e-ISSN 1734-9168
http://www.isez.pan.krakow.pl/en/folia-biologica.html

Folia Biologica (Kraków), vol. 73 (2025), No 1
https://doi.org/10.3409/fb_73-1.03

Comparative analysis of Bayesian methods and GBLUP for genomic evaluations in dairy cattle

Magdalena Kolenda, Hasan Önder, Dariusz Piwczyński, Burcu Kurnaz and Beata Sitkowska

Accepted April 08, 2025 *Published online April 28, 2025* *Issue online April 29, 2025*

Original article

Kolenda M., Önder H., Piwczyński D., Kurnaz B., Sitkowska B. 2025. Comparative analysis of Bayesian methods and GBLUP for genomic evaluations in dairy cattle. Folia Biologica (Kraków) **73**: 21-29.

In this study, the effects of Genomic Best Linear Unbiased Prediction (GBLUP) and Bayesian alphabet methods (A, B, C and Cp) were investigated on genomic predictions and indirect estimations in the Polish Holstein Friesian (PHF) dairy cattle population. The study analysed the milk yield data (MY, kg/lactation) and 13,481 single nucleotide polymorphism (SNP) genotype records from 534 Polish Holstein Friesian (PHF) dairy cattle raised on private farms in Poland. The quality control of the genotypic data included the removal of monomorphic loci and the exclusion of samples with SNP missing rates exceeding 10%. After the quality control, 493 animals and 13,250 SNPs were retained for the genomic prediction. Marker effects and genomic breeding values (GEBVs) were calculated using the Bayesian alphabet and GBLUP. The results indicated that, for the milk yield of PHF cows, the Bayes C method outperformed other approaches. This method achieved the highest prediction accuracy among the evaluated methods. Additionally, the Bayes C method required the shortest computational time, underscoring its efficiency.

Key words: Bayesian alphabet, genomic prediction, Holstein Friesian, milk yield.

Magdalena Kolenda✉, Dariusz Piwczyński, Beata Sitkowska, Bydgoszcz University of Science and Technology, Faculty of Animal Breeding and Biology, Department of Animal Biotechnology and Genetics, Bydgoszcz, Poland.
E-mail: kolenda@pbs.edu.pl
Hasan Önder, Burcu Kurnaz, Ondokuz Mayis University, Faculty of Agriculture, Department of Animal Science, Samsun, Türkiye.

Increasing the production and quality of animal-derived foods, which play a critical role in human nutrition, requires improvements in both the environmental conditions and genetic structures of farm animals (Abaci & Önder 2020). The primary goal of animal breeding is to increase the frequency of the desired alleles of genes associated with economically important traits in subsequent generations, thereby enabling farm animals to supply higher and better-quality production under controlled environmental conditions (Olfaz *et al.* 2019). Traditional animal breeding methods have been successfully used to estimate genetic parameters based on pedigree information. Unlike traditional animal breeding methods, advanced methods allow for a more accurate estimation of breeding values by incorporating variations in DNA sequences (Goddard & Hayes 2007). The Marker-Assisted Selection (MAS) has been a pioneering approach in integrating genomic information into breeding programmes and was first applied to dairy cattle in the 1980s. Its main benefit has been the extensive use of young bulls, based on combined pedigree and marker information. Since the early 21st century, many studies have employed

this method, especially in relation to dairy cattle breeding (Weller *et al.* 2017). With this approach, selection decisions are made based on a small number of markers that are individually associated with the target traits.

The markers used for MAS can be linked to the QTL, but in linkage equilibrium with it; while in linkage disequilibrium (LD), the QTL or the marker can be the QTL (Dekkers 2004). If the marker is in linkage equilibrium with the QTL. All QTL alleles in the founder animals are considered to be different and hence the number of QTL alleles whose effects must be estimated is further increased. Despite these difficulties, Boichard *et al*. (2012) showed how gains can be made, although a very large amount of genotyping was necessary. To overcome these difficulties, Meuwissen *et al*. (2001) proposed a variant of MAS that they called genomic selection. The key features of this method are that markers covering the whole genome are used, so that potentially all the genetic variance is explained by the markers. Furthermore, the markers are assumed to be in LD with the QTL, so that the number of effects per QTL to be estimated is small. Using simulations, they showed that the breeding value could be predicted with an accuracy of 0.85 from marker data sequences alone (Goddard & Hayes 2007). Therefore, the selection of animals for breeding can now be based on the cumulative effects of all markers across the genome (Meuwissen *et al.* 2001; De Koning *et al.* 2016).

