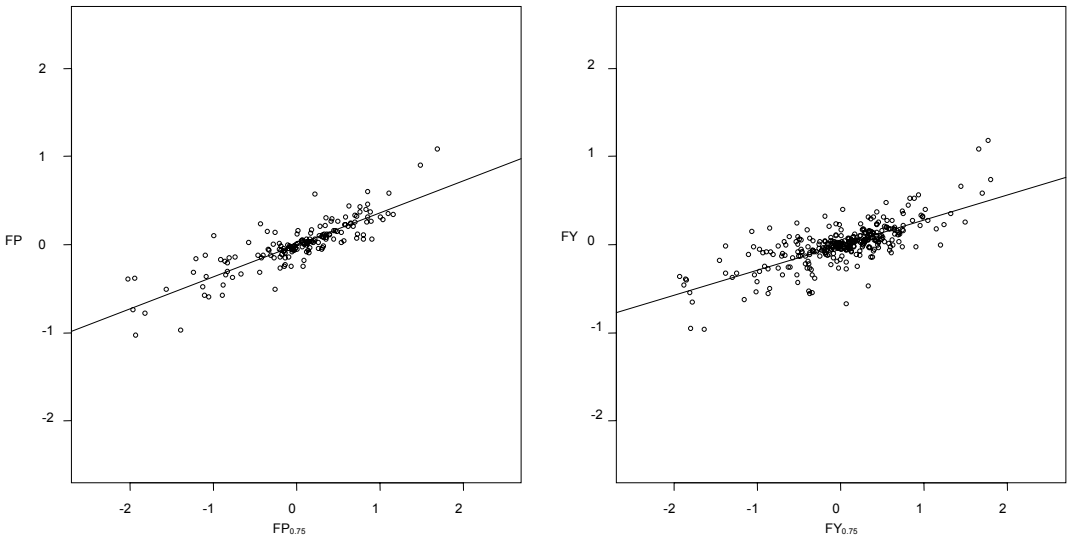

Figure 1. Genetic group solutions for the flock-period (left) and flock-year (right) grouping strategies with and without the 0.75 autocorrelation between groups within flocks.

ACKNOWLEDGEMENTS
This research was funded by Meat and Livestock Australia (MLA) project. SGA is also funded by Meat and Livestock Australia and Australian Wool Innovation, and SGA is made possible through the support of the Australian sheep industry.

REFERENCES
Brown D.J., Ball A., Huisman A.E., Swan A.A., Atkins K.D., Graser H.U., Banks R., Swan P. and Woolaston R. (2006) *Proc. 8th World Congr. Genet. Appl. Livest. Prod.* No 05-03.
Casellas J., Piedrafita J., and Varona L. (2006) *Proc. 8th World Congr. Genet. Appl. Livest. Prod.* 24-24.
Garcia-Cortes L.A., Cabrillo C., Moreno C. and Varona L. (2001) *Genet., Sel., Evol.* **33**:3.
Garcia-Cortes, L. A., Rico, M. and Groeneveld, E. (1998) *J. Anim. Sci.* **76**:441.
Kass R.E. and Raftery A.E. (1995) *Journal of the American Statistical Association* **90**:773..
Varona L., Garcia-Cortes L.A. and Perez-Enciso M. (2001) *Genet., Sel., Evol.* **33**:133.
Westell R.A., Quaas R.L. and Van Vleck L.D. (1988) *J. Dairy Sci.* **71**:1310.