Robust GMACE for Young Bulls – Application to Data

J.H. Jakobsen[1] and P.G. Sullivan[2]
[1]Interbull Center, SLU, Uppsala, Sweden
[2]Canadian Dairy Network, Guelph, ON, Canada

Abstract

A robust GMACE method was applied to genomically enhanced breeding values (GEBVs) of young bulls for the five key traits protein, stature, somatic cell score, longevity and female fertility (cc1). Eleven countries participated with GEBVs and conventional breeding values of progeny tested bulls were included from all countries currently subscribing to International Genetic Bull Evaluations. The robust GMACE procedure constrained ratios of genomic standard deviation relative to traditional standard deviation (SD-ratio) to be in the range of 0.80 to 1.20. This restriction changed the bull rankings and number of genomically tested young bulls on the top-100 lists especially for country-trait combinations far from the constraint of 0.80 to 1.20.

Key words: genomics, international evaluation, MACE, GMACE

Introduction

Global dairy cattle breeding has within the last couple of years moved into evaluation of AI bulls based on genomically enhanced breeding values (GEBVs). Young bulls with no progeny test do therefore get a GEBV with a higher reliability compared to an EBV based on pedigree information only. These young bulls are quickly becoming the cohort of largest interest for international trade and Interbull has therefore been asked to develop a genomic MACE procedure to compare GEBVs across country borders. Sullivan *et al.* (2011) suggested a procedure using only GEBVs of young bulls and EBVs of proven bulls. The procedure was applied to real data but showed in some cases a much higher genomic variance compared to the traditional variance.

Aims of the current study were: to test if more data could resolve the issue of extreme SD-ratios that are far from the expectation; to study the country-trait variation for each of the components (MS-deviation and MS-reliability) used to estimate the genomic variance; to apply a robust GMACE procedure rGM_ms(v) having similar variance ratios across countries and compare the ranking of bulls relative to rankings from the non-robust procedure GM_ms(v) and the GMACE procedure for young bulls (GM_yng).

Data

A data call for participation in the present study was sent to national evaluation centers passing the genomic validation test for protein (Nilforooshan *et al.*, 2011). Countries were invited to send GEBV data on all male animals for five key traits representing traits with high and low heritabilities and with large and small correlations among the countries: protein (pro), stature (sta), somatic cells (scs), direct longevity (dlo) and female fertility (cow conception trait #1; cc1) corresponding to the conventional data used in the August 2011 evaluation. Eleven countries (CAN, DEU, DFS, FRA, NLD, POL, USA, CHE, CHR, ITA, JPN) provided data.

The data were edited to include only GEBVs of bulls born since 2006 with no progeny test, and for all other bulls the conventional EBVs that would normally be included in MACE, per Interbull Code of Practice (www.interbull.org). The EBVs included were the same as used for the Interbull routine genetic evaluation in August 2011 for all countries and traits except France, which sent new conventional data for cc1 and dlo. Numbers of GEBV records used for the previous (P) and current (C) studies are listed in Table 1. Ranges of heritabilities and correlations are listed in Table 2.

Table 1. Number of genomic records of young bulls for protein yield, stature, somatic cell count, direct longevity and female fertility (cow conception one; cc1) included for the previous (P) study and for the current (C) study.

	Protein		Stature		Somatic Cell		Longevity		Fertility (CC1)	
	P	C	P	C	P	C	P	C	P	C
CAN	11372	13425	-	14157	11399	13443	-	13425	-	13793
DEU	11481	14143	11486	14247	11336	13850	11481	14864	11481	13795
DFS	1168	4276	1149	4254	1166	4285	1343	4563	764	3926
FRA	6051	7279	5936	7150	6051	7280	6299	7469	6238	7431
NLD	3795	4758	3712	4678	2883	4628	3716	4676	3794	4750
POL	337	296	244	250	336	295	-	-	-	303
USA	1102	11291	-	-	653	7368	653	7368	-	-
CHE	-	382	-	388	-	393	-	-	-	397
CHR	-	665	-	-	-	346	-	-	-	-
ITA	-	734	-	1494	-	1264	-	2298	-	-
JPN	-	653	-	667	-	668	-	-	-	-
Total	35306	57902	22527	47285	33824	53820	23492	54663	22277	44395

Table 2. Ranges of heritabilities and correlations for protein yield, stature, somatic cell count, direct longevity, and female fertility (cow conception one; cc1).

