Proc. Assoc. Advmt. Anim. Breed. Genet. 22:89-92

APPROXIMATING THE ACCURACY OF SINGLE STEP EBVS

L. Li, A.A. Swan and B. Tier

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

SUMMARY

To accompany the implementation of multi-trait Single Step Genomic BLUP (SS-GBLUP) in the BREEDPLAN and OVIS genetic evaluation systems, an algorithm to approximate accuracy with genomic information has been developed and is presented in this paper. Data from full terminal sire OVIS and Brahman BREEDPLAN runs were processed using this new method. Results demonstrate that the approximated accuracy of SS-GBLUP estimated breed values (EBVs) is highly correlated (R^2 >0.96) with exact accuracies in several small example analyses for both beef and sheep. SS-GBLUP EBV accuracies increase more for traits with a larger reference population and for traits with higher heritabilities. Animals with low pedigree-only (ABLUP) EBV accuracies benefit more from genomic information than animals with high ABLUP EBV accuracies.

INTRODUCTION

Single Step Genomic BLUP (SS-GBLUP, e.g. Legarra *et al*. 2014) was implemented in the Australian sheep and beef cattle evaluation systems OVIS and BREEDPLAN during 2016, simultaneously combining phenotypic, pedigree, and genomic information. Conceptually, SS-GBLUP is compatible to the existing pedigree BLUP models and is relatively straightforward to implement by replacing the traditional inverse pedigree relationship matrix ($\mathbf{A}^{-1}$) in the mixed model equations (MME) with $\mathbf{H}^{-1}$ (Christensen and Lund, 2010):

$$\boldsymbol{H}^{-1} = \boldsymbol{A}^{-1} + \begin{pmatrix} 0 & 0 \\ 0 & \boldsymbol{G}^{-1} - \boldsymbol{A}_{22}^{-1} \end{pmatrix}$$

where $\mathbf{G}$ and $\mathbf{A}_{22}$ are genomic and pedigree relationship matrices for genotyped animals, respectively. This make modification of models and software to estimate breeding values (EBVs) relative straightforward, although computational requirements can increase significantly.

Accuracies of EBVs are also an important output of genetic evaluation systems, and these have traditionally been approximated using "effective progeny numbers" (EPN) as a basis which accumulate information from animals' own performance, progeny, parents, and from correlated traits (Graser and Tier 1997). In this paper, we present a modification to this algorithm to account for EPN from genomic information, allowing the calculation of accuracies for SS-GBLUP EBVs. We also investigate the impact of genomic information on the improvement of accuracy of EBV for real examples.

MATERIALS AND METHODS

Algorithms to derive "genomic EPN". In order to ensure compatibility with the current accuracy algorithm, information from the genomic relationship matrix needs to be expressed in the form of an EPN for each animal. This "genomic EPN" must be accumulated with existing sources of EPN to derive approximations of the total accuracy for multi-trait SS-GBLUP analyses. The steps required are described below.

Step 1. Calculate a prediction error variance (PEV) using a series of single trait GBLUP pseudo-analyses. For each trait, we construct the MME for genotyped animals with additive genetic effects

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

considered in the model, ignoring all the other fixed and random effects. The diagonal of the inverse of the MME then represents the genomic PEV for the trait. Because the pedigree relationship matrix $\mathbf{A}_{22}$ for these animals has already been used to contribute accuracy from pedigree and performance information, and also because a proportion of **G** is used to build $\mathbf{H}^{-1}$ in SS-GBLUP, an adjusted PEV must be used to derive the contribution of genomic information to accuracy. This adjusted PEV for the i[th] animal is calculated as:

$$PEV_i^* = wt * PEV_i + (1 - wt)\sigma_a^2$$

where σ_a^2 is the additive variance, and *wt* is a tuning parameter (referred to as the "genomic PEV weight" below) determined empirically by comparing approximate accuracies calculated across a range of *wt* values from 0.1 to 1.0 with exact accuracies calculated by direct inversion of the SS-GBLUP mixed model equations for a range of examples reported below.

After PEV^* for each trait is calculated with appropriate values of *wt*, accuracy is calculated as:

$$acc = \sqrt{1 - PEV_i^*/g_{ii}\sigma_a^2}$$

where g_{ii} is the diagonal of **G** for the i[th] animal. This is assumed to be the gain in accuracy due to genomic information for genotyped animals.

