Proc. Assoc. Advmt. Anim. Breed. Genet: 24: 205-208

ESTIMATES OF ACCURACY FOR GENOMIC-ONLY BREEDING VALUES AND FLOCK PROFILES IN AUSTRALIAN MERINO SHEEP

L. Li, P.M. Gurman, A.A. Swan and D.J. Brown

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

SUMMARY

As uptake of genomic-only estimated breeding values (GEBVs) and Flock Profiles – defined as the average GEBV from a random sample of approximately 20 animals from a commercial flock – expands within the Australian sheep industry, the need for accurate and reliable results becomes increasingly critical. Currently, GEBVs are presented without accuracy estimates, making it difficult to judge the genomic connectedness of the genotype to the reference population. This study aims to develop and implement a robust methodology for estimating the accuracy of GEBVs and Flock Profiles. Building on existing genetic evaluation accuracy frameworks, we propose a new method that can be integrated into routine evaluations, providing users with confidence in the reliability of their results.

INTRODUCTION

Single-step genomic BLUP (SS-GBLUP) combines pedigree, phenotypic, and genomic information and is routinely used in genetic evaluations for the Australian sheep and beef cattle industries (Brown *et al.* 2018; Johnston *et al.* 2018). One of the key advantages of genomic selection is its ability to calculate genomic-only estimated breeding values (GEBVs) for individuals without phenotypes and pedigree information. This application has been implemented in the Australian Merino sheep industry by back-solving SNP effects from SS-GBLUP EBVs, and then calculating GEBVs for animals with only genotype data (Swan *et al.* 2018). Currently, the resulting GEBVs for the genotype-only animals are then averaged to establish a benchmark for commercial flocks, referred to as Flock Profiles, using a random sample of approximately 20 animals from a flock.

However, this process does not estimate the associated accuracy for these predictions. Some flocks are genetically more different from the genotypes of the reference population in MERINOSELECT, which is expected to result in GEBVs of lower accuracy. It is crucial for the interpretation of GEBVs that predictions are sufficiently accurate and, like ASBVs, are only reported if they reach reporting thresholds for accuracy. Additionally, having an accuracy estimate attached to Flock Profiles and individual animal GEBVs is vital as these products are further adopted. This study aimed to derive a methodology for approximating the prediction error variance (PEV) and accuracy for both GEBVs and Flock Profiles. Initially, the methods were validated using a small dataset by comparing the approximated PEV with true PEV. Subsequently, these methodologies were applied to recent data from the Merino sheep industry to assess the accuracy of both GEBVs and Flock Profiles for a subset of traits.

MATERIALS AND METHODS

Derivation of Accuracy for GEBVs. Single-trait SNP-BLUP models can approximate accuracies for the genomic component of the single-step BLUP model (Li *et al.* 2023). In short, consider a single trait SNP-BLUP model with a mixed model equation (MME) $\begin{bmatrix} \mathbf{X'X} & \mathbf{X'ZW} \\ \boldsymbol{W'}\mathbf{Z'X} & \boldsymbol{W'}\mathbf{Z'ZW} + \mathbf{I}\lambda \end{bmatrix} \begin{bmatrix} \boldsymbol{b} \\ \boldsymbol{s} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X'y} \\ \boldsymbol{W'Z'y} \end{bmatrix}$, where **y** is the vector of observations, **b** is the vector of

---

[*] A joint venture of NSW Department of Primary Industries and Regional Development and the University of New England

fixed effects, $\mathbf{s}$ is the vector of estimated SNP effects, $\mathbf{X}$ and $\mathbf{Z}$ are incidence matrices which map observations to fixed effects and individual animals' breeding values, respectively, $\mathbf{W}$ is the animals by markers matrix of centred and scaled marker genotypes and $\lambda = \sigma_e^2/\sigma_s^2$, i.e. the ratio of the residual variance and the marker variance. The prediction error covariance (PEC) for SNP effect, $\boldsymbol{C}^{\mathbf{22}}$, can be obtained by inverting the left-hand side of MME, which by block matrix inversion rules can be written as $\mathbf{C}^{\mathbf{22}} = [\boldsymbol{W}'\mathbf{Z}'(\mathbf{I} - \mathbf{P})\mathbf{Z}\boldsymbol{W} + \mathbf{I}\boldsymbol{\lambda}]^{-\mathbf{1}}$, where $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'$.

