Implementation of single-step evaluations for fitness traits in the German and Austrian Fleckvieh and Brown Swiss populations

J. Himmelbauer, H. Schwarzenbacher, C. Fuerst

ZuchtData EDV-Dienstleistungen GmbH, Dresdner Straße 89/B1/18, 1200 Vienna, Austria

Abstract

The official genomic evaluations for the German-Austrian-Czech Fleckvieh and German-Austrian Brown Swiss populations were implemented in 2011 and since then only bull genotypes were used in the calibration set of the two-step system. In April 2021, genomic breeding values from a routine single-step system for almost all traits in the breeding program were published for the first time. The single-step system includes roughly 300.000 Fleckvieh and 70.000 Brown Swiss genotyped animals, of which for some traits up to 285,000 and 58,000 had direct phenotypes for Fleckvieh and Brown Swiss, respectively. During the process of implementation of single-step for fitness traits, some important points came up that needed to be dealt with. One key aspect is the validation of the estimated breeding values (EBVs). The Linear-Regression-Method published by Legarra and Reverter (2018) shows some considerable advantages, for example to have large validation groups and validation groups of female animals, which are less strictly selected. Other points are scaling and singularity prevention of the genomic relationship matrix (G). Scaling the G to fit to the numerator relationship matrix (NRM) showed positive effects on the validation metrics. The application of singularity prevention methods has shown that there are effects on bias in Mendelian sampling in specific families.

Key words: single-step genetic evaluation, genomic breeding value, Mendelian sampling bias

Introduction

In 2011, the first official genomic EBVs were available for the German-Austrian-Czech Fleckvieh population and the German-Austrian Brown Swiss population (Edel et al., 2011). Since then, a two-step system was used for genomic evaluation where only genotypes of male animals were used in the calibration set. Due to decreasing genotyping-costs and several projects in Germany and Austria, more and more female animals have been genotyped in recent years.

In April 2021, genomic EBVs from a single-step-system were published for the first time for almost all traits in the breeding program for the German-Austrian-Czech Fleckvieh and German-Austrian Brown Swiss populations.

The objective of this work is to give an overview of the most important issues that came up in the process of implementation of the single-step system for fitness traits.

Materials and Methods

Since 2002, the genetic evaluation for the German-Austrian and later also the Czech Fleckvieh cattle population is carried out jointly, whereby the responsibilities for the different trait groups are distributed among the three evaluation centers in Baden-Württemberg, Bavaria and Austria. The same applies to the German-Austrian Brown Swiss population. This paper focuses on the genetic evaluation of fitness traits only, which is carried out by the evaluation center at ZuchtData located in Vienna, Austria.

Data

Genotype data

For Fleckvieh in total 335,335 genotypes were available for breeding value estimation in April 2021, of which 210,222 (62.7 %) were female.

A total of 73,498 genotypes were available for Brown Swiss, 43,062 (58.6 %) of them were from female animals. In Figure 1, you see the number of genotyped animals within each birth year for Fleckvieh (Figure 1a) and for Brown Swiss (Figure 1b).