Genomic selection (GS) has significantly transformed animal breeding activities. Unlike traditional breeding methods, GS is used to estimate the genetic breeding values of all animals with known genotypes, even before their phenotypes have been measured, by utilising marker effect coefficients (Ding *et al.* 2013). In contrast to GS, which relies on a limited number of markers, the calculation of Genomic Estimated Breeding Values (GEBVs) involves the genomic prediction (GP) of additive genetic traits of the animals under consideration for selection (Taylor *et al.* 2016; Abaci & Önder 2020). The genomic breeding value can be obtained through two approaches: direct and indirect. In the direct method, genomic breeding values are estimated in a single step using mixed equation models for individuals with known phenotypes and genotypes. In the indirect method, there are two steps. In the first step, marker effects are estimated using a training population. In the second step, breeding value estimates are obtained using only marker effects for the genotyped population and the population under selection (Calus *et al.* 2008; Hayes *et al.* 2009). To shorten the generation interval and to evaluate a large number of candidates for selection, it is essential to achieve the highest possible accuracy in the estimation of these breeding values (Brito *et al.* 2021).

Different statistical approaches using marker effects, including linear mixed models and Bayesian mixture models, have been applied in order to evaluate GEBVs. These methods, as described by Meuwissen *et al*. (2001), have been instrumental in assessing the accuracy and performance of GP estimations (Taylor *et al.* 2016). The advent of medium-density or high-density SNP chips (containing >50,000 markers and higher) in 2008 marked a turning point for genomic selection. This technology has since been successfully implemented in many countries. In nations with high milk production levels, the application of breeding selection based on GEBVs has brought about transformative advancements in the dairy industry (Hayes *et al.* 2009).

In Poland, Holstein Friesian cattle constitute the dominant dairy breed, contributing significantly to the national milk production. Implementing genomic selection in Poland has the potential to improve the genetic progress, increase the selection accuracy and enhance economically important traits, such as the milk yield, fertility, and udder health. In this study, we decided to utilise SNP data that breeders have already genotyped, to ensure that our findings are directly applicable to ongoing breeding programmes without requiring additional costly genotyping efforts.

The aim of the study is to evaluate the effects of Genomic Best Linear Unbiased Prediction (GBLUP) and Bayesian alphabet methods (A, B, C and Cp) on single-step genomic predictions and indirect estimations in the Polish Holstein Friesian dairy cattle population.

Material and Methods

As per Resolution No. 13/2016 of the National Ethics Committee for Animal Experiments (Poland) dated 17 June 2016, ethical approval was not required in order to collect the animal material for genotyping (Approval No: 13/2016).

Materials

The genetic material used for the study was collected during a routine estimation of breeding values. The genotyping was conducted by the Polish Federation of Cattle Breeders and Dairy Farmers (PFCB&DF, Warsaw, Poland) using customised EuroGenomics microarrays: EuroGenomics v3_POL; EuroGenomics v4_POL; EuroGenomics v5_POL;

EuroGenomics v6_POL; EuroGenomics v8b_POL; and EuroGenomics MD_POL, following the Infinium HD Illumina protocol. Over time, multiple microarray chips were employed, each differing in the number and composition of the included SNPs. To ensure consistency, a selection of SNPs was performed. The dataset included milk yield data (MY, kg/lactation) purchased from PFCB&DF and 13,481 SNP genotype records from 534 Polish Holstein Friesian (PHF) dairy cattle reared by private enterprises in Poland. In the quality control process for the genotype data, SNPs with a minor allele frequency (MAF) of at least 5% were retained, while animals with more than 10% missing SNPs were excluded from the analysis. Following the quality control, 493 animals and 13,250 SNPs were used for the genomic predictions (Lee *et al*. 2019; Kudinov *et al*. 2022).

Methods

For the predictions, the Markov chain and Monte Carlo were employed, running for 50,000 iterations of Gibbs sampling. Convergence was successfully achieved, with the first 5,000 cycles discarded as burn-in and excluded from the analysis (Abaci & Önder 2020).

Out of 493 animals, 400 were randomly allocated to the training population, while the remaining 93 comprised the test population. The lactation order (OL, categorised as 1, 2 and 3) was used as a fixed factor and the days in milk (DIM) as a random factor. All the data was collected from the same year and the herd effect was statistically not significant ($p>0.05$) As a result, it was removed from the final model to avoid unnecessary complexity.

The use of Bayesian methods for animal breeding was first introduced by Gianola and Fernando (1986). Marker effects and genomic breeding values were calculated using the Bayesian alphabet methods and GBLUP (Bayes C0), following the model presented by Meuwissen *et al*. (Gianola *et al.* 2009).

$$y = Xs + C\beta + W\alpha + e \quad (1)$$

Where: y is a vector of the phenotypes, **X** is the incidence matrix for fixed effects, s is the vector of fixed effects, **C** is the covariate design matrix, β is the vector of covariate effect, **W** is a known matrix of the numerical genotype scores for each marker (AA, AB, and BB for 10, 0, and -10, respectively), α is a marker for the additive effects vector and e is a random error vector with $e\sim N(0, I\delta^2_e)$, where **I** is the identity matrix and δ^2_e is the random error variance.