The PEC for GEBVs for all animals can therefore be calculated as $\boldsymbol{PEC_{GEBVs}} = \boldsymbol{W}\boldsymbol{C}^{\mathbf{22}}\boldsymbol{W}'\sigma_e^2$. In practice, only the diagonal elements of the PEC are required for individual animal calculations. After the derivation of PEC, the GEBV accuracy of individual animal can be calculated as $\mathbf{Acc} = \sqrt{\mathbf{1} - \frac{\mathbf{diag}(\boldsymbol{PEC_{GEBVs}})}{diag(\boldsymbol{WW'})\sigma_a^2}}$, where $diag(\boldsymbol{WW'})$ are the diagonal elements of the genomic relationship matrix and σ_a^2 is the additive genetic variance. It is worth noting that the PEV for GEBVs calculated by this algorithm used SNP information only. We consider these PEV as approximations for the PEV of EBV of these genotype-only animals if they are derived in the routine SS-GBLUP analysis directly. This model utilizes an H matrix with a lambda value to integrate genomic and pedigree relationship matrices to derive EBVs (Brown *et al.* 2018).

Derivation of PEV and Accuracy for Flock Profiles. The above formula for an individual animal's accuracy can be extended to describe the accuracy for a group of animals. Suppose the flock profile contains genotypes from *n* individual animals, with the centred and scaled matrix $\boldsymbol{W_{FP}}$ obtained as the subset of $\boldsymbol{W}$ for the animals in the profile. The PEC for the animals is $PEC_{FP} = \boldsymbol{W}_{FP}\boldsymbol{C}^{22}\boldsymbol{W}'_{FP}\sigma_e^2$. If we define a vector $\boldsymbol{v} = \frac{1}{n}\mathbf{1}'$ as an n-dimensional column vector with each weight as 1/n, then the accuracy for the flock profile is $Acc_{FP} = \sqrt{1 - \frac{\boldsymbol{v}'\boldsymbol{PEC_{FP}}\mathbf{v}}{\boldsymbol{v}'(\boldsymbol{W_{FP}W'}_{\boldsymbol{FP}})\mathbf{v}\sigma_a^2}}$. These calculations have been incorporated into the current snpEPN genomic accuracy software (Li *et al.* 2023).

Validation. A small test dataset from the MERINOSELECT analysis was extracted to validate the methodology as a subset of flocks and genotypes. The validation aimed to confirm that our PEV estimates are close to true values and achieve higher accuracy for Merino genotypes versus lower accuracy for non-Merino genotypes. This dataset comprised 106,000 animals, with 38,492 having genotypes for 61,146 SNPs. Approximately half of the genotyped animals were extracted from Merino flocks designated as the reference population, while the remaining genotypes were added from non-Merino flocks, which were expected to be genetically highly different from these reference genotypes for testing purposes. Three traits with different reference sizes were investigated: weaning weight (WWT, 15,279), yearling greasy fleece weight (YGFW, 8,379), and adult staple length (ASL, 1,186). Genetic parameters used in this validation were extracted from the MERINOSELECT with heritability of 0.29, 0.34 and 0.34 for WWT, YGFW and ASL, respectively. In this validation, the PEV of GEBVs from snpEPN were first compared with the PEV based on SS-GBLUP model (SS-PEV) with assumption that these genotype-only animals were included in the routine analysis. SS-PEV were calculated directly from the inverse of mixed model equations using an H relationship matrix with a lambda value of 0.95 to combine genomic and pedigree relationship matrices. Then, the accuracy of GEBVs was calculated and compared by breeds to demonstrate the influence of reference breeds on predictive accuracy.

Example data. The described methods were also applied to a MERINOSELECT analysis extracted from the October 1, 2024 run, including 6,359,892 animals, of which 523,845 were genotyped for 61,260 SNPs. Additionally, 78,852 genotypes were back-solved, including 27,670 genotypes from 996 test flocks. Traits investigated included birth weight (BWT), weaning weight (WWT), yearling weight (YWT), intramuscular fat (IMF), yearling eye muscle depth (YEMD), yearling greasy fleece weight (YGFW), and yearling fleece diameter (YFD). Heritability values used for accuracy approximation, along with the count of animals having both genotypes and phenotypes

(Nref) ranging from 7,357 for IMF to 396,307 for WWT for each trait, are detailed in Table 1.

