

## A COMPARISON OF GENOMICALLY ENHANCED BREEDING VALUES PREDICTED BY DIFFERENT SINGLE-STEP APPROACHES

Dawid Słomian<sup>1</sup>, Kacper Żukowski<sup>1</sup>, Joanna Szyda<sup>2,1\*</sup>

<sup>1</sup>Department of Cattle Breeding, National Research Institute of Animal Production, 32-083 Balice n. Kraków, Poland

<sup>2</sup>Department of Genetics, The Biostatistic Group, Wrocław University of Environmental and Life Sciences, Koźuchowska 7, 51-631 Wrocław, Poland

\*Corresponding author: joanna.szyda@upwr.edu.pl

### Abstract

Many countries are currently adopting the single-step model for national genetic evaluations of dairy cattle. The two most widely applied statistical formulations of the single-step model are Genomic Best Linear Unbiased Prediction (ssG-BLUP) and Single Nucleotide Polymorphism BLUP (ssSNP-BLUP), with the main difference being the handling of additive genetic covariance between individuals with genotypes. Using solvers available in MiXBLUP software, our study aimed to compare both models in terms of Genomic Enhanced Breeding Value (GEBV) prediction, bull rankings, and computational efficiency (memory consumption and computational time). The results did not show marked differences in the quality of GEBV prediction expressed by the metrics underlying the Interbull validation, except for ssG-BLUP, APY-based solvers with 3,000 core bulls. However, the ranking of the top 50 bulls differed between models, which has implications for the breeding industry and selection, since the top-ranking bulls are typically the most widely used. 39 and 31 of the top 50 bulls were common to all models for stature and foot angle, respectively. Regarding computational time, ssSNP-BLUP and ssG-BLUP with APY solver using 3,000 bulls were the fastest, and ssG-BLUP with GT solver was the slowest. The selection of core individuals for the APY solver was a crucial element that affected the prediction accuracy. GT or SNP-BLUP solvers can circumvent this issue, since no selection of core individuals is required.

**Key words:** APY, G-BLUP, GEBV, GT, single-step, SNP-BLUP

### Abbreviations

APY – algorithm for proven and young  
 BLUP – best linear unbiased prediction  
 DGV – direct genetic value  
 DRP – deregressed proofs  
 EDC – effective daughter contributions  
 GEBV – genomically enhanced breeding values  
 $h^2$  – heritability  
 MACE – multiple across country evaluation  
 MME – mixed model equations  
 RHS – right hand side  
 SNP – single nucleotide polymorphisms

Today, many countries are implementing the single-step model for their national genetic evaluations of dairy cattle. The main difference between the currently used multi-step and single-step models is the incorporation of genomic information available for calculating additive genetic covariances between individuals using phenotype, genotype, and pedigree information simultaneously. The single-step model incorporates the identical-by-descent or at least identical-by-state similarity between the genotyped fraction of the evaluated individuals, expressed by the similarity of their genotypes. Genotypic

information is typically expressed by single nucleotide polymorphisms (SNPs). Two equivalent statistical formulations of such a single-step model comprise the genomic best linear unbiased prediction (ssG-BLUP) (Aguilar et al., 2010; Christensen and Lund, 2010; Fragomeni et al., 2015) and ssSNP-BLUP (Liu et al., 2016; Strandén and Garrick, 2009). The only difference between these models lies in the formulation of the additive genetic covariance between individuals, in particular, between individuals with genotypes. In ssG-BLUP, the covariance matrix of breeding values of all evaluated individuals comprises the components related to non-genotyped individuals constructed using the pedigree information, and to genotyped animals constructed as a weighted sum of pedigree and SNP genotype information. In the ssSNP-BLUP, the genomic relationship between the additive genetic effects of SNPs is incorporated as a separate component in addition to a pedigree-based relationship. However, despite different parameterizations both models are equivalent mathematically (Gao et al., 2018; Koivula et al., 2012).

However, practical applications for the evaluation of large national populations of dairy cattle are computationally very challenging. Three solvers have been developed and are used for national genetic evaluations. The ssG-

BLUP model can be implemented either by the APY solver (Legarra et al., 2014) or the GT solver (Legarra et al., 2014). The APY (algorithm for proven and young) procedure does not use the entire genomic relationship matrix but only a part of it based on a predefined set of genotyped individuals (so-called core individuals) while the remainder, i.e. non-core genotyped individuals is fitted as genomically uncorrelated using a diagonal genomic relationship matrix. Therefore, the system for solving model equations only requires a computationally intensive inverse of the dense genomic submatrix for the core individuals and the submatrices corresponding to the covariance between core and noncore individuals. The GT procedure implements an inverse of the genomic relationship matrix as a function of the inverse of the pedigree-based numerator relationship matrix corresponding to genotyped individuals and the SNP genotype design matrix, calculated based on the Woodbury formula. In the ssSNP-BLUP (Gao et al., 2018), due to an alternative formulation of the covariance structure, solving the system of equations does not require inverting of the partial (APY) or full (GT) relationship matrices but only a computationally simple inverse of the SNP covariance matrix.

