Iranian Society of Animal Science

Journal of Livestock Science and Technologies

Shahid Bahonar University of Kerman

ISSN: 2322-3553 (Print) ISSN: 2322-374X (Online)

Paper type: Original Research

Comparing genomic prediction models for genomic selection of traits with additive and dominance genetic architecture

Seyed Javad Khorami, Farhad Ghafouri-Kesbi and Ahmad Ahmadi*

Department of Animal Science, Faculty of Agriculture, Bu-Ali Sina University, Hamedan, Iran

*Corresponding author,
E-mail address:
f.ghafouri@basu.ac.ir

Received: 19 Feb 2023,
Accepted: 10 May 2023,
Published online: 17 May 2023,

ORCID
Seyed Javad Khorami
0009-0000-3464-9244
Farhad Ghafouri-Kesbi
0000-0002-2219-055X
Ahmad Ahmadi
0000-0002-9690-8632

Abstract The purpose of this research was to compare different statistical methods such as GBLUP, BayesA, BayesB, BayesC, BayesL, Ridge regression, Boosting and SVM for genomic evaluation of traits with additive and dominance genetic architecture. A genome consisting of 5 chromosomes was simulated, with 1000 single nucleotide polymorphism markers (SNP) uniformly distributed on each chromosome. In two different scenarios, 50 and 500 quantitative trait loci (QTL) were considered and in each scenario of QTL number, 0.00, 10, 20, 50 and 100% of QTLs were given dominance genetic effect. The prediction accuracy, bias and reliability of genomic breeding values were used for analyzing the results and comparing the methods. The results showed that not separating the dominance effects from the additive effects lead to a decrease in the accuracy and reliability and an increase in the bias of the predicted genomic breeding values. In all examined scenarios of the QTL number and percentages of QTLs with dominance effect, the Bayesian methods had higher prediction accuracy and reliability and their predictions had the least bias. Boosting predicted the genomic breeding values with the lowest accuracy and reliability and highest bias. The performance of SVM and Ridge regression was better than Boosting, but lower than Bayesian methods and GBLUP. In terms of computing speed, GBLUP and Boosting were, respectively, the fastest and the slowest method. It can be concluded that to increase the efficiency of genomic selection, first, the dominance genetic effects need to be included in the model and, second, methods with the highest predictive performance should be used.

Keywords: dominance genetic effect, genomic evaluation, genetic architecture, QTL, SNP

Introduction

In animal breeding, the main goal is to increase profitability of production system through increasing the animal performance and reducing production costs (Amiri Roudbar et al., 2018). Most economic traits are multifactorial inheritance, that is, they are under the control of a large number of genes (polygenic inheritance) and are also influenced by the environmental factors (Falconer and MacKay, 1996). Selection as an efficient strategy has been adopted to improve these traits. Traditionally, selection of superior animals is based on their breeding values estimated with animal models using phenotypic records and pedigree information (Amiri Roudbar et al., 2017). This strategy has made significant progress in economic traits of livestock (Hill, 2008). Today, with the advance in DNA science and molecular techniques, a large number of genetic markers, especially single nucleotide polymorphism (SNP) markers have been identified by which it has become possible to estimate genomic breeding values (GBVs) of candidate animals by summing the effects of thousands of SNPs that cover their entire genome (Mohammadabadi et al., 2010). This method is termed "genomic selection" (Meuwissen et al., 2001). The advantages of genomic selection includes reducing the generation interval, increasing the accuracy of selection, reducing inbreeding rate, and increasing the annual genetic gain (Ahmadi et al., 2021). Currently, in most developed countries, traditional methods have been replaced with genomic selection. Genomic selection is done in two steps. First, a predictive model is developed for estimating the effect of SNPs in the reference population using both genotypic and phenotypic data and the effect of each SNP is estimated. This prediction equation is then used to estimate the breeding values of candidate animals whose phenotypic information is unknown (Meuwissen et al., 2001). Therefore, making decision to select superior individuals can be done at early ages without needing their phenotypic information, with a higher accuracy than the average accuracy of the breeding values of the parents. It is why application of genomic selection specially in dairy herds is increasing at a remarkable rate. Instead of waiting for the records of the daughters of a bull, which can last 5-6 years, young bulls can be selected at birth by genotyping and estimating their breeding values (Schefers and Weigel, 2012)

Several methods have been developed to estimate the effect of SNPs. These methods belong to three groups of parametric methods (such as GBLUP and Bayesian methods), quasi-parametric methods (such as reproducing kernel Hilbert space, RKHS) and non-parametric methods (such as machine learning methods). These methods have been compared in different scenarios of heritability levels, number of QTLs and distributions of QTL effects. However, in nearly all of the studies, only additive effects of genes were considered and other gene effects such as dominance and epistasis were not investigated (see for example Neves et al., 2012; Howard et al., 2014; Sahebalam et al., 2019, Ahmadi et al., 2021). Recent studies have shown that part of the phenotypic variation in the economic traits of domestic animals is caused by non-additive gene effects (such as dominance), which is significant for some traits (Ebrahimi et al., 2017; Sadeghi et al., 2019). In some studies, in addition to additive genetic effects, the dominance effects (Aliloo et al., 2016) led to an increase in the accuracy of genomic evaluation. A common finding of these studies is that if the relative share of the dominance genetic effects in the phenotypic variation of the trait is high and it has not been considered in the model, then, it lead to inaccurate and biased estimates of the genomic breeding values (Mohammadi and Sattaei Mokhtari, 2018).

