Using a single-step SNP BLUP maternal-effect model for calving trait genomic evaluation in German Holsteins

H. Alkhoder, Z. Liu, D. Segelke, and R. Reents
*IT Solutions for Animal Production (**vit**), Heinrich-Schroeder-Weg 1, D-27283 Verden, Germany*

Abstract

A multi-parity maternal-effect linear animal model has been used for conventional evaluation of calving traits in Germany. Bull EBV of direct and maternal effects of calving ease and stillbirth are submitted to Interbull bull MACE evaluation that uses a single-trait single-effect model. For a single-step evaluation of the calving traits, we applied a single-step SNP BLUP model to the national calving data integrated with foreign pseudo-phenotype data of bulls. We used phenotype, genotype and pedigree data from German official genomic evaluation in August 2021. About 25 million national calving records and deregressed EBV of c.a. 113,000 bulls with foreign data were evaluated. The number of calving cows or calves with national calving data exceeded 31 millions. A total of 1,003,041 genotyped animals were considered in the single-step evaluation without any approximation of genomic information. A genomic validation was performed for the single-step and the current multi-step model. SNP effects from the single-step model were shown to have less bias and were more correlated between the truncated and the full evaluation than those from the multi-step model. Using genomic EBV of validation bulls from the full and truncated evaluations, we conducted a linear regression validation test and obtained less satisfactory validation results for all the calving traits than the test-day or conformation traits. More inflation was seen in the direct than the maternal genetic effects, in particular the stillbirth direct effect with largest overestimation. The overestimation seemed to be more evident for the younger than the older validation bulls. Increasing the residual polygenic variance improved little in reducing the prediction inflation. A special bull reference population for the calving traits failed to markedly reduce the overestimation either. However, removing genotype data of older bulls, born before 2005, resulted in favorable validation results, though to a limited degree. Post-processing genomic EBV of young candidates seemed to be unavoidable for a routine implementation of the single-step model for the calving traits.

Key words: single-step SNP BLUP model, maternal-effect model, calving traits, dairy cattle

Introduction

A single-step SNP BLUP model (ssSNPBLUP, Liu et al. 2014) directly estimates the effects of SNP markers together with all other effects of the single-step model (SSM). The ssSNPBLUP model was successfully applied to conformation traits of German Holsteins (Alkhoder et al. 2022) and promising validation results were obtained for all the conformation traits. In addition, we also positively validated the ssSNPBLUP model for milk production traits and somatic cell scores analyzed with a multiple lactation random regression test-day model (Alkhoder et al. 2022a).

Since 2012, a multi-parity maternal-effect linear animal model has been applied to calving ease (CE) and stillbirth (SB) data for conventional calving evaluation in Germany (Liu et al. 2012). Direct genetic effect of a calf and maternal genetic effect of a calving cow are treated as correlated traits for either of the two calving traits in each of first three parities. Official breeding values of the calving traits, defined as CE and SB in first parity, have been submitted to the Interbull bull evaluation of the calving traits. A multi-step genomic model (MSM) based on deregressed EBV from the conventional national and bull MACE evaluations have been used for routine genomic evaluation in Germany Holsteins. In 2019 the multi-step genomic evaluation was upgraded to using a mixed cow and bull reference population (Liu et al. 2019) from the previous international bull reference population.

The objectives of this study were 1) to implement the ssSNPBLUP maternal-effect linear animal model to genomic evaluation of the calving traits, 2) to assess the accuracy and bias of the single-step evaluation via a genomic validation, and 3) to compare predicative ability of the single-step model SSM to the current multi-step genomic model MSM.

Materials and Methods

A single-step SNP BLUP maternal-effect linear animal model for national calving data

A SSM was applied to calving records of a cow in first three parities:

$$\mathbf{y} = \mathbf{X}_{\mathrm{h}}\mathbf{h} + \mathbf{X}_1\mathbf{f}_1 + \mathbf{X}_2\mathbf{f}_2 + \mathbf{u}_{\mathrm{d}} + \mathbf{u}_{\mathrm{m}} + \mathbf{e} \quad [1]$$

where **y** is a vector of Snell scores (Snell, 1964) for both traits CE and SB in first three parities of the calving cow, **h** is a vector of fixed effects of herd-year associated with the calving records, $\mathbf{f}_1$ is a vector of fixed effects of region-year-season of calving by calf sex, $\mathbf{f}_2$ is a vector of fixed effects of year-month-age of calving by calf sex, $\mathbf{u}_{\mathrm{d}}$ is a vector of direct genetic effects of calves of the calving cow, $\mathbf{u}_{\mathrm{m}}$ is a vector of maternal genetic effects in first three parities of the calving cow, **e** is a vector of random error effects. $\mathbf{X}_{\mathrm{h}}$, $\mathbf{X}_1$, and $\mathbf{X}_2$ are the incidence matrix for respective fixed effects.