Trait	Range of heritabilities	Range of correlations	No of countries
Protein Yield (pro)	0.136 – 0.508	0.751 – 0.949	28
Stature (sta)	0.370 - 0.630	0.697 – 0.991	21
Somatic Cell Count (scs)	0.062 – 0.433	0.753 - 0.972	27
Direct Longevity (dlo)	0.016 – 0.223	0.299 – 0.934	18
Female Fertility (cc1)	0.010 – 0.067	0.517 – 0.961	13

Sire-dam pedigrees were extracted from the Interbull database. Missing and conflicting birth years were resolved and pedigrees were traced as far back as possible, one trait at a time, starting from animals with national EBV or GEBV. Animals born before 1960 were set to missing, as were parents with unknown sire and dam and only one progeny, but keeping the information about breed, country and sex for the assigning of phantom parent groups.

Methods

Data were analyzed using three different procedures. Firstly the **GM_yng** procedure, which is the GMACE procedure described by VanRanden & Sullivan (2010), but applied to the data set composed of EBVs of progeny tested bulls and GEBVs of young genotyped bulls with no progeny test. Secondly, the GMACE on Mendelian Sampling procedure (**GM_ms(v))** was applied. This procedure was further described by Sullivan *et al.* (2011), but with the acronym VCNV. Lastly the **rGM_ms(v)** procedure was applied. This procedure was based on the idea from GM_ms(v) but with robust restrictions to SD-ratios deviating largely from the expectation of one. This procedure is further described in Sullivan & Jakobsen (2012).

Two groups of genomic data sharing were identified reflecting the situation per August 2011. These were North America [CAN; USA] and EuroGenomics [DEU; DFS; FRA; NLD]. Sharing was assumed to be 100% within each of the groups, 25% between the groups, and 25% for all pairings with and among the remaining countries that were not in any sharing group.

Results and Discussion

Global bull rankings are affected by the sire standard deviations and the genomic variance of the young bulls are expected to line up with the

sire variance of the proven bulls. This ratio is referred to as the SD-ratio. The SD-ratios for pro, sta, scs, dlo, and cc1 are listed in Table 3 for each of the populations participating with GEBV data for young bulls. Decimal differences of these valus compared to the values presented by Sullivan & Jakobsen (2012) were caused by a difference in deregression of national breeding values between the two studies.

Table 3. SD-ratios for protein (pro), stature (sta), somatic cell score (scs), direct longevity (dlo) and female fertility (cc1) for the different countries (COU) participating with GEBV data.

COU	pro	sta	scs	dlo	cc1
CAN	0.80	0.96	0.95	0.79	0.92
DEU	1.06	1.09	1.23	0.82	1.11
DFS	1.04	1.15	1.02	1.01	1.52
FRA	0.87	0.93	1.08	1.27	1.00
NLD	1.00	1.03	0.99	0.71	0.72
POL	1.62	3.24	1.12		1.04
USA	1.03		1.01	1.01	
CHE	0.79	1.09	1.54		1.93
CHR	1.05		0.98		
ITA	1.04	2.63	1.04	1.06	
JPN	1.06	1.00	0.93		

Some of the population-traits were far from the expectation of an SD-ratio of one. The variance of young bulls is computed from ratios of averages of MS^2 and the reliability of MS, where MS is a genomic estimate of Mendelian Sampling (described as $\widehat{M}$ in Sullivan and Jakobsen, 2012). These components were investigated to possibly find the cause of high SD-ratios. The average MS per country and trait is shown in Figure 1. The expected average MS per country and trait is zero and the values in Figure 1 indicate a deviation from zero for some country-trait combinations [ITA-sta; POL-sta; FRA-dlo; DFS-cc1; POL-cc1]. Three of these country-trait combinations coincide with the country-traits showing a high SD-ratio in Table 3.

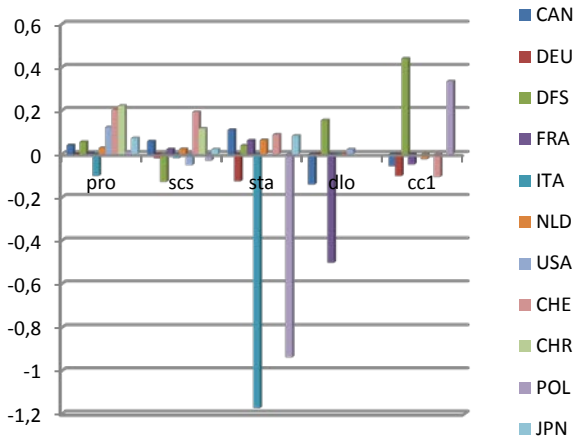

Figure 1. Average MS per country and trait.

The variation in MS reliability per country and trait was studied (Figure 2). The MS reliability was much lower for CHE-scs and CHE-cc1 compared to other countries. These low values can influence the genomic variance of the young bulls and further the SD-ratio.

