STØTTET AF
mælkeafgiftsfonden

Comparison of Model Reliabilities from Single-Step and Bivariate Blending Methods

M. Taskinen[1], E.A. Mäntysaari[1], M.H. Lidauer[1], T. Knürr[1], J. Pösö[2], G. Su[3], G.P. Aamand[4], and I. Strandén[1]

[1] MTT Agrifood Research Finland, Biotechnology and Food Research, Biometrical Genetics, Jokioinen, Finland
[2] FABA Co-op, Vantaa, Finland
[3] Department of Molecular Biology and Genetics, Aarhus University, Denmark
[4] NAV Nordic Cattle Genetic Evaluation, Aarhus, Denmark

Abstract

Model based reliabilities in genetic evaluation are compared between three methods: animal model BLUP, single-step BLUP, and bivariate blending after genomic BLUP. The original bivariate blending is revised in this work to better account animal models. The study data is extracted from the production trait evaluation of Nordic Red dairy cattle. Genotyped bulls with daughters are used as training animals, and genotyped bulls and producing cows as candidate animals. For simplicity, size of the data is chosen so that the full inverses of the mixed model equation coefficient matrices can be calculated. Model reliabilities by the single-step and the bivariate blending methods were higher than by animal model due to genomic information. Compared to the single-step method, the bivariate blending method reliability estimates were, in general, lower. Computationally bivariate blending method was, on the other hand, lighter than the single-step method.

Key words: genomic evaluation, reliability, single-step genomic evaluation

Introduction

There is increasing interest on estimating model reliability in genetic evaluation that use both genomic and pedigree information. Reliabilities of genomic enhanced estimated breeding values (GEBV) of individual animals exhibit large differences: range is from pedigree accuracy to accuracy of full progeny test. In addition to quantifying accuracy of the indices, the reliabilities are also used as weights in the international bull genomic evaluations. Often the GEBVs are calculated using mixed model equations (MME) with genomic relationship matrix ($\boldsymbol{G}$). Model reliability for basic genomic BLUP (GBLUP) is easily computed if $\boldsymbol{G}^{-1}$ can be formed (Strandén and Garrick, 2009). Matrix $\boldsymbol{G}$ has a size of number of animals genotyped *N*. The GBLUP coefficient matrix of MME can be inverted if $\boldsymbol{G}$ can be inverted, because MME has size *N*+1.

In the future, genomic evaluations will be mostly based on single-step BLUP (ssGBLUP by Aguilar et al. 2010; Christensen and Lund 2010). Then, the exact model based reliability estimation requires inverting a matrix of size all animals in the evaluations. Approximations have been suggested by Misztal et al. (2013) based on added genomic information into MME.

Nordic genomic evaluations use bivariate blending (Mäntysaari and Strandén, 2010) to combine direct genomic value (DGV) and traditionally estimated breeding values (EBV). Bivariate blending is based on a bivariate model having information from these models as two correlated traits, phenotypic "trait in interest" (EBV) and the estimated DGV. The DGV is considered to have 100% accuracy and a correlation of $\sqrt{R^2_{DGV}}$ with the EBV-trait. As

long as $\sqrt{R^2_{DGV}}$ is known, the reliability of GEBV from the bivariate blending is easily estimated using standard bivariate reliability approximation. This approximation can be calculated using any standard multivariate reliability approximation approaches (Tier and Meyer, 2004) etc.).

Our aim in this study was to compare model based reliability computed from the full inverse of MME using: animal model BLUP (AM-BLUP), single-step BLUP (ssGBLUP), and bivariate blending after GBLUP (bbGBLUP). Moreover, the original bivariate blending was revised to better account animal models.

