Global Interbull EBV in Domestic Single Step Genomic Evaluation

J. Přibyl*[1]*, J. Bauer*[1]*, J. Přibylová*[1]*, L. Vostrý , L. Zavadilová*[1]*, V. Čermák*[2]*, Z. Růžička*[2]*,
***J. Šplíchal*[2]*, M. Verner*[2]*, J. Motyčka*[3] *and L. Vondrášek*[3]**
[1] *Institute of Animal Science, Praha – Uhříněves, Czech Rep.*
[2] *Czech Moravian Breeding Corporation, Praha , Czech Rep.*
[3] *Holstein Cattle Breeders Association of the Czech Rep.*

Abstract

EBVs and GEBVs for milk production of young genotyped Holstein bulls were predicted using a conventional Animal Model, Ridge Regression, GBLUP, and ssGBLUP procedures. For prediction, the nation-wide database of domestic Czech production records was combined with deregressed proofs from Interbull files through 2008, which had been transformed by MACE to reflect domestic production conditions. 1,259 genotyped bulls had already been proven in 2008. Analyses were run that used Interbull values only for these genotyped bulls and used Interbull values for all available sires. Predictions were validated by comparing correlations of breeding value predictions with EBV and DYD after progeny test in 2012 of 140 young genotyped bulls and their associated reliabilities. Combining domestic data with Interbull EBVs improved prediction of both EBV and GEBV. Prediction by Animal Model (traditional EBV) using only the domestic database had 0.29 validated reliability of prediction; whereas combining the nation-wide domestic database with all available deregressed proofs for genotyped and non-genotyped sires from Interbull resulted in reliability of 0.34, compared to 0.36 when using Interbull data only. Highest reliabilities were for predictions from the ssGBLUP procedure using combined data, or with all available deregressed proofs from Interbull only, which reached validated reliabilities for GEBV predictions 0.53 and 0.54 respectively.

Key words: genomic breeding value, single-step prediction, animal model, validated reliability

Introduction

In small Holstein populations, a substantial proportion of matings are often to imported bulls or semen. In such cases, sires have low and only indirect genetic relationship to the domestic population. Interbull correlations of Czech Republic with others countries are approximately 0.85, resulting in reliability of estimated breeding values (EBV) after imports to about 72 % of values in the country of origin. Between countries with different climate and production conditions (Ireland, Israel, NZL for example), correlations are even lower, about 0.75.

These circumstances negatively influence genetic evaluations of animals and also international (MACE) comparisons. Typically, however, the criterion for selection is the production and rank of animals under domestic management and environmental conditions.

For prediction of genomic enhanced breeding value (GEBV), multi-step procedures using a variety of regression-based methodologies (Meuwissen *et al.*, 2001) and the GBLUP method using a genomic relationship matrix (VanRaden 2008) have been developed. Using daughter yield deviations (DYD) or deregressed proofs (DRP), direct genetic values (DGV) are calculated, which are then combined with parent average (PA) in a selection index to produce GEBV. Misztal *et al.* (2009), Christensen and Lund (2010), Aguilar *et al.* (2010), and Legarra and Ducrocq (2012) developed a single-step procedure ssGBLUP which effectively combines nation-wide production record databases with genomic information. This method avoids critical assumptions required by some others methods, and allows common rank of all genotyped and un-genotyped animals in a population.

Přibyl *et al.* (2012) used this methodology for the genetic evaluation of the Czech Holstein population. Despite using a small number of proven reference bulls, genotyping of proven and young bulls led to an increased correlation of the GEBVs of young bulls with their genetic prediction after progeny test. As mentioned, imported sires typically have a low genetic relationship to the domestic population. Therefore using information from Interbull EBVs could improve accuracy of prediction.

Gao *et al.* (2012) and Su *et al.* (2012) used as input data DRP of sires instead of production records in ssGBLUP. Přibyl *et al.* (2013) combined ssGBLUP nation-wide databases of production with Interbull DRP. Implanting the Interbull file converted by MACE into a scale reflecting Czech production conditions improved accuracy of prediction. To demonstrate possible benefits of combining data sources, the new issue from Interbull (2011) was used. Improvement depended on correlation of young genotyped bulls with Interbull database (unpublished results). The newer the data and younger the Interbull bulls, the higher the improvement in accuracy of prediction of young bulls under domestic conditions.

The aim of this study was evaluate methods of genetic prediction for young bulls by GEBV using both domestic and Interbull data from 2008.

Material and Methods

Production records from first lactations of Czech Holstein cows, Interbull milk yield EBVs of bulls and pedigree databases were used. Two overlapping data sets of domestic milk production performance data and one converted data set were used:

Domestic - 969,269 1st lactations, calving years 1991 - 2008, and 1,762,905 animals in the pedigree file.

