GMACE Variance Estimation

P.G. Sullivan
Canadian Diary Network, Guelph, ON, Canada

Abstract

Estimation of genomic variances for GMACE is similar to the estimation of genetic variances for MACE, and is based on REML equations that include a prediction error variance term. For GMACE, the prediction error variances must be approximated, and this approximation was improved in the present study to better reflect statistical covariance between the national GEBV of young bull and MACE EBV of parents. In previous estimation, some parental information was being excluded, allowing a simple approximation to work well, but the simple approximation did not work well after including all parental information. Use of any approximation can bias genomic variance estimates, and such bias would adversely affect conversions of genomic information from the evaluation scales of GEBV to non-GEBV countries. This bias was minimized by scaling, to eliminate across-country average difference between estimated variance from GMACE relative to MACE. This "MACE-neutral" scaling of genomic variance estimates does not affect bull comparisons among GEBV countries, because it does not alter relative genomic variances. However, it should improve comparisons between GEBV and non-GEBV countries. The new estimates of genomic variances were very similar to previous estimates from the same data. Some individual estimates changed, but the rankings of countries from high to low variance were nearly identical as before. As such, all GMACE bull rankings were nearly identical to previous rankings, with correlations higher than 0.997 for all 5 traits studied and all country scales. Standard deviations of GMACE predicted breeding values were also very similar, within 1% of the previous in almost all cases.

Key words: genomics, international evaluation, GMACE, robust, variance estimation

Introduction

An important step in the application of GMACE for young bulls is the estimation of genomic variances. This variance estimation has thus far relied on somewhat crudely estimated matrices of prediction error covariances (Sullivan and Jakobsen, 2012). The purposes of this study were to improve these prediction error covariance estimates, to test the impacts on genomic variance estimates and on robust GMACE predictions of young bull genetic merit.

Data

The data and edits for the present study are described in detail by Jakobsen and Sullivan (2012). In summary, there were five traits included: protein (pro), stature (sta), somatic cells (scs), direct longevity (dlo) and female fertility (cow conception trait #1; cc1). August 2011 national GEBV data from eleven populations (CAN, DEU, DFS, FRA, NLD, POL, USA, CHE, CHR, ITA and JPN), and EBV data from all countries participating in the August 2011 MACE service of Interbull were used for the present study. The total numbers of national GEBV on young genotyped bulls without daughter data, across all populations, were: 57902 for pro, 47285 for sta, 53820 for scs, 54663 for dlo, and 44395 for cc1.

Methods

Genomic variance estimates are based on Mendelian Sampling (MS) of young bulls that are not yet progeny-tested, and using predictions of MS ($\hat{M}$) defined as the difference between a bull's national GEBV and his parent-average based on MACE. The predictions and corresponding prediction error

variances ($V(\hat{M})$) are combined via REML (Sullivan, 1999) to estimate the required genomic variances for Robust GMACE model rGM_ms(v). An approximation is required for $V(\hat{M})$. The variable $\hat{M}$ is a linear function of the evaluations of a young GEBV bull and his parents (k'g = [1 -.5 -.5][animal sire dam]'), and $V(\hat{M})$ is a matrix product k'Hk, where H is the 3x3 matrix of prediction error covariances. Normally, the animal and parent evaluations would be solved simultaneously prior to calculating $\hat{M}$, which creates both correlations and prediction error correlations among the respective evaluations of the animal, sire and dam. However with the GM_ms family of models the bull's international genomic evaluation (GEBVi) and his MACE parent-average are computed separately, and are therefore not regressed towards each other as they normally would be in a standard genetic evaluation model.

A simple approximation was used previously, which assumes prediction error correlations between animal and parents from MACE (no GEBVs) do not change after replacing the bulls MACE solution with his national GEBV (Sullivan and Jakobsen, 2012). Although not optimal, this assumption guaranteed estimates in the parameter space, and across several traits and populations, resulted in estimates that were generally consistent with MACE variances.

It was recently discovered, however, that while progeny information was correctly absorbed into the effective-records matrix used to construct $V(\hat{M})$, the information from parents of the sire and dam was being ignored. This meant, for example, that dams were considered to have zero reliability unless they had proven sons, because the maternal grandsire information was ignored. Correcting the dam reliability had important effects on the prediction error correlations among the bull, sire and dam, and thus also the approximation of $V(\hat{M})$ and the resulting estimates of genomic variance, which became less consistent with the corresponding MACE variances. The assumption that prediction error correlations would not change after adding the bull's national GEBV seemed very unlikely, and thus new options to approximate $V(\hat{M})$ were explored.