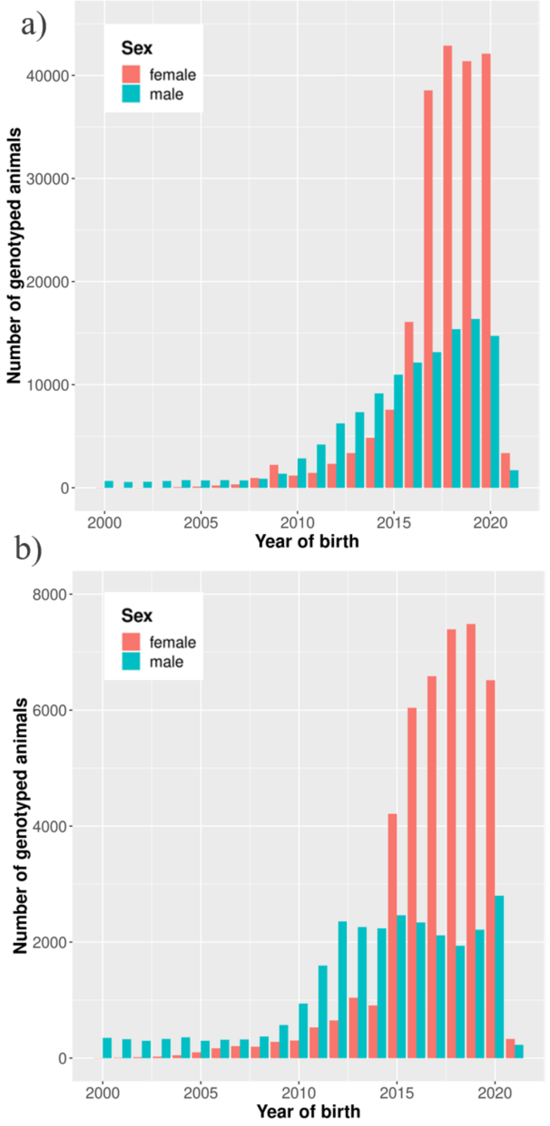

Figure 1. Number of genotyped animals for each birth year (2000-2021) for Fleckvieh (a) and Brown Swiss (b) (April 2021).

Especially in the last few years, due to decreasing genotyping-costs and several herd genotyping projects (FoKUHs in Austria, Braunvieh-Vision, FLEQS and Fleckficient in Baden-Württemberg und Bavaria) large numbers of female animals were genotyped. Currently, on average 10 % of heifers and young cows of the Austrian breeding population are genotyped.

Fitness traits and phenotypic data

The fitness traits for which genomic EBVs are estimated are longevity, fertility, calving ease, rearing losses, mastitis, early fertility disorders and ovarian cysts. The heritabilities of the traits are around 2-3 % and slightly higher for calving ease with around 6 % and longevity with around 11 % (see Table 1).

Table 1. Heritabilities of fitness traits for Fleckvieh and Brown Swiss.

trait	h^2 Fleck-vieh	Brown Swiss
longevity	10.6%	13.1%
fertility index	2.5%	2.7%
calving ease	6.0%	5.4%
rearing losses	2.0%	3.0%
mastitis	2%	3.0%
early fertility disorders	2.1%	2.3%
cystic ovaries	2.1%	1.0%

Table 2 and 3 provide an overview of the number of animals in the training sample for the single-step system for Fleckvieh and Brown Swiss. In contrast to the two-step system, there are no limits in EBV-reliabilities for bulls to enter the calibration set, which means that the number of bulls in the training sample even doubled for some traits (see column 4, table 2 and 3). In columns 5 to 7 of Table 2 and 3 the sums of progeny observations of the bulls in the training sample are shown. The increase from two-step to single-step is here still between 1.5 and 5 % dependent on breed and trait. For the three health traits, mastitis, early fertility disorders and ovarian cysts, there was no genomic evaluation until the implementation of the single-step system, due to the low numbers of proven bulls for these traits. Therefore, there is no comparison possible for these traits. In the single-step system all genotyped animals with own performances contribute to the calibration set. These numbers are shown in Table 2 and 3 in the last column.

Table 2. Number of animals in the calibration set for the single-step system for Fleckvieh compared to the two-step system (April 2021).

Trait	**Bulls**			**∑ Progeny with records of these bulls**			**Genotyped animals with records**
	Single-step	**Two-step**		**Single-step**	**Two-step**		
longevity	21,087	10,579	*(+99.33%)*	7,379,253	7,076,766	*(+4.27%)*	91,926
fertility	21,718	11,593	*(+87.34%)*	8,597,906	8,377,583	*(+2.63%)*	108,560
calving ease	27,534	14,969	*(+83.94%)*	19,839,311	19,546,678	*(+1.50%)*	248,514
rearing losses	27,854	13,222	*(+110.66%)*	23,609,769	23,159,283	*(+1.95%)*	285,173
mastitis	11,122	0	-	9,597,702	0	-	36,669
early fertility disorders	17,488	0	-	3,122,087	0	-	75,986
cystic ovaries	11,090	0	-	945,182	0	-	36,212

Table 3. Number of animals in the calibration set for the single-step system for Brown Swiss compared to the two-step system (April 2021).

