A weekly genomic evaluation of newly genotyped selection candidates based on a single-step genomic model

H. Alkhoder, Z. Liu and R. Reents

IT-Solutions for Animal Production (vit), Heinrich-Schroeder-Weg 1, 27283 Verden / Aller, Germany

Zengting.liu@vit.de

Abstract

A single-step SNP BLUP model was developed for routine genomic evaluation of German Holstein. The current weekly genomic evaluation of young selection candidates based on a multi-step SNP BLUP model needed to be upgraded to optimally use the effect estimates from the single-step model. For indirect genomic prediction of newly genotyped selection candidates, two alternative statistical methods were assessed, an exact GRV method and a summation method. Both methods calculated direct genomic values using the SNP effect estimates from the full evaluation in the same way, but they differed in the computation of residual polygenic effects for the young candidates. GEBV of the candidates from the two methods were then compared to those from a single-step evaluation using phenotypic, genotypic and pedigree data from April 2023. To investigate the accuracy and bias of the two weekly evaluation methods, all 1,318,720 genotyped Holstein animals were divided into a reference set containing 1,169,502 animals born before April 2022 and a validation set of 149,218 animals born after April 2022. For all 69 evaluated traits in the German dairy cattle evaluation, correlation of GEBV of the weekly evaluation with the full evaluation was unity for the exact GRV method and ranged from 0.996 to 1 for the summation method. The regression coefficient of GEBV the full evaluation on the weekly evaluation was 1 for the exact GRV method and ranged from 0.988 to 1.002 for the second summation method. The two statistical methods for the indirect prediction of young candidates were shown to be accurate and unbiased.

Key words: indirect prediction, genomic evaluation, single-step model, selection candidates

Introduction

An indirect prediction of genomic breeding values (GEBV) for newly genotyped selection candidates at a weekly basis provided a key service for routine genomic selection in German Holstein (Alkhoder et al. 2014). In contrast to a full genomic evaluation, based on either a multi-step model (MSM, Liu et al. 2011) or a single-step model (SSM, Liu et al. 2014), the weekly genomic evaluation does not have any new phenotypic records added to evaluate but only new genotypic data of typically young animals. Therefore, the SNP marker effect estimates from the latest full genomic evaluation can be used to calculate direct genomic values (DGV) of the newly genotyped animals. Under the model MSM, a parental average (PA) of conventional evaluation was estimated via a BLUP animal model and was then combined with DGV using the selection index approach to obtain genomic estimated breeding values (GEBV) of the young candidates. For the SSM model, Liu et al. (2014) showed that GEBV of a newly genotyped animal be equal to the sum of DGV and parental average of residual polygenic effect (RPG, Liu et al. 2011).

A full single-step genomic evaluation including genotypes of new animals provides the most accurate GEBV for the newly genotyped animals. However, it is infeasible to complete the full single-step evaluation with millions of genotyped animals for all trait groups during a weekend. Therefore, the single-step weekly evaluation must be

computationally fast while ensuring GEBV being as accurate as possible.

The aims of this study were 1) to compare two statistical methods for an indirect prediction of GEBV of newly genotyped animals; and 2) to investigate accuracy and bias of the indirect prediction methods via a validation study.

Statistical methods for indirect prediction of GEBV for candidates

For the single-step SNP BLUP (ssSNPBLUP) model with an RPG effect (Liu et al. 2014), GEBV of a genotyped animal is the sum of its two components DGV and RPG:

$$u = d + a \qquad [1]$$

where u is GEBV, d is DGV, and a is RPG of the animal. GEBV of a newly genotyped young candidate after the full single-step evaluation can be approximated based on estimates of all model effects from the latest full single-step evaluation (Liu et al. 2014):

$$u_c = d_c + a_c = d_c + \tfrac{1}{2}(a_s + a_d) \qquad [2]$$

where u_c is GEBV of the genotyped candidate, d_c is its DGV, a_c is its RPG, a_s a_d represent RPG of its sire and dam, respectively, that were evaluated in the latest full single-step evaluation with their own genotype data. Note that the models [1] and [2] are a univariate model or single-trait model, not like a multi-lactation random regression test-day model for a full single-step evaluation of milk production traits in German dairy cattle (Alkhoder et al. 2024).

The ssSNPBLUP model estimated RPG for all genotyped animals, with or without their own phenotypic data in the latest, full single-step evaluation. However, for the young, genotyped animal that was not included in the latest single-step evaluation, its RPG effect was assumed be equal to its expected value of parental average of RPG, $\frac{1}{2}(a_s + a_d)$.