Genomic selection methods predict the GBVs with different accuracy because different methods have different assumptions about the distribution of the marker effects, the selection of covariates and/or the genetic variances and (co)variances matrix. Different combinations of these assumptions modify the genetic variation explained by the markers, which directly reflects on the accuracy (Andrade et al., 2019). Therefore, inclusion of the non-additive genetic effects could affect the predictive performance of genomic selection methods. Hence, it is necessary to compare the predictive performance of different genomic selection methods in the presence of both the additive and dominance genetic effects. Therefore, this study was conducted to compare the accuracy, bias and reliability of the genomic breeding values predicted by parametric and non-parametric methods for traits with additive and dominance genetic architecture.

Materials and methods

Population and genome simulation

Genome and population were simulated with the package *hyperd* in R software (Technow, 2013). Five chromosomes formed the genome, on each of which 1000 SNPs were uniformly distributed. To simulate the population, first a base population including 50 males and 50 females was simulated and by using random mating for 50 generations, the LD between the marker and QTL was established. LD was calculated using the r^2 statistic of Hill and Robertson (1968) as follows:

$$r^2 = D^2/\mathrm{freq}(A_1) * \mathrm{freq}(A_2) * \mathrm{freq}(B_1) * \mathrm{freq}(B_2)$$

freq (A1) is the frequency of A1 allele in the population likewise for other alleles in the population. D is the deviation of parental genotypes from the recombinant genotypes and estimated using the haplotype frequencies as follow:

$$D=\mathrm{freq}(A_1\text{-}B_1)*\mathrm{freq}(A_2\text{-}B_2)\text{-}\mathrm{freq}(A_1\text{-}B_2)*\mathrm{freq}(A_2\text{-}B_1)$$

The size of the population in the 51st generation was increased to 1000 individuals. These animals had both genotypic and phenotypic information and formed the reference population. The phenotype of the animals was obtained through the sum of the genetic value of QTLs and an environmental component obtained from a normal distribution with a mean of zero and a standard deviation equal to the square root of the environmental variance. Then the generation 52 was generated from individuals of the reference population. Individuals in generation 52 only had genotypic information but no phenotypic information and labeled as validation population for which the genomic breeding values had to be predicted (Table 1).

Table 1. Parameters used for simulation program

Genome size	500 cM
Number of chromosomes	5
Number of marker	5000
Distribution of additive QTL effects	Gamma
Number of QTL	50, 500
Effective population size (*Ne*)	100
Heritability	0.5
Historical population	Generations 1-50
Reference population	Generation 51
Validation population	Generation 52

Scenarios examined

Number of QTL: The number of QTL was considered as the percentage of the total number of SNPs (1% and 10%, namely 50 and 500 QTLs)
Dominance genetic effects: To assign dominance effects to QTLs, different scenarios were considered: in the first scenario, all QTLs were given only an additive effect (scenario A). In the second scenario, all QTLs were given an additive effect and 10% of them were given a dominant effect (scenario A+10%D). In the third scenario, all QTLs were given an additive effect and 25% of them were given a dominance effect (scenario A + 25% D). In the fourth scenario, an additive effect was given to all QTLs and a dominance effect was considered for 50% of them (scenario A + 50% D). In the fifth scenario, all QTLs were given both additive and dominance effects (scenario A+100%D).

Genomic evaluation method

Genomic best linear unbiased prediction (GBLUP): The statistical model was as follows (VanRaden, 2008):

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{Xb} + \mathbf{e}$$

where **y** is the vector of phenotypic observations and **μ** is the overall mean. **X** matrix elements include codes 0, 1 and 2, which indicate the number of alleles related to each of the SNPs for each individual. **b** is the vector of genomic breeding values and **e** is the vector of residual effects. BGLR package (de los Campos and Perez-Rodriguez, 2020) was used for GBLUP analyses.
Bayesian method A (BayesA): The main assumption of BayesA is that of total number of loci underlying a quantitative trait, only a small numbers have large effects and remained others have small effects. BayesA was fitted using following model:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u} + \sum_{\boldsymbol{k=1}}^{\boldsymbol{k}} z_k\, a_k + \boldsymbol{e}$$

