Proc. Assoc. Advmt. Anim. Breed. Genet. 19:359-362

CHOICE OF PARAMETERS FOR REMOVAL OF INFLATION IN GENOMIC BREEDING VALUES FOR DAIRY CATTLE

B.L. Harris, D.L. Johnson and R.J. Spelman

LIC, Private Bag 3016, Hamilton, New Zealand

SUMMARY

A hybrid multiple step genomic evaluation procedure that uses a modified augmented relationship matrix which, simultaneously blends pedigree and genomic relationships, is outlined. The method allows the scale of the genomic predictions to be adjusted. The method was applied to an across breed genomic evaluation for protein yield and somatic cell score. The optimal scale values indicated that the un-scaled genomic matrix was 10-20% to large and the information coming from the parental index was 40-50% to large. It was also found that the scale of the parental index of the genotyped sires had a large impact on the inflation of the genomic breeding values but a smaller impacts on the accuracy of the genomic predictions.

INTRODUCTION

The use of genomic selection (Meuwissen *et al.* 2001) to increase the rates of genetic improvement is now widespread in livestock species. In dairy cattle, the use of genomic information from SNP panels has increased the published reliability of young unproven sires to close to sires graduating from a progeny test selection program. Interbull has published a validation test that determines the accuracy and bias of the dairy cattle genomic evaluation systems (Mäntysaari *et al.* 2010). The validation test compares a sire's subsequent daughter performance with his juvenile genomic breeding value (**BV**). Results from several validation tests have shown that in most cases the juvenile genomic BVs over estimate the daughter performance and are postively inflated (Mäntysaari *et al.* 2010). The cause of the inflation is unknown.

Genomic evaluation of dairy cattle generally uses a multiple step procedure (Hayes *et al.* 2009). The multiple step procedure uses the outputs from a traditional genetic evaluation as inputs to the estimation of genomic BV for genotyped animals. The inputs are either de-regressed breeding values (**DBV**) or daughter yield deviations (**DYD**). The genomic BVs are estimated for genotyped animals only. Then the genomic BVs are blended with parent average breeding values from the traditional genetic evaluation (Hayes *et al.* 2009). The blending process incorporates information in to the genomic BV from parents that were not genotyped and not in the genomic evaluation.

A single-step procedure for genetic evaluation has been proposed by Misztal *et al.* (2009) that includes the genomic information directly in to a traditional genetic evaluation. There are two benefits of this approach. Firstly, the genomic BVs are calculated directly from the phenotype records rather than from DYDs or DBVs. Secondly, all the pedigree information is used to calculate the genomic BV, which removes the need for blending. The single step method augments the pedigree-based relationship matrix by contributions from the genomic relationship matrix. A simplified inverse of the augmented relationship matrix has improved the feasibility of the single-step approach in genetic evaluations (Christensen and Lund 2010). Recently, Misztal *et al.* (2010) have enhanced the single step method by modifying the augmented relationship matrix to adjust for the scale of the genomic predictions. The adjustment to the scale provides a way to adjust for inflation of the GBVs.

The first aim of this study was to incorporate the modified augmented relationship matrix into a multiple step procedure. This would allow the modified augmented relationship matrix to be used to provide genomic evaluations for systems where it is currently computationally infeasible

to run a single step analysis, such as multiple trait test-day models. The second objective was to investigate the effects of modifications to the augmented relationship matrix on genomic BVs for protein yield and somatic cell score (**SCS**) with respect to inflation and accuracy in the New Zealand Holstein-Friesian (HF), Jersey (J) and HF x J crossbred joint genomic evaluation.

METHODS AND MATERIALS

Data. Genetic markers from the Illumina BovineSNP50 Beadchip were available for 5180 sires. There were 41,032 SNPs available per sire after editing. Traditional BVs were calculated from 157,502,869 test-day protein yields and somatic cell scores from 1986 to November 2010. The traditional BV evaluation had pedigree records on 21,417,977 animals recorded from 1960 onwards.