Domestic - 1,185,225 1st lactations, calving years 1991 - 2012, and 1,958,139 animals in the pedigree file.

Interbull - 98,037 EBVs through year 2008, average reliability 0.70 (> 0.28), converted by MACE on a national scale. Values were modified in order that variability of EBV of domestic proven bulls and of Interbull EBVs were similar. Estimated breeding values were deregressed (Rozzi *et al.,* 1990) and pseudo-data yield deviation (YD) and effective record contributions (ERC) were calculated, considering sire as animal with its own production:

$$ERC = ((1- h^2) / h^2)*(rel/(1-rel))$$

where rel = reliability of EBV.

Bulls were genotyped by Illumina BovineSNP50 BeadChip V2 (Illumina Inc., San Diego, CA), and data were edited for: MAF, Gscore, number of loci per bull, number of bulls per locus, substantial error of prediction of old bulls in the training set, large discrepancy of relationship A22 : G, and proportion of H.

After editing, 39,904 loci for 1,605 bulls, from which 1,259 were already proven in 2008 (training set), 140 young with 0 daughters in 2008 and > 50 daughters (average 67) in 2012, and 206 others bulls with a small number of daughters were used.

Data were evaluated by weighted analysis using several statistical procedures. Because ERC was used as the weight for individual records, for all domestic production records ERC was set equal to 1.

Ridge Regression (RR) according to the model:

$$y_j = \mu + \Sigma\delta_i \cdot g_i + e_j \quad ,$$

where
y_j = DRP of milk production for bull j ;
μ = common constant (contemporary group) as a fixed effect ;
δ_i = regression coefficient for locus i, with addition of diagonal matrix of constant value to the system of equations ;
g_i = value of alleles in locus i <0, 1, 2>;
e_j = random error.

GBLUP according the model:

$$y_j = \mu + an_j + e_j \quad ,$$

where
an_j = animal j with genomic relationship matrix **G.**

BLUP, and ssGBLUP according to the animal model:

$$y_{ij} = HYS_i + \beta_1 \cdot ca_j + \beta_2 \cdot ca_j^2 + \beta_3 \cdot do_j + \beta_4 \cdot do_j^2 + an_j + e_{ij}\ ,$$

where
y_{ij} = first lactation milk yield of cow, or DRP of milk production for bull;
HYSi = contemporary group within a herd in a 3-mo calving period (fixed effect);
β_1, β_2, β_3, and β_4 = regression coefficients;
ca_j and ca_j^2 = parameters for curvilinear regressions on calving age (fixed effect);
do_j and do_j^2 = parameters for curvilinear regressions on days open (fixed effect);
an_j = animal j with pedigree additive relationship matrix **A** in BLUP, or matrix **H** in ssGBLUP.

H is the pedigree additive relationship matrix **A** augmented by genomic relationship matrix **G.** Weights of 80% **G** and 20% additive pedigree relationship matrix only for genotyped animals **A22** were used for incorporation into **H.**

Matrix **G** was calculated according to deviations from the averages of observed allele frequencies and was standardized by division by the average value of the diagonal of **G** (Forni *et al.*, 2011), then shifted, so that the elements of the **A22** and elements of **G** would have the same average (Vitezica *et al.*, 2011).

The programme G-matrix (Su and Madsen 2011) was used for construction of the **G** relationship matrix, and the DMU package of Madsen and Jensen (2010) was used for genetic prediction.

Procedures for the various models for genetic prediction are summarized in Table 1.

Table 1. Prediction procedures.

Method	Calculated value	Sources of production 2008 **D**omestic	**I**nterbull	**D + I**
BLUP	EBV	D-EBV	I-EBV	DI-EBV
Ridge Regression	DGV		rI-DGV	
	GEBV*			rI-GEBV
GBLUP	DGV		gI-DGV	
	GEBV *			gI-GEBV
ssGBLUP**	GEBV	D-GEBV	I-GEBV	DI-GEBV

* GEBV = 0.8 DGV + 0.2 D-EBV .
** Genomic relationship **G** is weighted 80 % and pedigree relationship **A22** 20 %.

Domestic production records were used in BLUP and ssGBLUP genetic prediction procedures (Table 2); whereas in RR and GBLUP analyses, Interbull DRPs from 1,259 referenced bulls were used, which represented a total of 57,864 ERCs. These values were combined in an index with EBV estimates according to pedigree information from the domestic Holstein population. Of all DRPs available from Interbull, a total of 98,037 were used in BLUP and ssGBLUP procedures, and this database represents 785,276 ERCs. The combination of both domestic and Interbull databases identified 1,064,912 records (1,632,668 ERCs) that were analyzed by BLUP and ssGBLUP procedures. In these analyses, Interbull DRPs were used only when sires did not have daughters in the domestic population.

