Comparison of Breeding Values from Single-Step and Bivariate Blending Methods

M. Taskinen[1], E.A. Mäntysaari[1], G.P. Aamand[2], and I. Strandén[1].
[1]MTT Agrifood Research Finland, [2]NAV Nordic Cattle Genetic Evaluation

ABSTRACT: Breeding values in genetic evaluation are compared between conventional animal model BLUP, single-step BLUP, and revised bivariate blending after genomic BLUP. The data are extracted from the production trait evaluation of Nordic Red dairy cattle. Genotyped bulls with daughters are used as training animals, and younger genotyped bulls and producing cows as candidate animals. Breeding values estimated with the bivariate blending method are shown to match the single-step values. Computationally bivariate blending method is, however, lighter than the single-step method and, thus, can be used also with larger population sizes.
Keywords: genomic evaluation; breeding values; single-step genomic evaluation

Introduction

The genomic breeding values can be calculated using mixed model equations (MME) with genomic relationship matrix ($\mathbf{G}$) or with many other approach such as BayesA or BayesB (Meuwissen et al. (2001)). The basic genomic BLUP (GBLUP) breeding values are easily computed if $\mathbf{G}^{-1}$ can be formed (Strandén and Garrick (2009)). Here matrix $\mathbf{G}$ has a size of number of genotyped animals.

Commonly genomic information is further combined with the conventional breeding values estimation. This can be achieved by single-step BLUP (ssGBLUP by Aguilar et al. (2010); Christensen and Lund (2010)), simple selection index (VanRaden (2008)) or by bivariate blending (Mäntysaari and Strandén (2010)). ssGBLUP combines the genomic and traditional information sources in a unified linear mixed effects model setup. Use of ssGBLUP can be prohibitive if the number of genotyped animals is large because the method requires two dense inverted matrices of size number of genotyped animals.

Nordic genomic evaluations use bivariate blending to combine direct genomic value (DGV) and traditionally estimated breeding values (EBV). Bivariate blending attempts to mix the two pre-computed information sources but at the same time avoiding of any possible double counting of information. Bivariate blending is based on a bivariate model having information from the two sources 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 $\sqrt{R_{DGV}^2}$ with the EBV-trait. One of the main advances of bivariate blending over ssGBLUP is that it avoids need to include two dense matri-

ces of size number of genotyped animals in the mixed model equations. In Taskinen et al. (2013) a revised bivariate blending method was introduced and used in model reliability comparison.

In this study our aim was to further apply the revised bivariate blending method to breeding value estimation and to compare the breeding values of three methods: animal model BLUP (AM-BLUP), single-step BLUP (ssGBLUP), and bivariate blending after GBLUP (bbGBLUP).

Materials and Methods

Estimation of breeding values. Consider the model

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

where $\mathbf{y}$ is vector of observations, $\mathbf{b}$ is vector of fixed effects, $\mathbf{X}$ is design matrix, $\mathbf{u}$ is vector of random effects, $\mathbf{Z}$ is incidence matrix, and $\mathbf{e}$ is random residual vector. Assume $\text{Var}(\mathbf{u}) = \mathbf{V}_u$ and $\text{Var}(\mathbf{e}) = \mathbf{R}$. The MME are

$$\begin{bmatrix} \mathbf{X'R^{-1}X} & \mathbf{X'R^{-1}Z} \\ \mathbf{Z'R^{-1}X} & \mathbf{Z'R^{-1}Z} + \mathbf{V_u^{-1}} \end{bmatrix} \begin{bmatrix} \mathbf{\hat{b}} \\ \mathbf{\hat{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'R^{-1}y} \\ \mathbf{Z'R^{-1}y} \end{bmatrix}. \quad [1]$$

In AM-BLUP

$$\mathbf{V}_u^{-1} = \frac{1}{\sigma_u^2} \mathbf{A}^{-1} \quad [2a]$$

and in ssGBLUP

$$\mathbf{V}_u^{-1} = \frac{1}{\sigma_u^2} \left[\mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - (\mathbf{A_{22}})^{-1} \end{bmatrix} \right] \quad [2b]$$

where $\mathbf{A}$ is the pedigree based relationship matrix, $\mathbf{G}$ is genomic relationship matrix, $\mathbf{A_{22}}$ contains pedigree based relationships of genotyped animals, and σ_u^2 is the genetic variance. Let the residual covariance matrix $\mathbf{R}$ be diagonal with j^{th} diagonal element σ_e^2/w_j where σ_e^2 is the residual variance, and w_j is weight for the observation j.