Methods. A multiple lactation test-day animal model was used for the traditional BV calculation where each lactation was considered as a separate trait. The SCS model included the first 3 lactations per cow and protein yield included the first 6 lactations per cow.

The genomic evaluation was undertaken using a multiple step approach. First, DBVs where calculated from the protein yield and SCS test-day genetic evaluation models. The DBVs were calculated for animals with genomic data and for their immediate parents irrespective of whether the parents had genomic data or not. Second, the inverse of augmented relationship ($\mathbf{H}^{-1}$) was formed for animals with genomic data and their immediate parents as:

$$\mathbf{H^{-1}} = \mathbf{A^{-1}} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \tau\left((0.95\mathbf{G} + 0.05\mathbf{A_{22}})^{-1} - \varpi\mathbf{A^{22}}\right) \end{bmatrix}$$

where **A** is a pedigree-base relationship matrix, **G** is a genomic relationship matrix, $\mathbf{A}^{22}$ is the inverse of the pedigree-based relationship matrix ($\mathbf{A_{22}}$) for genotyped animals, ϖ is a weight factor for $\mathbf{A}^{22}$ and τ is weight factor for the scale of genomic relationship matrix. In a single breed analysis the genomic relationship matrix (**G**) is calculated from the SNP marker matrix so that the SNP markers have a mean of zero and a variance equivalent to the additive relationship matrix (VanRaden 2008). In an across breed analysis, the SNP marker matrix has to be adjusted so that the SNP markers have a mean of zero within breed and that the variances within and across breed are equivalent to the additive relationship matrix. Finally, the GBVs were calculated using the mixed model equations from VanRaden (2008) with the inverse of augmented relationship substituted for the inverse of the genomic relationship. Comparison of the proposed method (**method H**) with VanRaden's (2008) standard multiple step procedure (**method G**) was undertaken by setting both weighting factors to 1. To determine the effect of the weighting factors on the accuracy and the inflation of the GBVs a number of genomic evaluations were undertaken across a 2 dimensional grid of weighting factors. The weighting factor τ was varied from 0.5 to 1.5 in 0.1 steps and the weighting factor ϖ was varied from 0.1 to 1.0 in 0.1 steps. The accuracy and the level of inflation of the GBVs for each genomic evaluation were calculated by regressing the 2005 GBVs on the 2010 DBVs for 3 crops of young bulls. The accuracy was calculated as the square root of the regression r-square value.

RESULTS AND DISCUSSION

The GBV means and standard deviations for method H and method G were close to identical for sires with daughters for both traits. In contrast, the means for the juvenile sires without daughters were regressed more towards the breed means for both traits for method H. The

correlations among the GBVs for sires with daughters from method H and G were greater than 0.99 for both traits. The regression coefficients from regressing method H GBVs on method G GBVs were between were greater than 0.98 for sires with daughters for SCS and protein yield. The correlations among methods for juvenile sires were between 0.96 and 0.97 for both traits. The corresponding regression coefficients were between 0.94 – 0.99 and 0.96 – 0.98 for SCS and protein yield, respectively. The differences among the GBVs for juvenile sires between the two methods is a measure of the errors resulting from approximations in the blending process in method G.

Table 1. The inflation and accuracy from regressing 2005 protein yield genomic breeding values on 2010 deregressed breeding values for different values of the weighting factors.

Value for τ	Value for ϖ	Inflation			Accuracy		
		HF	J	X	HF	J	X
0.5	0.1	0.980	1.012	1.126	0.504	0.524	0.664
0.5	0.5	0.890	0.912	1.052	0.501	0.512	0.668
0.5	1.0	0.588	0.572	0.730	0.462	0.452	0.632
1.0	0.1	1.028	1.029	1.123	0.531	0.544	0.668
1.0	0.5	0.959	0.956	1.062	0.533	0.540	0.671
1.0	1.0	0.757	0.745	0.861	0.519	0.512	0.656
1.5	0.1	1.047	1.034	1.118	0.541	0.553	0.667
1.5	0.5	0.991	0.977	1.067	0.545	0.551	0.671
1.5	1.0	0.843	0.827	0.923	0.540	0.535	0.663