An Exact Method for GEBV of New Animals

In contrast to the single-step genomic BLUP model (ssGBLUP, Aguilar et al. 2010), the ssSNPBLUP provided direct estimates of SNP marker effects that can be used to calculate DGV of all newly genotyped animals.

The RPG effects of the newly genotyped animals can be estimated using RPG effect estimates of all genotyped animals in the latest single-step evaluation (Liu et al. 2016):

$$\widehat{\boldsymbol{a}}_c = \mathbf{A}_{\mathrm{cg}}\,\mathbf{A}_{\mathrm{gg}}^{-1}\boldsymbol{a}_{\mathrm{g}} \qquad [3]$$

where $\widehat{\boldsymbol{a}}_c$ is a vector of estimated RPG effects of all new genotyped candidates, $\boldsymbol{a}_{\mathrm{g}}$ is a vector of RPG effects of all genotyped animals in the latest single-step evaluation, $\mathbf{A}_{\mathrm{gg}}^{-1}$ is the inverse of pedigree relationship matrix for all the genotyped animals of the latest evaluation and $\mathbf{A}_{\mathrm{cg}}$ is the pedigree relationship matrix between the new candidates and the old, genotyped animals. This statistical method for indirect prediction of the RPG, together with the calculation of DGV, was termed as an *exact GRV method* (Vandenplas et al. 2023).

The RPG effects for the new selection candidates via Equation [3] are estimated by setting up the following equations:

$$\begin{bmatrix} \mathbf{A}^{00} & \mathbf{A}^{0\mathrm{g}} & \mathbf{A}^{0\mathrm{c}} \\ \mathbf{A}^{\mathrm{g}0} & \mathbf{A}^{\mathrm{gg}} & \mathbf{A}^{\mathrm{gc}} \\ \mathbf{A}^{\mathrm{c}0} & \mathbf{A}^{\mathrm{cg}} & \mathbf{A}^{\mathrm{cc}} \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{a}}_0 \\ \widehat{\boldsymbol{a}}_g \\ \widehat{\boldsymbol{a}}_c \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{A}_{\mathrm{gg}}^{-1}\widehat{\boldsymbol{a}}_g \\ \mathbf{0} \end{bmatrix} \qquad [4]$$

where $\widehat{\boldsymbol{a}}_0$ is a vector of RPG effects of for ancestors of the genotyped animals from the latest single-step evaluation and the new selection candidates. Solving the Equation [4] is technically equivalent to deregress the RPG effect estimates of the genotyped animals $\boldsymbol{a}_{\mathrm{g}}$ without using the inverse matrix $\mathbf{A}_{\mathrm{gg}}^{-1}$ but the Henderson's inverse of the pedigree relationship matrix for the three groups of animals:

$$\begin{bmatrix} \mathbf{A}^{00} & \mathbf{A}^{0\mathrm{g}} & \mathbf{A}^{0\mathrm{c}} \\ \mathbf{A}^{\mathrm{g}0} & \mathbf{A}^{\mathrm{gg}} & \mathbf{A}^{\mathrm{gc}} \\ \mathbf{A}^{\mathrm{c}0} & \mathbf{A}^{\mathrm{cg}} & \mathbf{A}^{\mathrm{cc}} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{00} & \mathbf{A}_{0\mathrm{g}} & \mathbf{A}_{0\mathrm{c}} \\ \mathbf{A}_{\mathrm{g}0} & \mathbf{A}_{\mathrm{gg}} & \mathbf{A}_{\mathrm{gc}} \\ \mathbf{A}_{\mathrm{c}0} & \mathbf{A}_{\mathrm{cg}} & \mathbf{A}_{\mathrm{gg}} \end{bmatrix}^{-1} \qquad [5]$$

The deregression process, without generating deregressed RPG effects for the genotyped animals g, estimates RPG effects of the ancestors denoted as group 0, that is equivalent to solving:

$$\widehat{\boldsymbol{a}}_0 = -(\mathbf{A}^{00})^{-1}\,\mathbf{A}^{0\mathrm{g}}\,\boldsymbol{a}_{\mathrm{g}}\,. \qquad [6]$$

From Equation [4], we can see that the RPG effects of the new candidates, $\boldsymbol{a}_c$, are estimated with the (deregressed) RPG effects of the genotyped animals via the pedigree [5].