Direct and maternal genetic effects were assumed to be correlated, though at a rather low level (Liu et al. 2012). There were a total 12 breeding values per animal: 2 calving traits x 3 parities x 2 genetic effects. Prior to the estimation of the model effects, original CE scores with four categories or SB scores with two categories were transformed to Snell scores within each level of the fixed effect ($\mathbf{f}_1$) of region-year-season of calving by calf sex.

If the calving cow was genotyped, her genomic estimated breeding values (GEBV) for the maternal genetic effects were decomposed into direct genomic values (DGV) and residual polygenic effects (RPG) according to the ssSNPBLUP model (Liu et al. 2014):

$$\mathbf{u}_{\mathrm{m}} = \mathbf{Z}_{\mathrm{m}}\mathbf{g}_{\mathrm{m}} + \mathbf{a}_{\mathrm{m}} \quad [2]$$

where $\mathbf{g}_{\mathrm{m}}$ is a vector of SNP effects for the maternal genetic effects for both calving traits in first three parities, $\mathbf{Z}_{\mathrm{m}}$ is the design matrix containing genotype data of the cow and $\mathbf{a}_{\mathrm{m}}$ is a vector of RPG effects for the two traits in the parities of the cow.

If one of the calves of the calving cow was genotyped, its GEBV of direct genetic effects consisted of the two components DGV and RPG:

$$\mathbf{u}_{\mathrm{d}} = \mathbf{Z}_{\mathrm{d}}\mathbf{g}_{\mathrm{d}} + \mathbf{a}_{\mathrm{d}} \quad [3]$$

where $\mathbf{g}_{\mathrm{d}}$ is a vector of SNP effects for the direct genetic effects for both calving traits in first three parities, $\mathbf{Z}_{\mathrm{d}}$ is the design matrix containing genotype data of the calf and $\mathbf{a}_{\mathrm{d}}$ is a vector of RPG effects for the traits in the parities for the calf.

A single-trait single-effect model for foreign phenotype of bulls

In contrast to the 2-effect model [1] for the national calving data, Interbull MACE evaluation applied a single-effect model for conventional MACE evaluation, in which direct and maternal genetic effects of the traits CE and SB were analyzed in separate MACE evaluations. German calving EBV for direct and maternal genetic effects of CE and SB in first parity were submitted to the four separate MACE evaluations and results of the four MACE evaluation were subsequently received back on the German country scale. An iterative matrix deregression method (Liu, 2011) was implemented to deregress MACE EBV of all bulls for each effect of either trait separately. Deregressed EBV (DRP) from the four MACE evaluation were used as pseudo-phenotype data for the bulls included in the MACE evaluation. A single-trait ssSNPBLUP model was applied to MACE DRP of the two genetic effects separately:

$$\boldsymbol{y}_{\mathbf{d}} = \boldsymbol{\mu}_{\mathbf{d}} + \mathbf{z}'\mathbf{g}_{\mathbf{d\_M}} + \boldsymbol{a}_{\mathbf{d\_M}} + \boldsymbol{e}_{\mathbf{d}} \qquad [4]$$

$$\boldsymbol{y}_{\mathbf{m}} = \boldsymbol{\mu}_{\mathbf{m}} + \mathbf{z}'\mathbf{g}_{\mathbf{m\_M}} + \boldsymbol{a}_{\mathbf{m\_M}} + \boldsymbol{e}_{\mathbf{m}} \qquad [5]$$

where $\mathbf{y}_{\mathbf{d}}$ and $\mathbf{y}_{\mathbf{m}}$ represent DRP of the bull for direct and maternal effect in either MACE trait, $\mathbf{g}_{\mathbf{d\_M}}$ and $\mathbf{g}_{\mathbf{m\_M}}$ are SNP effects for the direct and maternal genetic effects of the MACE trait respectively, $\boldsymbol{a}_{\mathbf{d\_M}}$ and $\boldsymbol{a}_{\mathbf{m\_M}}$ are the RPG of direct and maternal effects of the bull, and $\boldsymbol{e}_{\mathbf{d}}$ and $\boldsymbol{e}_{\mathbf{m}}$ are the corresponding residual effects with

$$\boldsymbol{var}(\boldsymbol{e}_{\mathbf{d}}) = \boldsymbol{\sigma}^{\mathbf{2}}_{\mathbf{e\_d}}/\boldsymbol{n}_{\mathbf{d}} \qquad [6]$$

$$\boldsymbol{var}(\boldsymbol{e}_{\mathbf{m}}) = \boldsymbol{\sigma}^{\mathbf{2}}_{\mathbf{e\_m}}/\boldsymbol{n}_{\mathbf{m}} \qquad [7]$$

where $\boldsymbol{\sigma}^{\mathbf{2}}_{\boldsymbol{e\_d}}$ and $\boldsymbol{\sigma}^{\mathbf{2}}_{\boldsymbol{e\_m}}$ are error variances of direct and maternal effects of the MACE trait, respectively, $\boldsymbol{n}_{\mathbf{d}}$ and $\boldsymbol{n}_{\mathbf{m}}$ are effective progeny contribution or effective daughter contribution (EDC) of the bull expressed on an animal-model basis.