Trait	**Bulls**			**∑ progeny with records of these bulls**			**Genotyped animals with records**
	Single-step	**Two-step**		**Single-step**	**Two-step**		
longevity	6,393	4,377	*(+46.06%)*	1,163,211	1,106,256	*(+5.15%)*	23,993
fertility	6,522	3,358	*(+94.22%)*	1,296,342	1,249,090	*(+3.78%)*	26,037
calving ease	7,587	5,641	*(+34.50%)*	2,685,828	2,606,727	*(+3.03%)*	54,359
rearing losses	7,662	3,732	*(+105.31%)*	3,129753	3,029,272	*(+3.32%)*	58,449
mastitis	3,160	0	-	116,390	0	-	14,226
early fertility disorders	5,356	0	-	447,204	0	-	19,334
cystic ovaries	3,138	0	-	114,401	0	-	13,979

For Fleckvieh between 36,000 animals for health traits and up to 285,000 animals for rearing losses are genotyped and have an own performance. For Brown Swiss the numbers are between 14,000 and 58,000, depending on trait.

Processing of G-Matrix

The genomic relationship matrix (G) is computed based on VanRaden's method one G, according to this formula

$$G = \frac{ZZ'}{2\sum p_i q_i}$$

(VanRaden, 2008). For p and q base allele frequencies estimated based on the subset of bulls which have progeny with observations. Inversion of G is done using the **A**lgorithm for **P**roven (core) and **Y**oung (noncore) (APY) (Misztal et al., 2015). The core is defined as all male animals with progeny with observations, which leads to around 30,660 core animals in Fleckvieh and around 12,980 core animals in Brown Swiss. To fit the G to the pedigree-based numerator relationship matrix (NRM) a scaling method based on the following formula is used

$$G_{sc} = \beta G_{VR} + \alpha$$

where G_{sc} denotes the scaled G and G_{VR} the VanRaden's method one G. α and β are the scaling parameters and are derived applying the approach of Christensen et al. (2012) on the core animals. For Fleckvieh the scaling parameters derived by this approach are $\alpha = 0.00461$ and $\beta = 1.01009$ and for Brown Swiss $\alpha = 0.04629$ and $\beta = 1.01675$ (April 2021).

To prevent singularity of the G, 0.001 is added to the diagonal of G.

Validation strategy

For validation and for final scaling of the EBVs (see ***Further aspects***) the **L**inear-**R**egression (LR)-Method published by Legarra and Reverter (2018) is used. This method is based on the comparison of EBVs based on a reduced and a full dataset. For the reduced dataset, four years of phenotype data are removed from the actual dataset. Concerning genotype data, no genotypes are removed for the reduced evaluation. The animals are divided into four validation groups: The first group are all bulls with progeny with performance records in the full and the reduced dataset. The second group are bulls with no progeny performance records in the reduced dataset but progeny with performance records in the full dataset. This group is comparable to the validation group of the GEBV test (Mäntysaari et al., 2010). The third group are non-genotyped cows without performance records in the reduced dataset but performance records in the full dataset. Finally, the fourth group are genotyped cows without performance records in the reduced dataset but performance records in the full dataset. For each validation group bias, dispersion, and correlation is calculated according to Legarra and Reverter (2018).

Further aspects

For most traits singe-step evaluations use direct phenotypes, exceptions are longevity, fertility, calving ease, early fertility disorders and ovarian cysts for Brown Swiss where yield deviations (YDs) were used due to the integration of information from MACE (for more details on the approach of integrating MACE information see Pimentel et al., 2021). Another exception is calving ease for Fleckvieh, where YDs are used due to faster convergence.