A Summation Method for GEBV of New Animals

GEBV of the newly genotyped selection candidates are computed using the Model [1], as with the exact GRV method. However, a simpler method is assumed here for calculating the RPG effects of the new candidates, namely a linear summation of RPG effects of all genotyped ancestors from the latest single-step evaluation. When both parents of a new candidate *c* were evaluated with own genotype data in the latest single-step evaluation, $a_c = \frac{1}{2}(a_s + a_d)$. Should only male animals be genotyped in a population, then $a_c = \frac{1}{2}(a_s + \frac{1}{2}(a_{mgs} + \frac{1}{2}(a_{smgd} + \cdots)))$, where a_{mgs} is RPG effect of maternal grandsire of the candidate, and a_{smgd} is RPG effect of sire of maternal granddam of the candidate. In practice, the RPG of the candidate a_c is calculated by processing the pedigree from the youngest candidate to its oldest genotyped relatives for the summation. Ancestors having no genotype data in the latest single-step evaluation were assumed to have RPG effect being 0 in this process. The summation method for computing RPG effects of the new candidates may be described as:

$$\hat{\boldsymbol{a}}_c = \mathbf{A}_{\mathrm{cg}}\,\mathbf{I}\boldsymbol{a}_{\mathrm{g}} \qquad [7]$$

where **I** is an identity matrix.

Data materials for a comparison of the indirect prediction methods

Genotypic, phenotypic and pedigree data from the April 2023 single-step evaluation were used to investigate accuracy and bias of the two indirect prediction methods. A total of 1,318,720 genotyped German Holstein population were divided into two groups: 1,169,502 '*reference animals'* born in March 2022 and earlier, and 149,218 '*genotyped candidates'* born from April 2022 onwards. The pedigree for all animals of the two groups contained 3,427,852 animals, including 2,109,132 non-genotyped ancestors.

In the single-step evaluation for German Holstein, a total of 69 single traits or indices of evaluated traits were evaluated. For instance, a total of 9 random regression coefficients of a multi-lactation random regression test-day model (Alkhoder et al. 2024) were combined into a single value, 305-day milk yield on a combined lactation basis. The weekly genomic evaluation was conducted for milk yield on the 305-day combined lactation basis instead of the 9 random regression coefficients.

Results & Discussion

Estimates of SNP markers for the 69 traits or indices were obtained from the latest single-step evaluation with data from April 2023. The RPG effects for the genotyped candidates were computed using the two statistical methods: the exact GRV method and the summation method. The program *predict_GEBV* of the MiX99 software suite (Strandén and Lidauer 1999) was used to compute GEBV of the young candidates with the exact GRV method (Vandenplas et al. 2023). Our own software for the summation method was developed in python. For all the 69 traits or indices, the GRV method took 38 minutes on 46 cores simultaneously and the peak RAM usage was 15.5 Gb.

Table 1 shows correlations of GEBV with DGV and RPG for the reference animals as well as regression slopes of GEBV on DGV or RPG for the genotyped animals in the full single-step evaluation April 2023 for all the 69 traits or indices. Similarly, the correlations and regression slopes are given in Table 2 for the genotyped selection candidates. In general, GEBV is higher correlated with DGV than RPG for either group of the genotyped animals. Regression slope values of GEBV on DGV are close to 1, on average, for both groups of the animals, whereas the average regression slopes of GEBV on RPG deviate more from 1.

Table 1: Correlations and regressions of GEBV, DGV and RPG estimates of the reference animals for all 69 traits or indices

	Average	Minimum	Maximum
Correlation of GEBV with DGV			
	0.969	0.935	0.986
Regression slope of GEBV on DGV			
	1.05	0.990	1.138
Correlation of GEBV with RPG			
	0.406	0.248	0.650
Regression slope of GEBV on RPG			
	1.64	0.886	3.169

Table 2: Correlations and regressions of GEBV, DGV and RPG estimates of the genotyped selection candidates for all 69 traits or indices

	Average	Minimum	Maximum
Correlation of GEBV with DGV			
	0.976	0.944	0.991
Regression slope of GEBV on DGV			
	1.024	0.974	1.118
Correlation of GEBV with RPG			
	0.304	0.064	0.589
Regression slope of GEBV on RPG			
	1.434	0.324	3.102

To validate GEBV of the weekly genomic evaluation, GEBV of the new candidates from the full single-step evaluation were correlated with their GEBV from the weekly genomic evaluation. Figure 1 shows the GEBV correlations of the selection candidates between any of the three evaluations: the weekly genomic evaluations with the exact GRV and the summation methods, and the latest full single-step evaluation. It can be seen in Figure 1 that the exact GRV method has a unity correlation with the latest single-step evaluation for all the 69 traits or indices. As far as the summation method for the weekly evaluation is concerned, its GEBV correlations with the single-step evaluation ranged from 0.9970 to 0.9999 with a mean of 0.9995. The GEBV correlations between the two methods for the weekly evaluation have an average of 0.9994.

