The German-Austrian Genomic Evaluation System for Fleckvieh (Simmental) Cattle

***C. Edel*[1]*, H. Schwarzenbacher*[2]*, H. Hamann*[3]*, S. Neuner*[1]*, R. Emmerling*[1] *and K.U. Götz*[1]**
1Bavarian State Research Center for Agriculture, 85586 Poing-Grub, Germany
2ZuchtData EDV Dienstleistungen, Wien, Austria
3 Landesamt für Geoinformation und Landentwicklung Baden-Württemberg, LGL

Abstract

In December 2010 a routine application for genomic breeding value estimation in the German-Austrian Fleckvieh population was established that is conducted monthly. Since August 2011 the evaluation system is ICAR proven and combined genomic breeding values (GEBV) have now become official. Genomic breeding values are estimated for a total of 45 traits. A special aspect of the current implementation is the division of labor between the three evaluation-centers in Bavaria, Austria and Baden-Württemberg that follows the joint alpine collaboration in conventional breeding value estimation. Starting from a central preparation step for genotypes, results are propagated to the partners for the estimation of genomic breeding values for their specific traits. A two step approach with method G-BLUP based on the use of a genomic relationship matrix is used for all traits. Results of the current validation show a substantial gain in realized reliabilities from genomic breeding values over the reliabilities of the simple parent-average. However, gains do not reach values reported for the Holstein population at a comparable size of the calibration sample, which is probably an effect of the significantly larger effective population size in Fleckvieh.

Introduction

The dual-purpose breed Fleckvieh is one of the major breeds in worldwide milk production. However, it’s main distribution and breeding area covers a quite limited number of regions, including the south of Germany, Austria, parts of Switzerland and the north of Italy. A joint genetic evaluation for the German and Austrian population was established in 2002. Based on cooperation between the breeding organizations and evaluation centers of both countries a division of labor was established, where the development of the evaluation-routine, as well as the responsibilities with respect to the routine-evaluation itself were divided among evaluation centers by trait groups. Since then Baden-Württemberg is responsible for traits associated with beef-production, the Austrian team has its focus on functional traits and the Bavarian team works on milk-production, conformation traits, somatic cell score and milkability. With this cooperation a high proportion of the international pure-bred population of Fleckvieh is covered. Breeding values are estimated jointly without the need for an additional combination step like MACE. This was seen as a favorable situation for the development of a genomic evaluation system following the same concept. Starting with an unofficial test release in December 2010 and followed by a monthly estimation of candidates, genomic breeding value estimation was established for dual-purpose Fleckvieh. The evaluation routine was validated by ICAR with respect to milk-production traits in June, 2011. The recalibration of the genomic system in August, 2011 based on the results of the conventional breeding value estimation, was followed by the first official release of genomic breeding values. Since then, combined genomic breeding values (GEBV) are the officially published breeding values of a genotyped animal.

Details

Within the cooperation currently about 12,400 animals are genotyped with the Illumina BovineSNP50 chip. Depending on the availability of phenotypes, trait specifically between 5,200 and 6,100 bull genotypes are used for the calibration of the genomic system. Historically the foundation for the now jointly used genotype pool was a cooperation between the German and Austrian breeders-associations for Fleckvieh (ASR, AGÖF), the scientific research program called 'Genotrack' focusing on the use of genomic information in animal breeding, the FBF (Förderverein Biotechnologieforschung e.V.), and the Bavarian Ministry for Agriculture represented by the Institute of Animal Breeding of the Bavarian State Research Center for Agriculture (ITZ). Each of these supporters contributed several hundreds of genotypes into the pool. This starting stock was augmented by 3,100 bulls currently under progeny testing in Germany and Austria. It is assumed that the calibration pool will be constantly growing by around 600 to 700 bulls that get first proofs from daughters per year.

As already stressed a remarkable aspect of the German-Austrian routine implementation is the division of labor established between countries and evaluation-centers. The flow of data within this cooperation can be summarized as follows:

- Genotypes are stored in two databases in Germany and Austria that are mirrored.
- Genotypes and supporting pedigree information are extracted and filtered by minimum criteria (e.g. at least 95% call-rate).
- Genotypes are validated and prepared in a central processing step localized at the ITZ. Validation includes standard marker based criteria (MAF, HWE) and checks on reasonable accordance of genotypes of related animals using direct comparisons (parent-offspring) and marker-based IBD-coefficients (Wang, 2002) for checks on grand-parent-grand-offspring pairs and within paternal and maternal half and full sib families. Offending genotypes are either deleted or the offending pedigree-links is corrected following a defined protocol. In the future this step might also include an imputation step based on all available validated genotypes.
- Validated and processed genotypes and intermediate results are propagated to the collaborating evaluation centers. Data supplied include a refined version of the pedigree of the project-animals, a conventional relationship-matrix based on this pedigree, a genomic relationship matrix G like described by VanRaden, 2008 ('type 1') using base allele frequencies calculated as outlined by Gengler *et al.* 2007. Base allele frequencies and rough frequency estimates based on all genotyped animals as well as detailed reports covering all results from the preparation and validation step are also supplied.