The purpose of this study was to compare G-BLUP, GT, and SNP-BLUP solvers in terms of model validation performance to predict breeding values, differences in predicted breeding values between solvers, as well as their computational efficiency and computational resources requirements, in the context of routine genetic evaluation.

## Material and methods

### Material

The analyzed set (Table 1) represented data from the Polish national genetic evaluation of Holstein cattle from December 2021 for stature, which represents a highly heritable trait ( $h^2=0.54$ ), and foot angle, representing a low heritability trait ( $h^2=0.09$ ). It comprised 1,098,611 cows with stature phenotypes and 1,098,766 cows with foot angle phenotypes born between 1992 and 2019 as well as 141,397 bulls (stature) and 117,482 bulls (foot angle) born between 1986 and 2017 with pseudo-phenotypes expressed by their deregressed proofs (DRP) from the multiple across country evaluation (MACE) carried out by Interbull (interbull.org). Genomic data in the form of

46,118 SNP markers were available for 134,960 individuals, including 70,134 cows with phenotypes born between 2009 and 2021, as well as 64,826 bulls born between 1985 and 2021, of which 26,471 represented young individuals without pseudo-phenotypes and 38,355 were bulls with MACE-DRP. In our study, the pedigree of animals with phenotype data was truncated after the fifth generation, resulting in 1,555,995 individuals and 33 genetic groups.

### GEBV prediction models

The following single-step single-trait models were considered for the prediction of GEBV.

The ssG-BLUP (Christensen and Lund, 2010):

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{W}_G \mathbf{a} + \mathbf{e}, \quad (1)$$

where:  $\mathbf{y}$  is the vector of dependent variables represented by cows' measured phenotypes and bulls' pseudo-phenotypes expressed by their MACE DRPs,  $\mathbf{b}$  represents a vector of fixed effects scored at cows' first parity including age at calving, lactation phase, and herd, while for bulls the corresponding fixed effects were represented by assigning common classes,  $\mathbf{a}$  represents a vector of breeding values, and  $\mathbf{e}$  is the vector of residuals. The underlying covariance structure of the random effects is given by  $\mathbf{a} \sim N(0, \mathbf{H}_G \sigma_a^2)$  and  $\mathbf{e} \sim N(0, \mathbf{R}\sigma_e^2)$ .  $\mathbf{H}_G$  is given by (Liu et al., 2016):

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} + \begin{bmatrix} \mathbf{A}_{12}\mathbf{A}_{22}^{-1}(\mathbf{G} - \mathbf{A}_{22})\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{A}_{12}\mathbf{A}_{22}^{-1}(\mathbf{G} - \mathbf{A}_{22}) \\ (\mathbf{G} - \mathbf{A}_{22})\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{G} - \mathbf{A}_{22} \end{bmatrix},$$

where:  $\mathbf{A}_{11}$ ,  $\mathbf{A}_{12}/\mathbf{A}_{21}$ , and  $\mathbf{A}_{22}$  are the components of the numerator relationship matrix constructed based on the pedigree corresponding to non-genotyped animals, the covariance between non-genotyped and genotyped animals, as well as between genotyped animals, respectively, while  $\mathbf{G}$  represents the genomic relationship matrix between genotyped animals.  $\mathbf{R}$  is a diagonal matrix containing 1.00 for cows with phenotypes or  $n_i$  for bulls with MACE DRPs, with  $n_i$  representing a difference in effective daughter contributions of  $i$ -th bull between the MACE and the national evaluation.  $\mathbf{X}$  and  $\mathbf{W}_G$  denote the corresponding design matrices.