Bull EDC ($\boldsymbol{n}_{\boldsymbol{NAT}}$) of national evaluation was computed using the Interbull standardized method and the procedure by Liu (2011) was applied to calculate MACE EDC ($\boldsymbol{n}_{\boldsymbol{MACE}}$) for all bulls. For each of the four MACE traits/effects, let y_{NAT} and y_{MACE} represent deregressed EBV of a bull from the national and MACE evaluation, respectively, with corresponding national EDC n_{NAT} and MACE EDC n_{MACE}. Pseudo-phenotype of this bull for model [4] or [5] was adjusted for the contribution of national progeny or daughters with:

$$\boldsymbol{y} = (\boldsymbol{n}_{\boldsymbol{MACE}}\boldsymbol{y}_{\boldsymbol{MACE}} - \boldsymbol{n}_{\boldsymbol{NAT}}\boldsymbol{y}_{\boldsymbol{NAT}}) \,/\, (\boldsymbol{n}_{\boldsymbol{MACE}} - \boldsymbol{n}_{\boldsymbol{NAT}}) \qquad [8]$$

with its corresponding weight changed to:

$$\boldsymbol{n} = \boldsymbol{n}_{\boldsymbol{MACE}} - \boldsymbol{n}_{\boldsymbol{NAT}}\ . \qquad [9]$$

If the bull had no daughters or progeny outside Germany, then his MACE data would not need to be integrated in to the national single-step evaluation. In case that this bull had no domestic daughters or progeny at all, his deregressed MACE EBV y_{MACE} and n_{MACE} MACE EDC would be used directly for the integration without the adjustments.

For the 12 national genetic effects as well as four traits/effects from the MACE data, a 30% of additive genetic variance was assumed to be contributed by the RPG effects. The SSM model [1] with integrated four MACE traits had a total of 16 genetic effects per animal.

Differences in modelling of the genetic effects between national and MACE evaluations

The model [1] for the national calving data contains two correlated genetic effects: direct and maternal effects. In contrast, the MACE data were modelled using a single-effect model: model [4] for direct genetic effect and model [5] for maternal genetic effect. The difference in modelling the national and international MACE calving data originated from the conventional bull MACE evaluation that analyzed separately the direct and maternal national DRP of bulls with a single-effect model. The different modelling of genetic effects in national and MACE evaluation complicated the integration of foreign pseudo-phenotype data DRP of bulls into the national animal model evaluation. The German MACE trait, defined as direct or maternal effect in first parity, should have a unity genetic correlation with the same effect in first parity in national evaluation. To avoid numerical instability in the SSM evaluation with the integrated MACE data, a multiplication factor of 0.97 was introduced to decrease the genetic correlations between the MACE trait and all the national traits. Our procedure of integrating the foreign MACE data of bulls guaranteed the use of pseudo-phenotype data of all the four MACE traits. In other words, we did not discard MACE data of either direct or maternal effect, and we did not model the single-effect MACE DRP of bulls with the 2-effect national calving model.

In comparison to other trait groups such as conformation traits (Alkhoder et al. 2022) and the test-day traits (Alkhoder et al. 2022a), calving traits showed largest differences between the current MSM and SSM models. As shown in model [1], SSM estimated SNP effects for the direct and maternal effects of the two correlated traits jointly, whereas MSM evaluated them separately for either effect of each trait. The SSM model [1] had identical modelling as the conventional evaluation (Liu et al. 2012) for the national calving data, i.e. treating CE and SB in first three parities as genetically distinct traits. However, the MSM model considered only the first parity of either trait at a time. For the calving traits, direct genetic effects were measured two years earlier than the maternal genetic effects. In addition, there was only a single birth record for the calf, but the calving cow may have up to three calving records.