To account for genetic distance between subgroups in the base population unknown parents are assigned to unknown parent groups (UPG). UPGs are defined based on the year of birth of the animals with unknown parent(s) and the sex of the unknown parent. To ensure that enough phenotype information for descendants of these groups is available, groups with very low phenotype information are merged. This procedure results in 13-15 UGPs depending on breed and trait.

To avoid systematic bias and overestimation of EBVs for young animals the EBVs are scaled using regression coefficients derived from the LR-validation method for the third and fourth validation group (cows with no phenotype in the reduced but in the full dataset).

The polygenetic variance taken into account is between 10 and 20 %, depending on breed and trait. To account for this polygenetic effect conventional EBVs estimated from pedigree are combined with EBVs from single-step to the final EBVs.

Software

For single-step evaluation the software package MiX99 Release XI/2019 version 19.1129 (MiX99 Development Team, 2019) was used. For all other steps that require special software solutions, other programs of the MiX99 software suite are used. This includes HGINV program, version 0.993 (Stranden and Mantysaari, 2020) for processing of the G, RelaX2 version 1.89 update 8/2019 for pedigree analysis (Stranden and Vuori, 2006) and for reliability estimation ApaX99p release XI/2019 version 19.1129 (MiX99 Development Team, 2019) and SNP_BLUP_REL program Nov 2019, version 0.68 (Stranden and Mantysaari, 2019).

For data manipulation steps mainly R (R Core Team, 2019), awk and bash scripts were used.

Results & Discussion

Validation using LR-method

As an example, the validation metrics for bias, dispersion and correlation are shown for the single-step system and the conventional evaluation based on pedigree in Table 4 and 5. In column 2 also the number of animals in each of the four validation groups are given. With the LR-method it is possible to get validation statistics not only for the small group of highly selected young bulls (1,712 for Fleckvieh and 290 for Brown Swiss) but also for the large groups of moderately selected female animals.

Due to the fact, that the comparison is based on EBVs, the genotype information is considered in the full dataset as a source of information, in contrast to a comparison that uses DYDs.

Especially for fitness traits with quite low heritabilities it is an advantage that no deregression is needed for the validation, as the calculation of DYDs or deregressed proofs is not so reliable for these traits (Edel et al., 2011).

An additional point concerning the validation is that it is possible to compare genetic trends, standard deviations of EBVs and correlations even for the youngest animals, if the full genotype information is used in the evaluation based on the reduced dataset. It has proved helpful to compare the development of, for example, the standard deviations over the years of birth, up to the youngest animals, for the full and the reduced data set and to check them for plausibility.

Scaling of G

The main objective of the applied scaling procedure is to better fit the G to the NRM. The method applied here corresponds to the scaling as suggested by Christensen et al. (2012) with the advantage that it is not necessary to set up the NRM for all genotyped animals. In tables 6 and 7 the statistics of VanRaden's method one G (G_{VR}), the G_{sc} and the NRM are compared.

Table 4. LR-validation metrics for longevity and breed Fleckvieh (April 2021). (Bias is given in genetic standard deviations)

Validation group	N	Bias		Dispersion		Correlation	
		conv.	single-step	conv.	single-step	conv.	single-step
Bulls	1,491	0.057	0.128	0.974	0.966	0.983	0.984
New bulls	1,712	-0.028	-0.113	0.894	0.905	0.486	0.798
New cows (no genotype)	1,441,839	0.010	0.007	0.954	0.919	0.686	0.758
New cows (with genotype)	79,645	0.040	0.034	0.942	1.010	0.630	0.855

Table 5. LR-validation metrics for longevity and breed Brown Swiss (April 2021).

