Deregressed genomic breeding values from single-step evaluations of test-day traits using all genotype data

H. Alkhoder, Z. Liu, and R. Reents

IT Solutions for Animal Production (vit), D-27283 Verden / Aller, Germany

Abstract

Single-step model has become the Golden Standard for routine genetic evaluation in dairy cattle. For various statistical analyses or genomic validation, (daughter) yield deviations or deregressed genomic breeding values may be considered as pseudo-phenotype that are more independent from early genomic prediction. The aims of this study were to assess GEBV deregression methods for cows and bulls, and to validate the deregressed GEBV via a reversibility test. A total of 13.5 million animals with phenotypic records, evaluated with a single-step model using the German genotypic and phenotypic data from April 2023, were considered in the cow GEBV deregression. Likewise, all bulls with daughters and all reference cows were included in the bull GEBV deregression. Both GEBV deregression processes used the same genotype data and pedigree file as the preceding single-step evaluation. Deregressed GEBV of the cows or the bulls were moderately or highly correlated with their GEB, respectively. For the four test-day traits, milk, fat and protein yields and somatic cell scores, the deregressed GEBV seemed to have a lower trend than their original GEBV. Equal GEBV were obtained in a special single-step evaluation using the deregressed GEBV as phenotypic data, in comparison to those GEBV from the original single-step evaluation. We obtained equal GEBV not only for the cows with test-day records and bulls with daughters but nearly equal also for young, genotyped candidates without own phenotypic records. The validation results confirmed that the GEBV deregression was a reversible process and the deregressed GEBV were proven to be correct.

Key words: Deregression, genomic breeding values, single-step model, test-day traits

Introduction

The pedigree-based deregression of estimated breeding values (EBV) by Jairath et al. (1998), also known as the matrix deregression (Calus et al. 2016), has been widely used in dairy cattle evaluations, for example, for generating deregressed bull EBV as input data in Interbull MACE evaluation. The current multi-step genomic model needed ‘pseudo-phenotypic data’, such as the deregressed EBV or proofs (DRP), for genomic evaluation and SNP effect estimation. Calus et al. (2016) confirmed that the matrix deregression method by Jairath et al. (1998) was more accurate than the other deregression methods. In 2020, a reversibility test was conducted on DRP of bull MACE EBV on country scale DEU and on DRP of cow national EBV for all trait groups evaluated in Germany (Liu and Masuda, 2021). We could successfully validate the correctness of the DRP for all the bulls included in MACE evaluation on DEU scale and for all the domestic cows with own phenotypic records across all the evaluated trait groups.

Liu and Masuda (2021) and Masuda et al. (2021) developed GEBV deregression methods for the single-step SNP BLUP model and the single-step GBLUP model, respectively. The aims of this study were to 1) deregress genomic estimated breeding values (GEBV) of the single-step model for four test-day traits in German Holstein separately for bulls with daughters and for cows with own test-day records, and 2) validate the deregressed GEBV for the two groups of animals with phenotypic data as well as for all genotyped animals including young candidates.

Materials and Methods

A single-step SNP BLUP model for GEBV deregression

Prior to deregressing GEBV of cows or bulls, four test-day traits, milk yield (MKG), fat yield (FKG), protein yield (PKG), and somatic cell scores (SCS), were evaluated separately using a single-step SNP BLUP multi-lactation random regression test-day model (Alkhoder et al. 2022). For a detailed description of the single-step model, see the paper by Alkhoder et al. (2023). We applied here a special single-trait single-step model to deregress GEBV from the preceding single-step evaluation:

$$\mathbf{y} = \mu\mathbf{1} + \mathbf{u} + \mathbf{e} \qquad [1]$$

where **y** is a vector of deregressed GEBV (dGEBV) of animals with own phenotype data, **1** is a vector of 1s, μ is a general mean, **u** is a vector of GEBV for the animals with own phenotype data, and **e** is a vector of residuals. The dGEBV **y** are unknown and will be estimated in the deregression process. It is assumed that

$$[var(\mathbf{e})]^{-1} = \mathbf{D}\sigma_e^{-2} = diag\{\varphi_i\}\sigma_e^{-2} \quad [2]$$

where **D** is a diagonal matrix containing effective daughter contribution (EDC) of bulls with daughters or effective record contribution (ERC) of cows with own phenotype records on the animal-model basis, φ_i , for animal i, $i = 1, \dots, n$, and n is the number of animals with phenotype data. σ_e^2 is residual variance. For more details about the deregression model [1], see the paper by Liu and Masuda (2021).