Phenotype, genotype and pedigree data

Phenotype, genotype and pedigree data were obtained from the official August 2021 evaluation for Germany dairy breeds. Calving data from year 2000 were included in the German routine conventional evaluation. A total of 25,379,991 calving records from 13,194,532 calving cows of the breeds, Holsteins, Red Dairy Cattle and Jersey, were analyzed together with 112,076 MACE Holstein bulls that had daughters or progeny outside Germany. The total number of national cows or calves with calving or birth data and integrated bulls with foreign daughters or progeny amounted to 31,279,129. All genotyped Holstein animals, including culled animals, were jointly evaluated with those animals with phenotype data, and the number of genotyped animals was 1,003,041 for the August 2021 evaluation. A maximum number of 20 generations was used to trace ancestors of the genotyped or phenotyped animals. The oldest bulls with daughters or cows with records were guaranteed to have at least three generations of ancestors. The pedigree file for the single-step evaluation contained 38,150,805 animals and 90 phantom parent groups that were defined according to breed, country of origin, four selection paths and birth year of animals with missing parents. The total number of estimated effects of the ssSNPBLUP model for the calving data was 615,8484,330, which represented the largest genetic evaluation system among all trait groups in Germany.

A genomic validation for the single-step SNP BLUP model

We followed the rules of Interbull trend validation test III and GEBV test (Mäntysaari et al. 2010) to assess the predictive ability of the ssSNPBLUP model for the calving traits. Last four years of calving records of the national cows were removed to simulate a genomic evaluation four years ago. Due to a lack of MACE EBV from a truncated MACE evaluation, bull MACE data from August 2021 were used for the validation study. Because of the two different genetic effects, validation bulls were defined separately as: youngest bulls born in 2015 through 2018 with progeny for the direct genetic effects, and youngest bulls born in 2013 through 2016 for the maternal genetic effects. The validation bulls must have daughters or progeny in at least 10 herds in Germany with a minimum EDC of 20. The youngest four birth years of the bulls were deleted from the MACE data, separately for the two genetic effects. As an extra step, national daughters or progeny of the validation bulls were removed from the national calving data, if there were any daughters or progeny still left in the truncated calving data set.

In contrast to the phenotyped animals, all genotyped animals from the full evaluation were also included in the truncated genomic evaluation. Having removed the data for the validation, 20,910,798 calving records from 26,020,283 cows or calves and 105,343 bulls with foreign data remained in the phenotype data for the validation study.

Table 1 shows the numbers of national calves or calving cows with calving records and bulls with integrated MACE data for the full and truncated single-step evaluations. Additionally, the numbers of genotyped animals with phenotype data were also given for the two genomic evaluations.

Table 1. Description of the calving phenotype and genotype data sets for a full and a truncated evaluation in German dairy breeds

National data	Full data set	Truncated set
Calving cows	13,194,532	11,036,077 (-2,158,455)
genotyped	339,215	85,843 (-253,372)
Calves	25,380,001	20,910,798 (-4,469,203)
genotyped	519,321	178,256 (-341,065)
Bulls with MACE data[§]		
Direct effect	104,545	97,834 (-6,711)
genotyped	31,810	26,510 (-5,300)
Maternal effect	77,530	71,334 (-6,196)
genotyped	29,376	24,398 (-4,978)

[§] Bulls must have daughters or progeny outside Germany

The current single-trait single-effect multi-step genomic model

To compare the accuracy and bias of the SSM model with the current multi-step genomic model, genomic evaluations under the MSM model were also performed using a full and truncated data set. As stated above, a mixed bull and cow reference population was introduced in 2019 to routine genomic prediction in German Holsteins (Liu et al. 2019). A single-trait single-effect SNP BLUP model was used for estimating the effects of SNP markers for the direct and maternal effects of the two calving traits CE and SB. It was assumed that RPG explained 30% genetic variance for all the calving traits, except stillbirth maternal with 20%.

Unlike the single-step model, phenotype, genotype and pedigree data for the multi-step evaluations were taken from the official German Holstein evaluation in April 2021. The mixed reference population included 296,897 or 203,439 cows for maternal effect or direct effect, respectively, for trait SB. The number of reference bulls was 37,285 for maternal effect and 35,780 for direct effect for trait SB.

Following the routine procedure of genomic validation for the MSM model, current evaluation of April 2021 of the reference cows or bulls were used, instead of a conventional evaluation corresponding to four years ago. Due to the relative short history of female genotyping in German Holsteins, the youngest three birth years of national bulls, born in 2014 through 2016, were selected as validation bulls. A total of 991 Holstein validation bulls were chosen with daughters in at least 10 herds in Germany. Regarding the genotyped and phenotyped cows, we removed reference cows of the youngest two birth years for the validation study. To ensure no daughters of the validation bulls remained in the truncated data set, all daughters of the validation bulls were deleted at an extra step from the truncated reference population as well.

In comparison to the validation study of the single-step model above, the genomic validation of the MSM model did not re-estimate breeding values of the reference cows using the truncated calving data set. It is important to keep in mind that the truncated data differed between the validation studies of the two models SSM and MSM.