Table 1. Number of animals in the analyzed data sets

Category	Sex	Number of animals
Phenotype data (Stature)	Females with phenotypes	1,098,611
	Males with MACE DRP	141,397
Phenotype data (Foot angle)	Females with phenotypes	1,098,766
	Males with MACE DRP	117,482
Genotype data	Females	70,134
	Males (bulls and candidates)	64,826
Pedigree data	Females	1,368,487
	Males	187,508

The ssSNP-BLUP (Gao et al., 2018):

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{W}_s\mathbf{a} + \mathbf{e}, \quad (2)$$

where  $\mathbf{y}$ ,  $\mathbf{b}$ ,  $\mathbf{a}$ , and  $\mathbf{e}$  are the same as defined above, but  $\mathbf{a}$  is parameterized as  $\mathbf{Z}\mathbf{g} + \mathbf{u}$  with  $\mathbf{g}$  being the vector of random SNP effects and  $\mathbf{u}$  being a vector of random additive residual polygenic effects. The covariance structure imposed on the residual effect ( $\mathbf{e}$ ) is the same as defined above, while for the additive genetic effect  $\mathbf{a}$  the distribution is defined as  $\mathbf{a} \sim N(0, \mathbf{H}_s \sigma_a^2)$  where the structure of  $\mathbf{H}_s$  is

$$\text{expressed by } \begin{bmatrix} \mathbf{TGT}' + \mathbf{D} & \mathbf{TG} & \mathbf{TZB} \\ \mathbf{GT}' & \mathbf{G} & \mathbf{ZB} \\ \mathbf{BZ}'\mathbf{T}' & \mathbf{BZ}' & \mathbf{B} \end{bmatrix} \text{ with } \mathbf{D} = (\mathbf{A}^{II})^{-1},$$

$\mathbf{T} = \mathbf{A}_{12}\mathbf{A}_{22}^{-1}$  and  $\mathbf{B}$  representing a diagonal matrix of the

form  $\mathbf{I} \frac{1}{\sum_{i=1}^N 2p_i(1-p_i)}$  with  $p_i$  denoting the allele frequency of the  $i$ -th SNP,  $N$  is the number of SNPs, and  $\mathbf{A}^{II}$  is the upper

left block corresponding to  $\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}^{-1}$ .  $\mathbf{X}$  and

$\mathbf{W}_s$  denote the corresponding design matrices (Gao et al., 2018). Note that the variance components and the proportion of residual additive polygenic variance were not estimated, instead their values corresponding to the parameters used in the Polish national genetic and genomic evaluation (Table 2).

Table 2. Variance components underlying the analyzed phenotypes

Trait	$\sigma_a^2$	$\sigma_e^2$	$\sigma_u^2$
Stature	5.50	4.63	20%
Foot angle	0.11	1.06	20%

### Solving the GEBV prediction equations

Solutions for the effects fitted in the above models were obtained using the MiXB LUP 3.0 software (Liu et al., 2014; Vandenplass, 2022) that solves the following equation:  $\mathbf{D}^{-1}\mathbf{M}^{-1}\mathbf{C}\mathbf{x} = \mathbf{D}^{-1}\mathbf{M}^{-1}\mathbf{p}$  where  $\mathbf{C}$  represents the coefficient matrix corresponding to the Mixed Model Equations (MME) for solving (1) or (2),  $\mathbf{x}$  is the vector of model parameters, and  $\mathbf{p}$  is the RHS of MME, while  $\mathbf{M}$  and  $\mathbf{D}$  respectively represent the first-level and the second-level preconditioning matrices. The computations were performed on a dedicated computing server running the Linux Red Hat operating system with 260GB of RAM, 16 Intel Xeon CPUs with 2.20GHz, and 600GB of hard disk space. For a large national dairy population, such as the one used in our study, the numerically severe challenge is to obtain the inverse of the  $\mathbf{H}_c$  and  $\mathbf{H}_s$  matrix, in particular of its component related to genotyped individuals – the dense sub-matrix  $\mathbf{G}$ . In our study, the following approaches were considered.

The GT approach (Lourenco et al., 2020; Mäntysaari et al., 2020), in which the  $\mathbf{G}^{-1}$  matrix is represented

by  $\frac{1}{w}\mathbf{A}_{22}^{-1} - \frac{1}{w}\mathbf{T}'\mathbf{T}$ , with  $\mathbf{T} = \mathbf{L}^{-1}\mathbf{Z}'\mathbf{A}_{22}^{-1}$  where  $w$  denotes the proportion of a residual polygenic variance and  $\mathbf{L}$  is defined by  $\mathbf{Z}'\mathbf{A}_{22}^{-1}\mathbf{Z} + w\mathbf{I} = \mathbf{L}\mathbf{L}'$ .