where **y** is the vector of phenotypic observations, **X** is an incidence matrix associating observations to fixed effects in ***β***, ***u*** is the vector of polygenic effects, *k* denotes the number of SNPs, z_k is an N × 1 vector of genotypes at SNP *k*, $a_k$is the additive effect of that SNP, and ***e*** is a vector of residual effects. The prior for ***u*** is constant, the prior for δ_u^2 is assumed to follow normal distribution**,** *N*(**0**, Aδ_u^2) where A is the numerator-relationship matrix and δ_u^2 is additive genetic variance apart from that explained by SNPs. The *BayesA* was run using package BGLR in R (de los Campos and Perez-Rodriguez, 2020).
Bayesian method B (BayesB): In BayesB, it is assumed that only parts of the loci explain the entire genetic variance, and many loci do not contribute to genetic variance. BayesB can be written as follows:

$$\boldsymbol{y_i} = \boldsymbol{\mu} + \sum_{\boldsymbol{j=1}}^{\boldsymbol{k}} \boldsymbol{x_{ij}\beta_j\delta_j} + \boldsymbol{e_i}$$

where ***y*** is the phenotype of the animal *i*, *μ* is the mean, *k* is the number of marker loci, *x* is the genotype of the marker at the locus *j* (i_{th} allele) which is encoded as 0, 1 and 2 (number of copies of the SNP allele carried by the i_{th} animal). β_j is the effect of allelic substitution at position *j* and δ_j which coded as 0 and 1 indicates the absence (with probability π) or the presence (with probability 1- π) of the locus *j* in the model (Meuwissen et al., 2001). To implement Bayesian method B, BGLR package (de los Campus and Perez-Rodriguez, 2020) was used.
Bayesian method C: (BayesC): BayesC is a special type of Bayes B, and in some articles it is also referred to as Bayes Cπ. In BayesC, the normal distribution is used instead of the t distribution to model the marker effects, and as a result, the posterior distribution will also be normalized (Baneh et al., 2017).

Bayesian Lasso (BayesL): In BayesL, marker effects are assigned to markers using the bi-exponential prior probability density function (Baneh et al., 2017) as follows:

$$DE\left(\beta_j \middle| \gamma^2, \sigma_\varepsilon^2\right) = \int N(\beta_j \,|0, \sigma_\varepsilon^2, \tau_j^2)Exp(\tau_j^2| \frac{\gamma^2}{2})$$

The lambda parameter has a gamma distribution with the lambda shape parameter and the *t* scale.

Ridge regression BLUP (RidgeR): In RidgeR, the predicted GEBVs are obtained by the summing of all the marker effects of an individual. Marker effects were estimated using the following mixed model:

$$\mathbf{y} = \mathbf{1}_n\mu + \mathbf{Zg} + \boldsymbol{e}$$

where **y** is the vector of observed phenotypes, $\mathbf{1}_n$ is a column vector of *n* ones and *μ* is a common intercept, **Z** is the design matrix for the random marker effects; **g** is the vector of random marker effects. In RidgeR, the residuals and marker effects follow normal distributions with constant variance, i.e., $e \sim N(0, \boldsymbol{I}\sigma_e^2)$ and $g \sim N(0, \boldsymbol{I}\sigma_g^2)$, where **I** is an identity matrix. The R package BGLR (de los Campos and Perez-Rodriguez, 2020) was used to run RidgeR.
Boosting: Boosting is based on the idea that it is easier to find and average many rough rules of thumb, than to find a single, highly accurate prediction rule (Hastie et al., 2009). In fact it is a numerical optimization technique for minimizing the loss function by adding, at each step, a new tree that best reduces the loss function. It can model interactions between predictive variables (SNPs) and is capable of variable selection. In addition, it is robust to outliers, missing data and numerous correlated and irrelevant variables. The following model was used to fit Boosting algorithm (Hastie et al., 2009).

$$f(x) = \sum_{m=1}^{M} \beta_m\, b(x; \gamma_m)$$

where β_m , *m* =1, 2,…, *M* are the basis expansion coefficients, and *b*(*x*, γ) are simple functions of the multivariate argument *x*, with a set of parameters γ =(γ_1, γ_2,…, γ_M). The R package *gbm* (Ridgeway, 2013) was used to run Boosting. Tuning parameters in Boosting are number of tree (*ntree*), tree depth or tree complexity (*tc*) and shrinkage rate or learning rate (*lr*). We specified a series of values for each parameter with R coding. The model which provided the least error was: *ntree* =1500, *tc* = 7 and *lr* =0.02. These tuning parameters then used for analyzing the data in all the scenarios. Support Vector Machines (SVM): The SVM introduced in 1992 by Boser et al. (1992) belong to the general category of kernel methods and has been widely used in bioinformatics due to its high accuracy, ability to deal with high-dimensional data such as gene expression, and flexibility in modeling diverse sources of data (Schölkopf and Smola, 2005). For quantitative responses, a kind of SVM which termed Support Vector Regression (SVR) is used. The SVR uses linear models to implement nonlinear regression by mapping the input space (the marker dataset) to a feature space of a different dimension (lower in the case of GS) using a nonlinear kernel function followed by linear regression in this feature space. Using Radial kernel, SVR was fitted by the following model (Hastie et al., 2009):