Step 2. Propagate genomic accuracy to un-genotyped ancestors and descendants so that the impact of genomic information on close relatives is included. Propagation is performed upwards first (to ancestors) and then downwards (to descendants). If an un-genotyped animal has its parents and progeny genotyped, accuracy is calculated from the progeny, except for the case where only one progeny and both parents are genotyped, in which accuracy is calculated from the parents. The accuracy of un-genotyped parents with genotyped progeny is given by:

$$acc = \overline{acc} \times (1 - 0.5^n)$$

where $\overline{acc}$ is the average accuracy over n genotyped progeny for the sire or dam. The accuracy of the un-genotyped progeny is given by:

$$acc = \sqrt{(acc_{sire}^2 + acc_{dam}^2)/4}$$

Step 3. Accuracy for genotyped animals and progeny and parents of genotyped animals is transformed to the equivalent number of effective progeny as:

$$EPN = \delta \times acc^2/(1 - acc^2)$$

where $\delta = (4 - h^2)/h^2$ and h^2 is the heritability of the trait.

Step 4. For each animal with genomic EPN derived from the above single trait analyses, multiple trait EPN are derived by constructing multiple trait MME with additive genetic effects as follows: 1) Accumulating the residual matrices based on the common minimal EPN across traits based on the phenotypes observed into a trait by trait matrix; 2) The additive genetic co-variance matrix is added to the accumulated residual matrix; 3) Multiple trait PEV are then calculated by inverting this matrix and then converted to EPN following the procedures above.

Step 5. Because EPN due to genomic information for each animal are confounded with EPN arising from phenotypic own-performance information, the final step is to calculate the difference between the genomic EPN of an animal and the EPN arising from its own phenotype, as calculated from the current algorithm. Only when this difference is positive is the genomic EPN accumulated with EPN from all other sources to derive the final accuracy.

Note that when calculating final EBV accuracies following the formula in equation above, rather than using g_{ii} for genotyped animals, we use $\lambda g_{ii} + (1 - \lambda)a_{ii}$, following the specification of the

H matrix used for SS-GBLUP, where λ is a weighting factor for genomic and pedigree information as described by McMillan and Swan (2017).

Selection of genomic PEV weight. Data from OVIS and BREEDPLAN runs were used to investigating the genomic prediction error variance weight. Traits considered for sheep were intramuscular fat (IMF, $h^2 = 0.56$) and shear force at day 5 (SF5, $h^2 = 0.32$), with the data including 11,416 genotyped animals from terminal sire evaluation. Traits considered for beef were beef 600 day weight (FWD, $h^2 = 0.49$) and days to calving (DTC, $h^2 = 0.08$), with the data including 5,847 genotyped animals from the Brahman BREEDPLAN evaluation. These data were analysed repeatedly with the new accuracy algorithm, fitting a range of genomic PEV weights from 0 to 1 in increments of 0.1. The approximate accuracies derived were then compared to exact accuracies calculated from PEVs derived by inversion of the mixed model coefficient matrix for each data set, and varying the value of λ used to construct the **G** matrix from 0 to 1 in increments of 0.1.

Application to industry data. The new SS-GBLUP accuracy algorithm was applied to data from full sheep terminal sire evaluation and full Brahman BREEDPLAN runs. The numbers of genotyped animals were 11,832 in sheep and 7,166 in beef data.

RESULTS AND DISCUSSION

Selection of the genomic PEV weight. Based on comparison of approximate accuracy calculated over a series of genomic PEV weights to exact accuracy with a series of λ values, results showed that the means and standard deviations of true accuracies increased with λ from 0 to 1. When λ = 0.5, the value currently chosen to run SS-GBLUP analyses in OVIS and BREEDPLAN, the closest genomic PEV weight for the new accuracy algorithm was 0.3 for all sheep and beef traits, based on comparison of means and standard deviations. Across all four traits, high R-squared values (>0.95) and regression coefficients (from 0.96 to 1.1) were observed for the regression of approximate accuracies with genomic PEV weight = 0.3 on true accuracies with λ = 0.5, indicating a genomic PEV weight of 0.30 is appropriate to tune the genomic prediction error variances for the current implementation of SS-GBLUP.