Table 1. Heritability (h^2) and number of animals with both genotypes and phenotypes (Nref) for live weight, carcass and fleece traits extracted from the MERINOSELECT analysis

Trait	BWT	WWT	YWT	IMF	YEMD	YGFW	YFD
h^2	0.22	0.34	0.41	0.52	0.31	0.34	0.58
Nref	29,775	396,307	273,915	7,357	165,088	197,280	212,179

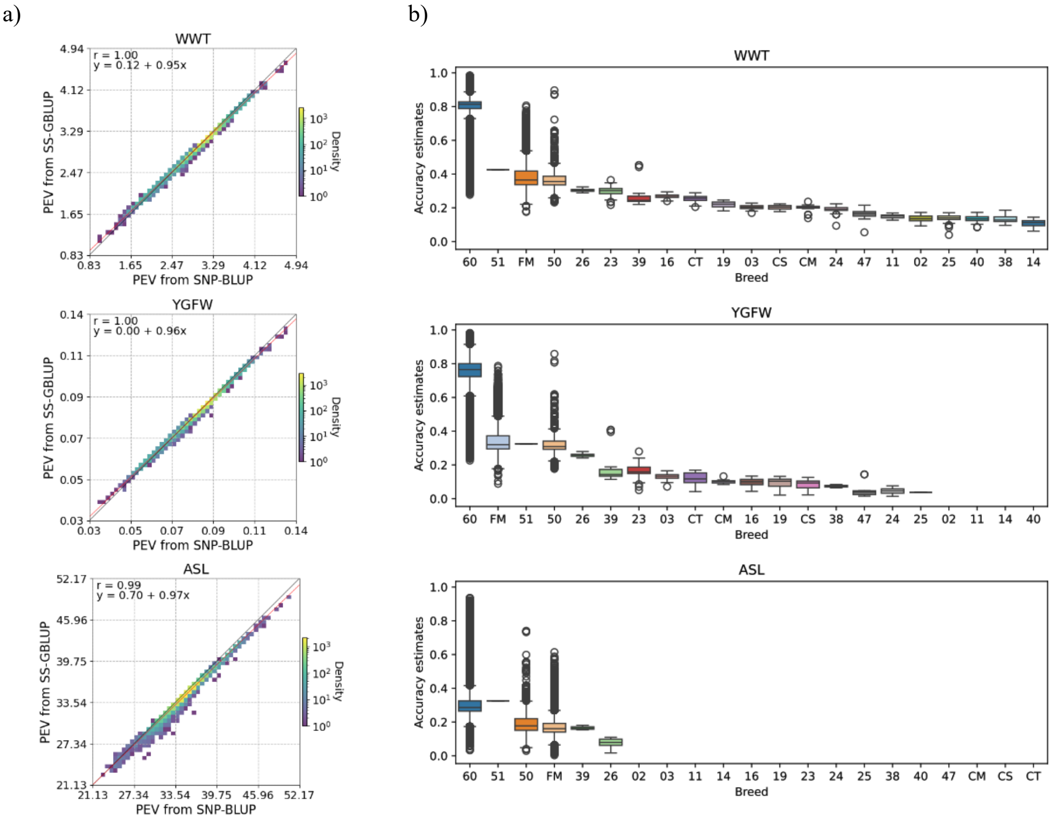

Figure 1. a) PEV from SNP-BLUP versus PEV from SS-GBLUP; b) Accuracy estimates for all animals by breeds for weaning weight (WWT), yearling greasy fleece weight (YGFW), and adult staple length (ASL). Note: Corresponding breed names for the codes listed in b) can be found at: https://www.sheepgenetics.org.au/resources/breeders-guide/animal-identification/.