The mathematical model of the marker-based methods with polygenic effects was used as follows (Abaci & Önder 2020):

$$y = \mu 1_n + \mathbf{X}s + C\beta + \sum_j W_j \alpha_j \delta_j + Zu = e \quad (2)$$

Where: y is an N×1 vector of the phenotypes where N is the number of individuals, μ is the overall mean, 1_n is a vector of ones, **X** is an incidence matrix for constant effects (order of lactation), s is a constant effects vector, **C** is a covariate design matrix, β is an effect vector of a covariate, W_j is an N×1 vector of the genotypes at SNP j which was coded (10, 0, -10), α_j is the random allele substitution effect for SNP j, δ_j is a 0/1 indicator variable which equals 1 if SNP j is included in the model and zero otherwise, **Z** is the associated design matrix, u is a vector with random polygenic effects of all the individuals with Var(u) = $A\delta^2_u$ (A is the numerator relationship matrix and δ^2_u is the polygenic variance), and e is a vector of the random residuals $e \sim N(0, I\delta^2_e)$, where **I** is the identity matrix and δ^2_e is the random error variance.

Bayes A Method

With Bayes A, it is assumed that most SNPs have a minor individual effects on the trait, while only a few exhibit a moderate to high impact. The prior distribution of marker variance follows a scaled inverted chi-square distribution $\chi^{-2}(v, S)$, where S is the scale parameter and v is the number of the degree of freedom. The prior distribution of the marker substitution effect (α_j) is assumed to be normal, with a mean of zero and a variance of $\delta^2_{\alpha j}$. To fit all markers into the model, all δ_j are set as 1. The marginal SNP distribution follows the Student's t-distribution, allowing for a higher probability of moderate to large SNP effects compared to a normal distribution and pi=0 (Meuwissen *et al.* 2001; Hayes *et al.* 2009; Abaci & Önder 2020).

Bayes B Method

The distribution of genetic variances across *loci* shows that only a few *loci* exhibited genetic variance, while many *loci* showed none. However, the prior density of the Bayes A method lacked a density peak at $\delta^2_{gj} = 0$, where its probability is infinitesimal. Method Bayes B, therefore, uses a prior distribution with a high density π at $\delta^2_{\alpha j} = 0$ and has an inverted chi-square distribution for $\delta^2_{\alpha j} > 0$ (Meuwissen *et al.* 2001; Abaci & Önder 2020).

Bayes C Method

A general consensus reveals that the full-conditional posterior distribution of a locus-specific variance adds only one to the degrees of freedom in Bayes A and Bayes B, when compared with its prior assumption, while in Bayes Cπ it increases the value by the number of markers that have effects in each iteration (Zhu *et al.* 2016), so the shrinkage of SNP effects largely depends on the scale parameter. To overcome this limitation, the proposed method is Bayes C, which involves estimating a single variance common to all SNPs, thereby reducing the influence of the scale parameter. In Bayes C, π is assumed to be known and specified by the user (Abaci & Önder 2020).

Bayes Cp Method

The Bayes Cp method assumes a mixture model for marker effects, where the elements of the Xs vector were calculated for each animal using the following formula:

$$Xs = \sum_{i=1}^{N} (\alpha_j \delta_j I_j) \quad (3)$$

Where: α_j is the genotype of the j^{th} marker, j is the effect of the j^{th} marker, and I_j is the indicator variable. Unlike Bayes C, an additional feature of Bayesian Cp is a prior distribution assigned to π (Neves *et al.* 2012; Abaci & Önder 2020).

GBLUP Method

The Genomic Best Linear Unbiased Prediction (GBLUP) method fits the model using all SNPs and assumes that each SNP contributes equally to the total genetic variance. This method is equivalent to Bayes C when Pi = 0 and, to simplify matters, is called Bayes C0 (Abaci & Önder 2020).

Accuracy

For an indirect prediction, marker effects were determined by using phenotypic and genotypic recordings from 400 animals (training population), to determine the accuracy of the examined methods for predicting the genomic breeding value. Subsequently, estimated genomic breeding values for the test population, consisting of 93 animals, were calculated using only genotypic data (Abaci & Önder 2020). The GEBVs of the individuals in the test population were predicted using the following formula:

$$GEBV = \sum_{j=1}^{k} z_{ij} \bar{\alpha}_j \quad (4)$$

where: GEBV is the genomic estimated breeding value for an individual animal *i* in the test population, k is the total number of markers (13,250), z_{ij} is the genotype of an individual *i* for marker j, and $\bar{\alpha}_j$ is the posterior mean effect of marker j.