Validation group	N	Bias		Dispersion		Correlation	
		conv.	single-step	conv.	single-step	conv.	single-step
Bulls	273	0.021	0.038	0.993	0.987	0.992	0.993
New bulls	290	0.028	-0.051	0.805	0.956	0.561	0.865
New cows (no genotype)	164,555	0.031	0.014	0.974	0.984	0.814	0.870
New cows (with genotype)	16,835	0.007	0.013	0.928	0.987	0.736	0.914

Add table 4 and 5:

"Bulls"	*...bulls with progeny with phenotype in the full and the reduced dataset*
"New bulls"	*...bulls with no progeny phenotype in the reduced but in the full dataset*
"New cows (no genotype)"	*...not genotyped cows with no phenotype in the reduced but in the full dataset*
"New cows (with genotype)"	*...genotyped cows with no phenotypes in the reduced but in the full dataset*

After scaling, the G is better balanced with the corresponding NRM. Although the scaling parameters α and β are very close to 0 and 1, the scaling has a noticeable effect on the validation statistics.

Table 6. Statistics of G_{VR}, G_{sc} and NRM for Fleckvieh (April 2021).

	G_{VR}	G_{sc}	NRM
Diagonal			
Mean	1.015	1.030	1.024
SD	0.024	0.024	0.014
Minimum	0.942	0.956	1.000
Maximum	1.297	1.315	1.179
Off-diagonal			
Mean	0.009	0.010	0.053
SD	0.025	0.026	0.031
Minimum	-0.073	-0.070	0.000
Maximum	0.825	0.838	0.700

Table 7. Statistics of G_{VR}, G_{sc} and NRM for Brown Swiss (April 2021).

	G_{VR}	G_{sc}	NRM
Diagonal			
Mean	1.024	1.078	1.065
SD	0.031	0.031	0.024
Minimum	0.762	0.812	1.000
Maximum	1.306	1.365	1.277
Off-diagonal			
Mean	0.031	0.046	0.137
SD	0.051	0.072	0.044
Minimum	-0.123	-0.078	0.000
Maximum	0.860	0.921	0.789

In Figure 2, the bias in genetic standard deviations of the estimated breeding values for the different validation groups is shown for the trait longevity and Fleckvieh based on data from December 2020. The results in grey are based on the unscaled G and the results in red are based on the scaled G. Due to scaling the bias was slightly reduced. In Figure 3, the results for the dispersion are presented in the same way, also for trait longevity and breed Fleckvieh as of December 2020. The dispersion slightly improves due to scaling. The validation statistics shown in Figure 2 and 3 are based on single-step EBVs before accounting for polygenetic effects and final scaling, so they are not based on final published EBVs.

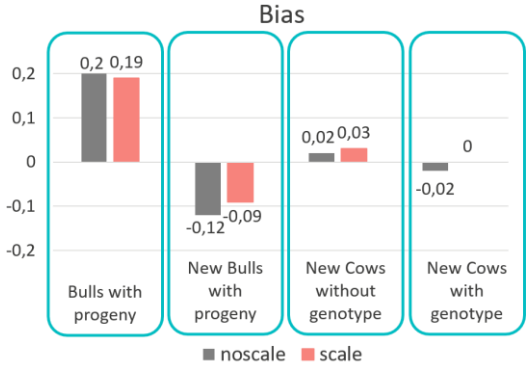

Figure 2. Bias in genetic standard deviations of the four validation groups for longevity and Fleckvieh (December 2020).

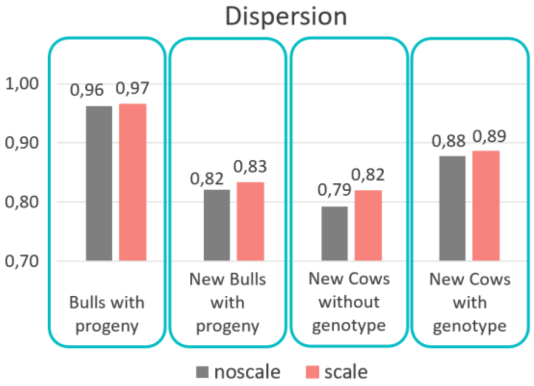

Figure 3. Dispersion of the four validation groups for longevity and Fleckvieh (December 2020).

In a test run, it was shown that using the core animals to derive the scaling parameters α and β seems to be slightly superior than using a random sample of genotyped animals with the same number of animals (results not shown).