Results & Discussion

Diverse test runs of the ssSNPBLUP calving evaluation were performed with the software package MiX99 (Strandén and Lidauer, 1999), in which the ssSNPBLUP by Liu et al. (2014) was implemented in a special way (Mäntysaari, personal communication). Furthermore, we applied the software package MiXBLUP (Ten Napel et al. 2020) to the same calving and pedigree data. Per round of iteration, both software packages required comparable CPU time. MiXBLUP needed a maximum RAM of 140Gb for the full evaluation, and MiX99 required 378Gb for the same data set. We analyzed only the solutions from MiX99, because identical effect estimates were obtained with the two software packages.
For the single-step evaluation of the calving traits with the software MiX99 (Strandén and Lidauer, 1999), a total clock time per round of iteration was for the full data set 2.12 minutes on a Linux server with 2x24 cores of Intel® Xeon® Gold 6252 CPU @2.10GHz. Reasonably accurate solutions could be obtained with approximately 5000 rounds of iteration.

SNP effect estimates

The Liu-Goddard ssSNPBLUP model estimated SNP effects for the 12 national calving effects and 4 MACE traits. Estimates of the SNP markers for direct or maternal effects of the two traits, as defined as the official published breeding values were chosen for the further analyses. Table 2 shows observed correlations of the SNP effect estimates between the full and truncated evaluations. The SNP effect correlation ranged from 0.90 to 0.93 for the SSM model, a little lower than those of the MSM model, which can be explained by the data difference between the full and truncated evaluations for the two models. For the evaluation with the full data set, SNP effect estimates of the two genomic models were correlated from 0.76 for SB direct (SBd) effect to 0.85 CE direct effect (CEd).

We regressed the SNP effect estimates from the full evaluation on the truncated evaluation and presented the regression coefficients in Table 3. It can be clearly seen that the SSM model has regression coefficient being close to 1, whereas a little inflation in SNP effect variance is indicated for the MSM model by the slightly lower regression coefficients.

Table 2. Correlations of SNP effect estimates between the full and truncated evaluations for the calving traits

	Both models (full data set)	Single-step model	Multi-step model
Calving ease in first parity			
Direct	0.85	0.93	0.94
Maternal	0.82	0.92	0.94
Stillbirth in first parity			
Direct	0.76	0.90	0.94
Maternal	0.83	0.93	0.95

Table 3. Regression coefficients of SNP effect estimates of the full on the truncated evaluation

Genomic model	Single step	Multi-step
Calving ease, Direct	0.97	0.95
Maternal	0.99	0.93
Stillbirth, Direct	0.97	0.94
Maternal	0.98	0.94

Correlation and dispersion of genomic prediction

For the validation bulls, GEBV from the early, truncated genomic evaluation, as young candidates, were compared to GEBV from the later, full genomic evaluation as sires of numerous daughters or calves. The correlation of the GEBV from the two genomic evaluations indicates the accuracy of genomic prediction. The regression coefficient of GEBV from the later, full evaluation on the early, truncated evaluation (Legarra and Reverte, 2018) reveals the prediction dispersion of the genomic models. Table 4 shows the results of the linear regression of

GEBV between the two evaluations for the two genomic models. The model R^2 value or squared correlation of GEBV between the full and truncated evaluation suggest higher consistency or prediction accuracy of GEBV of the SSM than MSM model. The current MSM model does not utilize information such as phenotype data of non-genotyped relatives, therefore, the lower R^2 values than the SSM model are indeed expected. Among the four traits/effects, stillbirth direct SBd has significantly lower R^2 value than the other three traits or effects for either of the two genomic models. This may be argued by the fact that no dead calves were genotyped, and the genotyped calves or cows did not represent a non-selective reference population. However, inclusion of the genotype data of reference bulls can reduce the impact of the selective genotyping of the calves on the genomic prediction of the trait SBd. Because the truncated evaluation and the full evaluation were not expressed on the same cow base population, estimates of the regression intercept (b_0) were difficult to interpret and are thus not presented in Table 4. The regression slope b_1 indicates whether a genomic prediction in inflated, if $b_1 < 1$ or underestimated, if $b_1 > 1$. From Table 4 we can clearly see a significant inflation of the genomic prediction by the SSM model for all the four calving traits. GEBV of maternal effects were less inflated than the direct genetic effects, with SBd effect being most overestimated $b_1 = 0.70$. Due to a post-processing adjustment in the current MSM model, regression coefficients are close to 1, suggesting unbiased genomic prediction in the current routine genomic evaluation. It seems a posterior adjustment may be required for the SSM evaluation of the calving traits.

We took a deeper look at the regression coefficient b_1 of the SSM model for validation bulls born in different years. Figure 1 shows the validation regression coefficients stratified by birth years of the validation bulls.