Due to the significant influence of prior values for residual and genetic variances on the results of Bayes A, the results from Bayes C were used as the prior values for Bayes A.

The theoretical accuracy (R^2) was calculated using the formula $(1-(PEV/\delta_u^2))$, where PEV is the prediction error variance and δ_u^2 is the polygenic variance. For the calculation of the PEV, the method suggested by Wang *et al.* (2015) was used. The prediction accuracy was evaluated using the linear regression (LR) and Pearson correlation methods (Abaci & Önder 2020; Cesarani *et al.* 2021; Mancin *et al.* 2021). To compare the correlation coefficients and the regression coefficients, Fisher's z-test was used (Werts *et al.* 1976). The Bayesian methods (Bayes A, Bayes B, Bayes C, Bayes Cπ and Bayes C0 (GBLUP)) were implemented using the online GenSel package in the Cy-Verse cyberinfrastructure, within the Discovery Environment web interface (https://de.cyverse.org/).

Results

Table 1 presents the descriptive statistics for the milk yield (MY, kg/lactation) based on the lactation number (LN) with the days in milk (DIM) as a covariate. The analysis of variance showed that the LN had a statistically significant ($p < 0.01$) effect on the milk yield with a covariate DIM correction. Days in milk had a statistically significant ($p < 0.01$) impact on the MY, similarly to when the lactation order (1, 2 and 3) was used as a fixed factor and the days in milk as a random factor in the genomic prediction analysis.

Table 1

Descriptive statistics of the milk yield (MY, kg/lactation) according to the lactation number (LN)

Lactation number	Milk yield (kg)*
1	13,130.61 ± 116.96[b]
2	15,311.05 ± 231.20[a]
3	15,174.06 ± 269.64[a]
Significance	<0.001

*: Covariates appearing in the model are evaluated at the values of the days in milk (DIM) = 369.6077; a,b: in the columns, values marked with different letters show the statistical difference (P<0.01).

Table 2

Results of the variance components

	Bayes A	Bayes B	Bayes C	Bayes Cπ	GBLUP
Residual Variance	2,377,380	2,451,950	2,331,910	2,323,240	2,420,610
Genetic Variance	2,847,840	2,747,280	2,773,180	2,786,310	2,668,390
Heritability	0.55	0.53	0.54	0.55	0.52
Computing time (sec)	1352	1914	391	788	1081
π	0.00	0.95	0.95	0.53	0.00
R^2	0.62	0.62	0.62	0.62	0.61

Table 2 presents the residual variance, genetic variance, heritability and computation time, Pi value and the theoretical accuracy R^2 (for the direct prediction) related to the genetic parameters obtained using different methods.

The Bayes A method provided the highest heritability estimate, while GBLUP gave the lowest heritability. The computing times per model ranged between 391 seconds for Bayes C and 1,914 seconds for Bayes B. The highest genetic variance was observed for Bayes A, whereas GBLUP showed the lowest genetic variance. The highest coefficient of the determination value for a single-step prediction was achieved using Bayes, with the lowest observed for GBLUP.

It was observed that the breeding values obtained by GBLUP and Bayes Cπ methods (Table 3) had the highest positive correlation (1.000). However, the correlation between the milk yield (MY, kg/lactation) and all estimation methods was low (Table 3).

The accuracy (*r*) and deviations (*b*) of the different methods were calculated and are presented in Table 4. There were no statistical differences obtained for the accuracy and deviations among the models. The highest accuracy (correlation) between the breeding values was obtained from the Bayes C method and MY, while the lowest accuracy value was observed for the GBLUP method. While the highest deviation was calculated for GBLUP, the lowest deviation was observed for Bayes A. The results indicate that the Bayes C method demonstrated the best accuracy for estimating breeding values (Table 4).

Table 3

Correlations between the breeding values calculated using different methods

Methods	Bayes B	Bayes C	Bayes Cπ	GBLUP	MY
Bayes A	0.994	0.999	1.000	1.000	0.638
Bayes B		0.997	0.993	0.992	0.632
Bayes C			0.999	0.998	0.640
Bayes Cπ				1.000	0.641
GBLUP					0.636

MY: Milk yield

Table 4

Accuracy and deviations of the different methods, with standard error values in parentheses

Methods	Accuracy	Deviations
Bayes A	0.905 (0.034)	1.596 (0.085)
Bayes B	0.899 (0.035)	1.60 (0.090)
Bayes C	0.908 (0.035)	1.61 (0.089)
Bayes Cπ	0.906 (0.035)	1.61 (0.089)
GBLUP	0.895 (0.035)	1.64 (0.093)

Table 5

Pearson correlation (upper diagonal) and Spearman rank correlation (lower diagonal) values of the different methods for marker effects

	Bayes A	Bayes B	Bayes C	Bayes Cπ	GBLUP
Bayes A		0.727	0.945	0.991	0.987
Bayes B	0.985		0.854	0.653	0.629
Bayes C	0.995	0.992		0.919	0.902
Bayes Cπ	0.998	0.982	0.994		0.997
GBLUP	0.997	0.981	0.993	0.998	

The highest Pearson correlation of the marker effects was estimated between Bayes Cπ and GBLUP, while the lowest was between Bayes B and GBLUP. The highest Spearman rank correlation of the marker effects was found between Bayes Cπ and GBLUP, while the lowest correlation was calculated between Bayes B and GBLUP (Table 5).