Singularity prevention of G

In order to prevent G from singularity and to improve the convergence behavior in the evaluation it is quite common to modify the diagonal of the G, e.g. by adding a small value. We analysed the effects of the singularity prevention in two test runs, where 0.01 and 0.001 was added to the diagonal of G, respectively. The correlation of EBVs from these test runs was higher than 0.99. Nevertheless, for some traits considerable effects on the bias in Mendelian samplings in the families of elite bulls with many genotyped progeny were found. When 0.01 was added to the diagonal of G, genotyped progeny of bulls with many genotyped progeny showed a tendency for positively biased Mendelian samplings, while the sires of these progeny showed a tendency for underestimation of their EBVs. If 0.001 was added, the Mendelian samplings of these families were on average much closer to zero and all extreme outliers disappeared. This is shown in Figures 4 and 5. In Figure 4, the Mendelian sampling for bull families is plotted on the y-axis and in the x-axis the birth year of the bull is plotted for the case where 0.01 was added to the diagonal of G. The Mendelian sampling of non-genotyped progeny is shown on the left side of the plot, for genotyped progeny on the right side of the plot. Each box represents the distribution of Mendelian samplings in bull families, for which the bull was born in a given year. The extreme outliers with a Mendelian sampling deviation of more than 3 EBV-points (with 12 being the standard deviation) are labelled with the number of progeny of the respective sire (N). Figure 5 shows the same plot for adding 0.001 to the diagonal of G. All the outliers in Figure 4 are bull families with more than 1,500 genotyped progeny. Figure 5 shows that there are hardly any extreme outliers when adding only 0.001 to the diagonal of G.

Conclusions

The implementation of Single-Step clearly improves the genomic evaluation for the German-Austrian-Czech Fleckvieh and German-Austrian Brown Swiss populations. We have found that even small scalings to fit the G to the NRM have noticeable positive effects on the validation metrics. Singularity prevention by adding a small value on the diagonal of G can lead to inconsistencies in large bull families, by causing positive biases in Mendelian samplings for bulls with many genotyped progeny. Concerning validation of the single-step system the LR-Method has some advantages over the standard Interbull GEBV-Test, e.g. validation metrics are also available for large female animal groups.

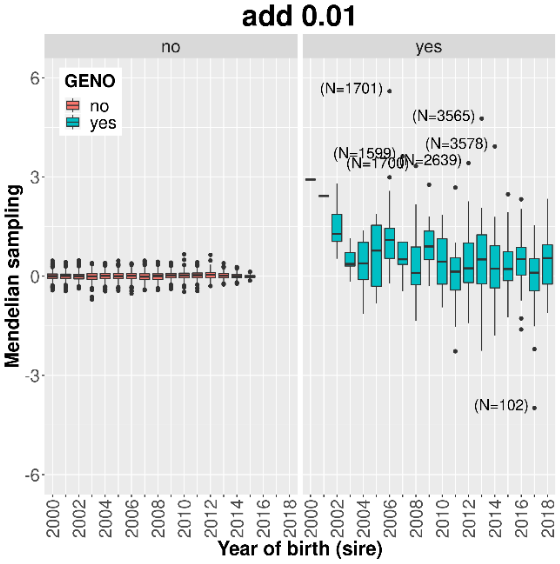

Figure 4. Mendelian sampling of bull families grouped by birth year of the bulls for adding 0.01 to the diagonal of G. The Mendelian sampling deviation is given in EBV-points, with 12 being the standard deviation. Non-genotyped progeny on the left and genotyped progeny on the right side. Outliers are labeled with number of progeny of the bull family. (trait: longevity, breed: Fleckvieh, December 2020)