Phenotypic, genotypic and pedigree data were taken from the routine evaluation in April 2023 for German dairy breeds Holstein, Red Dairy Cattle, and Jersey. Table 1 describes the data sets for the single-step evaluation as well as for the following step of GEBV deregression for all cows with phenotypic records of the three breeds. All the cows with own test-day records included in the original single-step evaluation were considered in the cow GEBV deregression process, too. A total of 1,318,780 genotyped Holstein animals were included in the deregression process as in the original single-step evaluation. Consequently, the same pedigree file containing 21,850,276 animals was used in the cow GEBV deregression process as in the preceding single-step evaluation.

Table 1. Description of the data sets for the single-step evaluation and cow GEBV deregression

Frequency	Single-step evaluation	Cow GEBV deregression
Genotyped animals	1,318,780	1,318,780
Phenotyped animals	13,528,444	13,528,444
Phenotypic input data	263,673,267 test-day records	13,528,444 GEBV
Genotyped or phenotyped animals	14,402,662	14,402,662
Animals in pedigree	21,850,276	21,850,276

For deregressing GEBV of bulls with daughters, sires of the phenotyped cows were treated as animals with own phenotypic records. In addition, genotyped cows with own phenotypic data must be considered also as phenotyped animals, because the genotyped sires of the cows no longer represented the full genomic reference population when the genotyped cows were available. To guarantee the complete phenotypic and genotypic information content of the reference population to be utilized in the GEBV deregression process as the preceding single-step evaluation, all the genotyped cows with own test-day records were also added to the list of animals with phenotypic data for the deregression process of the bull GEBV. The total number of bulls with daughter phenotypic information and the genotyped cows with test-day records amounted to 664,548. To avoid double counting the reference cows’ contribution to their sires, EDC of the sires was adjusted for the contribution by their reference daughters. As in the deregression of cow GEBV process, all genotype data of 1,318,780 Holstein animals were also considered in the deregression process for GEBV of the bulls.

A reversibility test for validating the GEBV deregression

A validation study of the GEBV deregression was conducted to see if the deregression of GEBV was a reversible process or in other words if equal GEBV could be obtained from a special single-step evaluation with their dGEBV as input ‘phenotypic’ data. In case of the cow GEBV deregression, a single-trait single-step model [1] was applied to the dGEBV of all the cows with test-day records. The aim of this validation was to see if equal GEBV could be obtained from the special single-step evaluation as from the original single-step evaluation using the test-day records for all the cows.

To validate the GEBV deregression for the bulls with daughters, we used dGEBV of all the sires of the cows as input phenotypic data for the special single-step evaluation under Model [1]. In addition, dGEBV of the reference cows were used also in the validation process, because the genotyped bulls with daughters did not represent the complete reference population due to the high number of genotyped cows with test-day records. Like the validation of the cow GEBV, it was to be verified if equal GEBV of the bulls were obtained from the special single-step validation evaluation as from the preceding, original single-step evaluation.

Genotyped, young candidates were included both in the special single-step evaluations for validating the GEBV deregression and in the preceding, original single-step evaluation based on the test-day data. Though these candidates did not have own phenotypic records, we would like to know if they received equal GEBV from the two single-step evaluations.

Results & Discussion

The deregression of the cow or bull GEBV from the original single-step evaluation was conducted using the software suite MiX99 (Strandén and Mäntysaari, 2010). Two separate deregression processes were performed for the cows with test-day records and for the bulls with daughters. The original single-step evaluation for the four test-day traits was described already in the paper by Alkhoder et al. (2022). Overall, the GEBV deregression using the single-trait single-step model [1] required a little less computing time and less memory than the original single-step evaluation with test-day records.

Deregressed GEBV of cows with test-day records

For the cow GEBV deregression, a total of 11.8 million Holstein cows with test-day records were considered in the multi-breed evaluation system in Germany. Figure 1 shows observed correlation between dGEBV and GEBV of milk yield for the Holstein cows in green line. The number of cows born each year was shown in grey bars on the secondary Y-axis. The correlation is about 0.84 from the oldest cow birth year 2000 to 2013 and decreases gradually afterwards. As a result of incomplete or missing lactations in the last three birth years, the correlation between dGEBV and GEBV drops markedly between the single-trait model with dGEBV as input data and the original multi-lactation single-step model with test-day yields.