Table 4. Linear regression of GEBV of the full on the truncated evaluation of validation bulls

	No. of validation bulls	Slope b_1	Model R^2 value
Single-step genomic model			
Calving ease, Direct	1611	0.79	0.71
Maternal	1710	0.89	0.75
Stillbirth, Direct	1618	0.70	0.57
Maternal	1662	0.90	0.79
Multi-step genomic model			
Calving ease, Direct	1961	1.04	0.54
Maternal	955	0.99	0.49
Stillbirth, Direct	1970	1.02	0.38
Maternal	935	0.91	0.55

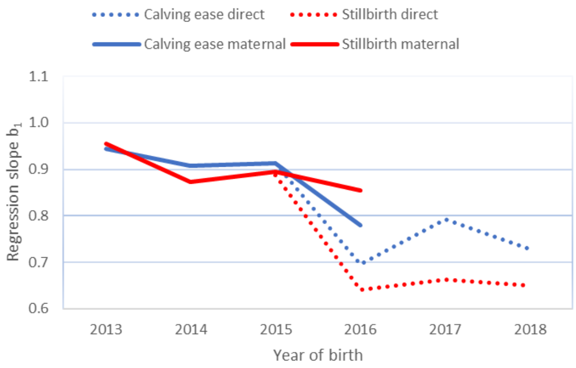

Figure 1. Regression coefficients of genomic validation stratified by birth years of validation bulls for the single-step model

It can be seen in Figure 1 that the regression slopes for the maternal genetic effects are about 0.9, except for the youngest birth year of the validation bulls having clearly lower b_1 value. In comparison to the maternal effects, regression slope b_1 of the direct genetic effects was 0.9 only for the oldest birth year of the validation bulls and the three younger birth years have much smaller b_1 value. This might be hypothesized that most reference animals, i.e., the calving cows, have phenotype data up to three parities for the estimation of maternal genetic effects, whereas the data for the direct genetic effects, birth information of the calves, are limited to one year. In general, we observe a lower level of b_1 values, i.e., more inflation of GEBV, for the calving traits than the test-day traits (Alkhoder et al. 2022a) and the conformation traits (Alkhoder et al. 2022).

Genetic trends of direct and maternal effects of the calving traits

Genotyped AI bulls belonged to a highly selected group of animals and usually received most attention by breeding organizations. Figure 2 shows genetic trends in genotyped AI bulls for calving ease direct effect CEd. The numbers of AI bulls by birth year are identical to those reported in a previous study by Alkhoder et al. (2022), with only the youngest birth year added. GEBV of the AI bulls are expressed as relative breeding values (RBV) and standardized using the genetic standard deviation of RBV. Higher value of RBV means biologically favorable. We can see in Figure 2 that the genetic trends of the two models are essentially identical before the implementation of genomic selection in German Holsteins in 2009. The SSM model, solid black line MU_CEd_SS for the full evaluation or dashed black line MU_CEd_SS-val for the validation evaluation, has a higher genetic trend than the corresponding MSM model, solid red line MU_CEd_MS for the full and dashed red line MU_CEd_MS-val for the validation, for younger AI bulls born in 2010 and after. GEBV difference between the full and truncated evaluation seems to be marginal for either of the genomic models.

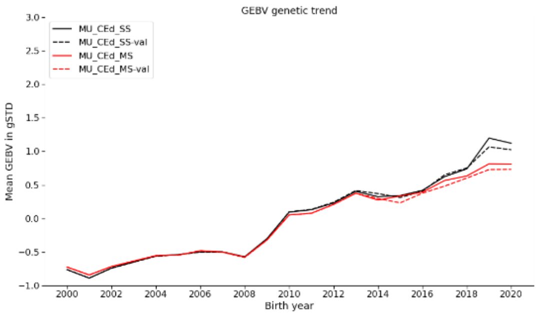

Figure 2. Genetic trends in genotyped German AI bulls for direct effect of calving ease of the genomic models with two data sets

In analog to the trends in the direct effect CEd, genetic trend of maternal effect of stillbirth (SBm) in the AI bulls is shown in Figure 3. Similar genetic trends are observed for the two genomic models with slightly higher trend of the SSM. Genetic trend from the truncated evaluation is a little greater than from the full evaluation for the two genomic models.

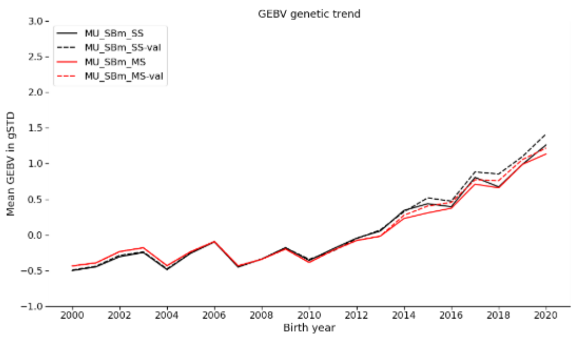

Figure 3. Genetic trends in genotyped German AI bulls for maternal stillbirth effect of the two genomic models with two data sets