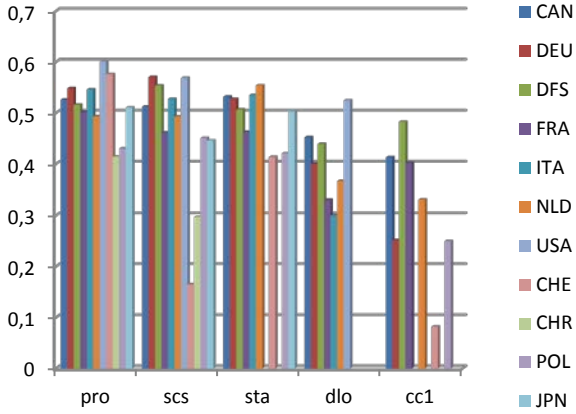

Figure 2. MS reliability per country and trait,

The first aim of the current study was to investigate if a data call requesting complete data would resolve the issue of large SD-ratios as presented by Sullivan *et al.* (2011). Some of the issues were resolved but some remained and the robust GMACE procedure rGM_ms(v) was

herefore applied, where the average MS was subtracted from all MS before conversion among country scales and the SD-ratio was constrained to the interval between 0.80 and 1.20.

Data were analyzed with the GM_yng, the GM_ms(v) and the rGM_ms(v) models and number of young genotyped bulls on the top-100 lists for each of the traits, each of the countries and each of the models are shown in Figures 3 to 7. The most remarkable result was the change in number of young genotyped bulls on the top-100 list when changing the model from GM_ms(v) to rGM_ms(v). The restrictions on SD-ratios strongly affected the rankings for populations with an SD-ratio far from one. The validation results presented by Sullivan & Jakobsen (2012) further indicated that the procedure including the robust restrictions is clearly outperforming the procedure without restrictions.

In order to study the behavior of the global genetic trend for the joint data of progeny tested bulls and young genotyped bulls with no progeny test, the global trends were obtained for pro, sta, scs, dlo and cc1 on DFS scale (Figure 8). The graph shows a smooth continuation in trends from proven bulls to young genotyped bulls. Although smooth continuation in trends were also seen on other country scales, the DFS scale was chosen here because all DFS traits are published on an RBV scale with a mean of 100 and a SD of 10.

Conclusions

More complete data resolved some but not all of the issues raised at the previous pilot about extreme SD-ratios. A robust GMACE procedure (rGM_ms(v)) was therefore developed to constrain SD-ratios within the interval of 0.80 to 1.20. Number of bulls on the top-100 lists were compared for individual countries when applying the GM_yng, GM_ms(v) and the rGM_ms(v) procedures. Applying the rGM_ms(v) procedure reduced the number of young genotyped bulls on the top 100 lists compared to GM_ms(v) especially for the country-traits with very large SD-ratios. The reductions in number of young bulls on these top-100 lists was expected due to the reductions in the SD-ratios.

Acknowledgements

The national evaluation centers that kindly provided data for this study are greatly acknowledged for their participation.

References

Nilforooshan, M.A., Zumbach, B., Jakobsen, J., Loberg, A., Jorjani, H. & Dürr, J. 2011. Validation of national genomic evaluations. *Interbull Bulletin 42*, 56-61.

Sullivan, P.G. & Jakobsen, J.H. 2012. Methods for international genomic evaluation of young bulls. Interbull Workshop, Feb. 2-3. Verona. Italy. (in preparation)

Sullivan, P.G., Zumbach, B., Dürr, J. & Jakobsen, J.H. 2011. International genomic evaluation of young bulls. *Interbull Bulletin 44*, 87-94.

VanRaden, P.M. & Sullivan, P.G. 2010. International genomic evaluation methods for dairy cattle. *Gen. Sel. Evol. 42*, 7.