Figure 2 shows average dGEBV and GEBV of the Holstein cows born between 2000 to 2020 for trait milk yield. The trends in both dGEBV and GEBV are similar, with a little lower trend in dGEBV (the solid line in red) than GEBV (the dotted line in black). On average, the difference between dGEBV and GEBV changes from about 50kg in year 2000 to -50kg in birth year 2020, representing 15% genetic standard deviations over 20 years. The dGEBV of the cows have much larger (error) variance than their GEBV, with an average within-year ratio, var(GEBV)/var(dGEBV), being 0.23.

For the 11.8 Holstein cows, similar results were obtained for the other 3 traits as for milk yield.

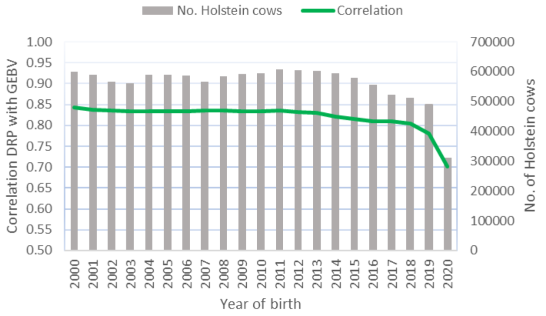

Figure 1. Correlation between deregressed GEBV and GEBV of milk yield for Holstein cows with test-day records

Deregressed GEBV of bulls with daughters

A total of 24,016 Holstein bulls that had daughters in 10 or more herds in Germany were selected for evaluating dGEBV of the bulls. Figure 3 shows the correlation between dGEBV and GEBV of the Holstein bulls born in 1998 through 2018. With the introduction of genomic selection in Germany in 2009, the correlation decreases steadily from 0.97 to 0.90. Bulls born in the last 3 years, 2016 to 2018, have much lower correlation, because their daughters have missing or incomplete lactations.

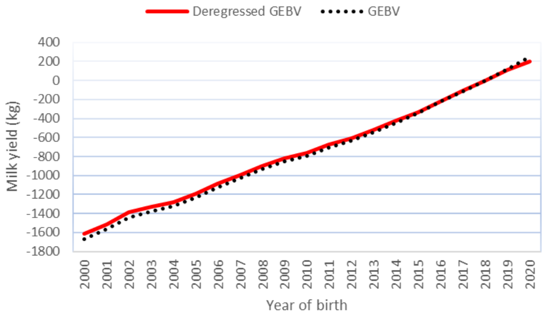

Figure 2. Averages of deregressed GEBV and GEBV of milk yield for Holstein cows with test-day records

As far as the trends are concerned, dGEBV of the bulls are shown to have almost equal averages by birth year, except the last 3 birth years (Figure 4). For the youngest bulls born in the last 3 years having daughters with missing or incomplete lactations, dGEBV of these bulls deviate evidently from their GEBV.

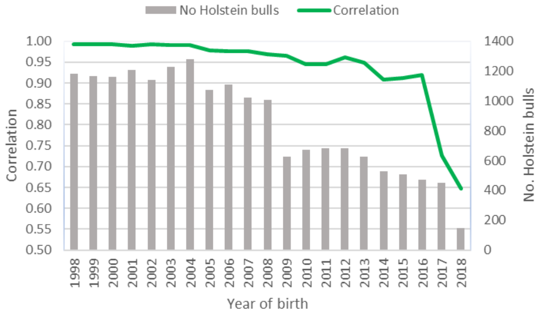

Figure 3. Correlation between deregressed GEBV and GEBV of protein yield for Holstein bulls with at least 10 herds in Germany

Deregressed GEBV of the bulls have higher (error) variance, with a ratio of standard deviation of GEBV over standard deviation of dGEBV being 0.91 for all birth years till 2016. Deregressed GEBV of the bulls born in last two years have significantly larger (error) variance than their GEBV.