Discussion

When evaluated without considering the breed, the heritability estimates for milk yield fall between 0.19 and 0.45 (Abaci & Önder 2020). Kudinov *et al.* (2022) reported a heritability value of 0.21 for the milk yield in Russian dairy cattle. Lee *et al.* (2019), by using GBLUP for the milk yield for Holstein dairy cattle, indicated a heritability value of 0.28, while Mancin *et al.* (2021), also using GBLUP, reported a value of 0.25 for the Rendena breed. Meuwissen *et al.* (2001) estimated the heritability value at 0.50 by using Bayes A and Bayes B with simulated data. Mucha *et al.* (2019) reported a 0.33 heritability value for Holstein dairy cattle. For milk cheese-making traits, the heritability value for the Montbéliarde breed was estimated at 0.70 (Sanchez *et al.* 2022). Cesarani *et al.* (2021) reported a heritability value of 0.23 for the milk yield in Italian Buffalo. By contrast, the heritability values observed in our study were higher than those reported in most previous studies (0.523 and 0.546). This high heritability may be attributed to the structure of the dataset, which included 13,481 SNP and demonstrated a high genetic variance versus the residual variance. However, this cannot be attributed directly to the SNP number. The results of Stanton-Geddes *et al.* (2013) showed that the heritability estimates with 250,000 and 25,000 SNPs were very similar to those obtained with more than 5 million SNPs. Conversely, with 2,500 SNPs, the heritability values were lower and had a higher variance than those with at least 25,000 SNPs. The heritability estimates were slightly lower when common SNPs were used. Additionally, it may indicate evidence of a weak selection pressure being applied to the population.

Reliability (R^2) has been observed at the levels of 0.42 and 0.61 (training: 80%) using Bayes C by Kudinov *et al.* (2021, 2022), 0.67 using GBLUP (training: 78%) by Sanchez *et al.* (2022), 0.17 using ssGBLUP by Lee *et al.* (2019) and 0.35 using GBLUP (training: 80%) by Ma *et al.* (2015). The literature shows that the reliability of the estimations ranges from 0.17 to 0.67. The results of our study showed reliability values between 0.6108 and 0.6240 (training: 81%), which are aligned with the data reported in the literature on single-step prediction methods. These differences cannot be attributed to the training set percentage, as Fernández-González *et al.* (2023) mentioned that regarding the optimal training set size, the maximum accuracy was obtained when the training set was the entire candidate set. They pointed out that 50-55% of the candidate set was enough to reach 95-100% of the maximum accuracy in the targeted scenario. These differences might have occurred due to SNP variance.

In our study, the accuracy was reported at the level of 0.895 for GBLUP and 0.908 for Bayes C. Comparatively, Lee *et al.* (2019) found 0.35, Ding *et al.* (2013) reported 0.36 and Cesarani *et al.* (2021) estimated the accuracy for Italian Buffalo at 0.82. Mancin *et al.* (2021) estimated an accuracy of 0.653 using GBLUP. Boichard *et al.* (2012) predicted a reliability of 0.56 for French Holstein, Montbéliarde and Normande breeds. Ding *et al.* (2013) estimated reliability values as 0.72 using Bayes B and as 0.76 using GBLUP. Meuwissen *et al.* (2001) reported accuracies of 0.732 using Bayes A and 0.848 using Bayes B; whereas Mucha *et al.* (2019) reported an accuracy of 0.72.

Lee *et al.* (2019) reported that the mean accuracy of the direct genomic values for milk production traits was comparable between Bayes B and GBLUP methods. Similarly, Mancin *et al.* (2021) indicated

BGLUP was the model with the best overall result, showing a higher accuracy than Phenotypic BLUP along with optimal values of the bias and dispersion parameters. The findings of our study showed higher accuracies compared to most studies in the literature, except for the study by Sungkhapreecha *et al*. (2021), who reported higher values. It is also worth noting that the year of publication plays a critical role: older studies tend to show lower accuracies compared to newer ones, which is likely due to improvements in the data quality and computational methods.