The APY approach (Legarra et al., 2014) that divides the genotyped individuals into a core and non-core sub-groups for which the inverse is handled differently in such a way that the exact inverse is computed only for the core animal sub-group ( $\mathbf{G}_c$ ) and the covariance between core and non-core individuals, while the part of the matrix corresponding to the non-core genotyped animals ( $\mathbf{G}_n$ ) is handled as a diagonal matrix:

$$\mathbf{G}^{-1} \approx \begin{bmatrix} \mathbf{G}_c^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} -\mathbf{G}_c^{-1}\mathbf{G}_{cn} \\ \mathbf{I} \end{bmatrix} \mathbf{M}_n^{-1} [-\mathbf{G}_{nc}\mathbf{G}_c^{-1} \quad \mathbf{I}]$$

where the subscripts  $n$  and  $c$  represent non-core and core individuals respectively, and  $\mathbf{M}_n$  is a diagonal matrix. In our study, four approaches to the selection of core animals were considered: APY3000top – where 3,000 genotyped bulls with the highest effective daughter contributions (EDC) were selected as core individuals, APY3000random – where the 3,000 core individuals were selected randomly from the genotyped population and the corresponding versions termed APY15000top, APY10000random, and APY15000random implementing 10,000 and 15,000 core individuals respectively. APY10000random proposed by Misztal et al. (2014) formed a basis to assess differences between scenarios implementing a smaller (APY3000) as well as a larger (APY15000) selection of core animals.

The ssSNP-BLUP approach (Gao et al., 2018) does not meet the numerical burden of the models based on  $\mathbf{G}$ , since in terms of the modelling of genomic information only the inverse of the diagonal matrix  $\mathbf{B}$  is required:

$$\mathbf{H}_S^{-1} = \begin{bmatrix} \mathbf{A}^{11} & \mathbf{A}^{12} & \mathbf{0} \\ \mathbf{A}^{21} & \mathbf{A}^{22} + \left(\frac{1}{w} - 1\right)\mathbf{A}_{22}^{-1} & -\frac{1}{w}\mathbf{A}_{22}^{-1}\mathbf{Z} \\ \mathbf{0} & -\frac{1}{w}\mathbf{Z}'\mathbf{A}_{22}^{-1} & \frac{1}{1-w}\mathbf{B}^{-1} + \frac{1}{w}\mathbf{Z}'\mathbf{A}_{22}^{-1}\mathbf{Z} \end{bmatrix},$$

where  $\begin{bmatrix} \mathbf{A}^{11} & \mathbf{A}^{12} \\ \mathbf{A}^{21} & \mathbf{A}^{22} \end{bmatrix}$  represent the blocks corresponding to

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}^{-1}.$$

### Validation of GEBV

For the validation of GEBV prediction, bulls born after 2013 (7,296 bulls for stature; 6,200 bulls for foot angle) and cows born after 2016 (96,772 cows for stature; 96,771 cows for foot angle) were removed from the phenotype vector ( $\mathbf{y}$ ) of equations (1) and (2). Based on such truncated data sets, the GEBVs of bulls with EDC greater than 20 were predicted by models (1) and (2). Prediction accuracy was assessed by a weighted linear regression (Mäntysaari et al., 2010):  $\mathbf{GEBV}_f = b_0 + b_1\mathbf{GEBV}_t + e$ , with  $\mathbf{GEBV}_f$  representing the vector of GEBVs predicted based on the full data set with all available individuals while  $\mathbf{GEBV}_t$  contains the GEBVs predicted based on the truncated data set. For  $i$ -th bull,

weights were defined as  $\frac{EDC_i}{EDC_i + k}$  with  $k = \frac{4 - h^2}{h^2}$ .

The unbiased evaluation is then represented by  $b_1$  equal to 1.00.

The above linear regression equation was fitted using the *lm* function and Pearson correlation coefficients, used to assess differences between GEBVs across scenarios were calculated using R software (www.rstudio.com).

### Results

#### Model validation

The validation data set for stature and foot angle comprised 1,727 and 1,725 bulls with EDC > 20, respectively. For stature, generally, no marked differences in the estimated slope of the linear regression were observed between the models, varying between 0.94 (APY3000top) and 1.02 (APY10000random). Furthermore, except for APY3000, the differences between R<sup>2</sup> corresponding to each model were small. For stature, the GT approach achieved the highest R<sup>2</sup> of 0.83, while the lowest R<sup>2</sup> corresponded to 3,000 core animal scenarios with 0.57 for APY3000top and 0.60 for APY3000random. Similar results were observed for

foot angle, albeit with overall lower R<sup>2</sup>. GT and SNP-BLUP achieved the highest value of 0.76, while the lowest of 0.60 was attributed to APY3000random and 0.62 for APY3000top. Details of the validation results are summarized in Table 3. Pearson correlations (Figure 1) between GEBVs predicted for the validation bulls from the full and the truncated datasets were calculated for all pairs of models. For stature, they demonstrated a good agreement only between SNP-BLUP and GT, as expressed by correlations of 0.996. Interestingly, the lowest correlations of 0.810 were observed between both models implementing APY3000 (i.e. APY3000random and APY3000top), as well as between APY3000top and APY15000top. For foot angle correlations are generally higher, and the pattern is very similar, with the correlation between SNP-BLUP and GT reaching 0.999 and the lowest correlation of 0.870 observed between both APY3000 scenarios, between GT and APY3000top, as well as between GP and APY3000random. Notably, for both traits, the correlations within APY3000top and within APY3000random are markedly lower than for the other models.