While GEBVi and MACE parent solutions are not solved simultaneously, it is assumed in rGM_ms(v) that average values of $\hat{M}$ should be zero, and these values are thus constrained to ensure a zero average, by subtracting the computed average from each individual $\hat{M}$. Applying this constraint is similar to the approach of allowing MACE parent averages to change, to reflect the contributions of new progeny data (i.e. the national GEBV) added to the system of equations. We can assume for the purpose of estimating $V(\hat{M})$, that MACE parent averages would not change much more than the average change already imposed by this constraint on average $\hat{M}$. Under this assumption, the matrix of prediction error covariances can be directly derived by inverting the matrix W=[D+X]. Matrix D is a zero matrix with the bull's genomic edc added to the diagonal of the given country. The X matrix is from MACE equations for the bull, sire and dam, after absorbing all other animals. Matrix X is already approximated in methods to derive MACE reliabilities (Mark and Sullivan, 2006). An additional adjustment is proposed below, to remove potential bias in the genomic variance estimates due either to errors in assumptions above, or in the approximation of matrix X.

The order of matrix W is 3 times the number of countries included in MACE (e.g. for pro W has dimension 84x84). The matrix needed for $V(\hat{M})$ is only a 3x3 subset of W^{-1}, corresponding with the animal, sire and dam in the given country. However, the full matrix W, including all countries, must be formed and inverted when estimating variances for an individual country. Many matrix inversions are required (e.g. 28 times per bull for pro), but matrix W is small enough that genomic variance estimation is still very quick.

Genomic variance estimates are important in 2 contexts: 1) regarding the relative estimates among countries with genomic data, and 2) regarding the relative estimates between genomic and non-genomic countries. In the

first context, genomic variance estimates can be rescaled (multiplicatively) by any constant value without affecting any of the bull comparisons either within or among the genomic countries. In the second context, genomic variances should on average be very similar to MACE variances to ensure fair comparisons between the genomic and non-genomic countries. In this 2nd context genomic variances can be rescaled so they are "MACE-neutral". Such rescaling is also important to remove systematic bias in genomic variance estimates that can be caused by errors in the approximation of V($\hat{M}$), as discussed above.

Thus the following 3 changes were implemented for the present study:

1. V($\hat{M}$) was updated to consider effective records from ancestors of the bull's parents.
2. V($\hat{M}$) was approximated from W^{-1}, as described above.
3. Genomic variance estimates were re-scaled to be "MACE-neutral", by dividing each genomic SD estimate by the weighted average of the ratios of genomic SD relative to MACE (genomic SD / MACE SD). Weights were numbers of bulls per country used for genomic variance estimation.

The new genomic variance estimates were compared with those previously reported from this same data set, by comparing the relative SD ratios (genomic/MACE) from the 2 studies. Young bull predictions of genetic merit were compared between the 2 studies to measure the impacts of changing the genomic variance estimation procedure.

Results and Discussion

Ratios of genomic relative to MACE estimates of genetic standard deviation (Table 1) were very similar with the new methods relative to recently reported estimates with previous methods (Sullivan and Jakobsen, 2012). The dlo (longevity) estimates were affected relatively more than the estimates for other traits, but mainly because previous estimates were the least consistent with MACE for trait dlo (weighted average SD ratio of 0.86). If not for the "MACE-neutral" rescaling of new estimates, genomic variances would generally have increased for all traits with the newer approximation of V($\hat{M}$). Rescaling did not seem necessary with the previous methods (except for dlo), but this was probably due to an offsetting combination of different biases in the previous approximation of V($\hat{M}$).

Correlations between international predicted breeding values (GEBVi) using either the current or the previous variance estimates were greater than 0.999 on all country scales for all traits except dlo, for which the correlations were all greater than 0.997. These very high correlations indicate very little re-ranking among the young genomically-tested bulls, due to the change in methods for estimating genomic variances.