Genomic evaluations. In the standard methodology for DGVs (VanRaden (2008)) the MME includes only animals with genotypes and, following the notation in [1] and [2b], the matrices $\mathbf{A}$ and $\mathbf{A_{22}}$ are the same. The right hand side in [1] are either deregressed measures of produc-

tion or summarized phenotypic records, e.g.. daughter yield deviations. While the weights w_j in original EBV or ssGBLUP estimation relate to observations, in MME for DGV, the weights attempt to account the total information included in phenotypes. Alternative form for ssGBLUP can be obtained if equations of all non-genotyped animals in [1] are absorbed into equations of genotyped animals. After the absorb $\mathbf{Z'R^{-1}Z}$ and $\mathbf{Z'R^{-1}y}$ terms are written as $\mathbf{Z'MZ}$ and $\mathbf{Z'My}$, with matrix $\mathbf{M}$ as described in text books (e.g. Searle (1966)). In bbGBLUP we approximate the matrix $\mathbf{M}$ with a reduced diagonal matrix $\mathbf{R_{bb}}$ having the information in weights w_{bb} which gives exactly the same $\mathbf{u}$ as originally obtained by combining [1] and [2a] for genotyped animals. Now, if the same weights w_{bb} are used for standard DGV estimation, the resulting DGV should approximate the ssGBLUP well.

Our approach to derive the w_{bb} follows closely the methodology to approximate the EBV reliabilities from animal model. There the principle is to find arbitrary diagonal $\mathbf{R}$' such that $(\mathbf{R'^{-1}} - \frac{1}{\sigma_u^2}\mathbf{A^{-1}})^{-1}$ has the same diagonals as the inverse of original MME [1]. The reversed reliability approximation is based on the original EBV reliabilities of genotyped animals. By iteratively applying any method (e.g. Tier and Meyer (2004)) for animal model reliability approximation, the w_{bb} are solved such that they retain the original reliabilities. For this study we used the method of Harris and Johnson (1998).

The w_{bb} are considered as effective record contributions (ERC) and can thereafter be used in deregression of the EBVs of the genotyped animals. This yields into an approximation of $\mathbf{Z'My}$ of MME of the absorbed ssGBLUP for the genotyped animals.

Bivariate Blending. The revised bivariate blending method is performed in five steps.

Step 1: Get reliabilities r_{EBV}^2 from AM-BLUP and calculate ERCs for genotyped animals from the reliabilities and the pedigree using the procedure described above.

Step 2: Calculate deregressed proofs of the AM-BLUP EBVs for genotyped animals with AM-BLUP EBVs as observations and ERCs of Step 1 as weights.

Step 3: Calculate DGV and reliabilities r_{DGV}^2 from genomic evaluation with DRPs of Step 2 as observations and ERCs of Step 1 as weights.

Step 4: Calculate reliability increase due to genotype information. The relative increase in model accuracy from AM-BLUP due to GBLUP for genotyped animals can be estimated as

$$v_G = \frac{ERC_G}{\lambda} = \frac{r_{DGV}^2}{1 - r_{DGV}^2} - \frac{r_{EBV}^2}{1 - r_{EBV}^2}$$

where possible negative values are replaced by zeros. Hence, according to simple selection index principles the accuracy of the added value due to DGV is

$$r_a = \sqrt{\frac{v_G}{v_G + 1}}.$$

Step 5: Setup bivariate blending model by a single trait random regression AM-BLUP

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

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 the DGV are from GBLUP which were calculated with approximate bbGBLUP weights w_{bb}. Values in the design matrices $\mathbf{K}$ and used weights depend on the type of the observation. When the observation is the original 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., ERC. 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 squared accuracy due to genomic information from Step 4, and the weight is a large value (e.g. 1,000). The variances are $Var(\mathbf{u_i}) = \sigma_u^2\mathbf{A}$, $i = 1, 2$ where σ_u^2 is from AM-BLUP. After solving $\mathbf{u_1}$ and $\mathbf{u_2}$ bivariate blending DGVs (DGV_{bb}) are calculated as linear combination

$$DGV_{bb} = r_a\,\mathbf{u_1} + \sqrt{1 - r_a^2}\,\mathbf{u_2}$$

where r_a values of non-genotyped animals are set to large value, for example, 0.95.

Data. Study data were extracted from the production trait evaluation of Nordic Red dairy cattle. We assumed heritability $h^2 = 0.50$. After edits, 38,194 SNPs were used from the BovineSNP50 chip.

Group of 1,055 genotyped bulls born 2001-2005 were used as training animals (Training bulls). Daughters (with 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 1,223 genotyped cows with records (Candidate cows) and group of 607 genotyped bulls (Candidate bulls), both born 2006-2011, were used as candidate animals.

Results and Discussion

Breeding value comparison. The three methods (AM-BLUP, ssGBLUP, and bbGBLUP) were implemented and breeding values were estimated for the three animal groups (Training bulls, Candidate cows, and Candidate bulls).

