Understanding Genomic Evaluations from Various Evaluation Methods and GMACE

R. Mrode[1,] M. Coffey[1] and D.P. Berry[2]

[1]*Scottish Agricultural College, Sir Stephen Watson Building, Bush, Penicuik, EH26 0PH.* [2]*Teagasc, Moorepark Dairy Production, Co. Cork, Ireland*

Abstract

In an attempt to understand evaluations based on single nucleotide polymorphisms (SNPs), equations for SNP effects expressed in terms of SNP deviations were examined for BLUP and Bayesian methods. The SNP deviations from Bayesian methods were higher and more varied compared to BLUP. It seems that Bayesian methods were able to more differentially and appropriately weight computed SNP deviations by estimating different variances for each SNP. The influence of parent average (PA) on GMACE evaluations for young bulls in the importing country when a bull has only a genomic breeding value (GEBV) in the exporting country was also examined. The GMACE proof of such young bulls could be reduced by up to 63% in the importing country when the PA is low.

Introduction

The use of simplified equations showing the contributions of various sources of information to the evaluations of animals are important to the understanding of evaluations. VanRaden and Wiggans (1991) first presented such equations for animal model evaluations. These equations have been extended to multi-trait across country evaluations (Mrode and Swanson, 1999) and random regression models (Mrode and Swanson, 2004). With the rapidly increasing advances in genomic evaluations using SNPs, this paper examined SNP deviations (marker deviations, VanRaden, 2008) under various models using real data and the associated weights. Secondly the influence of parent average on the GMACE proofs of young bulls in the importing country when such bulls have only genomic breeding values (GEBVs) and no daughters in the exporting country was examined assuming varying genetic correlations between countries.

Materials and Methods

Genotypic and phenotypic data editing

The genotypic data used for study consisted of 54,001 (SNPs) from the Illumina Bovine50 Beadchip on 755 Holstein-Friesian AI sires with daughters in Ireland. After data edits, 42287 SNPs were available on each bull. The study is part of the on going validation exercise of software on http://www.genomicselection.net. Bull daughter yield deviations (DYDs) for 305-day milk yield and the associated reliabilities, evaluated by the Irish Cattle Breeding Federation in the January 2009 domestic genetic evaluations were available. Parental contribution to the reliability of each DYD was removed using the approach of Harris & Johnson (1998) for PTAs and this adjusted reliability will be used as weights in the next stage of this study. SNPs were coded as 0 (homozygote), 1 (heterozygote) and 2 (homozygote).

Genomic evaluations

Using a linear model, the equation for SNP effects (VanRaden, 2008) is

$$\hat{\mathbf{u}} = (\mathbf{Z'R^{-1}Z} + \mathbf{I}\alpha)^{-1} \mathbf{Z'R^{-1}(y- X\beta)} \qquad (1)$$

For the jth SNP, this can be expressed as

$$\hat{u}_j = (z_j'r^{-1}z_j + \alpha)^{-1} z_j'r^{-1} z_j (yd_j)$$

$$\hat{u}_j = wt_j (yd_j)$$

where yd_j is the SNP deviation for the jth SNP, that is, data information for that SNP

corrected for all effects apart from the SNP and the SNP deviation can be defined as $yd_j = z_j\text{'}r^{-1}(y - z_i\hat{u}_i - x\beta)$, $i \neq j$ and $wt_j = (z_j\text{'}r^{-1}z_j + \alpha)^{-1} z_j\text{'}r^{-1} z_j$. The direct genomic breeding value ($\hat{a}_i$) of animal i therefore is

$$\hat{a}_i = \sum_j z_j wt_j (yd_j)$$

In the case of Bayesian methods, there is an additional component as a result of sampling from the conditional posterior distribution of **u**, such that

$$\hat{u}_j = wt_j (yd_j) + N(\hat{u}_j, (z_j\text{'}r^{-1}z_j + \alpha_j)^{-1} \sigma^2 e) \quad (2)$$