The deviation, defined as the regression coefficient, was estimated as 0.80 by Kudinov *et al*. (2022) and 0.99 by Kudinov *et al*. (2021) using the GBLUP method. Lee *et al*. (2019) reported a deviation of 1.75; while Ma *et al*. (2015), using GBLUP, described it as 0.75 for Jersey cattle. Meuwissen *et al*. (2001) indicated that the deviations were 0.827 for Bayes A and 0.946 for the Bayes B method. Mucha *et al*. (2019) reported the highest deviation, with a value of 3.83. Lee *et al*. (2019) argued that Bayes B tends to show a relatively lower bias compared to the GBLUP method for Korean Holstein populations. Consistent with these findings, our study confirmed that the Bayes methods generally demonstrate a smaller bias than the GBLUP method.

When the correlations between the breeding values estimated using different methods were analysed, the highest correlation was observed between the milk yield and Bayes Cπ (0.641), while the lowest correlation (0.636) was between the true milk yield and the GBLUP method. The correlations among the evaluated methods were generally high and positive. The correlations of the marker effects were further analysed in this study using the Pearson correlation, while the ranking of the marker effects was evaluated using the Spearman rank correlation. The lowest Pearson correlation coefficient for the marker effect value was observed between Bayes B and GBLUP (0.629), whereas the highest was between Bayes Cπ and GBLUP (0.997). The lowest Spearman rank correlation coefficient for the marker effect rankings was observed between Bayes B and GBLUP (0.981), while the highest was found between Bayes Cπ and GBLUP, as well as between Bayes Cp and Bayes A (0.998). The Pearson correlations were lower than the Spearman rank correlations, indicating that while the marker effect values show lower correlations across these methods, the weight of the markers in the models is more consistent across these methods.

Our results showed that the Bayes C method was the most effective for single-step genomic predictions and the indirect estimations of the milk yield of Polish Holstein Friesian dairy cattle, exhibiting the highest reliability and accuracy among the examined methods. Moreover, the Bayes C method required the shortest computation time. However, Abacı and Önder (2020) reported that the Bayes A and Bayes B methods were preferred over Bayes C, Bayes Cπ and GBUP methods. In theory, Bayes C and Bayes Cπ are considered to be more robust than the Bayes A and B methods. Furthermore, Bayes A is highly sensitive to prior variance values, which can significantly influence the results (Neves *et al*. 2012). Many researchers have employed the GBLUP method, especially for single-step predictions (Boichard *et al*. 2012; Ding *et al*. 2013; Cesarani *et al*. 2021; Naserkheil *et al*. 2021; Steyn *et al*. 2021; Sungkhapreecha *et al*. 2021b; Kudinov *et al*. 2022; Sanchez *et al*. 2022; Önder *et al*. 2023). Still, our study showed that the GBLUP method was the least accurate among those tested. In conclusion, our study suggests that the Bayes C method is highly suitable for both single-step genomic predictions and indirect estimations of dairy cattle milk yields.

Conclusions

The results of this study can be applied for predicting the genomic milk yield of dairy cattle. The results suggest that the use of GBLUP should be replaced with the use of the Bayes C method, even though there was no statistical difference among the Bayes methods. Bayes C is only marginally more accurate than the other methods, and a proper cross-validation would have shown that this is indeed the case because of its statistical properties. However, further advancements in genomic prediction methods are needed, particularly in relation to statistical theory and genomic technology, to fully realise the potential of genomic selection.

Funding

This project was financially supported by the Minister of Science under the 'Regional Initiative of Excellence' (RID/SP/0017/2024/01) programme. This article/material has been supported by the statutory activity, BN-WHiBZ-4/2022. This study was supported by the Ondokuz Mayıs University Scientific Research Office with Project Number PYO. ZRT.1901.22.001.

Author Contributions

Research concept and design: M.K., H.Ö.; Collection and/or assembly of data: M.K., B.S.; Data analysis and interpretation: M.K., H.Ö., B.K.; Writing the article: M.K., H.Ö., B.K.; Critical revision of the article: M.K., H.Ö., D.P., B.S.; Final approval of article: M.K., H.Ö., B.S.

Conflict of Interest

The authors declare that there is no conflict of interest.