Subsequent steps are not completely uniform across traits and evaluation-centers. A rough guidance may be:

- Project partners prepare phenotypes and weights for the traits under their responsibility. In general DYD (VanRaden and Wiggans, 1991) are used as a phenotype. However, especially with functional traits or in some specific multivariate settings the calculation of DYD is not straightforward. In these cases deregressed proofs (Garrick *et al.*, 2009) are calculated and used in the genomic evaluation.
- Project partners have also defined a set of rules for data inclusion. With DYD as phenotype, data inclusion is depending on whether a DYD is well defined or not. With respect to milk-production traits within the framework of random regression models, a bull's DYD function is considered as well defined if at least one daughter has information beyond the 250th day in milk in first lactation.
- Trait specific models are set up, containing between 10% and 25% polygenic variance assumed to be not captured by markers and represented by the conventional relationship matrix.
- Solving of the system is done by direct inversion. Reliabilities are calculated using the prediction-error variances derived. Inverted matrices are stored for monthly candidate runs between recalibrations of the system. Within the routine-evaluation the system is never solved for SNP-effects.
- Systems are again solved to derive the so called subset-EBV, replacing the matrix **G** by the conventional relationship-matrix.
- Combination of the direct genomic values (DGV) and the conventional breeding values (EBV) is done via an index as described by VanRaden *et al.*, 2009 or

some modifications hereof as described by Edel *et al.,* 2010. No penalty is given to the theoretical reliability of the DGV in this index as proposed by other authors. The reliability published with the GEBV is the theoretical reliability of the index without further modifications.

- Results are sent to the Austrian evaluation center (ZuchtData). ZuchtData calculates the total merit index based on GEBV and does final post-processing. Calculation of all indexes published (total merit, milk-production, fitness and beef-production) is based on GEBV using the same procedures as used with conventional EBV.
- Results are transferred to the data-bases. Results for very young animals are delivered back to initiators of the genotyping and the related breeding organization. Results for AI-bulls are published broadly.

As seen from this list the method applied in routine-evaluation is a standard two-step approach using G-BLUP (VanRaden, 2008). More sophisticated methods that were intensively tested in the past including BayesA, BayesB (Meuwissen *et al.,* 2001) and the Bayesian Lasso (Park and Casella, 2008) consistently failed to show their theoretical superiority up to now (Gredler *et al.,* 2010 and own unpublished results). A one step approach as proposed for example by Misztal *et al.,* 2009 might be an interesting alternative in the future.

Validation

A part of the now established genomic routine evaluation is a validation process that is comparable although slightly different to the Interbull GEBV-Test guidelines. After each recalibration of the genomic system that is done after conventional breeding-value estimation, a shifting age group of animals is declared as validation group. The calibration of the genomic system is repeated omitting the phenotypes of the validation animals in the prediction. Currently the ‘split date’ is the 1. August 2003. Phenotypes of bulls born after that date are discarded and DGV are estimated for them. These DGV are blended with conventional parent averages (PA) that are individually extracted for each animal from historical conventional breeding value estimations. This means the information content of this PA is quite close to a situation where a male calf is genotyped and judged by PA or GEBV as a potential future AI-bull. Finally, DGV and GEBV of these validation animals are compared to their current conventional daughter based EBV if this EBV has an information content of at least 20 effective daughter equivalents (similar to what is proposed for the Interbull GEBV-Test). Table 1 summarizes some results drawn from the validation in August 2011. In this validation between 3,466 (beef-production index) and 4,074 (milk yield) animals were part of the calibration group and between 825 (fitness-index) and 2,228 (paternal calving ease) animals were part of the validation group.

Results and Discussion

Squared validation correlations (in all cases to current EBV) are substantially higher for GEBV than PA. Gains observed were in the range of 0.10 to 0.21 for most traits and even higher with indexes combining several traits e.g. the total merit index. Published reliabilities of GEBV in some traits show a tendency to be higher than so called ‘observed’ or realized reliabilities (VanRaden, 2009). However, accounting for the approximate nature of both reliability concepts the agreement seems to be satisfying and no general conclusion is drawn from this observation.