Table 3. Validation of GEBV predictions

Model variants	$\hat{b}_0$	SE( $\hat{b}_0$ )	$\hat{b}_1$	SE( $\hat{b}_1$ )	R <sup>2</sup>
Stature, 1,727 validation bulls, h <sup>2</sup> = 0.54					
SNP-BLUP	-3.90	0.32	1.01	0.01	0.77
GT	-4.40	0.27	1.01	0.01	0.83
APY3000top	-1.81	0.44	0.94	0.02	0.57
APY3000random	-2.20	0.44	0.99	0.02	0.60
APY10000random	-3.74	0.36	1.02	0.02	0.72
APY15000top	-2.59	0.32	0.96	0.01	0.75
APY15000random	-2.32	0.33	0.96	0.01	0.73
Foot angle, 1,725 validation bulls, h <sup>2</sup> = 0.09					
SNP-BLUP	-2.03	0.18	1.03	0.01	0.76
GT	-1.96	0.18	1.04	0.01	0.76
APY3000top	-0.68	0.21	0.97	0.02	0.62
APY3000random	-0.52	0.26	1.03	0.02	0.60
APY10000random	-2.04	0.21	1.02	0.02	0.69
APY15000top	-2.15	0.18	1.02	0.01	0.75
APY15000random	-2.10	0.19	1.03	0.02	0.73

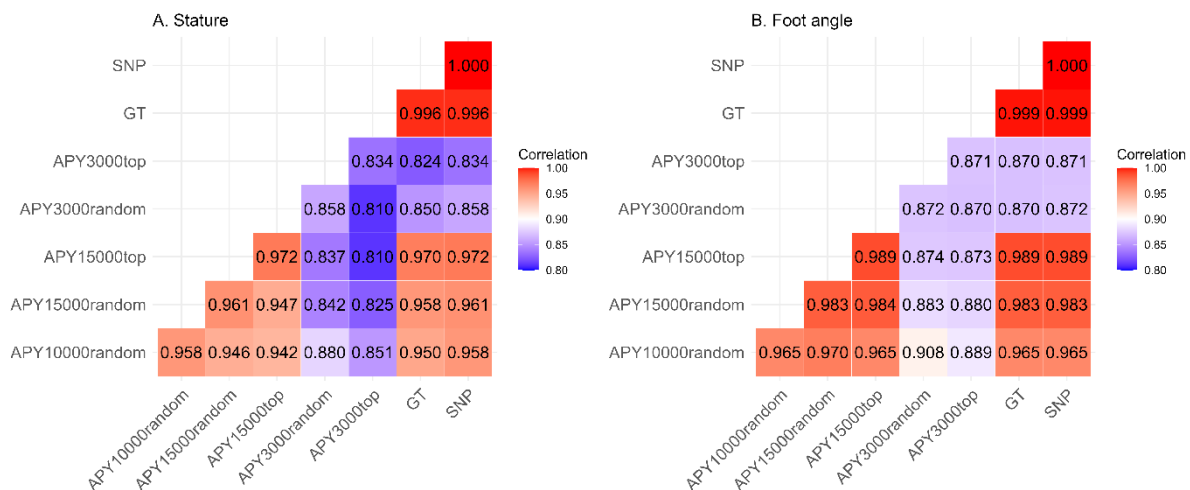


Figure 1. Pearson correlation coefficients of GEBVs predicted by different model variants

**GEBV predictions**

Figure 2 shows the differences between the GEBVs predicted by ssSNP-BLUP compared to the GEBVs predicted by the six ssG-BLUP implementations. For the difference in GEBV between SNP-BLUP and GT, all genotyped animals were considered, while for all comparisons involving APY, only the core individuals were used. The original solutions were rescaled for each model by subtracting the mean of a cow base population to provide GEBVs comparable across all models. For stature, the differences were generally small across bull birth years (2A), however, for foot angle a different pattern emerged (2B) as for all comparisons, except GT and the most informative APY15000top, the differences were unstable across bull birth years. Furthermore, we compared correlations of GEBVs predicted by the full and the truncated models for individuals representing base cows (i.e. cows with phenotype records) and bulls (i.e. bulls, which have a minimum