The second term on the right hand side tends towards zero averaged over all samples after the burn in period. The estimates for the second term in this study averaged from all samples after the burn in period were $10^{-0.5}$ and 0.0008 for BayesA and BayesB respectively. The proportion of SNP effects set to zero in BayesB was 0.34. To gain an understanding of SNP estimates from the various models SNP deviations and associated weights were computed for BayesA, BayesB and BLUP and compared. For the Bayesian methods, the MCMC chains were run for 80000 cycles, the first 24000 of which were discarded as the burn in period. In the case of BayesB, 20 Metropolis-Hastings cycles were undertaken within each MCMC chain. These analyses were carried out using software at the http://www.genomicselection.net. The genetic variance used for BLUP was the sire genetic variance for 305-day milk yield divided by the number of SNPs and the residual variance was that estimated from BayesA. One of the major draw backs of Bayesian methods is the sometimes enormous computing time. However, the genetic variances and residual variance estimated using Bayesians methods could be used in BLUP. The efficiency of using such genetic parameters in BLUP evaluations was also examined.

GMACE

Secondly, the prediction of young bulls which have only GEBV and no daughters in the exporting country was examined in an importing country using a small data set consisting of 16 animals. The GMACE evaluations in the importing country for these bulls is a function of only two components, parent average and Mendelian sampling, given the genetic parameters, so results will still be relevant in spite of the small data set. Just as in MACE, the GMACE proof of a young bull with no progeny can be obtained as

$$\hat{a}_{bull} = W_1 PA + W_2 CD \quad (3)$$

where **CD** is the de-regressed proof corrected for country effects, W_1 and W_2 are the product of $(Z'R^{-1}Z + G^{-1}\alpha_{bull})^{-1}$ and $2G^{-1}\alpha_{par}$ and $Z'R^{-1}Z$ respectively and $\alpha_{bull} = 2\alpha_{par}$ with $\alpha_{par} = 8/11, 8/15, 2/3$ or $1/2$ if both sire and maternal grandsire (mgs), only mgs, only sire or no parents are known respectively. When a bull has a proof only in country i and not in j, its GMACE proof in country j, from equation [3] reduces to:

$$\hat{a}_j = PA_j - (g_{ij}/g_{ii})(\hat{a}_i - PA_i),$$

where g_{ij} and g_{ii} are elements of **G**

The influence of genetic correlations was examined on two bulls with GEBVs in the exporting country but with parent averages in the importing country about one standard deviation below and half standard deviation above the country mean respectively. Three genetic correlations between countries were investigated: 0.95, 0.85 and 0.65. Heritability was assumed to be 0.36 in country 1 (importing country) and 0.30 in the exporting country. The bulls with GEBVs were assumed to have a genomic reliability of 0.67 in the exporting country and the corresponding daughter equivalents were computed. The percentage of common genotyped (c12) bulls in both countries was assumed to be 80 percent.

Also using equation (3) the relative weights on PA and CD were investigated in MACE and GMACE for a bull with daughters in both countries with c12 assumed to be 0.3, 0.7 or 0.95 and the genetic correlation assumed to 0.95.

Results and Discussion

Genomic evaluations

The summary statistics of estimates of SNP effects and SNP deviations and associated weights for the BLUP, BayesA and BayesB are presented in Table 1. The estimates of genetic variances for the SNPs were higher and more variable for BayesB compared with BayesA probably because BayesB sets 34% of the SNP effects to zero. Consequently, the SNP deviations and associated weights were higher for BayesB compared with the other two methods; the absolute mean of the SNP deviations and associated weights were lowest for BLUP. It seems that Bayesians methods were able to more differentially and appropriately weight computed SNP deviations by estimating different variances for each SNP. However weights on SNP deviations were lowest with BLUP and given that alpha term is constant, this weight seems to be more dependent on allele frequencies. The correlation between the weights and allele frequencies was 0.99 for BLUP, 0.40 for BayesA and -0.05 for BayesB. However, the alpha term in BayesB had a strong positive correlation (0.78) with SNP deviations but this correlation was zero for BayesA. The correlation between SNP deviations and SNP effects was 0.37 for BayesA, 0.33 for BayesB but was 0.99 for BLUP. Interestingly, the correlation between the SNP genetic variances and SNP effects was -0.96 for BayesA and -0.91 for BayesB, indicating that SNP effects tended to decrease as the genetic variances of the SNPs increased

The simple correlations of SNP effects from BayesA and BayesB with those from BLUP using BayesA and BayesB parameters were 0.99 and 0.91 respectively. Corresponding rank correlations were 0.99 and 0.96. The regressions of SNP effects from BayesA and BayesB on BLUP were 0.790 and 0.413 respectively. Thus the predictive ability of BLUP using parameters from BayesB was rather poor. The correlation of BLUP SNP effects with those from BayesA and BayesB was 0.30.