$$f(x) = \beta_0 + h(x)^T\beta$$

where the basic functions, $h(x)^T$, is a linear (or nonlinear) transformations of one (or more) predictor variables (x), are additively combined with the vector of weights (β) (Hastie et al., 2009). Important tuning parameters in SVR are cost parameter (λ) and gamma which were predefined with a tuning function letting the parameters take a range of different values and identifying the value that corresponds to the best model performance assessed by cross-validation. The outputs of the tuning function were: λ =2 and gamma= 0.01. Using these values for λ and gamma, SVR was run using the *R* package "e1071" (Meyer et al., 2013).

Analysis of genomic breeding values (Legarra and Reverter, 2018)

Prediction accuracy: This criterion was calculated as the Pearson's correlation between predicted and true (simulated) breeding values.

$$\text{Pearson's correlation} = \frac{cov(true, predicted)}{\sqrt{var(predicted)var(true)}}$$

Bias: This criterion was calculated as the difference between the average predicted breeding values and true breeding values.

$$Bias = \overline{predicted} - \overline{true}$$

Reliability: This is the slope of the regression of predicted breeding values on true breeding values.

$$Reliability = \frac{cov(predicted, true)}{var(true)}$$

Each scenario of number of QTL and percentage of QTLs with dominance effect was analyzed 10 times and the average of 10 replications was reported.

Results

In all methods, with the increase in the percentage of QTLs with dominance effect from 0.00% (scenario A) to 100% (scenario A+ 100%D), the prediction accuracy decreased, bias increased and reliability decreased. In the 500 QTLs scenario, decrease in accuracy was between 20% (BayesA, BayesB) to 24% (RidgeR) and in the 50 QTLs scenario, it was between 20% (BayesL) to 25% (BayesB) (Table 2). In addition, by increasing the percentage of QTLs with dominance effect from 0.00% (scenario A) to 100% (scenario A+ 100%D), the bias increased between 25% (BayesA) and 29% (Boosting) in the 500 QTLs scenario and between 18% (BayesA) and 22% (Boosting) in 50 QTLs scenario (Table 3). The reliability of GBVs decreased between 0.20 (RidgeR) to 38% (Boosting) in 500 QTLs scenario and from 0.20 (BayesB and RidgeR) to 29% (Boosting) in 50 QTLs scenario by increasing the percentage of QTLs with dominance effect from 0.00% (scenario A) to 100% (scenario A+ 100%D) (Table 4).

Table 2. Prediction accuracy of genomic breeding values in different scenarios of QTL number and percentage of QTLs with dominance effect

	A	A+10%D	A+25%D	A+50%D	A+100%D
500 QTLs					
GBLUP	0.76	0.74	0.71	0.63	0.61
BayesA	0.78	0.75	0.73	0.67	0.62
BayesB	0.80	0.79	0.75	0.69	0.64
BayesC	0.77	0.77	0.75	0.67	0.61
BayesL	0.79	0.77	0.77	0.69	0.63
RidgeR	0.78	0.72	0.72	0.64	0.59
Boosting	0.64	0.64	0.63	0.62	0.54
SVM	0.73	0.72	0.68	0.64	0.57
50 QTLs					
GBLUP	0.80	0.75	0.71	0.64	0.60
BayesA	0.82	0.79	0.75	0.66	0.62
BayesB	0.84	0.78	0.73	0.68	0.63
BayesC	0.79	0.78	0.74	0.70	0.62
BayesL	0.78	0.76	0.76	0.69	0.62
RidgeR	0.76	0.74	0.70	0.66	0.58
Boosting	0.65	0.66	0.62	0.60	0.53
SVM	0.76	0.74	0.71	0.64	0.59

In the scenario of 500 QTLs, the Bayesian methods had higher accuracy compared to other methods in almost all scenarios. The GBLUP, RidgeR and SVM were ranked next. Boosting had significantly lower accuracy than other methods. In the 50 QTLs scenario, the difference between the methods regarding the accuracy of prediction was more significant compared to 500 QTLs scenario. Here too, Bayesian methods had higher accuracy, followed with the GBLUP, RidgeR, SVM and Boosting.

Regarding bias in estimates of the genomic breeding values, in 50 and 500 QTLs scenarios, the Bayesian methods had lower bias compared to other methods in almost all scenarios of %QTLs with dominance effects. Genomic breeding values predicted by Boosting had maximum bias and bias of SVM and RidgeR were intermediate among other estimates.