The standard deviations of GEBVi affect the rankings of young genomically-tested bulls relative to progeny-tested bulls. Relative (i.e. proportional) changes in standard deviation of GEBVi after updating the genomic variance estimates are shown in Table 2. For countries that did not contribute national genomic data for GMACE of the trait, only 1 set of results (the average) is shown, because the results for those countries were very similar on all scales of evaluation. Standard deviations of GEBVi were generally affected by only small amounts, and specific changes for each country were consistent with the changes in genomic variance estimates (Table 1).

The new methods to estimate genomic variance are recommended for the September 2012 RGMACE test run.

Acknowledgements

Appreciations extend to countries providing data for the August 2011 GMACE pilot run, and for helpful review and discussion by members of the GMACE working group; Jette Jakobsen, Gerben deJong, Zengting Liu and Paul VanRaden.

References

Jakobsen, J.H. & Sullivan, P.G. 2012. Robust GMACE for young bulls – application to data. *Interbull Bulletin 45,* 8-12.

Mark, T. & Sullivan, P.G. 2006. Multiple-trait multiple-country genetic evaluations for udder health traits. *J. Dairy Sci. 89,* 4874-4885.

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

Sullivan, P.G. 1999. REML estimation of heterogeneous sire (co)variances for MACE. *Interbull Bulletin 22,* 146-148.

Table 1. Ratio of genetic standard deviation estimates (genomic/MACE), using the current (new) and the previous (old) genomic variance estimation methods.

	Protein		Stature		SCS		Longevity		Fertility	
Country	New	Old	New	Old	New	Old	New	Old	New	Old
CAN	0.83	0.82	0.92	0.96	0.90	0.95	0.87	0.77*	0.89	0.91
CHE	0.64*	0.63*	1.08	1.08	1.48*	1.45*	-	-	1.96*	1.88*
CHR	0.87	0.81	-	-	0.95	0.94	-	-	-	-
DEU	1.10	1.07	1.04	1.07	1.14	1.23*	0.97	0.82	1.13	1.09
DFS	1.25*	1.20	1.19	1.21*	0.94	1.00	1.15	0.99	1.28*	1.33*
FRA	0.90	0.87	0.91	0.92	1.04	1.08	1.14	0.94	0.96	0.99
ITA	1.06	1.04	1.09	1.13	1.02	1.07	1.30*	1.09	-	-
JAP	1.06	1.02	0.97	0.99	0.91	0.93	-	-	-	-
NLD	1.06	1.02	1.00	1.04	0.94	0.99	0.84	0.71*	0.71*	0.71*
POL	1.93*	1.82*	2.00*	2.00*	1.09	1.13	-	-	0.95	0.90
USA	1.00	1.01	-	-	0.93	1.01	1.10	1.01	-	-
Initial Weighted Average[z]	1.04	0.98	1.09	1.03	1.12	1.06	0.99	0.86	1.19	1.00
Final Weighted Average[y]	1.00	0.98	1.00	1.03	1.00	1.06	1.00	0.86	1.00	1.00
Final Simple Average[x]	1.01	0.99	1.04	1.07	1.00	1.05	1.04	0.92	1.02	1.01

*Estimates that are truncated to the limits of range [0.80,1.20].

[z]Before re-scaling to be "MACE-neutral".

[y]After re-scaling, which was applied only to the New estimates.

[x]After re-scaling the New estimates and after truncating extreme New and Old estimates to [0.80,1.20].

Table 2. Ratio of predicted breeding value standard deviations, SD(GEBVi), by country scale of evaluation, for all GEBVi from robust GMACE rGM_ms(v), using the current and previous genomic variance estimates (ratio = current / previous).

Country	Protein	Stature	SCS	Longevity	Fertility
CAN	1.00	1.00	1.00	0.98	0.99
CHE	1.00	1.01	1.03	-	1.00
CHR	1.02	-	1.04	-	-
DEU	1.00	1.00	1.00	1.01	1.01
DFS	0.99	1.01	1.00	1.01	1.00
FRA	1.00	1.01	1.01	1.01	1.00
ITA	1.00	1.00	1.00	0.99	-
JAP	1.01	1.00	1.02	-	-
NLD	1.01	1.00	1.00	0.98	1.00
POL	0.99	1.01	1.01	-	1.01
USA	0.99	-	0.99	0.99	-
Other[z]	0.99	1.01	1.03	0.96	1.00
Number of bulls	42057	41669	42350	42672	39943

[z]Average results for the countries that did not contribute national GEBVs as input to GMACE.