GMACE without Variance Estimation

P. G. Sullivan[1] and J. H. Jakobsen[2]
[1]Canadian Dairy Network, Guelph, ON, Canada
[2]Interbull Centre, Box 7023, SE-750 07 Uppsala, Sweden

Abstract

Genomic variances have been estimated and used in GMACE since 2011, to adjust for differences among countries in the scaling of young bull genomic evaluations relative to progeny-tested bulls. Interbull has implemented validation tests for national genomic evaluations, which countries must pass in order to participate in GMACE, and the sharing of data and knowledge among countries for genomic evaluations has also increased. Each of these factors can improve consistency of genomic results among countries, and may reduce the need for genomic variance adjustments in GMACE. Cross-validation tests have been used previously to compare GMACE results when using versus not using genomic variance adjustments, and have shown clear advantages for including genomic variance adjustments. When repeated on current data for the present study, however, the cross-validation results no longer showed this clear advantage. Genomic variance adjustments were helpful for some traits and countries but not for others. On balance across all traits and countries, there was no longer a clear advantage either way. The international sharing of data and knowledge, combined with genomic validation tests of Interbull are likely helping to reduce differences among countries in the relative scaling of genomic versus progeny-test evaluations within the same country.

Key words: genomics, international evaluation, GMACE, variance estimation, cross-validation

Introduction

Prior to this study, genomic variances had been estimated and used in all applications of GMACE for young bulls (Sullivan and Jakobsen, 2012). The main purpose was to limit potential problems due to differences between countries, in the relative variance of young genomic bull evaluations versus progeny-tested bull proofs. Many countries have observed higher than expected variances for the young genomic bulls, and different approaches have been used to address the situation at the national level. In some cases, different adjustments are used by trait within the same country. A genomic validation test was developed to identify and prevent use of the more problematic national genomic data in GMACE. This test has limited sensitivity, however, it is based on genomic predictions that exclude most recent data, and although applied to most it is not required for all traits evaluated in GMACE. Cross-validation tests for GMACE have previously shown better results when including a genomic variance estimation step (Sullivan *et al*, 2011; Sullivan and Jakobsen, 2012).

As the sharing of both knowledge and data among countries continues to grow, it can be expected that national genomic evaluation results will become more consistent among countries, and that the need for genomic variance estimation in GMACE may decrease. The purpose of the present study was therefore to re-apply the cross-validation tests and assess current benefits of including a genomic variance estimation step in GMACE.

Material and Methods

Eleven countries participated with young bull GEBV data for as many as 37 different traits for the GMACE implementation run in December 2013. These countries were Australia (AUS), Canada (CAN), Switzerland Red Holstein (CHR), Germany (DEU), Denmark-Finland-Sweden (DFS), France (FRA), Great Britain (GBR), Italy (ITA), the Netherlands (NLD), Poland (POL), and the United States (USA).

The 3 applications of GMACE using different inputs of genomic reliability (G, GP.5

and GP in Sullivan and Jakobsen, 2014) were repeated for the present study, but in each case without estimating genomic variances. These applications without genomic variances are denoted as M, MP.5 and MP, where the letter M rather than G refers to the use of MACE variances rather than estimated genomic variances within GMACE.

Eliminating the adjustments for genomic variances significantly reduced the impacts of changes in GMACE results with different input reliabilities, because most of these impacts were due to changes in genomic variance estimates. The remaining impacts were limited to only those bulls with national GEBV from more than one country, and through relatively small changes in the relative weighting of those national GEBV for the bull. These remaining impacts were minimal, and cross-validation results were nearly identical for M, MP.5 and MP. Thus we present results here for only MP.5, in comparison with the GP.5 and G results from Sullivan and Jakobsen (2014). These models are summarized in Table 1. Models GP.5 and MP.5 used the partially-regressed national reliabilities as input, and are otherwise the same GMACE models as were denoted rGM_ms(v) and GM_ms, respectively, in Sullivan & Jakobsen (2012).

Table 1. GMACE models of evaluation.

Model	Description
G	Estimate genomic variances and use provided reliabilities as input
GP.5	Estimate genomic variances and use a .5-regressed reliability as input (5*predicted + .5*provided)
MP.5	Use traditional MACE variances and a .5-regressed reliability as input (5*predicted + .5*provided)

Cross-validation tests were conducted by country. Implied assumptions were that GMACE evaluations based on input data from only foreign scales were unbiased predictors of the excluded (i.e. cross-validated) national GEBVs for the given country, and that the national GEBV were also unbiased. Correlations and regressions of national GEBV on GMACE predictions from foreign data were used as test statistics, and these were generally very similar for the three models. The main differences observed were in the relative scaling of evaluations among the different countries. To demonstrate these differences, we focused our model comparisons on estimates of Top Bull bias; the relative difference between a GMACE prediction of +3 standard deviations ($\boldsymbol{z} = \bar{X} + 3\sigma$) and the expected value for $\boldsymbol{z}$ based on cross-validation regression of national GEBV on GMACE: $\boldsymbol{w} = \hat{a} + \hat{b}x$, at the value $x = \boldsymbol{z}$. We define, as in Sullivan and Jakobsen (2012):

$$\text{TopBias} = 100\% * (\boldsymbol{z} - \boldsymbol{w}) / \boldsymbol{w}$$

Regression equations estimated with different levels of truncation selection on the x-variable (GMACE predictions) were similar as expected, since regression estimates are not biased by selection on x. Therefore, to focus on key bulls of interest we used the regression equation for bulls with a GMACE prediction above +1σ to estimate TopBias (of a +3σ bull).