Table 3. Bias of genomic breeding values in different scenarios of QTL number and percentage of QTLs with dominance effect

	A	A+10%D	A+25%D	A+50%D	A+100%D
500 QTLs					
GBLUP	765.4	793.4	832.4	899.0	929.3
BayesA	778.8	805.4	827.1	894.2	975.6
BayesB	746.3	774.4	791.6	866.4	924.3
BayesC	744.0	753.5	784.2	831.3	901.6
BayesL	756.1	764.7	791.2	854.8	943.3
RidgeR	732.2	747.7	780.6	843.2	936.9
Boosting	947.9	966.5	994.3	1089.5	1224.6
SVM	815.6	836.3	867.7	931.4	1031.8
50 QTLs					
GBLUP	694.6	721.7	734.1	786.3	835.2
BayesA	687.8	717.4	741.3	774.9	811.5
BayesB	659.6	683.5	707.4	736.6	789.3
BayesC	685.6	705.3	722.0	777.4	803.4
BayesL	693.4	711.3	739.5	778.2	839.6
RidgeR	701.2	732.5	754.7	793.5	841.6
Boosting	875.7	894.2	936.7	978.8	1076.9
SVM	734.7	745.8	789.4	835.9	890.5

For reliability of GBVs, similar result was observed in a way that while Bayesian methods provided GBVs with higher reliability, predictions of SVM and Boosting had minimum reliability. For all methods studied, in the 50 QTLs scenario, the accuracy and reliability were higher and bias was smaller than those observed in 500 QTLs scenario.

The computing time of different methods is shown in Figure 1. As shown, the GBLUP with 0.5 minute was the fastest method and Boosting with 9.53 minutes was the slowest method. Although the predictive performance of the Bayesian methods was higher than other methods, they were relatively slow. RidgeR and SVM methods were in the middle.

Table 4. Reliability of genomic breeding values in different scenarios of QTL number and percentage of QTLs with dominance effect

	A	A+10%D	A+25%D	A+50%D	A+100%D
500 QTLs					
GBLUP	0.63	0.61	0.58	0.52	0.45
BayesA	0.65	0.65	0.62	0.55	0.48
BayesB	0.64	0.63	0.61	0.56	0.42
BayesC	0.62	0.60	0.57	0.54	0.43
BayesL	0.60	0.58	0.55	0.50	0.42
RidgeR	0.64	0.63	0.63	0.57	0.51
Boosting	0.53	0.50	0.47	0.41	0.33
SVM	0.59	0.57	0.55	0.48	0.41
50 QTLs					
GBLUP	0.64	0.63	0.60	0.55	0.49
BayesA	0.63	0.62	0.58	0.55	0.52
BayesB	0.66	0.65	0.62	0.57	0.53
BayesC	0.64	0.62	0.59	0.56	0.49
BayesL	0.61	0.59	0.57	0.55	0.50
RidgeR	0.62	0.62	0.60	0.57	0.52
Boosting	0.55	0.53	0.45	0.48	0.39
SVM	0.60	0.60	0.57	0.51	0.45