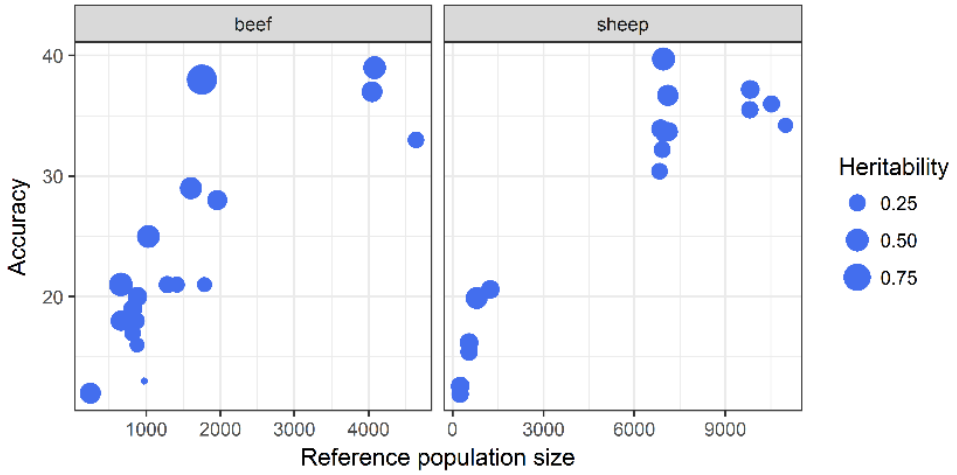

Figure 1. Relationship between average accuracy for genotyped animals arising from the genomic relationship matrix, number of genotyped animals recorded in the reference population, and heritability (size of points) for different sheep and beef traits.

The impact of genomic information on accuracy. The relationship between average accuracies for genotyped animals arising from the genomic information, number of genotyped animals recorded in the reference population, and heritability for sheep and beef traits are shown in Figure 1. The average accuracies of genotyped animals as calculated in Step 1 above varied from 0.12 to 0.40 in both sheep and beef across different traits. The accuracies were positively related to the number of animals with records and heritability for each trait.

Figure 2 shows the distribution of average accuracy improvement for SS-GBLUP relative to ABLUP for beef and sheep. For animals with low starting ABLUP accuracies (<30%), the SS-GBLUP accuracy was on average 18% points higher for sheep (ranging from 11 to 24% points), and on average 13% points higher for beef (ranging from 3 to 29% points). For medium starting accuracies (30 to 50%), the improvements were 6% (2 to 8%) for sheep and 4 (1 to 9%) for beef, while very little improvement in accuracy was observed for high starting accuracies (>50%). These trends confirm expected benefits of accuracy from genomic information.

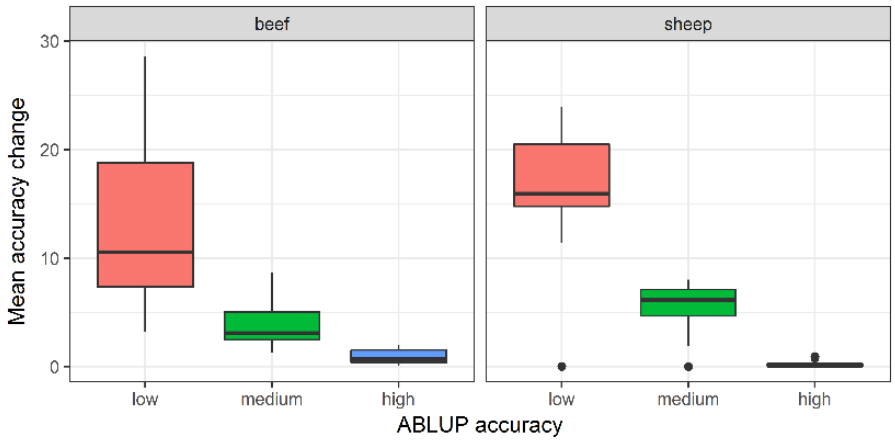

Figure 2. Distribution across beef and sheep traits of improvement of SS-GBLUP accuracies over ABLUP accuracies within bands of ABLUP accuracy from low (<30), medium (30 – 50) and high (>50).

CONCLUSIONS

An algorithm to approximate SS-GBLUP EBV accuracies was developed, and shown to be consistent with exact accuracies in several small example analyses for beef and sheep. SS-GBLUP EBV accuracies increase more for traits with a larger reference population (numbers of animals phenotyped and genotyped), and for traits with higher heritabilities. Animals with low pedigree-only (ABLUP) EBV accuracies gain more improvement in accuracy from genomic information than animals with high ABLUP EBV accuracies.

ACKNOWLEDGMENTS

This project was funded by the MLA through projects B.SGN.0127 and B.BFG.0050

REFERENCES

Christensen O.F., and Lund M.S. (2010) *Gen. Sel. Evol.* **42**: 2.
Graser H.-U and Tier B. (1997) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **12**: 547.
Legarra A., Christensen O.F., Aguilar I. and Misztal I. (2014) *Livest. Sci.* **166**: 54.
McMillan, A.J. and Swan, A.A. (2017) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **21:** these procedings.