Table 1. Overview over some results from routine validation, August 2011. ‘R^2-PA’ is average published reliability of PA, ‘R^2-Gain’ is increase in squared correlation to current EBV when using GEBV instead of PA, ‘pub. R^2’ is average published reliability of GEBV.

Trait	R^2-PA	R^2-Gain	pub. R^2
milk yield	42	21	58
fat yield	38	17	61
protein yield	36	16	53
milk prod. index	38	17	61
Longevity	24	13	51
somatic cell score	37	15	59
daughter preg. rate	23	13	49
pat. calving ease	37	14	56
fitnessindex	30	21	61
beef prod. index	30	24	57
stature	35	20	57
overall udder score	34	14	55
feet and legs score	34	12	52
total merit index	38	24	62

As long as the computational burden is manageable we prefer individual reliabilities for DGV and GEBV over a uniform measure derived from cross-validation procedures. Individual reliabilities in that way might be seen as pure comparative measures of how well the genomic breeding value of a candidate can be estimated from the system. Probably because of a still comparatively small calibration group in Fleckvieh we see a quite large variation in individual reliabilities ranging from 45 in some exceptional cases up to 75 for a trait like milk yield. This point might be particularly important with a growing number of candidates where sires or even paternal grandsires are not part of the calibration group because of no or still insufficient information from daughter performances.

Gains from genomic prediction for Fleckvieh as presented here are in general lower as values reported for Holstein cattle at a comparable size of the calibration group (VanRaden, 2009; Olson *et al.*, 2011). This is potentially an effect of a substantially larger effective population size in Fleckvieh that results in lower linkage disequilibrium exploitable by genomic selection procedures (Pryce *et al.*, 2011). This might also be the reason why procedures like BayesB or the Bayesian Lasso up to now have failed to show superiority in Fleckvieh as already mentioned. Fleckvieh might therefore benefit to a larger extent from higher density chips as is currently reported for Holstein.

References

Edel, C., Emmerling, R. & Götz, K.-U. 2010. A modification of VanRaden’s index for the blending of genomic breeding values. *Proc. 9th World Congr. Genet. Appl. Livest. Prod.*

Garrick, D.J., Taylor, J.F. & Fernando, R.L. 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet. Selec. Evol. 41*, 55.

Gengler, N., Mayeres, P. & Szydlowski, M. 2007. A simple method to approximate gene content in large pedigree populations: application to the myostatin gene in dual-purpose Belgian Blue cattle. *Animal 1,* 21-28.

Gredler, B., Schwarzenbacher, H., Egger-Danner, C., Emmerling, R. & Sölkner, J. 2010. Accuracy of genomic selection in dual purpose Fleckvieh cattle using three types of methods and phenotypes. *Proc. 9th World Congr. Genet. Appl. Livest.Prod.*

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., Aguilar, I., Johnson, D., Legarra, A., Tsuruta, S. & Lawlor, T.J. 2009. A unified approach to utilize phenotypic, full pedigree, and genomic information for a genetic evaluation of Holstein final score. *Interbull Bulletin 40,* 240-244.

Olson, K.M., VanRaden, P.M., Tooker, M.E. & Cooper, T.A. 2011. Differences among methods to validate genomic evaluations for dairy cattle. *J. Dairy Sci. 94,* 2613-2620.

Park, T. & Casella, G. 2008. The Bayesian Lasso. *J. Americ. Stat. Assoc. 103,* 681-686.

Pryce, J.E. Gredler, B., Bolormaa, S., Bowman, P.J., Egger-Danner, C., Fuerst, C., Emmerling, R., Sölkner, J., Goddard, M.E. & Hayes, B.J. 2011. Genomic selection using a multi-breed, across country reference population. *J. Dairy Sci. 94*, 2625-2630.

VanRaden, P.M. & Wiggans, G.R. 1991. Derivation, calculation, and use of national animal model information. *J. Dairy Sci. 74,* 2737-2746.

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

VanRaden, P.M., Van Tassell, C.P., Wiggans, G.R., Sonstegard, T.S., Schnabel, R.D., Taylor, J.F. & Schenkel, F.S. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci. 92,* 16-24.

Wang, J. 2002. An estimator for pairwise relatedness using molecular markers. *Genetics 160,* 1203-1215.