of one daughter). The dashed line divides the plot into animals defined as old and young in the validation process. For stature, regardless of the model, correlations between full and truncated data sets, for old animals, were very close to unity. For young candidates, we observed a declining trend with the lowest correlations calculated for APY-3000random (Figure 3 A). For foot angle, similar results were obtained, except for APY3000random for females, where we observed a decreasing correlation trend as early as from 2004. From 2008, correlations dropped below 0.9 (Figure 3 B). Finally, we compared the overlap of bulls with the top 50 GEBV predictions resulting from different models (Figure 4). The number of bulls common across all models for stature (Figure 4 A) was 39, while the highest number of exclusive bulls (3) was observed in the top 50 list resulting from the APY15000top. Regarding foot angle (Figure 4 B), the number of common bulls was even lower, being 31, while 8 were exclusive for the APY3000random.

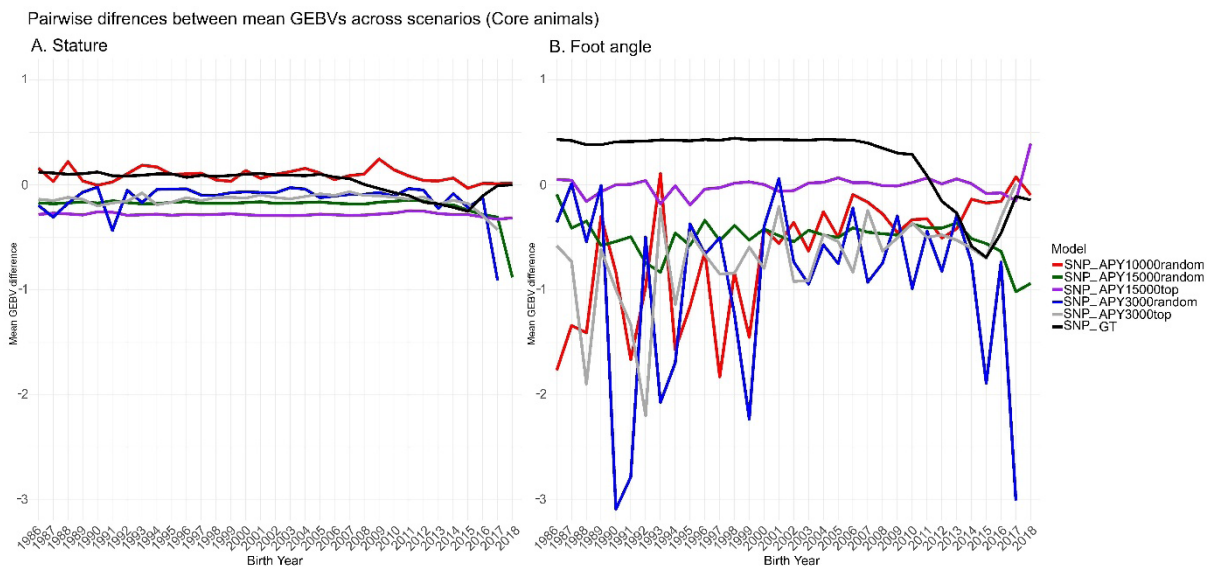


Figure 2. Mean GEBV difference divided by the genetic standard deviation between SNP-BLUP and GT, as well as SNP-BLUP and APY models implementing different scenarios of core individual selection. (A) stature, (B) foot angle

