Development of Genomic GMACE

P.G. Sullivan[1] and P.M. VanRaden[2*]
[1]Canadian Dairy Network, Guelph, ON, Canada
[2]Animal Improvement Programs, USDA Agricultural Research Service, Beltsville, MD, USA

Abstract

The use of genomics to enhance national genetic evaluation systems of dairy cattle is quickly becoming standard practice. The current MACE procedure used by Interbull may not accommodate these new "genomically-enhanced" national evaluations. An important assumption in MACE may no longer be valid in the genomics era, the assumption that national evaluations used as input were derived from independent data sets. Genomic predictions are limited by the amount of data currently available within individual countries. Groups of countries may share data, genotypes and/or predicted marker effects to improve genomic predictions, and this will create the need for a modified MACE system, which can account for non-zero residual correlations among genomic predictions from these countries. A system modified for genomics (GMACE) is presented with examples and preliminary results from an application to simulated Brown Swiss data for 9 countries.

Keywords: genomics, international evaluation, MACE, GMACE

Introduction

The use of genomics to enhance national genetic evaluations of dairy cattle is quickly becoming standard practice around the world. The current MACE procedure used by Interbull to combine national evaluations (e.g. EBVs) from member countries may not accommodate these new, "genomically-enhanced" national evaluations (e.g. GEBVs). An important assumption of MACE is that the input national evaluations were derived from independent data, and this may no longer be true in the genomics era.

Genomic predictions are limited by the amount of data available within individual countries. Groups of countries can therefore share data, genotypes and/or prior information about selected markers or genes to get better genomic predictions (VanRaden *et al.,* 2009). A consequence of this cooperation among countries, however, is that input data for MACE may no longer be generated from independent national data sets. The purpose of this paper is to present a genomic GMACE model that eliminates the requirement in MACE that input data are independent.

Methods

Most genetic evaluation systems for dairy cattle, with or without genomics, are based on the linear mixed model. Data are linked to model effects by incidence matrices (e.g. $\mathbf{Z}$) and effects are solved using mixed model equations (MME) of Henderson, (1950), which involve products of the incidence matrices and residual covariances ($\mathbf{Z'R^{-1}Z}$). Sires can have progeny averages in multiple countries, but since each is for an independent group of daughters, the matrix blocks between a pair of countries (*i* and *j*), $\mathbf{Z_i'R^{-1}Z_j=0}$ in the MME of MACE. With sharing of information for genomics, however, a sire's genomic value may be predicted from the same information in multiple countries, in which case $\mathbf{Z_i'R^{-1}Z_j\neq 0}$ in the MME of GMACE. The main challenge for GMACE is to quantify and separate the independent from the shared information within a sire's daughter averages (de-regressed proofs) among countries.

We consider the case of a single-trait per country in this paper, but the methods can also be extended to multiple traits per country (e.g. Sullivan *et al.,* 2005). We also ignore country

means to simplify the presentation, but the means are included in the GMACE model as in MACE (Schaeffer, 1994). With genomics we need to extend from a MACE sire model to a GMACE animal model (van der Linde *et al.*, 2005), because genotyping of females is already happening and is increasing in frequency.

Let **D** be a diagonal matrix of residual variances for de-regressed animal EBV ($\mathbf{D} = \left[\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}\right]^{-1}$). Matrix **E** is block-diagonal by animal, with all diagonals the same as in **D** and off-diagonals reflecting residual covariances from shared data for genomics. Further, let δ represent the progeny equivalents in each average without genomics and δ_g the additional progeny equivalents with genomics included. The MME for GMACE are then:

$$(\mathbf{E}^{-1} + \mathbf{A}^{-1} \otimes \mathbf{T}^{-1})\hat{\mathbf{g}} = (\mathbf{E}^{-1})\mathbf{y} \qquad [1]$$

Matrices **A** and **T** contain animal relationships and genetic covariances among traits, respectively. For comparison, the MME for regular MACE are:

$$(\mathbf{D}^{-1} + \mathbf{A}^{-1} \otimes \mathbf{T}^{-1})\hat{\mathbf{g}} = (\mathbf{D}^{-1})\mathbf{y} \qquad [2]$$

The additional progeny equivalents from genomics can be calculated separately by country using domestic reliabilities before and after adding genomic information to the national evaluation system. A simple way to do this is with the following formula, applied separately for each animal:

$$\delta_g = k\left[\frac{REL_g}{(1-REL_g)} - \frac{REL}{(1-REL)}\right].$$

A disadvantage with this approach is an upward bias that is most severe for younger animals, which are of the key animals of interest in genomic selection strategies (Schaeffer, 2006). The bias comes from increased reliability due to genotyped relatives, which should theoretically be removed from δ_g. Otherwise, when GMACE applies similar reliability increases through the relationship matrix A, the contributions of genotyped relatives will be double-counted. A more precise approach would be to re-compute domestic reliabilities under the GMACE model using only domestic data, and to iteratively modify δ_g until the re-computed reliabilities match the domestic *RELg*.