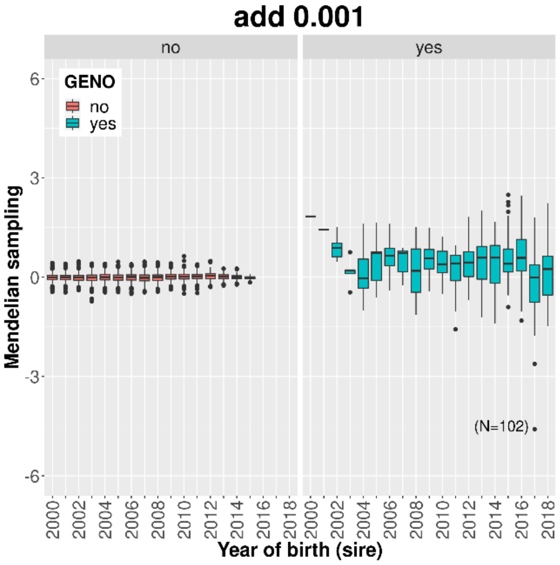

Figure 5. Mendelian sampling of bull families grouped by birth year of the bulls for adding 0.001 to the diagonal of G. The Mendelian sampling deviation is given in EBV-points, with 12 being the standard deviation. Non-genotyped progeny on the left and genotyped progeny on the right side. Outliers are labeled with number of progeny of the bull family. (trait: longevity, breed: Fleckvieh, December 2020)

Acknowledgments

We gratefully acknowledge the Federal Ministry of Agriculture, Regions and Tourism in Austria for funding the project „FoKUHs".

References

Christensen, O.F., Madsen, P., Nielsen, B., Ostersen, T., Su, G., 2012. Single-step methods for genomic evaluation in pigs. Animal 6: 10, 1565-1571. URL https://www.sciencedirect.com/science/article/pii/S1751731112000742?via%3Dihub.

Edel, C., Schwarzenbacher, H., Hamann, H., S. Neuner, S., Götz, K.U., Emmerling, R., 2011. The German-Austrian genomic evaluation system for Fleckvieh (Simmental) cattle. Interbull Bulletin 44, 152–156. URL https://journal.interbull.org/index.php/ib/article/view/1209.

Legarra, A., Reverter, A., 2018. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. Genetics Selection Evolution 50:53.

Mäntysaari, E., Liu, Z., VanRaden, P., 2010. Interbull validation test for genomic evaluations. Interbull Bulletin No 41, 17-22. URL https://journal.interbull.org/index.php/ib/article/view/1134/1125.

Misztal, I., Fragomeni, B.O., Lourenco, D.A., Tsuruta, S., Masuda, Y., Aguilar, I., Legarra, A., and Lawlor, T.J., 2015. Efficient inversion of genomic relationship matrix by the algorithm for proven and young (APY). Interbull Bulletin 49, 111–116. URL https://journal.interbull.org/index.php/ib/article/view/1387.

MiX99 Development Team, 2019. MiX99: A software package for solving large mixed model equations. Release XI/2019. Natural Resources Institute Finland (Luke). Jokioinen, Finland. URL: http://www.luke.fi/mix99.

Pimentel, E., Edel, C., Krogmeier, D., Emmerling, R., Götz, K.U., 2021. Effects of use of external information in Single-Step evaluations for linear type traits in Brown Swiss. Interbull Bulletin 56.

R Core Team, 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Stranden, I., Vuori, K., 2006. RelaX2: pedigree analysis program. Proceedings of the 8th World Congr. Genet. Appl. Livest. Prod., Aug 13-18, Belo Horizonte, MG, Brazil, Communication 27-30. URL https://www.cabdirect.org/cabdirect/abstract/20063170112.

Stranden, I., Mäntysaari, E., 2019. SNP_BLUP_REL program, Nov 2019, version 0.68. Natural Resources Institute Finland (Luke).

Stranden, I., Mäntysaari, E., 2020. HGINV program, Nov 2020, version 0.993. Natural Resources Institute Finland (Luke).

VanRaden P.M., 2008. Efficient Methods to Compute Genomic Predictions. Journal of Dairy Science, 91, 4414–4423. URL https://www.sciencedirect.com/science/article/pii/S0022030208709901.