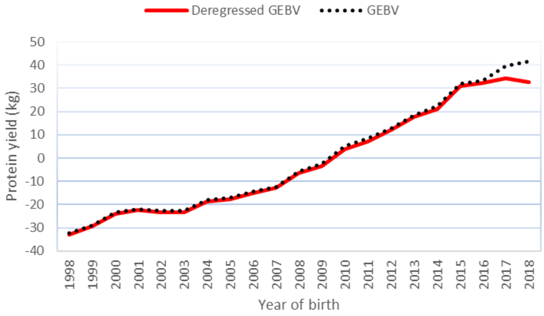

Figure 4. Averages of deregressed GEBV and GEBV of protein yield for the Holstein bulls by birth year

Validation results on deregressed GEBV of cows with test-day records

Due to different cow base populations for the Holstein breeds Black and White (B&W) and Red and White (R&W), 12.6 million B&W Holstein female animals were chosen to compare their single-step GEBV using dGEBV and test-day yields as ‘phenotypic data.’ The B&W female animals include all B&W cows with test-day records, their female ancestors, and young genotyped female animals without own test-day data yet.

It can be seen in Figure 5 that the two sets of GEBV of the B&W female animals, estimated using dGEBV as ‘phenotypic data’ and using original test-day milk yields, are above 0.99 for all birth years, except for the birth years 2021 and later. Even the young, genotyped females without own test-day milk yields have a correlation of GEBV higher than 0.98.

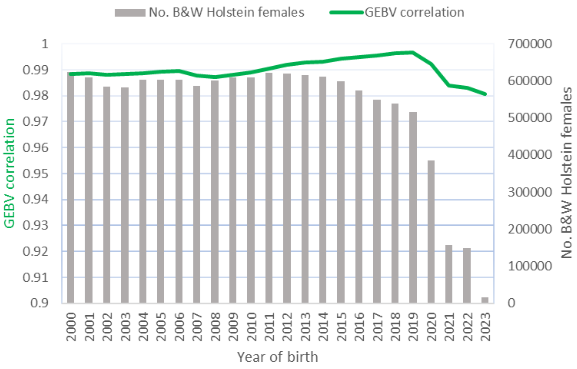

Figure 5. Correlation of single-step GEBV using deregressed GEBV and test-day yields of milk yield for all B&W Holstein females

Equal genetic trends were found in the single-step GEBV using dGEBV and test-day milk yields of the cows. In addition, the two sets of GEBV had equal variances for all birth years of the Holstein females, except that the female candidates born after 2020 had about 1% to 2% lower standard deviations of the GEBV using dGEBV as ‘phenotypic records’ than using test-day milk yields.

The validation results for the trait milk yield were seen also for the other remaining test-day traits. In summary, the special single-step evaluation with dGEBV of the cows as phenotypic records gave identical GEBV, as those obtained from the original single-step evaluation using test-day data, for all the Holstein females as well as for all other groups of animals in the single-step evaluation. The identical GEBV from the two single-step evaluations indicated the dGEBV of all the cows were accurately calculated.

Validation results on deregressed GEBV of bulls with daughters

As stated above, all the reference cows were considered in the bull GEBV deregression process to guarantee the complete genomic reference population being used in the deregression of GEBV of the bulls with daughters. To compare GEBV from the special single-step evaluation with dGEBV of the bulls and all the reference cows to those from the original single-step evaluation, we selected 10,770 B&W German Holstein AI bulls born from 1998 through 2022, including young genomic AI bulls born in 2019 and after.

Figure 6 shows GEBV correlation of protein yield for 10,770 B&W German AI bulls between the special single-step evaluation with dGEBV as phenotypic records and the original single-step evaluation with test-day protein yields. It can be clearly seen that the GEBV correlation is unity for all the birth years of bulls with daughters. However, for the young AI bulls born in 2019 and later the GEBV correlation is decreased to 0.985. The slightly lower GEBV correlation for the young AI bulls suggests that the multi-lactation test-day single-step model made a little different genomic prediction than the single-trait single-step model with one dGEBV as ‘phenotypic records’. As far as variance of the GEBV of the two single-step models are concerned, young genomic AI bulls have slightly higher GEBV variance with dGEBV as input data of the special single-step evaluation than the original single-step evaluation (Figure 7).

From Figures 6 and 7 we can draw a conclusion that the deregression of GEBV for bulls with daughters seems to be correct.

Conclusions

Deregressed EBV or deregressed GEBV have appealing statistical properties for diverse applications. The current multi-step genomic model was relied on the deregressed conventional EBV as ‘pseudo-phenotypic’ data.