Correlations of GEBV between the truncated and full evaluations

GEBV of the genotyped Holstein AI bulls were correlated between the full and truncated evaluations under either of the genomic models. Figure 4 shows the GEBV correlations by birth year of the bulls for trait calving ease maternal (CEm). The within-model GEBV correlations are, on average, higher for the SSM (blue line, CEm_SS_SS-val) than for the MSM model (green line, CEm_MS_MS-val), particularly for the transition bulls, born in 2013 through 2016, that had daughters in the full evaluation but were young candidates in the truncated validation evaluation. For youngest AI bulls born in 2019 and 2020, both genomic models show a GEBV correlation about 0.94 on average. Between the two genomic models, GEBV correlations of the AI bulls (orange line, CEm_SS_MS) in different birth years vary a little around 0.9.

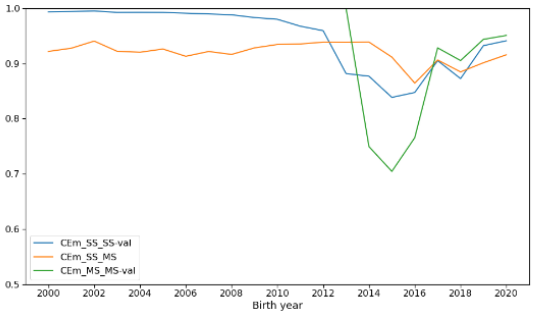

Figure 4. GEBV correlations of the genotyped Holstein AI bulls between the truncated and full evaluation for calving ease maternal effect

Regressions of GEBV of the full on the truncated evaluation

Because of the whole-herd genotyping of female calves in German Holsteins, genotyped female animals, with or without own phenotype data, represented an unselected sample of animals. GEBV of the female animals from the full evaluation were regressed on GEBV from the truncated evaluation. Regression coefficients/slopes within birth years of the female animals are shown in Figure 5.

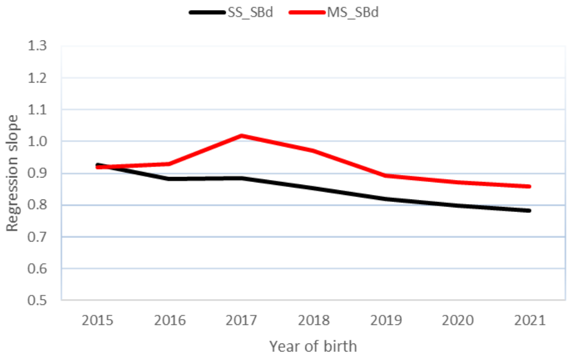

Figure 5. Regression coefficients of GEBV of the genotyped female animals from the full on the truncated evaluation for the two models

The regression coefficients of the SSM for trait SBd (black line, SS_SBd) are between 0.8 and 0.9 with lower value for youngest female candidates. In comparison, regression coefficients for the MSM model (red line, MS_SBd) are higher thanks to a post-processing procedure implemented in the current multi-step genomic evaluation that reduced the inflation of genomic prediction.

Impact of a higher residual polygenic variance

The lower regression coefficients of the genomic validation of the SSM in Table 4 indicated an overestimation of GEBV for young animals. To reduce the inflation of GEBV of the SSM, a special test run was conducted, in which the proportion of residual polygenic variance was increased from 30% to 40%. Little improvement in regression coefficient b_1 or model R^2 value were found using the higher RPG variance than the default scenario of 30%. GEBV difference between the two scenarios were marginal for all the animals.

Impact of a bull reference population

Like in most countries, dead calves were not genotyped routinely in Germany. Because the most reference animals were genotyped female animals that were born alive, a question arose if the selective nature of phenotype stillbirth data of the genotyped female animals might bias the genomic prediction of stillbirth direct SBd. Unlike the female animals, bulls with progeny in Germany or abroad had both dead and live calves, we would like to know if a bull reference population could improve the genomic prediction of stillbirth. Genotype data of all female animals with phenotype data, in total 517,704, were deleted, so that only genotyped bulls with progeny contributed to the SNP effect estimation. Genotype data of the other genotyped female animals, e.g., female candidates, were kept in this test evaluation. In addition, genotype records of all male animals were kept as well. The same RPG variance of 30% was assumed as the default scenario. The total number of genotyped animals remained in the full and truncated evaluation was 485,416. In comparison to the default scenario of 30%

RPG variance, model R^2 value of the genomic validation, based on the validation bulls, decreased by 0.01 for all traits, except SBd with an increase of 0.01. The regression slopes differed also only marginally with a change of 0.01. For stillbirth direct SBd, the b_1 value increased from 0.70 to 0.71. In summary, the impact of changing to the bull reference population on genomic prediction was small.