Procedures were validated by calculating correlations among predictors of genetic merit for 140 young bulls that had no daughter records in 2008 but > 50 daughter records in 2012, that is, their EBVs and DYDs after progeny test (Szyda *et al.*, 2008, 2011).

Table 2. Size of data for prediction 2008.

	Records	WeightsERC	Method
Domestic	969,269	969,269	D-EBV
			D-GEBV
Interbull for	1,259	57,864	rI-DGV
genotyped bulls	970,528§§	240,145	rI-GEBV
	1,259	57,864	gI-DGV
	970,528§§	240,145	gI-GEBV
Interbull for all bulls	98,037	785,276	I-EBV
			I-GEBV
D+I all§	1,064,912	1,632,668	DI-EBV
			DI-GEBV

§ From Interbull file, only bulls with no domestic daughters.
§§ Including pedigree information from domestic population.

Average validated reliabilities were computed from correlation of prediction with DYD as follows:

$$VRel = r_{P,DYD}^{2} / rel_{DYD} \quad ,$$

where
$r_{P,DYD}$ = correlation of predicted method with DYD after progeny test;
rel_{DYD} = reliability as affected by number of progeny, corresponding with DYD.

Results & Discussion

Merging domestic production records with Interbull files notably increased volume of input data for genetic evaluation (Table 2).

Results were expressed as deviations from a base population of 2,116 proven sires, each having at least 60 daughters in 2008. Average value (EBV/GEBV) of prediction of young bulls has deviation in a case of evaluation the domestic population 657, and 672 kg of milk for BLUP, and ssGBLUP methods, respectively; whereas from combined data these averages were 651 and 640 kg, respectively. Average values of prediction were in a good agreement with results based upon progeny test, in which average EBV for this group of young bulls was 629 kg of milk (Table 3). In average, predictions of breeding values of young bulls for all methods were overestimated about 1.7 to 6.8 %.

Correlations of predictions with EBV after progeny test (EBV12) were markedly higher than with DYD (DYD12) (Table 3). D-EBV were EBVs of young animals, reflecting the response of pedigree of (imported) young bulls in a domestic condition. Predictions of young bulls according to this “usual” BLUP-Animal Model analysis were correlated with EBV12 by 0.59 and with DYD12 by 0.47. Corresponding validated reliability (VRel) was 0.29. Predictions derived from ssGBLUP of domestic data reached VRel of 0.48.

Prediction with DGV by RR and GBLUP, which were according to Interbull DRP for genotyped bulls only, were correlated to EBV12 by 0.60, and 0.59 respectively, and correlated to DYD12 by 0.57. Differences in correlations to EBV12 and to DYD12 were much lower than when using BLUP and ssGBLUP and domestic databases. Corresponding VRel were 0.42 and 0.41, respectively. After combination with pedigree values, reliabilities reached VRel 0.47. This is close to value obtained on the domestic population using ssGBLUP.

Predictions of EBV using the BLUP method including all Interbull DRP versus using combined data reached VRel of 0.36 and 0.34, respectively, which were markedly higher than from domestic population data only (Table 3). Predictions by GEBV with ssGBLUP from Interbull and combined data had VRel values of 0.54 and 0.53, respectively.

Values achieved on Interbull and combined data were similar.

In combined data, only Interbull sires that did not have domestic daughters were used for BLUP and ssGBLUP procedures. In methods I-EBV and I-GEBV, all available data from Interbull were used, including contributions from the Czech population. Therefore sources of information were similar in both cases. The Interbull database contained 785,276 ERC (Table 2) connected directly to sires, which had substantial predicative ability, greater than a population of cows of similar size. On the other hand, the Interbull database was generated under production conditions not closely similar to those of the Czech domestic herds.

Conclusions

Combining genetic evaluation of domestic data with Interbull EBVs transformed by MACE into domestic production conditions improved prediction both of EBV and GEBV.

The ssGBLUP method enabled use of daughter’s production records and/or DRPs both for genotyped and un-genotyped sires in joint genetic evaluation.

Generally, the most reliable genetic predictions, according to repeated calculations, were produced by the ssGBLUP procedure utilizing combined data. Differences in accuracy of prediction between ssGBLUP in combined data and ssGBLUP using only Interbull data were small.

Table 3. Average genetic predictions for 140 young bulls, correlations of predictions with results after progeny test, and validated reliabilities (VRel).