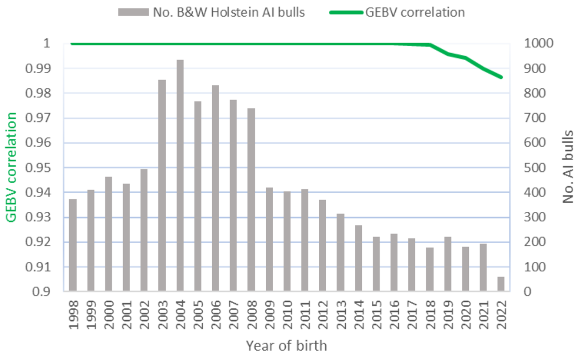

Figure 6. Correlation of single-step GEBV using deregressed GEBV and test-day yields for protein yield of all B&W German Holstein AI bulls

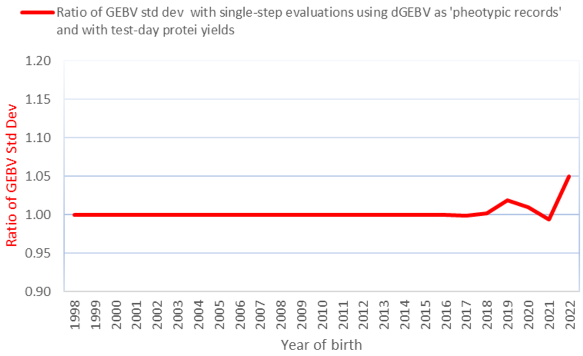

Figure 7. Ratio of standard deviations of GEBV of the B&W German Holstein AI bulls from the single-step evaluation with deregressed GEBV as input data over standard deviations of GEBV from the original single-step evaluation with test-day protein yields

We extended the matrix deregression method to deregress GEBV of the single-step evaluation. The single-step GEBV deregression method was assessed successfully for the four test-day traits in German Holstein. For all cows with test-day records, deregressed GEBV were moderately correlated with their GEBV. For bulls with daughters, degression of their GEBV must include GEBV of all genotyped cows with phenotypic records, because the genotyped bulls no longer represented the complete reference population which must be guaranteed in the bull GEBV deregression process as in the original single-step evaluation. For both cows with test-day records and bulls with daughters, deregressed GEBV had lower genetic trends, especially for cows with lactation in progress or missing lactations and for bulls having daughters with lactation in progress or missing lactations. Deregressed GEBV had higher variance than GEBV for the cows or bulls, in particular the cows' deregressed GEBV being much more variable than the bulls'. Equal GEBV were obtained from a special single-step evaluation with the deregressed GEBV as phenotypic data, compared to GEBV from the original single-step evaluation. This confirmed that the GEBV deregression was a reversible process. Not only the cows with test-day records and bulls with daughters received equal GEBV from the special single-step evaluations as from the original single-step evaluation, but also young, genotyped candidates obtained almost identical GEBV from the two single-step evaluations. In comparison to GEBV used as dependent variable in GEBV validation, the deregressed GEBV were more independent from the early GEBV of validation animals, therefore the deregressed GEBV may be more suited as the dependent variable of the GEBV test.

Acknowledgments

We thank Dr. Ismo Strandén and Dr. Esa Mäntysaari for their help in using MiX99 for the GEBV deregression.

References

Alkhoder, H., Liu, Z., Segelke, D., and Reents, R. 2022. Application of a single-step SNP BLUP random regression model to test-day yields and somatic cell scores in German Holsteins. Interbull Bulletin 57:74-83.

Alkhoder, H., Liu, Z., and Reents, R. 2023. The marker effects of a single-step random regression model for four test-day traits in German Holstein. J. Dairy Sci. (accepted)

Calus, M.P.L., Vandenplas, J., J. ten Napel, and F. Veerkamp. 2016. Validation of simultaneous deregression of cow and bull breeding values and derivation of appropriate weights. J. Dairy Sci. 99:6403-

6419. https://doi.org/10.3168/jds.2016-11028.

Jairath, L., Dekkers, J.C.M., Schaeffer, L.R., Liu, Z., Burnside, E.B., and Kolstad, B. 1998. Genetic evaluation for herd life in Canada. J. Dairy Sci. 81:550-562. https://doi.org/10.3168/jds.S0022-0302(98)75607-3.

Liu, Z., and Masuda, Y. 2021. A deregression method for single-step genomic model using all genotype data. Interbull Bulletin 56: 41-51.

Masuda, Y., Liu, Z., and Sullivan, P. 2021. Deregression formula for single-step genomic BLUP. Interbull Bulletin 56:142-146.

Strandén, I., and Mäntysaari, E. A. 2010. A recipe for multiple trait deregression. Interbull Bulletin 42:21-24.