In Figure 1 the correlations of the estimated breeding values of AM-BLUP and ssGBLUP are visualized as a scatter plot.
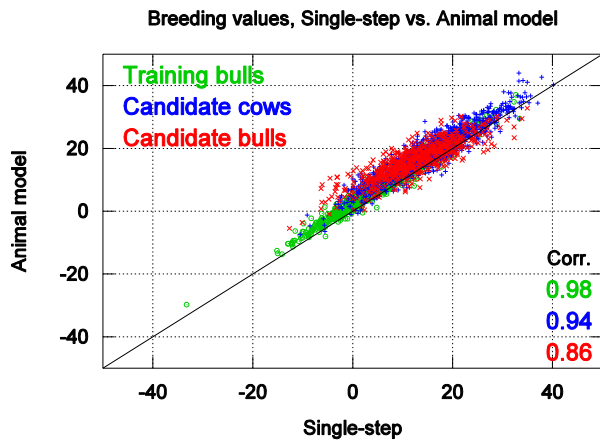

Figure 1: Animal model EBVs (AM-BLUP) plotted against single-step genomic evaluations (ssGBLUP)

Each mark in the figure represents individual EBV from ssGBLUP plotted against AM-BLUP. Training bulls are marked with green circles, Candidate cows blue pluses, and Candidate bulls red crosses. The figure illustrates how breeding value estimates of the training bulls are on average lower than the EBVs or GEBVs of the candidate bulls. The correlation between the EBVs and GEBVs is highest on training bulls and lowest in cows, which as plotted here, have both the genotypes and own records.

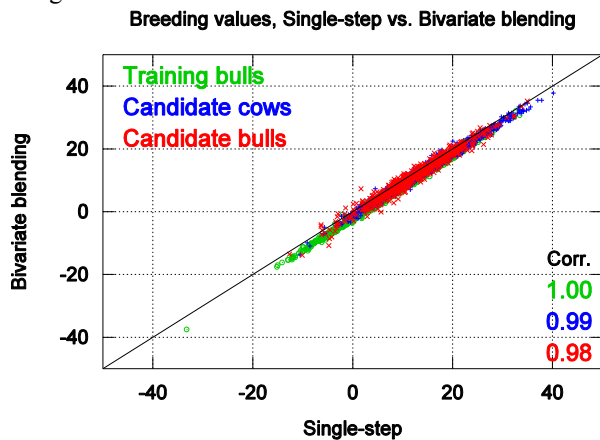Next, the ssGBLUP and bbGBLUP are compared in Figure 2.

Figure 2: The bivariate blending GEBVs (bbGBLUP) displayed against ssGBLUP GEBVs

Correlation between the two EBV is now high in training bulls and candidate bulls. This is also expected because the bivariate blending is done using DGV estimates based on DRPs derived using ERCs from the reversed reliability approximation. Thus, the DGV are supposed to resemble closely the ssGBLUP results. On average the bbGBLUP were lower than the ssGBLUP. Differences were larger in training bulls. However, for the use of evaluations this has no effect.

In our study we estimated the DGVs to be used in the bivariate blending using genomic evaluations with $\mathbf{G}^{-1}$ matrix. This, however, is not possible if the population of genotyped animals is much larger than the number of genetic markers genotyped. The DGVs can be well estimated with SNP based models (Strandén and Garrick (2009)), or even Bayesian variable selection models. The only prerequisite is the system to compute the reliabilities of DGV. In a case of SNP models, some sensitivity analyses are recommended to test how critical the accuracy of the reliability estimates is.

Conclusion

Breeding values of three models were compared with Nordic Red dairy cattle data. The estimated breeding values of the revised bivariate blending method matched the single-step breeding values relatively well. Thus, computationally bbGBLUP was lighter than ssGBLUP in breeding value estimation due to better sparsity. Furthermore, bbGBLUP can be implemented with standard software used for conventional AM-BLUP.

Literature Cited

Aguilar, I., Misztal, I., Johnson, D. L. et al. (2010). J. Dairy Sci. 93:743-752.

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

Harris, B. and Johnson, D. (1998). J. Dairy Sci. 81:2723-2728.

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Genetics 157:1819–1829.

Mäntysaari, E. A. and Strandén, I. (2010). Proc. 9th WCGALP.

Searle, S. R. (1966). Matrix Algebra for the Biological Sciences.

Strandén, I. and Garrick, D. J. (2009). J. Dairy Sci. 92:2971-2975.

Taskinen, M., Mäntysaari, E. A. Lidauer, M., et al. (2013). Interbull Bulletin 47:246-251.

Tier, B. and Meyer, K. (2004). J. Anim. Breed. Genet. 121:77-89.

VanRaden, P. M. (2008). J. Dairy Sci. 91:4414-4423.