Materials and Methods

Estimation of Reliability

Consider the model

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Zu} + \boldsymbol{e}$$

where $\boldsymbol{y}$ is vector of observations, $\boldsymbol{b}$ is vector of fixed effects, $\boldsymbol{X}$ is design matrix, $\boldsymbol{u}$ is vector of random effects, and $\boldsymbol{e}$ is random residual vector. Assume $\mathrm{Var}(\boldsymbol{u}) = \boldsymbol{V_u}$ and $\mathrm{Var}(\boldsymbol{e}) = \boldsymbol{R}$. Let inverse of the coefficient matrix of the MME be

$$\boldsymbol{C}^{-1} = \begin{bmatrix} \boldsymbol{C}^{b,b} & \boldsymbol{C}^{b,u} \\ \boldsymbol{C}^{u,b} & \boldsymbol{C}^{u,u} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X'R^{-1}X} & \boldsymbol{X'R^{-1}Z} \\ \boldsymbol{Z'R^{-1}X} & \boldsymbol{Z'R^{-1}Z} + \boldsymbol{V_u^{-1}} \end{bmatrix}^{-1}.$$

In AM-BLUP

$$\boldsymbol{V_u^{-1}} = \frac{1}{\sigma^2_u}\boldsymbol{A^{-1}}$$

and in ssGBLUP

$$\boldsymbol{V_u^{-1}} = \frac{1}{\sigma^2_u}\left[\boldsymbol{A^{-1}} + \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{G^{-1}} - \boldsymbol{(A_{22})^{-1}} \end{bmatrix}\right]$$

where $\boldsymbol{A}$ is the pedigree based relationship matrix, $\boldsymbol{G}$ genomic relationship matrix, $\boldsymbol{A_{22}}$ contains pedigree based relationships of genotyped animals, and σ^2_u is the genetic variance. Let the residual covariance matrix $\boldsymbol{R}$ be diagonal with j^{th} diagonal element σ^2_e/w_j where σ^2_e is the residual variance, and w_j is weight for observation j. In AM-BLUP and ssGBLUP the weight is effective daughter contribution (EDC) of corresponding deregressed proof (DRP).

Model reliability for animal i is calculated as

$$r_i^2 = 1 - \frac{\{\boldsymbol{C}^{u,u}\}_i}{\sigma^2_u}$$

where $\{\boldsymbol{C}^{u,u}\}_i$ is the diagonal element corresponding animal i.

Bivariate Blending

The revised bivariate blending method is performed in three steps.

Step 1: Get reliabilities r^2_{EBV} from AM-BLUP.

Step 2: Calculate reliability increase due to genotypes. First, estimate the EDC for all genotyped animals. For bulls, this is based on non-genotyped daughters, and for cows, the EDC is $\frac{\sigma^2_e r^2_0}{\sigma^2_u(1-r^2_0)}$ where r^2_0 is the Interbull reliability for cows own performance (Strandén, I., et al., 2000). With these EDC as weights, the model reliabilities r^2_{DGV} for DGV are calculated from GBLUP. The relative increase in evaluation accuracy from AM-BLUP due to GBLUP for genotyped animals can be estimated as

$$\nu_G = \frac{EDC_G}{\lambda} = \frac{r^2_{DGV}}{1-r^2_{DGV}} - \frac{r^2_{EBV}}{1-r^2_{EBV}}$$

where $\lambda = (1-h^2)/h^2$ and h^2 is the heritability. Hence, according to simple selection index principles the accuracy of the added value due to DGV is

$$r_a = \sqrt{1 - \frac{1}{\nu_G + 1}}.$$

Step 3: Setup bivariate blending model by a single trait random regression AM-BLUP

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{K_1u_1} + \boldsymbol{K_2u_2} + \boldsymbol{e}$$

where the observations of "trait" and the DGVs are observations of the same dependent variable. In our case, the observations of "trait" are deregressed proofs (DRP) as in AM-BLUP and DGV are from GBLUP which have

different weights. Values in the design matrices $\boldsymbol{K}$ and used weights depend on the type of the observation. When observation is the same DRP as in AM-BLUP, the regression covariables are

$$[k_1 \quad k_2] = [1 \quad 0]$$

and the weight is the same as in AM-BLUP, i.e., EDC. When the observation is DGV from GBLUP, the regression covariables for animal *i* are

$$[k_1 \quad k_2] = \left[\sqrt{r_{a,i}^2} \quad \sqrt{1 - r_{a,i}^2}\right]$$

where $r_{a,i}^2$ is the increase in accuracy due to genomic information from Step 2, and the weight is a large value (1000). The variances are $Var(\boldsymbol{u}_i) = \sigma_u^2 \boldsymbol{A}$, $i = 1, 2$ where σ_u^2 is from AM-BLUP. After fitting the model, the solutions in $\widehat{\boldsymbol{u}}_1$ have GEBV.