GMACE

The percentage change in the GMACE evaluations of the two bulls in the importing country at the different genetic correlations was computed relative to GMACE evaluations for these bulls assuming a genetic correlation of unity between both countries. For bull A with a parent average of about one standard deviation below country mean, the decrease in the GMACE proof in the importing country were 3, 15 and 63% when the genetic correlation was 0.95, 0.85 and 0.65 respectively. However for bull B with half standard deviation above the country mean, the GMACE proof increased by 6% when the genetic correlation was either 0.95 or 0.85 but decreased by 2% when the correlation was 0.65. The increase in GMACE evaluations was due to the fact that the contributions from the high parent average increased as the genetic correlation decreased from unity while the contributions from Mendelian sampling decreased. It seems the PA could have some large influence in the GMACE of foreign bulls in importing country for bulls with only GEBV in exporting country.

In general, GMACE resulted in a decrease on the relative weights on PA as the proportion of common genotyped bulls in common was above 30% (Table 2) and it seems to be more so for the country with the higher heritability.

Conclusion

SNP deviations and weights were higher with the Bayesian methods compared with BLUP. It seems that Bayesian methods were able to more differentially and appropriately weight computed SNP deviations by estimating different variances for the SNPs. The correlation between SNP effects from BayesA and BLUP using BayesA parameters was close to unity. The parent average in the importing country could have a substantial effect (3 to 63%) on the GMACE evaluation of bulls in the importing country when bulls have only GEBVs in the exporting country. GMACE resulted in less emphasis on parent average relative to MACE when proportion of genotyped bulls in common was above 30%.

References

Harris, B. & Johnson, D. 1989. Information source reliability method applied to MACE *Interbull Bulletin 17,* 31-36.

Mrode, R.A. & Swanson, G.J.T. 1999. Simplified equations for evaluations of bulls in the Interbull international evaluation system. *Livest. Prod. Sci. 62,* 43-52.

Mrode, R.A. & Swanson, G.J.T. 2004. Calculating cow and daughter yield deviations and partitioning of genetic evaluations under a random regression model. *Livest. Prod. Sci. 86,* 253-260.

VanRaden, P.M. & Wiggans, G.R. 1991. Derivation, calculation and use of national animal model information. *J. Dairy Sci. 74,* 2737-2746.

VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci. 91,* 4414-4423.

Table 1. Mean (standard deviations in brackets), minimum (Min) and maximum (Max) of SNP effects (kg, SNP), genetic variances (kg^2, Gvar), SNP deviations (kg, yd) and weights (WTs) for various models.

	BayesA			BayesB			BLUP		
	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max
SNP	-0.006 (0.29)	-56.6	2.62	-0.019 (1.18)	-204	11.2	0.007 0.10)	-0.688	0.517
Gvar	0.695 (5.32)	0.629	1092.5	1.105 (7.44)	0.882	1175.8	1.333		
SYD	20.53 (1686)	-38045	11076	-359.52 (5415)	-171596	36170	92.2 (1240)	-7818	6309
WTs	0.515 (0.022)	0.478	0.987	0.876 (0.017)	0.837	0.999	0.108 (0.061)	0.0004	0.219

Table 2. Relative weight on parent average (PA) and de-regressed proof corrected for country effects (CD) in MACE and GMACE at different proportions of genotyped bulls in common (c12).

Country		MACE	GMACE		
			c12		
			0.30	0.70	0.95
1	PA	0.38	0.35	0.25	0.08
	CD	0.62	0.65	0.75	0.92
2	PA	0.46	0.45	0.40	0.31
	CD	0.53	0.54	0.60	0.69