RESULTS AND DISCUSSION

Validation results. The PEV of GEBVs for genotyped animals estimated from SNP-BLUP was very close to the PEV based on SS-GBLUP model, with correlations near 1 for all traits (Fig. 1a), indicating that the algorithms implemented were functioning as expected. The accuracy of EBVs for all animals across traits by breeds are shown in Fig. 1b. The Merino breed (60) consistently displayed higher accuracy for all traits, with a wider spread indicating variability within the breed. Other breeds less related to the reference flocks generally presented much lower accuracies, especially for YGFW and ASL, where most breeds had accuracy less than 0.2 (YGFW) and close to zero (ASL).

This may be due to the reference animals predominantly being from Merino breeds, with the reference size reducing from 15,279 for WWT to 1,186 for ASL.

Example data results. Figure 2 shows two boxplots illustrating the accuracy of GEBVs and Flock Profiles across all traits. The GEBVs show a wide range of accuracies, with BWT and IMF showing much lower accuracy (~0.4-0.5) and greater variability. In contrast, other traits display higher accuracies (>0.8) with narrower spreads, indicating more accurate and consistent predictions. These results correspond to the size of the reference population, as detailed in Table 1. Not surprisingly, the accuracy of Flock Profiles (Figure 2b) presents a pattern like GEBVs but with generally higher and more uniform accuracies. These high accuracy levels for both GEBVs of genotype-only animals and Flock Profiles are highly promising for the implementation of genomic selection in the Australia Merino sheep industry. However, these high accuracy estimates require further validation. This could be achieved, for instance, by comparing the accuracy of testing flocks with the linkage metrices between genomic reference populations and testing flocks.

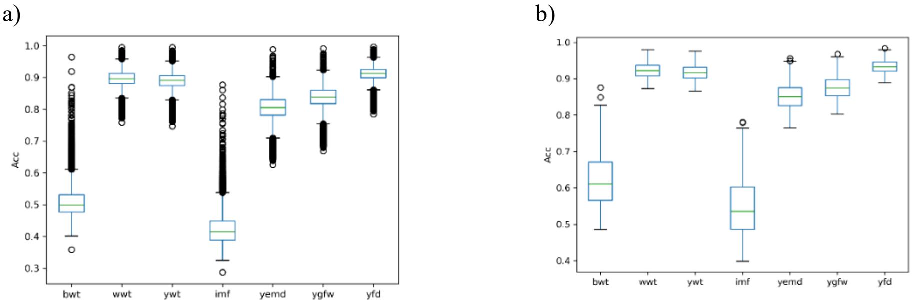

Figure 2. Accuracy by traits for a) GEBVs; b) Flock Profiles

CONCLUSIONS

A methodology for estimating the accuracy of GEBVs and Flock Profiles has been developed and incorporated into the current snpEPN program. This methodology was validated using a small dataset showing high correlations between PEV derived from SNP-BLUP and PEV from SS-GBLUP. Very high accuracy (>0.8) was observed for both GEBVs and Flock Profiles within a real Merino sheep dataset for traits with substantial-sized reference populations, such as WWT, YWT, YEMD, YGFW and YFD. However, lower accuracy (0.4-0.5) was noted for traits with smaller reference populations, such as BWT and IMF. By providing accuracy estimates for GEBVs and Flock Profiles, the Merino sheep industry can benefit from a more transparent and reliable genetic evaluation system. This methodology can be implemented in other sheep breeds and beef cattle.

ACKNOWLEDGEMENTS

This research was funded by Meat and Livestock Australia project L.GEN.2501. The authors acknowledge the contribution of data from Australian sheep breeders.

REFERENCES

Brown D.J., Swan A.A., Boerner V., Li L., Gurman P.M., McMillan A.J., van der Werf J.H.J., Chandler H.R., Tier B. and Banks R.G. (2018) *Proc. World Cong. Genet. Appl. Livest. Prod.* **11**: 460.

Johnston D.J., Ferdosi M.H., Connors N.K., Boerner V., Cook J., Girard C.J., Swan A.A. and Tier B. (2018) *Proc. World Cong. Genet. Appl. Livest. Prod.* **11**: 269.

Li L., Gurman P., Swan A.A. and Tier B. (2023) *Anim. Prod. Sci.* **63**: 1086.

Swan A.A., Gurman P.M., Boerner V., Brown D.J., Clark S., Gore K., Granleese, T. and van der Werf J.H.J. (2018) *Proc. World Cong. Genet. Appl. Livest. Prod.* **11**: 1.