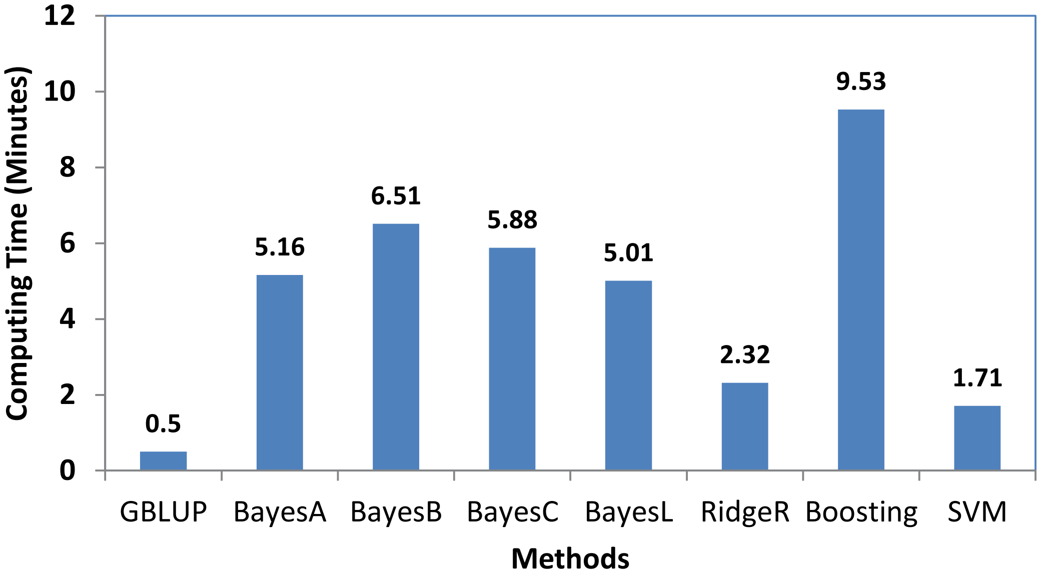

Figure 1. Computing time for the studied methods

Discussion

Dominance effects and genomic evaluation

In most studies on genomic evaluation, whether using real or simulated data, only the additive genetic effects of genes were considered and non-additive effects such as epistasis and dominance were ignored (Neves et al., 2012; Abdullahi Arpanahi et al., 2013; Wang et al. 2013; Zhang et al., 2017; Sahebalam et al., 2019; Ahmadi et al., 2021; Sahebalam et al., 2022). In the present study, in different scenarios, different percentages of QTLs were given dominance but a purely additive model was used, i.e., no attempt was made to separate the additive effects from dominance effects. The results showed that, if there are dominance effects but not separated from additive genetic effects, a decrease in the accuracy, an increase in bias and a decrease in reliability of genomic breeding values should be expected, which was in agreement with Aliloo et al. (2016). Using computer simulation, Mohammadi and Sattaei Mokhtari (2018) showed that by separating the dominance effects from the additive effects, the accuracy of genomic evaluation increased from 0.63 to 0.69 in the BayesA, and from 0.65 to 0.67 in the BayesL.

Comparison of methods

When comparing different methods for genomic evaluation, several factors should be considered such as the prediction performance, computing time, and memory requirement (Ahmadi et al., 2021). A method with accuracy close to one, bias close to zero and reliability close to one is desirable for genomic evaluation (Macedo et al., 2020). According to the current results, the Bayesian methods had higher prediction accuracy in most of the examined scenarios, and at the same time, their estimates of genome breeding values had the least bias and maximum reliability. On the other hand, predictions of Boosting had the lowest accuracy and reliability and highest bias. Therefore, if the dominance genetic effects contribute to the phenotypic variation of the trait, but not included in the statistical model, or if no information is available about the contribution of the dominance genetic effects to the phenotypic variation of the trait, parametric methods such as Bayesian methods and GBLUP should be preferred to non-parametric methods. This result has been reported when only additive effects of genes were considered in genomic evaluation and a purely additive model was used. For example, Moradi et al. (2017) compared the GBLUP and BayesB parametric methods, the semi-parametric method RKHS and the non-parametric methods Random Forest in the genomic evaluation of traits with additive genetic architecture and reported that GBLUP and BayesB performed better than semi-parametric and non-parametric methods. Ghasemi (2019) compared the performance of SVM, GBLUP and BayesB methods aiming at introducing a method with the highest prediction accuracy for genomic evaluation of threshold traits. In general, in almost all the examined scenarios of threshold number, QTL number and distribution of QTL effects, BayesB and GBLUP had higher prediction accuracy than SVM method, which is supported by our findings. They suggested that the SVM method should not be used for the genomic evaluation of threshold traits. Howard et al. (2014) with a simulation study, reported that if the genetic architecture of the trait is based on additive genetic effects, parametric methods (Bayesian methods) outperformed non-parametric methods (SVM and RKHS), but by including the dominance genetic effects in the model and separating them from additive effects, the accuracy of genomic prediction of non-parametric methods was higher than parametric methods. Salehi et al. (2021) compared the predictive performance of the GBLUP and BayesB with the two nonparametric methods BagBLUP and Random Forest. In the purely additive scenario, the BayesB method had the highest prediction accuracy (0.73) and the GBLUP, BagBLUP, and Random Forest methods ranked next with 0.63, 0.63, and 0.54, respectively. But in the additive-dominance-epistasis scenario, Random Forest and BayesB ranked first with prediction accuracy of 0.46 and 0.45, respectively, and BagBLUP and GBLUP methods with prediction accuracy of 0.37 and 0.36 were in the next ranks. Mohammadi and Sattaei Mokhtari (2018) simulated a trait with additive and dominance genetic architecture and compared the accuracy of Bayesian methods (BayesA, BayesB and BayesL) with RKHS method and reported that while BayesB outperformed other methods in terms of accuracy and bias, by increasing the ratio of dominance genetic variance to phenotypic variance, the accuracy increased with higher rate in RKHS method. As a result, when genetic architecture includes only additive effects of genes, parametric methods specially Bayesian methods have been proved to predict GBVs with higher accuracy and less bias. But when genetic architecture includes both additive and non-additive effects, it seems that one particular method cannot be introduced as the superior method and, therefore, a series of parametric, non-parametric and semi-parametric methods should be tested to select the best method to analysis data. More research is needed in this area.