References

Abaci S.H., Önder H. 2020. Comparison of BLUP and Bayesian methods for different sizes of training population in genomic selection. Turk. J. Vet. Anim. Sci. **44**: 994-1002. https://doi.org/10.3906/vet-2001-52

Boichard D., Guillaume F., Baur A., Croiseau P., Rossignol M.N., Boscher M.Y., Druet T., Genestout L., Colleau J.J., Journaux L., Ducrocq V., Fritz S., Boichard D., Guillaume F., Baur A., Croiseau P., Rossignol M.N., Boscher M.Y., Druet T., Genestout L., Colleau J.J., Journaux L., Ducrocq V., Fritz S. 2012. Genomic selection in French dairy cattle. Anim. Prod. Sci. **52**: 115-120. https://doi.org/10.1071/AN11119

Brito L.F., Bedere N., Douhard F., Oliveira H.R., Arnal M., Peñagaricano F., Schinckel A.P., Baes C.F., Miglior F. 2021. Review: Genetic selection of high-yielding dairy cattle toward sustainable farming systems in a rapidly changing world. Animal **15**: 100292. https://doi.org/10.1016/J.ANIMAL.2021.100292

Calus M.P.L., Meuwissen T.H.E., De Roos A.P.W., Veerkamp R.F. 2008. Accuracy of genomic selection using different methods to define haplotypes. Genetics **178**: 553-561. https://doi.org/10.1534/GENETICS.107.080838

Cesarani A., Biffani S., Garcia A., Lourenco D., Bertolini G., Neglia G., Misztal I., Macciotta N.P.P. 2021. Genomic investigation of milk production in Italian buffalo. Ital. J. Anim. Sci. **20**: 539-547. https://doi.org/10.1080/1828051X.2021.1902404

De Koning D.-J., Meuwissen T.H.E., Hayes B.J., Goddard M.E. 2016. Meuwissen *et al*. on Genomic Selection. Genetics **203**: 5-7. https://doi.org/10.1534/GENETICS.116.189795

Dekkers J.C. 2004. Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. J. Anim. Sci. **82 E-Suppl**. https://doi.org/10.2527/2004.8213_SUPPLE313X

Ding X., Zhang Z., Li X., Wang S., Wu X., Sun D., Yu Y., Liu J., Wang Y., Zhang Y., Zhang S., Zhang Y., Zhang Q. 2013. Accuracy of genomic prediction for milk production traits in the Chinese Holstein population using a reference population consisting of cows. J. Dairy Sci. **96**: 5315-5323. https://doi.org/10.3168/JDS.2012-6194

Fernández-González J., Akdemir D., Isidro y Sánchez J. 2023. A comparison of methods for training population optimization in genomic selection. Theoretical and Applied Genetics **136**: 1-20. https://doi.org/10.1007/S00122-023-04265-6/FIGURES/5

Gianola D., De Los Campos G., Hill W.G., Manfredi E., Fernando R. 2009. Additive Genetic Variability and the Bayesian Alphabet. Genetics **183**: 347-363. https://doi.org/10.1534/GENETICS.109.103952

Gianola D., Fernando R.L. 1986. Bayesian Methods in Animal Breeding Theory. J. Anim. Sci. **63**: 217-244. https://doi.org/10.2527/jas1986.631217x

Goddard M.E., Hayes B.J. 2007. Genomic selection. Journal of Animal Breeding and Genetics **124**: 323-330. https://doi.org/10.1111/J.1439-0388.2007.00702.X

Hayes B.J., Bowman P.J., Chamberlain A.J., Goddard M.E. 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. J. Dairy Sci. **92**: 433-443. https://doi.org/10.3168/JDS.2008-1646

Kudinov A.A., Koivula M., Strandén I., Aamand G.P., Mäntysaari E.A. 2021. Single-step genomic predictions of a minor breed, concurrently with the national genomic evaluations of main breeds. Interbull Bulletin **56**: 174-179.

Kudinov A.A., Mäntysaari E.A., Pitkänen T.J., Saksa E.I., Aamand G.P., Uimari P., Strandén I. 2022. Single-step genomic evaluation of Russian dairy cattle using internal and external information. Journal of Animal Breeding and Genetics **139**: 259-270. https://doi.org/10.1111/JBG.12660

Lee Y.M., Dang C.G., Alam M.Z., Kim Y.S., Cho K.H., Park K. Do, Kim J.J. 2019. The effectiveness of genomic selection for milk production traits of Holstein dairy cattle. Asian-Australas J. Anim. Sci. **33**: 382-389. https://doi.org/10.5713/AJAS.19.0546

Ma P., Lund M.S., Nielsen U.S., Aamand G.P., Su G. 2015. Single-step genomic model improved reliability and reduced the bias of genomic predictions in Danish Jersey. J. Dairy Sci. **98**: 9026-9034. https://doi.org/10.3168/JDS.2015-9703

Mancin E., Tuliozi B., Sartori C., Guzzo N., Mantovani R. 2021. Genomic prediction in local breeds: The rendena cattle as a case study. Animals **11**: 1815. https://doi.org/10.3390/ANI11061815/S1

Meuwissen T.H.E., Hayes B.J., Goddard M.E. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics **157**: 1819-1829. https://doi.org/10.1093/GENETICS/157.4.1819