Data 2008	Mean milk kg +	Method	EBV 2012	DYD 2012	Validated reliability
Domestic	657	D-EBV	0.59	0.47	0.29
	672	D-GEBV	0.70	0.61	0.48
Interbull for		rI-DGV	0.60	0.57	0.42
genotyped		rI-GEBV	0.67	0.61	0.47
bulls		gI-DGV	0.59	0.57	0.41
		gI-GEBV	0.66	0.61	0.47
Interbull for		I-EBV	0.62	0.53	0.36
all bulls		I-GEBV	0.70	**0.65**	**0.54**
D + I all	651	DI-EBV	0.63	0.51	0.34
	640	DI-GEBV	**0.73**++	0.64	0.53
Data 2012	629	D-EBV			

+ Difference of EBV / GEBV from basis of 2,116 sires each with at least 60 daughters in 2008.

+ +When using for response variable GEBV12 means GEBV by ssGBLUP in year 2012, the highest correlation is for DI-GEBV with value 0.75.

Acknowledgements

We are grateful to Dr. Per Madsen, University of Aarhus, Denmark, and to Prof. Ignacy Misztal, University of Georgia, USA, for methodical support and for provision of computer programs. Special thanks are due to W. D. Hohenboken (Corvallis, OR) for editing the English of the paper and for helpful comments. Supported by the Ministry of Agriculture of the Czech Republic Prague; (Projects no. MZe 0002701404 and QI111A167).

References

Aguilar, I., Misztal, I., Johnson, D.L., Legarra, A., Tsuruta, S. & Lawlor, T.J. 2010. Hot topic:A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of

Holstein final score. *J. Dairy Sci. 93,* 743–752.

Christensen, O.F. & Lund, M.S. 2010. Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol. 42:2.*

Forni, S., Aguilar, I. & Misztal, I. 2011. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. *Genet. Sel. Evol. 43:1.*

Gao, H., Christensen, O.F., Madsen, P., Nielsen, U.S., Zhang, Y., Lund, M.S. & Su, G. 2012. Comparison on genomic predictions using three GBLUP methods and two single-step blending methods in the Nordic Holstein population. *Genet. Sel. Evol. 44:8.*

Legarra, A. & Ducrocq, V. 2012. Computation strategies for national integration of phenotypic, genomic, and pedigree data in a single-step best linear unbiased prediction. *J. Dairy Sci. 95,* 4629-4654.

Madsen, P. & Jensen, J. 2010. *DMU—A package for analysing multivariate mixed models.* Version 6, release 5.0. *http://dmu.agrsci.dk.*

Meuwissen, T.H.E., Hayes, B.J. & Goddard, M.E. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics 157,* 1819–1829.

Misztal, I., Legarra, A. & Aguilar, I. 2009. Computing procedures for genetic evaluationincluding phenotypic, full pedigree, and genomic information. *J. Dairy Sci. 92,* 4648–4655.

Přibyl, J., Haman, J., Kott, T., Přibylová, J., Šimečková, M., Vostrý, L., Zavadilová, L., Čermák, V., Růžička, Z., Šplíchal, J., Verner, M., Motyčka, J. & Vondrášek, L. 2012. Single-step prediction of genomic breeding value in a small dairy cattle population with strong import of foreign genes. *Czech J. Anim. Sci. 57,* 151–159.

Přibyl, J., Madsen, P., Bauer, J., Přibylová, J., Šimečková, M., Vostrý, L. & Zavadilová, L. 2013. Contribution of domestic production records, Interbull estimated breeding values, and single nucleotide polymorphism genetic markers to the single-step genomic evaluation of milk production. *J. Dairy Sci. 96,* 1865-1873.

Rozzi, P., Schaeffer, L.R., Burnside, E.B. & Schlote, W. 1990. International evaluation of Holstein-Friesian dairy sires from three countries. *Livest. Prod. Sci. 24,* 15–28.

Su, G. & Madsen, P. 2011. *User's Guide for Gmatrix. A program for computing Genomic relationship matrix. http://dmu.agrsci.dk.*

Su, G., Madsen, P., Nielsen, U.S., Mäntysaari, E.A., Aamand, G.P., Christensen, O.F. & Lund, M.S. 2012. Genomic prediction for Nordic Red Cattle using one-step and selection index blending. *J. Dairy Sci. 95,* 909–917.

Szyda, J., Ptak, E., Komisarek, J. & Żarnecki, A. 2008. Practical application of daughter yield deviations in dairy cattle breeding. *J. Appl. Genet. 49,* 183–191.

Szyda, J., Żarnecki, A., Suchocki, T. & Kaminski, S. 2011. Fitting and validating the genomic evaluation model to Polish Holstein-Friesian cattle. *J. Appl. Genet. 52,* 363–366.

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

Vitezica, Z.G., Aguilar, I., Misztal, I. & Legarra, A. 2011. Bias in genomic predictions for populations under selection. *Genet. Res. (Camb.) 93,* 357–366.