Final bbGBLUP model reliabilities are estimated from this random regression AM-BLUP. This can be done using the inverse of the MME, or with approximation based on iteration of individual-sire-dam triplets.

Data

Study data were extracted from the production trait evaluation of Nordic Red dairy cattle. For simplicity DRP were assumed. Note that actual phenotypic data (DRP) or DGV were not used, only the EDCs and pedigree. We assumed heritability $h^2 = 0.50$. After edits, 38194 SNPs were used from the BovineSNP50 chip.

Group of 1055 genotyped bulls born 2001-2005 were used as training animals (Training bulls). Daughters (w. records) for the training bulls were searched and from them, 40 daughters were sampled for 522 “top” bulls, and 10 daughters for 533 “average” bulls, giving up to 26060 daughters. The “top” bulls were those having more than average number of daughters originally. Group of 1223 genotyped cows with records (Candidate cows) and group of 607 genotyped bulls (Candidate bulls), both born 2006-2011, were used as candidate animals.

Pedigree for all above animals was traced to two generations so that the total number of animals was limited to 73579 from which 67648 cows had records.

Results and Discussion

The three methods (AM-BLUP, ssGBLUP, and bbGBLUP) were implemented and model reliabilities were estimated for the three animal groups (Training bulls, Candidate cows, and Candidate bulls).

Figure 1 has scatter plot of the model reliabilities from AM-BLUP and ssGBLUP.

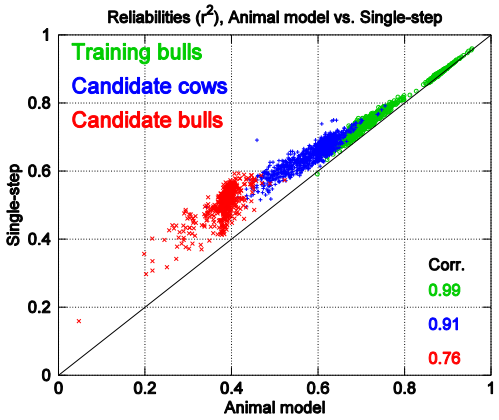

Figure 1. Model reliability correlation of AM-BLUP (x-axis) and ssGBLUP (y-axis).

In Figure 1 each mark represents values of ssGBLUP model reliabilities of individual animals plotted against AM-BLUP model reliabilities. If the reliabilities of the two approaches are the same, the mark lies on the diagonal of the image. Training bulls have green circles, Candidate cows blue pluses, and Candidate bulls red crosses.

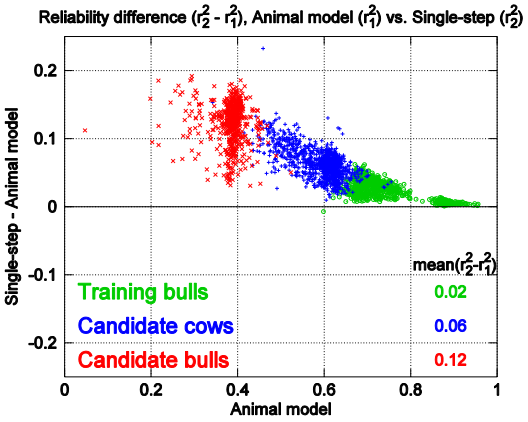

Figure 2. Model reliability differences between AM-BLUP and ssGBLUP.

Figure 2 shows differences in the model reliabilities by AM-BLUP and ssGBLUP. Now, position of each mark in the image is determined by the AM-BLUP model reliability $r^2_{AM,i}$ (x-axis) and y-axis displays the difference $r^2_{SS,i} - r^2_{AM,i}$ of the two methods for each animal.