Mucha A., Wierzbicki H., Kamiński S., Oleński K., Hering D. 2019. High-frequency marker haplotypes in the genomic selection of dairy cattle. J. Appl. Genet. **60**: 179-186. https://doi.org/10.1007/S13353-019-00489-9/TABLES/6

Naserkheil M., Mehrban H., Lee D., Park M.N. 2021. Evaluation of genome-enabled prediction for carcass primal cut yields using single-step genomic best linear unbiased prediction in Hanwoo cattle. Genes (Basel) **12**: 1886. https://doi.org/10.3390/GENES12121886/S1

Neves H.H.R., Carvalheiro R., Queiroz S.A. 2012. A comparison of statistical methods for genomic selection in a mice population. BMC Genet. **13**: 1-17. https://doi.org/10.1186/1471-2156-13-100/COMMENTS

Olfaz M., Tirink C., Önder H. 2019. Use of CART and CHAID Algorithms in Karayaka Sheep Breeding. Kafkas Univ. Vet. Fak. Derg. **25**: 105-110. https://doi.org/10.9775/kvfd.2018.20388

Önder H., Sitskowska B., Kurnaz B., Piwczyński D., Kolenda M., Şen U., Tırınk C., Çanga Boğa D. 2023. Multi-Trait Single-Step Genomic Prediction for Milk Yield and Milk Components for Polish Holstein Population. Animals 2023, Vol. 13, Page 3070 **13**: 3070. https://doi.org/10.3390/ANI13193070

Sanchez M.P., Tribout T., Fritz S., Wolf V., Laithier C., Brochard M., Boichard D. 2022. Opportunities for genomic selection of cheese-making traits in Montbéliarde cows. J. Dairy Sci. **105**: 5206-5220. https://doi.org/10.3168/JDS.2021-21558

Stanton-Geddes J., Yoder J.B., Briskine R., Young N.D., Tiffin P. 2013. Estimating heritability using genomic data. Methods Ecol. Evol. **4**: 1151-1158. https://doi.org/10.1111/2041-210X.12129

Steyn Y., Gonzalez-Pena D., Bernal Rubio Y.L., Vukasinovic N., DeNise S.K., Lourenco D.A.L., Misztal I. 2021. Indirect genomic predictions for milk yield in crossbred Holstein-Jersey dairy cattle. J. Dairy Sci. **104**: 5728-5737. https://doi.org/10.3168/JDS.2020-19451

Sungkhapreecha P., Misztal I., Hidalgo J., Lourenco D., Buaban S., Chankitisakul V., Boonkum W. 2021.a Validation of single-step genomic predictions using the linear regression method for milk yield and heat tolerance in a Thai-Holstein population. Vet. World **14**: 3119. https://doi.org/10.14202/VETWORLD.2021.3119-3125

Sungkhapreecha P., Misztal I., Hidalgo J., Lourenco D., Buaban S., Chankitisakul V., Boonkum W. 2021.b Validation of single-step genomic predictions using the linear regression method for milk yield and heat tolerance in a Thai-Holstein population. Vet. World **14**: 3119. https://doi.org/10.14202/VETWORLD.2021.3119-3125

Taylor J.F., Taylor K.H., Decker J.E. 2016. Holsteins are the genomic selection poster cows. In: Proceedings of the National Academy of Sciences of the United States of America (Vol. 113, pp. 7690-7692). National Academy of Sciences. https://doi.org/10.1073/PNAS.1608144113/ASSET/EE78C606-0200-454A-9C90-44A21AADE5E8/ASSETS/GRAPHIC/PNAS.1608144113FIG01.JPEG

Wang T., Chen Y.P.P., Goddard M.E., Meuwissen T.H., Kemper K.E., Hayes B.J. 2015. A computationally efficient algorithm for genomic prediction using a Bayesian model. Genetics Selection Evolution **47**: 1-16. https://doi.org/10.1186/S12711-014-0082-4/TABLES/10

Weller J.I., Ezra E., Ron M. 2017. Invited review: A perspective on the future of genomic selection in dairy cattle. J. Dairy Sci. **100**: 8633-8644. https://doi.org/10.3168/JDS.2017-12879

Werts C.E., Rock D.A., Linn R.L., Joreskog K.G. 1976. Comparison of correlations, variances, covariances, and regression weights with or without measurement error. Psychol. Bull. **83**: 1007-1013. https://doi.org/10.1037/0033-2909.83.6.1007

Zhu B., Zhu M., Jiang J., Niu H., Wang Y., Wu Y., Xu L., Chen Y., Zhang L., Gao X., Gao H., Liu J., Li J. 2016. The Impact of Variable Degrees of Freedom and Scale Parameters in Bayesian Methods for Genomic Prediction in Chinese Simmental Beef Cattle. https://doi.org/10.1371/journal.pone.0154118