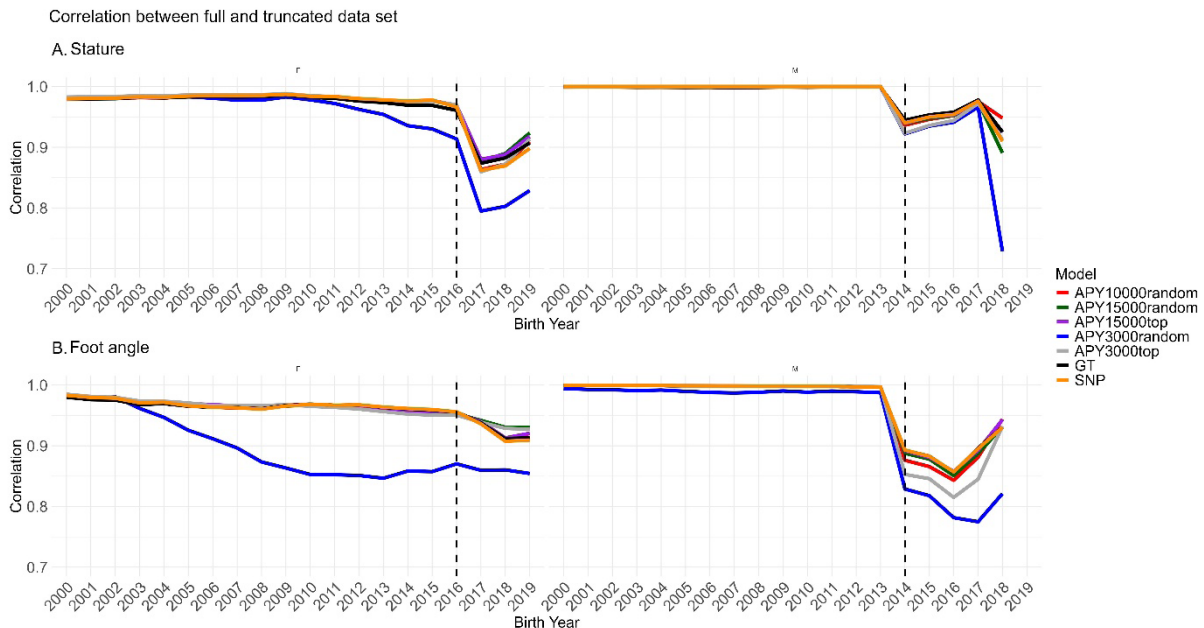


Figure 3. Correlation between full and truncated data set for cows with phenotype records and bulls with a minimum of one daughter. (A) stature, (B) foot angle



and APY3000 were the fastest, closely followed by APY10000 random. The wall clock time of GT was the longest, twice the time of APY15000top. The peak memory consumption by SNP-BLUP was on the order of ten times lower than for the remaining solvers. In the case of iterations, SNP-BLUP required the most iterations (673 for stature and 1,027 for foot angle) and an average of 2.3 seconds per iteration. The lowest number of 335 iterations and an average of 0.18 seconds per iteration were used for stature by APY3000top and 499 and 0.18 seconds for foot angle by APY3000random.

## Discussion

Many studies have been conducted related to comparing single-step genomic prediction models with conventional two-step approaches. However, the literature on comparisons within the single-step modelling frame is scarce. Koivula et al. (2012) considered genomic prediction models similar to (1) and (2). However, they addressed only the genotyped part of the population. Similarly to the results of our study, these authors observed very high correlations between models for predicted bull GEBVs or DGVs (direct genetic values) and similar validation results. Gao et al. (2018) also considered the validation performance of single-step models (Masuda et al., 2016). Although no marked differences were observed, the authors indicated that for APY-based solvers, not only the number but also the selection of core individuals was a crucial step influencing the prediction accuracy (Masuda et al., 2016; Misztal et al., 2016, 2014; Strandén and Garrick, 2009). Except worse prediction performance of APY3000random for cows, no marked differences due to the composition of the core animal set were observed in our study. Still, Macedo et al. (2022) demonstrated low robustness in predictions with ssG-BLUP model towards differential handling of missing parent information. Moreover, our results demonstrated that a large number of core individuals is recommended, provided computational resources, especially RAM, are available, to obtain stable prediction with ssG-BLUP. Similarly to the results of Misztal et al. (2014), we also did not observe a marked gain in the prediction quality of GEBV when using more than 10,000 core individuals.

Regarding the computational aspects, we observed very large differences in RAM consumption between SNP-BLUP and other solvers. GT needed over 10 times more RAM and fewer iterations than SNP-BLUP, as was also demonstrated by Vandenplas et al. (2023). Two possible ways to circumvent the problem of the optimal choice of core individuals are either the use of GT or SNP-BLUP solver-based prediction. The application of GT comes with the price of high memory requirements and long computing times. SNP-BLUP does not consume much memory and is computationally more efficient. Still the approaches did not yield identical rankings of top 50 bulls. Although 62% (foot angle) and 75% (stat-

ure) bulls were common in rankings across scenarios, there was still a considerable number of individuals that were ranked scenarios specific.

## Conclusion

Regarding the prediction of GEBV on the active population scope, no marked differences between solvers (except the APY with only 3,000 core individuals) were observed, still GT and SNP-BLUP solvers resulted in the highest prediction accuracy, while the effectiveness of the APY model depends on the size of the set of core animals: small cores lead to a noticeable decrease in accuracy, while cores with large numbers minimize these differences. Another factor influencing the choice of the solver is its computational efficiency expressed by the computation time and memory resources required, which was also indicated by Koivula et al. (2012). However, it should be kept in mind that the ranking of the top bulls is not identical between models, which has implications for the breeding industry in terms of semen pricing and selection since, typically, the top-ranking bulls are the most widely used.

Considering the abovementioned aspects, the results of our study indicate that GT and SNP-BLUP solvers offer more modelling and computational advantages.