GEBV of the candidates from the latest single-step evaluation were regressed on those from the weekly genomic evaluation based on either of the method: the exact GRV or summation method. Figure 2 shows the regression slope values of the two weekly evaluation methods for all the traits or indices. The regression slope values of the exact GRV method ranged from 0.9987 to 1.0008 with an average of 0.9998. In comparison, the summation method has a regression slope value between 0.9872 to 1.0018 with a mean of 0.9981 for the 69 traits or indices.

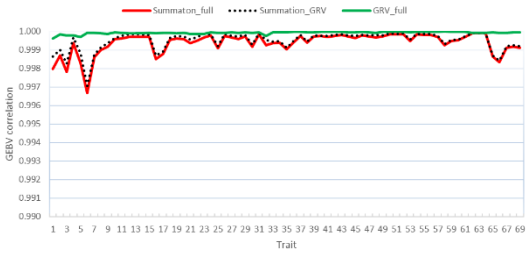

Figure 1. Correlations of GEBV of the candidates using the exact GRV and summation methods with the latest single-step evaluation for all the traits or indices.

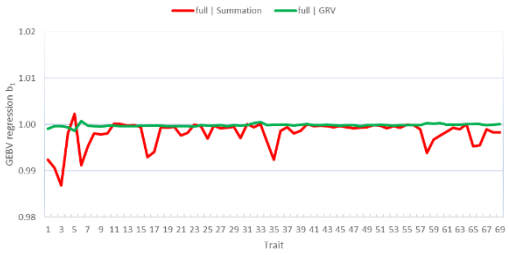

Figure 2. Regression of the latest single-step GEBV of the candidates on that of the exact GRV or summation method for all the traits or indices.

GEBV bias of the weekly genomic evaluation

In addition to the GEBV correlations and regressions of the weekly genomic evaluation methods, GEBV bias, defined as GEBV of the weekly evaluation minus that of the latest single-step evaluation, was investigated for all the selection candidates.

Figure 3 shows the frequency distribution of the GEBV biases of all the 149,218 candidates for trait no. 3 which was under the highest selection pressure among all the 69 traits or indices. A total of 87% or 67% of all the candidates had no bias, i.e., GEBV of the weekly evaluation being equal to that of the

single-step evaluation, for the exact GRV or the summation method, respectively. The distribution of the GEBV bias was symmetric around zero for both weekly evaluation methods. However, for the summation method about 5.4% of all the candidates had a downward bias of lower than -20% of genetic standard deviations of the trait no. 3. The downward bias was caused by the fact that some ancestors of the candidates did not have genotypic data in the latest single-step evaluation, and the summation method assumed RPG of those ancestors being zero. Due to the high selection pressure on this trait, those ancestors might have had an RPG greater than zero, if they had been genotyped.

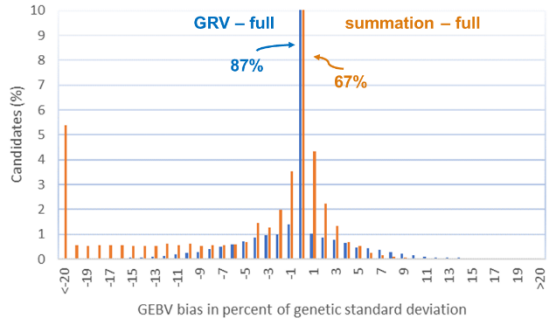

Figure 3. Distribution of GEBV bias of all the candidates using the two weekly evaluation methods for the trait no. 3 under the highest selection pressure.

To further investigate the impact of ancestors having no own genotype data in the full single-step evaluation, we selected a trait with little selection pressure on it, trait no. 4. The distribution of GEBV bias is shown for all the candidates in Figure 4. In contrast to Figure 3, no candidates have a downward GEBV bias for this trait as for the trait no. 3.

Compared to the summation method, the exact GRV method did not have the group of candidates showing a downward GEBV bias, because the GRV method estimated RPG of those ancestors based on RPG of all the genotyped animals in the full single-step evaluation.