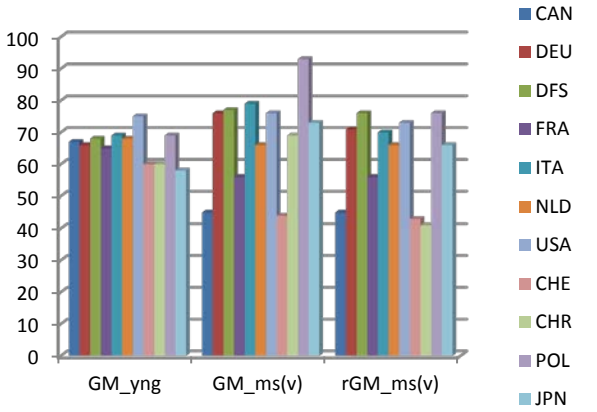

Figure 3. Number of bulls on top-100 list for protein per country and model

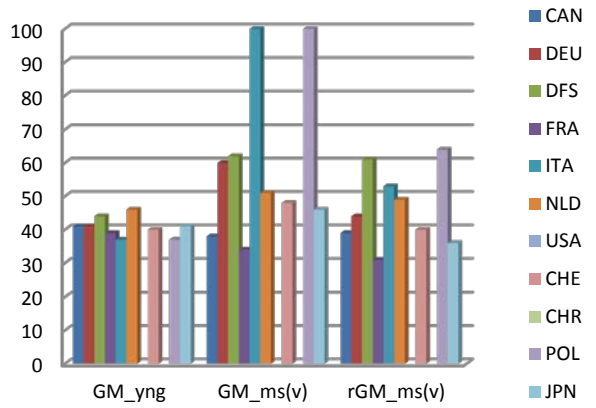

Figure 4. Number of bulls on top-100 list for stature per country and model

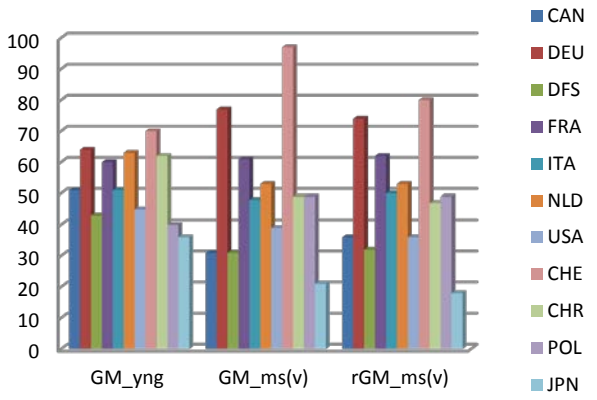

Figure 5. Number of bulls on top-100 list for somatic cell per country and model

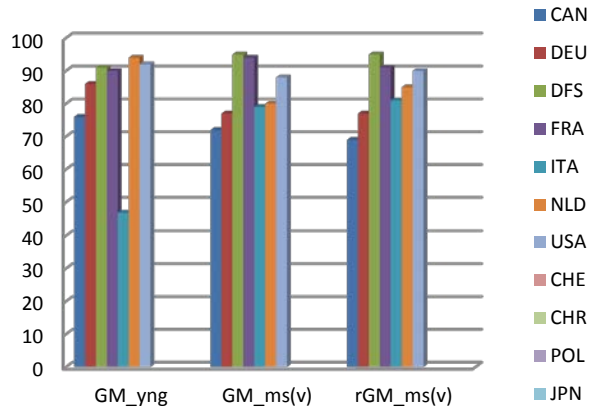

Figure 6. Number of bulls on top-100 list for direct longevity per country and model

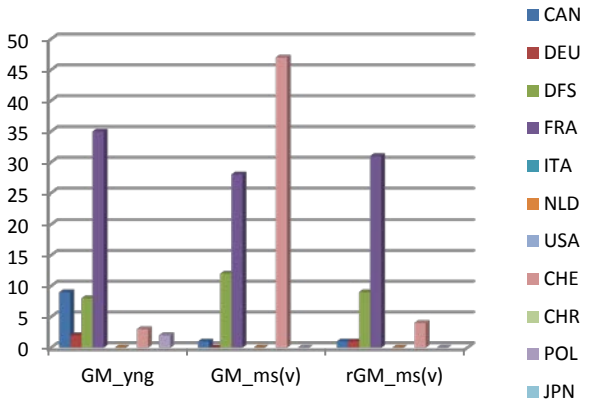

Figure 7. Number of bulls on top-100 list for female fertility (CC1) per country and model

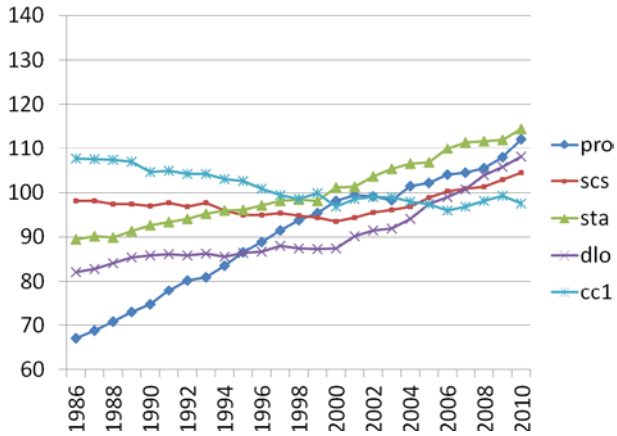

Figure 8. Global genetic trend for protein (pro), somatic cell (scs), stature (sta), longevity (dlo) and fertility (cc1) on DFS scale RBV(100, 10) for combined international (G)EBVs