For a given animal, residual variance for country i (E_i) is equal to $R_i/(\delta + \delta_g)$, and residual covariances between countries (E_{ij}) are a function of the proportion of total progeny equivalents from genomics ($\gamma = \frac{\delta_g}{\delta+\delta_g}$), the amount of shared genomic information (c) between countries and the genetic correlation (r_g) between countries:

$$E_{ij} = r_g c\sqrt{\gamma_i \gamma_j E_i E_j}$$

The genetic correlation acts as an upper limit for the residual correlation because genomic information is predicted in each country from individual gene and marker effects, which are expected to differ among environments in the same way that polygenic estimates differ between environments, due to genotype-by-environment interactions. The residual correlation reaches this maximum when all progeny information is from genomics only (γ=1) and exactly the same data are used for genomic predictions in both countries (c=1). Conversely, residual correlations are very small for proven bulls with many progeny ($\gamma \rightarrow 0$) and for pairs of countries that share very little data for genomic predictions (c→0). Residual correlations are zero, as with regular MACE, if either country has no genomic information (c=γ=0) or if no data are shared for genomic predictions (γ=0).

Step 1. GMACE de-regression

Each set of national evaluations must first be de-regressed to remove covariances among animal solutions, which will be imposed again in step 2, the GMACE evaluation (re-regression). The de-regression step is the same in GMACE as in MACE, and involves solving for **y** in the diagonalized equations:

$$(\mathbf{D}^{-1} + \mathbf{A}^{-1} \otimes \mathbf{T}_*^{-1})\hat{\mathbf{g}}_* = (\mathbf{D}^{-1})\mathbf{y} \qquad [3]$$

Data from each country are processed separately, the same way national proofs were computed. Thus residual correlations between countries are not relevant. Matrix $\mathbf{T}_*$ contains diagonals from **T** and zero off-diagonals.

Step 2. GMACE evaluation

For GMACE, model [1] is applied instead of [2] to the de-regressed proofs (**y**) obtained from model [3]. A simple test for MACE software is to use model [3] instead of [2] in the evaluation step to ensure that the process of de-regression is "reversible". This is useful because the only difference between [2] and [3] is the use of **T** instead of $\mathbf{T}_*$. For GMACE however, there is also the difference of **E** in [1] versus **D** in [3]. Intuitively, reversibility is a desirable feature of MACE, but it is not yet clear if this feature is maintained for GMACE.

Step 3. GMACE reliabilities

Reliabilities for single-trait MACE are approximated by methods described in Harris and Johnson (1988), and a modified version of these methods is used for multiple-trait MACE, as described in Mark and Sullivan (2006). Additional modifications will be needed to approximate reliabilities for GMACE, as neither of the above methods takes account of non-zero residual correlation between countries.

Results for a Single Sire

A simple example was used to demonstrate the effect of non-zero residual correlations in GMACE. GMACE reliabilities and EBVs were computed for a single sire with progeny and genomic data in 3 of 4 countries. To simplify interpretation, the GMACE reliabilities were converted to progeny equivalents, assuming heritability=0.30 in all countries. Genetic correlations among all countries were 0.90.

Results were identical if δ_g =0 or c=0, and are exactly what would result from regular MACE. In this case independent information from 3 countries is combined via MACE to achieve higher levels of reliability, with progeny equivalents increasing from 20 to 42 for each of the countries with data, and from 0 to 30 for country D (Table 1). Input EBVs were scaled up by MACE from 3.00, 3.25 and 3.50 to higher values; 4.25, 4.31 and 4.36 (Table 2) reflecting the increased reliabilities.

If genetic tests were essentially the same in the 3 countries, i.e. based on the same genotypes and performance data (i.e. c=1), then GMACE output would be equal to the input, both in terms of progeny equivalents and national EBVs. The country without data would receive a converted average EBV from the first 3 countries, and progeny equivalents equal to that of a single-country EBV conversion. Double counting is avoided for both the conversion of progeny equivalents (reliability) and EBV variance. At the same time EBVs from all countries are considered for country D. In this example the three countries had similar contributions to country D. However, in practice, countries with higher correlations to country D, or with relatively higher domestic progeny equivalents would contribute relatively more than other countries.

Results for a Population of 9 Countries

Simulated data (VanRaden, 2009) were used to test new GMACE software and to compare results against some alternative approaches. Population structures for Brown Swiss in 9 countries were used to model a simulated world population, with true breeding values available to compute empirical reliabilities. Some comparative results for young bulls are shown in Table 3. Young animals gain the most from international evaluation and from the inclusion of genomic information.