The genomic information in ssGBLUP increased the model reliabilities of candidate bulls by 12 %-units. The correlation was high 0.76. The increase was less in producing cows, and in bulls with daughters. The correlation between reliability estimates, as expected, was higher for animals with more reliable AM-BLUP evaluations.

Next, AM-BLUP and bbGBLUP are compared in Figure 3.

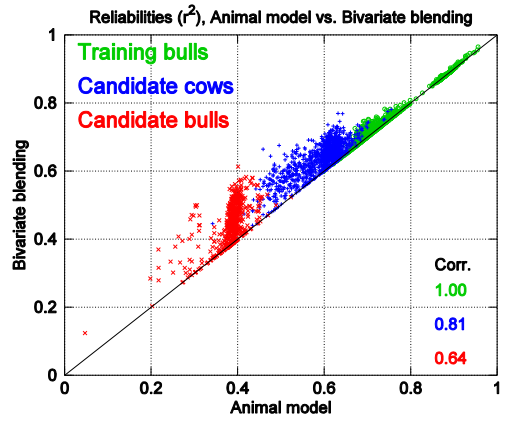

Figure 3. Model reliability correlation of AM-BLUP (x-axis) and bbGBLUP (y-axis).

AM-BLUP reliabilities are still on the x-axis but now y-axis had bbGBLUP reliabilities. As with the first comparison the bbGBLUP reliabilities are higher than those by AM-BLUP but now there are more individual reliabilities that have not changed so much.

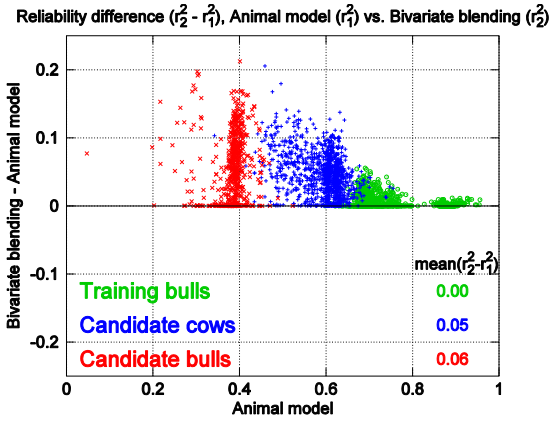

Figure 4. Model reliability differences between AM-BLUP and bbGBLUP.

This can be seen also from Figure 4 where the differences of the reliabilities of the two models have larger coverage in the image compared to Figure 2.

Finally, the reliabilities of ssGBLUP and bbGBLUP are in Figure 5.

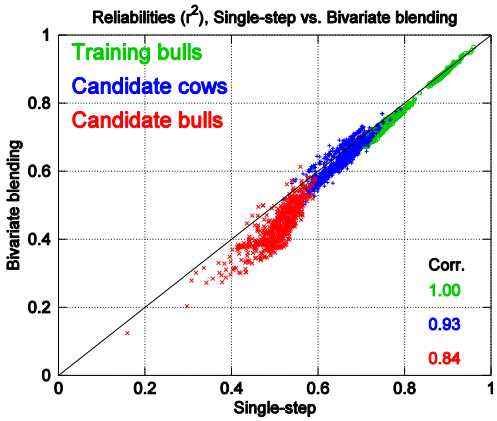

Figure 5. Model reliability correlation of ssGBLUP and bbGBLUP (y-axis).

The bbGBLUP gave almost the same reliabilities for the reference animals. For the candidate cows and bulls, the bbGBLUP gave slightly lower estimates, but the correlations were high for the cows (0.93) and relatively high for the bulls (0.84).

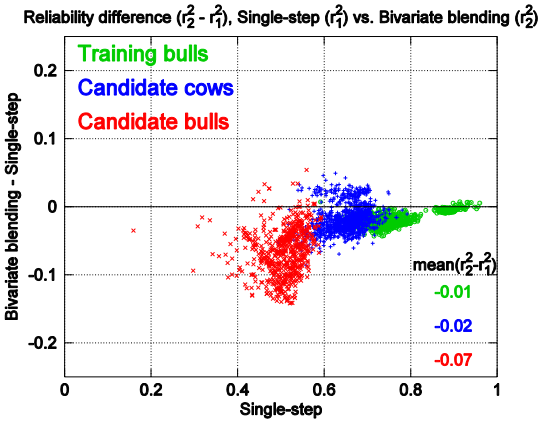

Figure 6. Model reliability difference between ssGBLUP and bbGBLUP (y-axis).