Impact of removing genotype data of older bulls

At the beginning of genomic selection, many inferior progeny-tested bulls were not genotyped due to a lack of semen available. The involuntary selective genotyping was shown to have a negative impact on genomic prediction in protein yield in German Holsteins (Alkhoder et al. 2022b). Genotyped data of all bulls born before 2005 were removed in this special test evaluation. This represented the only difference to the default scenario. Genomic validation results were compared to the default scenario to quantify the impact of the removal of genotype data of the older bulls. A decrease in model R^2 value was observed, ranging from -0.02 for SBd to -0.04 for CEd. However, an improvement in reducing genomic inflation was seen in an increase of b_1 value: +0.02 for CEm and SBd. For the other two traits CEd and SBm, equal b_1 value was obtained in this test scenario as the default one. It appeared that removing genotype data of the older bulls had a positive, though limited, impact on the genomic prediction of the calving traits.

Conclusions

The single-step model for the calving traits was proven to be one of the most complex and technically challenging models for genomic evaluation in dairy cattle, because the evaluation model included two correlated genetic effects and the number of estimated effects was ranked the highest among all trait groups in German single-step evaluations. The single-effect MACE model for bull calving evaluation complicated the integration of the foreign phenotype data of bulls into the national maternal-effect model with two correlated genetic effects. Our procedure of integrating the bull MACE data allowed us to use the foreign bull phenotype of all the four MACE traits. Comparing the SSM to the current MSM model, we observed the largest model differences in genomic evaluation for the calving traits than all the other trait groups.

The SNP effect estimates from the SSM model had less bias than the MSM model, according to the regression coefficients of SNP effects of the full on the truncated evaluation. The b_1 value of the SSM SNP effects was close to 1 for any of the calving traits, suggesting an unbiasedness of the SNP effects, the b_1 value from the MSM model was slightly lower than the SSM model. Based on GEBV of the validation bulls, less satisfactory validation results were obtained for the calving traits than for test-day or conformation traits and the validation results were less optimal for the direct than the maternal genetic effect, with least satisfactory results for the stillbirth direct effect. More inflation in GEBV was observed in the younger than the older validation bulls, particularly for the direct genetic effects.

Increasing the residual polygenic variance from 30% to 40% of the total genetic variance improved little in reducing the overestimation. Imposing a bull reference population, in contrast to the current single-step reference population, failed to improve the predictive ability of the genomic model either. However, removing genotype data of older bulls, born before 2005, resulted in a positive impact on the genomic prediction, i.e. reducing the prediction inflation, though to a limited degree. Therefore, we recommend removing genotype data of the older bulls in routine genomic evaluation.

For routine single-step genomic evaluation of the calving traits, a post-processing GEBV of young candidates may seem to be unavoidable.

Acknowledgements

Dr. Esa Mäntysaari and Dr. Ismo Strandén from the Finish research institute LUKE and Dr. Jeremie Vandenplas and Dr. Jan ten Napel from University of Wageningen are greatly thanked for their help in using the software packages MiX99 and MiXBLUP, respectively. The authors have stated no conflicts of interest.

References

Alkhoder, H., Z. Liu, D. Segelke, and R. Reents. 2022. *J. Dairy Sci.* 105:3306-3322. https://doi.org/10.3168/jds.2021-21145.

Alkhoder, H., Z. Liu, D. Segelke, and R. Reents. 2022a. Interbull Annual Meeting in Montreal, Canada, May 30 to June 1, 2022.

Alkhoder, H., Z. Liu, D. Segelke, and R. Reents. 2022b. EAAP annual meeting, Porto, Portugal, September 5 to 9, 2022.

Legarra, A., and Reverte, A. 2018. *Genet. Sel. Evol.* 50:53.

Liu, Z. 2011. Use of MACE Results as Input for Genomic Models. *Interbull Bulletin* 43.

Liu, Z., E. Pasman, F. Reinhardt, and R. Reents. 2012. *Interbull Bulletin* 46:133-142.

Liu, Z., M. E. Goddard, F. Reinhardt, and R. Reents. 2014. *J. Dairy Sci.* 97:5833-5850.

Liu, Z., H. Alkhoder, Ch. Kubitz, K. Stock, J. Heise, et al. 2019. *Interbull Bulletin* 55:35-45.

Mäntysaari, E. A., Liu, Z., and VanRaden., P. M. 2010. *Interbull Bulletin* 41:10–14.

Snell, E.J. 1964. *Biometrics*. 20:592-607.

Strandén, I. and M. Lidauer. 1999. *J. Dairy Sci.* 82:2779-2787.

Ten Napel, J., J. Vandenplas, M. Lidauer, I. Stranden, M. Taskinen, E. Mäntysaari, M. P. Calus, and R. F. Veerkamp. 2020 MiXBLUP user's guide.