Application of MACE with national EBVs as input did not affect reliability of young domestic bulls on the U.S. scale, but increased reliabilities on all foreign scales. Adding genomics, separately within each country (GEBV) had a bigger affect than MACE on the U.S. scale and on the scales of the larger populations. However, MACE was more helpful than genomics for the smaller foreign populations. Adding both genomics and international methods (mt-GEBV or GMACE) had the biggest impacts, and gave higher

reliabilities on all scales, including the smallest populations.

The mt-GEBV results are from a 9-trait simultaneous genomic analysis, which represents the theoretical ideal among the models considered here. GMACE with the appropriate value of c=0 for these data gave almost equivalent reliabilities as mt-GEBV. Additional studies are needed to test GMACE in situations where countries share varying levels of common data for their national genomic predictions (c>0).

Acknowledgments

The authors thank members of the Interbull genomics task force for ideas and encouragement.

References

Harris, B. & Johnson, D. 1998. Information source reliability method applied to MACE. *Interbull Bulletin 17,* 31-36.

Henderson, C.R. 1950. Estimation of genetic parameters. *Ann. Math Stat. 21,* 309.

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

Schaeffer, L.R. 1994. Multiple-country comparisons of daiy sires. *J. Dairy Sci. 77,* 2671-2678.

Schaeffer, L. R. 2006. Strategy for applying genome-wide selection in dairy cattle. *J. Anim. Breed. Genet. 123,* 218–223.

Sullivan, P.G., Wilton, J.W., Schaeffer, L.R., Jansen, G.J., Robinson, J.A.B. & Allen, O.B. 2005. Genetic evaluation strategies for multiple traits and countries. *Livest. Prod. Sci. 92,* 195-205.

van der Linde, C., de Roos, A., Harbers, A. & de Jong, G. 2005. MACE with sire-mgs and animal pedigree. *Interbull Bulletin 33,* 3-7.

VanRaden, P.M. 2009. Future animal improvement programs applied to global populations. Interbull Bulletin 40, xx-xx.

VanRaden, P.M., Wiggans, G.R., Van Tassell, C.P., Sonstegard, T.S. & Schenkel, F.S. 2009. Benefits from cooperation in genomics. *Interbull Bulletin 39,* 67-72.

Table 1. Progeny equivalents in and out of GMACE.

Progeny equivalents input to GMACE			Country A $\delta + \delta_g$	Country B $\delta + \delta_g$	Country C $\delta + \delta_g$	Country D 0
δ	δ_g	c	Progeny Equivalents resulting from GMACE			
20	0	-	42.3	42.3	42.3	30.4
0	20	0.0	42.3	42.3	42.3	30.4
0	20	0.5	25.4	25.4	25.4	19.4
0	20	1.0	20.0	20.0	20.0	14.3
100	20	0.0	172.2	172.2	172.2	63.4
100	20	0.5	163.1	163.1	163.1	61.4
100	20	1.0	155.3	155.3	155.3	59.5

Table 2. EBVs in and out of GMACE.

			Country A	Country B	Country C	Country D
National EBVs input to GMACE			3.00	3.25	3.50	-
Progeny equivalents input to GMACE			$\delta + \delta_g$	$\delta + \delta_g$	$\delta + \delta_g$	0
δ	δ_g	c	EBVs resulting from GMACE			
20	0	-	4.25	4.31	4.36	4.15
0	20	0.0	4.25	4.31	4.36	4.15
0	20	0.5	3.61	3.70	3.80	3.57
0	20	1.0	3.00	3.25	3.50	3.13
100	20	0.0	3.32	3.46	3.59	3.33
100	20	0.5	3.30	3.44	3.58	3.32
100	20	1.0	3.27	3.42	3.57	3.30

Table 3. Reliability for 120 young Brown Swiss bulls from the United States on the evaluation scales of 9 countries based on national or international evaluations with and without genomics.

		Traditional		Genomics				
Country	Proven Bulls	EBV	MACE	GEBV	mt-GEBV	GMACE* (c=0.0)	GMACE (c=0.5)	GMACE (c=1.0)
Germany	4398	4	12	64	69	68	67	44
Switzerland	2175	14	19	65	73	70	71	54
Italy	1383	1	13	34	64	60	56	23
United States	728	20	20	55	70	69	68	53
Slovenia	278	0	12	6	55	58	57	38
France	230	2	17	21	66	67	66	48
Canada	134	1	17	9	61	59	58	39
Netherlands	101	2	16	6	58	59	57	36
New Zealand	34	1	0	1	26	30	26	16

*Countries did not share data for genomic predictions, so c=0 was expected to perform the best.