HF = Holstein Friesian, J = Jersey and X = Holstein Friesian x Jersey Crossbred Sires

The inflation and accuracy results for different levels of τ and ϖ are summarised in Tables 1 and 2 for protein yield and SCS, repsectively. Changes to τ while keeping ϖ constant had small impacts on both the inflation and accuracy. Whereas, changes to ϖ while keeping τ constant had larger impacts on both the inflation and accuracy of the GBVs. The optimal values of ϖ and τ for protein yield and SCS were derived by maximising the accuracy while attempting to keep the inflation between 0.95 and 1.05 for all breeds. The optimal value of τ for protein yield was 1.1 and SCS was 1.2. The optimal value of ϖ for protein yield was 0.6 and SCS was 0.5. The optimal values for τ indictae that the genomic matrix was 10-20% too large in terms of the scale. The scale of the parental index of the young genotyped sires had a large impact on the GBV inflation (parameter ϖ) . For both protein yield and SCS reducing the scale of the parental index reduced the inflation in the GBVs. It was evident that choosing single values for τ and ϖ across breeds is a comprise with the Jersey sires having a greatest level of inflation. The ϖ optimal values indicate that the information coming from the parental index in the GBV should be reduced by 40% to 50%, compared to an un-scaled genomic evaluation. The results in this study are similar to the results reported by Misztal *et al.* (2010). Misztal *et al.* (2010) studied to final score data

Table 2. The inflation and accuracy measures from regressing 2005 somatic cell score genomic predictions on 2010 deregressed breeding values for different values of the weighting factors.

Value for τ	Value for ω	Inflation			Correlation		
		HF	J	X	HF	J	X
0.5	0.1	0.874	1.061	1.108	0.449	0.524	0.580
0.5	0.5	0.866	1.045	1.092	0.451	0.528	0.582
0.5	1.0	0.856	1.027	1.072	0.452	0.532	0.585
1.0	0.1	0.843	1.005	1.048	0.453	0.536	0.587
1.0	0.5	0.827	0.979	1.020	0.454	0.540	0.589
1.0	1.0	0.806	0.948	0.985	0.454	0.544	0.590
1.5	0.1	0.780	0.910	0.942	0.453	0.548	0.591
1.5	0.5	0.745	0.863	0.889	0.451	0.551	0.591
1.5	1.0	0.697	0.804	0.821	0.447	0.553	0.588

HF = Holstein Friesian, J = Jersey and X = Holstein Friesian x Jersey Crossbred Sires

from 10.5 million USA Holstein cows. They reported a regression coefficient of 0.75 when no modifications were made to the augmented relationship matrix. Misztal *et al.* (2010) found that reducing the fraction of information from genomics and parents both by 50% resulted in zero inflation in the genomic BVs and very little change in the accuracy.

The hybrid multiple step approach outlined in this study removes need to blend genomic and parent average BVs, as well as, providing a mechanism to reduce the inflation in GBVs for juvenile sires. However, the choice of optimal augmentation parameters will be more challenging in across breed genomic evaluations compared to single breed evaluations. With higher density SNP panels becoming available, further research is required to quantify the inflation and scale parameters for these new panels.

REFERENCES

Christensen O. F. and Lund M. S. (2010) *Genet. Sel. Evol.* **42**:2

Mäntysaari E, Liu Z. and VanRaden P. (2010). "Proceedings of the interbull international workshop: Genomic Information in Genetic Evaluations" Paris, France

Meuwissen T. H. E., Hayes B. J. and Goddard M. E. (2001) *Genetics* **157**:1819

Misztal I., Legarra A. and Aguilar I. (2009). *J. Dairy Sci.* **92**:4648

Misztal I., Aguilar I., Tsuruta S., Legarra A. and Lawlor T. J. (2010) "Choices of parameters for single-step genomic evaluation for type". Heraklion, Greece: EAAP

Hayes B. J., Bowman P. J., Chamberlain A. J. and Goddard M. E. (2009) *J. Dairy Sci.* **92**:433

VanRaden P. M. (2008) *J. Dairy Sci.* **91**:4414