One of the factors that affect the efficiency of genomic evaluation methods is the computing time. In this research, GBLUP and Boosting were the fastest and slowest methods, respectively. Ghafouri-Kasbi et al. (2016) used Boosting, Random Forest, SVM and GBLUP for genomic evaluation. In their study, GBLUP, SVM, Random Forest and Boosting were ranked first, second, third and fourth with 10 minutes, 15 minutes, 75 minutes and 600 minutes respectively, which is in consistent with the results of the present study. Our results showed that even though Bayesian methods had high accuracy, they were relatively slow. It can be a serious limitation for these algorithms, especially when a large dataset need to be analyzed. Since genotyping cost is decreasing with a significant rate, the number of SNPs in the SNP chips used in genomic prediction is increasing. Meanwhile, by genotyping more animals, the size of the reference populations has been increased. Therefore, the size of the genotypic matrix whose dimensions are equal to the number of individuals × number of SNP will be increased exponentially. Therefore, computing time and memory requirement of genomic prediction methods should be improved. Dealing with large datasets, methods that have high accuracy and perform calculations in a shorter time will be preferred (Ahmadi et al., 2021).

Conclusions

In conclusion, when an additive model was used, increasing the percentage of QTLs with the dominance effect led to a decrease in the accuracy and reliability and an increase in bias of GBVs. When QTLs had only additive effects and even when QTLs had additive and dominance genetic effects, the Bayesian methods were superior to other methods. Boosting and SVM did not show a decent performance. Regarding the computing time, GBLUP and Boosting were the fastest and slowest methods, respectively.

Acknowledgements

We would like to thank two anonymous referees for their valuable comments on a previous version of this manuscript.

References

Abdollahi-Arpanahi, R., Pakdel A., Nejati-Javaremi, A., Moradi Shahre Babak, M., 2013. Comparison of different methods of genomic evaluation in traits with different genetic architecture. *Journal of Animal Production* 15, 65-77.

Ahmadi, Z., Ghafouri-Kesbi, F., Zamani, P., 2021. Assessing the performance of a novel method for genomic selection: rrBLUP method6. *Journal of Genetics* 100, 24.

Aliloo, H., Pryce, J.E., González-Recio, O., Cocks, B.G., Hayes, B.J., 2016. Accounting for dominance to improve genomic evaluations of dairy cows for fertility and milk production traits. *Genetic Selection Evolution* 48, 8.

Andrade, L.R.B., Sousa, M.B., Oliveira, E.J., Resende M.D.V., Azevedo, C.F., 2019. Cassava yield traits predicted by genomic selection methods. *PLoS One* 14, e0224920.

Amiri Roudbar, M., Mohammadabadi, M.R., Mehrgardi, A.A., Abdollahi-Arpanahi, R., 2017. Estimates of variance components due to parent-of-origin effects for body weight in Iran-Black sheep. *Small Ruminant Research* 149, 1-5.

Amiri Roudbar, M., Abdollahi-Arpanahi, R., Mehrgardi A.A., Mohammadabadi, M.R., Taheri Yeganeh, A., Rosa, G.J.M., 2018. Estimation of the variance due to parent-of-origin effects for productive and reproductive traits in Lori-Bakhtiari sheep. *Small Ruminant Research* 160, 95-102

Baneh, H., Nejati, Javaremi A., Honarvar, M., Rahimi, G.H., 2017. Genomic evaluation of threshold traits with different genetic architecture using Bayesian approaches. *Research on Animal Production* 8, 149-154.

Boison, S.A., Santos, D.J.A., Utsunomiya, A.H.T, Carvalheiro, R., Neves, H.H.R., O'Brien, A.M.P., Garcia, J.F., Sölkner, J., Da Silva, M.V.G.B., 2015. Strategies for single nucleotide polymorphism (SNP) genotyping to enhance genotype imputation in Gyr (Bos indicus) dairy cattle: Comparison of commercially available SNP chips. *Journal of Dairy Science* 98, 4969-4989.

Boser, B., Guyon, I.M., Vapnik, V., 1992. A training algorithm for optimal margin classifiers. In Proceedings of the Fifth Annual Workshop on Computational Learning Theory. Pittsburgh, USA.

de los Campos G., Perez-Rodriguez, P., 2020. BGLR: Bayesian generalized linear regression. https://cran.r-project.org/web/packages/BGLR/BGLR.pdf

Doublet, A.C., Croiseau, P., Fritz S., Michenet, A., Hozé, C., Danchin-Burge, C., Laloë D., Restoux, G., 2019. The impact of genomic selection on genetic diversity and genetic gain in three French dairy cattle breeds. *Genetic Selection Evolution* 51, 52.

Ebrahimi, K., Dashab, G.R., Faraji-Arough, H., Rokouei, M., 2018 Estimation of additive and non-additive genetic variances of body weight in crossbreed populations of the Japan Quail. *Poultry Science* 1, 46-55.

Falconer, D.S., Mackay, T.F.C., 1996. Introduction to Quantitative Genetics, 4th ed. Longman Group. Harlow, Essex, UK.

Fernando, R.L., Grossman, M., 1989. Marker-assisted selection using best linear unbiased prediction. *Genetic Selection Evolution* 2, 467.