Results and Discussion

Model GP.5 or MP.5 are both reasonable alternatives to model G. The cross-validation results summarized across all countries were similar for all three models (Table 2). Model MP.5 has the advantage that it removes the requirement to estimate genomic variances, while GP.5 has the advantage of a robust component to limit the impacts of potential problems with national input data.

Correlations between national GEBV and GMACE predictions from only foreign GEBV were generally consistent with the underlying genetic correlations for the given trait (shown for model MP.5 in Table 3). For example, the relatively lower GEBV correlations for direct stillbirth and fertility correspond with relatively lower genetic correlations among countries for these same traits. The GEBV correlations are also lower if GEBV for GMACE are only available from one foreign country and if the foreign country does not exchange genotypes with the national country of interest. For example, correlations are lower for A01 and B01 than for A10 and B10, and are higher when there are 2 or more foreign GEBV available for GMACE (A20, B20, A11, etc).

Results from MP.5 and GP.5 were distributed to the member countries, for review against G. Model MP.5 was ultimately chosen for the ongoing application of GMACE at this stage. The addition of robust steps, such as genomic variance estimation, have some appeal and further research has been recommended in this area. In particular, the potential differences between parent averages at the national and international levels are an area of key interest, as these differences can impact both genomic variance estimates and GMACE results, with or without inclusion of a genomic variance estimation step.

Sullivan, P.G., Zunbach, B., Durr, J.W. & Jakobsen, J.H. 2011. International genomic evaluations for young bulls. *Interbull Bulletin 44*, 87-94.

Sullivan, P.G. & Jakobsen, J.H. 2012. Robust GMACE for young bulls – methodology. *Interbull Bulletin 45*, 3-7.

Sullivan, P.G. & Jakobsen, J.H. 2014. GMACE pilot #4: Adjusting the national reliability input data. *Interbull Bulletin 48*, 40-45.

Acknowledgements

Discussion and suggestions from the GMACE working group were very much appreciated.

References

Table 2. Bias (as a % of local GEBV) in the GMACE prediction of a top (+3 standard deviation) bull, for selected countries and 95% confidence ranges (L95 to U95) across all countries[y].

Country (n-Protein)	Model	Protein		Stature		Somatic Cell		Direct Stillbirth		Fertility CC1		Mastitis	
		gSD[z]	Bias	gSD	Bias	gSD	Bias	gSD	Bias	gSD	Bias	gSD	Bias
CAN (35708)	G	84	-3	83	-3	81	-2	101	39	85	1	81	-1
	GP.5	85	-3	83	-3	82	-1	109	43	90	1	83	0
	MP.5	100	-1	100	3	100	8	100	37	100	2	100	9
USA (38756)	G	93	1	97	-3	96	1					96	1
	GP.5	93	1	97	-3	94	0					95	0
	MP.5	100	2	100	-5	100	0					100	-1
NLD (5957)	G	107	6	117	8	103	7	277*	24	105	46	87	21
	GP.5	103	5	114	8	103	7	78*	19	80	33	82	18
	MP.5	100	7	100	6	100	11	100	17	100	52	100	40
FRA (13579)	G	93	3	104	7	102	15			96	33	109	13
	GP.5	93	3	101	7	100	14			101	36	103	9
	MP.5	100	8	100	11	100	23			100	37	100	15
DEU (14430)	G	118	8	118	6	121*	7	94	-12	130*	8	121*	0
	GP.5	116	7	114	5	122*	7	86	-13	109	1	123*	2
	MP.5	100	3	100	2	100	-1	100	-9	100	-4	100	-6
DFS (854)	G	121*	11	121*	11	127*	12	108	40	149*	11	130*	24
	GP.5	116	10	120	12	124*	13	110	44	158*	9	120	27
	MP.5	100	5	100	8	100	4	100	33	100	4	100	17
All Countries		L95	U95	L95	U95	L95	U95	L95	U95	L95	U95	L95	U95
Confidence Range of TopBias	G	-18	31	-4	14	-6	21	-27	71	-7	50	-11	31
	GP.5	-18	32	-5	15	-5	21	-33	75	-11	50	-10	29
	MP.5	-16	25	-6	12	-8	22	-25	61	-12	54	-19	42

[z]gSD is the genomic estimate of SD as a percentage of the genetic SD used for MACE.

[y]one country was excluded for cc1 due to an extreme bias > 250% for all models.

*gSD truncated in these cases to the edges of the range [80-120].

Table 3. Correlations (*100) between national GEBV and MP.5 GMACE using only foreign GEBV.

Data[z]	Protein	Stature	Somatic Cell	Direct Stillbirth	Fertility CC1	Mastitis
A 0 1	80	87	78	69	78	77
B 0 1	78	86	77	74	83	77
A 1 0	93	95	92		86	92
B 1 0	86	85	82	85	77	78
A 0 2	83	90	81	67	83	81
B 0 2	85	87	83		81	83
A 2 0	94	96	94			94
B 2 0	89	91	90		81	84
A 1 1	93	94	91		81	92
B 1 1	84	88	82	76	82	81
A 2 2	91	95	91			91
B 2 2	91	89	90		80	84
A 3 2	95	95	92			93
A 3 3	95	94	92			93

[z]A (CAN, GBR, ITA, USA) and B (DEU, DFS, FRA, NLD) are the 2 main consortia that share data for national genomic evaluations. The numbers for "Data" that follow A|B are the number of foreign countries, with a GEBV included in GMACE for the bull, from the same and different consortia, respectively. For example, A 3 1 is the group of bulls with a GEBV from all 4 countries in consortia A and 1 country in consortia B.