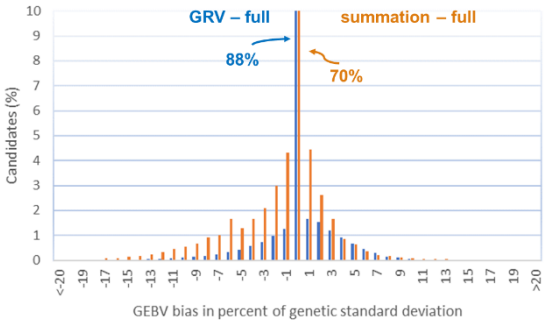

Figure 4. Distribution of GEBV bias of all the candidates of the two weekly evaluation methods for trait no. 4.

For the selection candidates of the weekly evaluation, GEBV differences between the exact GRV method and the full single-step evaluation were small but not non-existent. This may be contributed by several factors. Firstly, both weekly genomic evaluations assumed a single trait model, whereas a multi-trait model was used for all the 69 traits or indices in the full single-step evaluation. Secondly, the two weekly genomic evaluation methods estimated parental average of RPG effect for the selection candidates, while their RPG effects were estimated in the full single-step evaluation using all available genotypic and phenotypic data of all animals.

The same procedure of the Interbull genomic reliability method can be followed for approximating genomic reliabilities for the weekly genomic evaluation as for the full single-step evaluation, except that conventional reliability values of all the animals can be calculated from effective daughter contributions of bulls and effective record contributions of cows, which have been obtained from the latest, full single-step evaluation, instead of processing original phenotypic data.

Conclusions

Two statistical methods were assessed for the weekly genomic evaluation of newly genotyped selection candidates, based on the effect

estimates of the single-step model from a latest, full single-step evaluation. As a validation of the weekly genomic evaluation methods, GEBV of young selection candidates born in the last year were compared to the latest, full single-step evaluation containing those selection candidates. For all 69 traits or indices, GEBV of the selection candidates estimated using the two weekly genomic evaluation methods, the exact GRV and summation methods, were fully correlated with those from the single-step evaluation. Regression slopes of the single-step GEBV of the selection candidates on those of the weekly evaluation were all close to 1 for all the traits or indices. According to the distribution of GEBV bias to the single-step evaluation among the selection candidates, the exact GRV method resulted in equal GEBV as the full single-step evaluation. However, the summation method led to a downward bias for 5% of candidates whose partial ancestors had no own genotypic data in the latest, full single-step evaluation. Whenever possible, the exact GRV method should be preferred to the summation method for routine weekly evaluations. Both statistical methods of the weekly genomic evaluation were computationally efficient and feasible for a genomic evaluation of newly genotyped animals of German Holsteins during weekend.

Acknowledgments

Dr. Ismo Strandén from Luke, Finland, is thanked for providing the special program *predict_GEBV* for this project.

References

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* 93:743–752.

Alkhoder, H., Liu, Z., Bach, Th., Pasman, E., and Reinhardt, F. 2014. A continuous genomic evaluation system for German Holsteins. *Interbull Bulletin* 48:100-104.

Alkhoder, H., Liu, Z., and Reents, R. 2024. The marker effects of a single-step random regression model for four test-day traits in German Holstein. *J. Dairy Sci.* 107:410-424.

Liu, Z., Seefried, F., Reinhardt, F., Rensing, S., Thaller, G., and Reents, R. 2011. Impacts of both reference population size and inclusion of a residual polygenic effect on the accuracy of genomic prediction. *Genet. Sel. Evol*. 43:19.

Liu, Z., Goddard, M.E., Reinhardt, F., and Reents, R. 2014 A single-step genomic model with direct estimation of marker effects. *J. Dairy Sci.* 97:5833-5850.

Liu, Z., Goddard, M. E., Hayes, B. J., Reinhardt, F., and Reents, R. 2016. Technical note: Equivalent genomic models with a residual polygenic effect. *J. Dairy Sci*. 99:2016-2025.

Strandén, I., and M. Lidauer. 1999. Solving large mixed linear models using preconditioned conjugate gradient iteration. *J. Dairy Sci*. 82:2779–2787.

Vandenplas, J., ten Napel, J., Darbaghshahi, S.N. Evans, R., Calus, M.P.L., Veerkamp, R., Cromie, A., Mäntysaari, E.A., and Strandén, I. 2023. Efficient large-scale single-step evaluations and indirect genomic prediction of genotyped selection candidates. *Genet Sel Evol* 55, 37.