## CRedit authorship contribution statement

Dawid Słomian: data curation, formal analysis, investigation, methodology, validation, visualization, writing – original draft. Joanna Szyda: conceptualization, methodology, project administration, supervision, writing – original draft.

## Declaration of competing interest

The authors declare no competing interests.

## Funding

This study was supported by a grant from the Ministry of Agriculture and Rural Development (DŹW. pp.862.1.2023).

## References

- Aguilar I., Misztal I., Johnson D.L., Legarra A., Tsuruta S., Lawlor T.J. (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.
- Christensen O.F., Lund M.S. (2010). Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol.*, 42: 2.
- Fragomeni B.O., Lourenco D.A.L., Tsuruta S., Masuda Y., Aguilar I., Legarra A., Lawlor T.J., Misztal I. (2015). Hot topic: Use of genomic recursions in single-step genomic best linear unbiased predictor (BLUP) with a large number of genotypes. *J. Dairy Sci.*, 98: 4090–4094.
- Gao H., Koivula M., Jensen J., Strandén I., Madsen P., Pitkänen T., Aamand G.P., Mäntysaari E.A. (2018). Short communication: Genomic prediction using different single-step methods in the Finnish red dairy cattle population. *J. Dairy Sci.*, 101: 10082–10088.
- Koivula M., Strandén I., Su G., Mäntysaari E.A. (2012). Different methods to calculate genomic predictions – Comparisons of

- BLUP at the single nucleotide polymorphism level (SNP-BLUP), BLUP at the individual level (G-BLUP), and the one-step approach (H-BLUP). *J. Dairy Sci.*, 95: 4065–4073.
- Legarra A., Christensen O.F., Aguilar I., Misztal I. (2014). Single Step, a general approach for genomic selection. *Livest. Sci.*, 166: 54–65.
- Liu Z., Goddard M.E., Hayes B.J., Reinhardt F., Reents R. (2016). Technical note: Equivalent genomic models with a residual polygenic effect. *J. Dairy Sci.*, 99: 2016–2025.
- Liu Z., Goddard M.E., Reinhardt F., Reents R. (2014). A single-step genomic model with direct estimation of marker effects. *J. Dairy Sci.*, 97: 5833–5850.
- Lourenco D., Legarra A., Tsuruta S., Masuda Y., Aguilar I., Misztal I. (2020). Single-step genomic evaluations from theory to practice: using SNP chips and sequence data in BLUPF90. *Genes*, 11: 790.
- Macedo F.L., Astruc J.M., Meuwissen T.H.E., Legarra A. (2022). Removing data and using metafounders alleviates biases for all traits in Lacaune dairy sheep predictions. *J. Dairy Sci.*, 105: 2439–2452.
- Mäntysaari E.A., Koivula M., Strandén I. (2020). Symposium review: Single-step genomic evaluations in dairy cattle. *J. Dairy Sci.*, 103: 5314–5326.
- Mäntysaari E.A., Liu Z., VanRaden P. (2010). Interbull validation test for genomic evaluations. *Interbull Bulletin*, 17.
- Masuda Y., Misztal I., Tsuruta S., Legarra A., Aguilar I., Lourenco D.A.L., Fragomeni B.O., Lawlor T.J. (2016). Implementation of genomic recursions in single-step genomic best linear unbiased predictor for US Holsteins with a large number of genotyped animals. *J. Dairy Sci.*, 99: 1968–1974.
- Misztal I. (2016). Inexpensive computation of the inverse of the genomic relationship matrix in populations with small effective population size. *Genetics*, 202: 401–409.
- Misztal I., Legarra A., Aguilar I. (2014). Using recursion to compute the inverse of the genomic relationship matrix. *J. Dairy Sci.*, 97: 3943–3952.
- Strandén I., Garrick D.J. (2009). Technical note: Derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit. *J. Dairy Sci.*, 92: 2971–2975.
- Vandenplas J., ten Napel J., Darbaghshahi S.N., Evans R., Calus M.P.L., Veerkamp R., Cromie A., Mäntysaari E.A., Strandén I. (2023). Efficient large-scale single-step evaluations and indirect genomic prediction of genotyped selection candidates. *Genet. Sel. Evol.*, 55: 37.
- Vandenplas J., Veerkamp R.F., Calus M.P.L., Lidauer M.H., Strandén I., Taskinen M., Schrauf M.F.S.G., ten Napel J. (2022). MiXBLUP 3.0 – software for large genomic evaluations in animal breeding programs. *Proc. 12th World Congress on Genetics Applied to Livestock Production (WCGALP)*, Rotterdam, the Netherlands: Wageningen Academic Publishers, pp. 1498–1501.

Received: 15 II 2025

Accepted: 8 VIII 2025