From Figure 6 it can be seen that most of the candidate animals had higher reliability value from the ssGBLUP, but there are also animals, especially some candidate cows, that have higher reliabilities from bbGBLUP.

The bbGBLUP method seemed to give slightly lower reliabilities than ssGBLUP. In the ssGBLUP, the information is used comprehensively from the data, pedigree and genotypes. In the bbGBLUP, the information from the genotypes might have been underestimated, because in Step 2 we did not use the same amount of total EDC in estimation of reliability of DGV, as was used in estimation of reliability of EBV in Step 1. The amount of observation information for genotyped animals is therefore larger in ssGBLUP. This could be corrected by absorbing information from all non-genotyped animals to genotyped animals in Step 2.

In the original version of bbGBLUP (Mäntysaari and Strandén 2010) the double counting of information was reduced by subtracting some EDC from the genotyped bulls included in the reference population. Here we use the relative increase in information due to genomic EDC_G/λ. For animal that already has high reliability due to relatives, the relative increase is smaller. This will remove the earlier double counting that existed when the relationships were implicitly modeled by $\boldsymbol{G}$-matrix and $\boldsymbol{A}$ -matrix.

The computing times for all the methods were relatively short because of the small data size. The size of MME was larger here in case of bbGBLUP, and, therefore, the inverse time was longer. However, in practice, the inverse of ssGBLUP MME is much more difficult to approximate than inverse of MME from a simple animal model without $\boldsymbol{G}$ matrix.

Conclusions

Model reliabilities of three models were compared with Nordic Red dairy cattle data. Model reliabilities by ssGBLUP were higher than by AM-BLUP due to genomic information. Similarly, bbGBLUP reliabilities were also higher than those by AM-BLUP. Compared to ssGBLUP, bbGBLUP reliability estimates were, in general, lower because added value due to the genotype information might have been underestimated.

Computationally bbGBLUP was lighter than ssGBLUP in reliability calculation due to better sparsity. Also, bbGBLUP can be implemented with standard software used for AM-BLUP.

Acknowledgements

The genotypes were provided by Nordic genomic selection project (VikingGenetics, Aarhus University, NAV Nordic Cattle Genetic Evaluation, FABA, Svensk Mjölk (Växa Sverige)). The data were provided by NAV and FABA.

References

Aguilar, I., Misztal, I., Johnson, D.L., Legarra, A., Tsuruta, S. & Lawlor, T.J. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of

Holstein final score. *J. Dairy Sci.* 93, 743-752.

Christensen, O.F. & Lund, M.S. 2010. Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol.* 42:2.

Misztal, I., Tsuruta, S., Aguilar, I., Legarra, A., VanRaden, P.M. & Lawlor, T.J. 2013. Methods to Approximate Reliabilities in Single-Step Genomic Evaluation. *J. Dairy Sci.* 96, 647–654.

Mäntysaari, E.A. & Strandén, I. 2010. Use of bivariate EBV-DGV model to combine genomic and conventional breeding value evaluations. *Proc. 9th World Congress on Genetics Applied to Livestock Production*, Leipzig, Germany.

Strandén, I. Lidauer, M., Mäntysaari, E.A. & Pösö, J. 2000. Calculation of Interbull Weighting Factors for the Finnish Test Day. *Interbull bulletin* 26, 78-79.

Strandén, I. & Garrick, D.J. 2009. Technical note: Derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit. *J. Dairy Sci.* 92, 2971-2975.

Tier, B. & Meyer, K. 2004. Approximating prediction error covariances among additive genetic effects within animals in multiple-trait and random regression models. *J. Anim. Breed. Genet.* 121, 77-89.