Ghafouri-Kesbi, F., Rahimi-Mianji, G., Honarvar, M., Nejati-Javaremi, A., 2016. Predictive ability of random forests, boosting, support vector machines and genomic best linear unbiased prediction in different scenarios of genomic evaluation. *Animal Production. Science* 57, 229-236.

Ghasemi, M., 2019. Genomic evaluation of threshold traits considering different number of threshold using some parametric and non-parametric statistical methods. M.Sc. Thesis. Bu-Ali Sina University, Hamedan, Iran.

Hastie, T.J., Tibshirani, R., Friedman, J., 2009. The elements of statistical learning. Springer Publishing. New York, USA.

Hill, W.G., 2008. Estimation, effectiveness and opportunities of long term genetic improvement in animals and maize. *Lohman Information* 43, 3-19.

Hill, W.G., Robertson, A., 1968. Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics* 38, 226-231.

Howard, R., Carriquiry, A.L., Beavis, W.D., 2014. Parametric and nonparametric statistical methods for genomic selection of traits with additive and epistatic genetic architectures. *Genetics* 4, 1027-1046.

Legarra, A., Reverter, A., 2018. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. *Genetic Selection Evolution* 50, 53.

Macedo, F.L., Christensen, O.F., Astruc, J.M., Aguilar, I., Masuda, Y., Legarra, A., 2020. Bias and accuracy of dairy sheep evaluations using BLUP and SSGBLUP with metafounders and unknown parent groups. *Genetic Selection Evolution* 52, 47.

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.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, K., 2013. Misc publication of the Department of Statistics (e1071), TU Wien. Available at: http://cran.r-project.org/web/packages/e1071/index.html

Mohammadi, Y., Sattaei Mokhtari, M., 2018. Genomic selection accuracy parametric and nonparametric statistical methods with additive and dominance genetic architectures. *Research on Animal Production* 8, 161-167.

Mohammadabadi, M.R., Torabi, A., Tahmourespoor, M., Baghizadeh, A., Esmailizadeh Koshkoie, A., Mohammadi, A., 2010. Analysis of bovine growth hormone gene polymorphism of local and Holstein cattle breeds in Kerman province of Iran using polymerase chain reaction restriction fragment length. *African Journal of Biotechnology* 9, 6848-6852.

Moradi, M., Abdollahi-Arpanahi, R., Hemati, B., Lavvaf, A., 2017. Comparison of parametric and resampling methods in genetic evaluation of quantitative traits with different genetic structure. *Journal of Animal Production* 19, 1-12.

Neves, H.H.R., Carvalheiro, R., Queiroz, S.A., 2012. A comparison of statistical methods for genomic selection in a mice population. *BMC Genetics* 13, 100.

Sadeghi, S.A.T., Rokoue, M., Vafaye Valleh, M., Abbasi, M.A., Faraji-Arough, H., 2019. Estimation of additive and non-additive genetic variance component for growth traits in Adani goats. *Tropical Animal Health and Production* 52, 733-742.

Sahebalam, H., Gholizadeh, M., Hafezian, H., Ebrahimi, F., 2022. Evaluation of Bagging approach versus GBLUP and Bayesian LASSO in genomic prediction *Journal of Genetics* 101, 19.

Sahebalam, H., Gholizadeh, M., Hafezian H., Farhadi, A., 2019. Comparison of parametric, semiparametric and nonparametric methods in genomic evaluation. *Journal of Genetics* 98, 102.

Salehi, A., Bazrafshan, M., Abdollahi-Arpanahi, R., 2021. Assessment of parametric and non-parametric methods for prediction of quantitative traits with non-additive genetic architecture. *Annals of Animal Science* 21, 469-484.

Schefers, J.M., Weigel, K.A., 2012. Genomic selection in dairy cattle: Integration of DNA testing into breeding programs. *Animal Frontiers* 2, 4-9.

Schölkopf, B., Smola, A., 2005. Support Vector Machines. *Encyclopedia of Biostatistics*, 1 5328-5335.

Technow, F., 2013. hypred: simulation of genomic data in applied genetics. Available at: http://cran.r-project.org/web/packages/hypred/index.html.

VanRaden, P.M., 2008. Efficient methods to compute genomic predictions. *Journal of Dairy Science* 91, 4414-4423.

Wang, C.L., Ding, X.D., Wang, J.Y., Liu, J.F., Fu, W.X., Zhang, Z., Yin, J., Zhang, Q., 2013. Bayesian methods for estimating GEBVs of threshold traits. *Heredity* 110, 213-219.

Zhang, A., Wang, H., Beyene, Y., Semagn, K., 2017. Effect of trait heritability, training population size and marker density on genomic prediction accuracy estimation in 22 bi-parental tropical maize populations. *Frontiers in Plant Science* 8, 1916.