Available online at www.scholarsresearchlibrary.com

Scholars Research Library

European Journal of Zoological Research, 2013, 2 (5):79-86
(http://scholarsresearchlibrary.com/archive.html)

ISSN: 2278–7356

A simulation study of heritability and marker effect on accuracy of breeding value in animal breeding

Reza Behmaram[1*], Ali Asghar Aslaminejad[1], Mehdi Aminafshar[2] and Mojtaba Tahmoorespur[1]

[1]Department of Animal Science, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran
[2]Department of Animal Science, Faculty of Agriculture and Natural Resources, Science and Research Branch, Islamic Azad University, Tehran, Iran

---

ABSTRACT

Genomic selection (GS) can increase genetic gain per generation through early selection. GS is expected to be particularly valuable for traits that are costly to phenotype and expressed late in the life cycle of long-lived species. Alternative approaches to genomic selection prediction models may perform differently for traits with distinct genetic properties. Accuracies were investigated by simulation for a typical dairy cattle breeding setting. A genome consisting 3 chromosomes each with 100 cM in length was simulated. In order to create sufficient linkage disequilibrium after 50 generations of random mating in a finite population (Ne = 100), population was expanded to obtain intended population size (500 male and 500 female). Three measures of heritability (0.05, 0.30, 0.80) and four different numbers of markers (100, 200, 400, 800) were considered. Each simulation was replicated 10 times and results were averaged across replications. Six generations with only genotypes were generated to investigate the accuracy of breeding value over time. Accuracies without phenotypes ranged from 0.21 for threshold traits to 0.73 and from 0.24 to 0.74 for continuous traits. Accuracies were found sufficiently high to implement dairy selection schemes without progeny testing in which case a data time-lag of two to three generations may be present. Accuracies were also relatively high for low heritable traits, implying that genomic selection could be especially beneficial to improve the selection on, e.g. health and fertility. The results showed that using genomic selection can be useful for threshold traits which include some of important traits in animal breeding.

Keywords: Accuracy, breeding value, genomic selection, threshold traits.

INTRODUCTION

Estimation of breeding values requires a matrix describing the additive relationship between individuals in the population. If pedigree information has been collected over multiple generations, the additive relationship matrix can be constructed from this information. However, in many livestock populations, this information may be unavailable, incomplete, or contain errors.

Information from high-density marker maps and high-throughput genotyping can be utilized in new selection methods. An alternative to constructing the additive relationship matrix from pedigrees is to use marker information

to infer relationships. Attempting to do this from a limited number of markers can result in bias and inaccurate estimates of genetic parameters [2].

One problem of using marker assisted selection(MAS) is the limited variance explained by the detected quantitative traits loci(QTL). Meuwissen et al. [21] created a first step towards predicting a total genetic value using a genome-wide dense map of highly informative markers. The method was termed genomic selection, and the idea was to estimate the effects of all genes or chromosomal segments simultaneously. The effect of these segments is summed to predict the total breeding value. By treating the markers or haplotypes as random effects, the limitation of estimating large numbers of haplotypes effects from a limited number of animals can be managed.

They compared different methods for predicting breeding values based on haplotype effects, and reported accuracies in the range between 0.79 to 0.85. (where accuracy is the correlation between true breeding value (TBV) and estimated breeding value (EBV), and the reliability is the square of this result).

In several livestock species including cattle, chicken, and pig tens of thousands of SNP markers are now available. This could substantially reduce the costs of animal breeding, and accelerate genetic gain per year by reducing the generation interval [12]. GS could also be efficient for low heritable traits [15, 23].

Although most of the prediction methods focus on continuous traits, however, some of important traits in animal production, such as litter size of large mammals, degree of calving difficulty and resistance to disease, present a discrete(categorical) distribution of phenotypes, and are often termed threshold traits. Obviously, the GS methods proposed for continuous traits, cannot be adequately applied for such kind of traits.

In this research, a simulation experiment was performed to evaluate the accuracy of breeding values in continuous and threshold traits with different number of markers including low, moderate and high heritabilities.

MATERIALS AND METHODS

Simulation

A genome consisting 3 chromosomes each with 100 cM in length with 100, 200, 400, 800 equally spaced single nucleotide polymorphism(SNP) (each 1 cM) and a total number of (30, 60, 120, 240) QTLs (that scattered on chromosomes randomly) was generated for each individual.

This small genome size was chosen to decrease calculation time. Both SNP and QTL were assumed to be biallelic with equal initial allelic frequencies. For these simulations, gene substitution effects for each QTL were assigned randomly from a standard normal distribution, a ~ N (0, 1). QTLs covered total genetic variance and individual true breeding values. Only additive genetic effect was considered.

An effective population size of 100 individuals was simulated, of which 50 were male and 50 were female. This structure was followed by 50 generations of random mating, implying that each individual had on average two offspring in the next generation (variance of family size was two).

The paternal and maternal haplotypes for each individual were generated based on Haldane mapping function to generate recombinant haplotypes. Sires and dams in the base generation were assumed to be unrelated.

Fifty generations of random mating were practiced to generate sufficient linkage disequilibrium (LD) between loci. Two LD measurements, r^2 and D' , were used to calculate LD in generation 50, as average of all synthetic marker loci. Markers with a minor allele frequency of < 0.05 were discarded.

After the first 50 generations, 6 additional generations (51 to 56) were simulated. Population was expanded to obtain intended population size in generation 51. Population size was constant until generation 56. For population size, 1000 individuals with equal number of males and females in each of the last 6 generations were simulated.

Only females of generations 51 through 55 (500 females in each generation) had trait phenotype and, thus, were included in the training set according to different scenarios.

To investigate the effect of generation distance between training set and validation set on accuracy of GEBVs, females from different generations (distant and recent generations) were included in training set. The validation data contained individuals from generation 56. For simplification, no selection was considered. Population structure and parameters used in the simulation are presented in Table 1.

Analysis model

For calculation of GEBV, the simple mixed model estimation method suggested by Meuwissen et al. [21] was used assuming that all loci explained and equal amount of variance (That is, the variance per locus σ^2m, is σ^2m = σ^2a /n where σ^2a is the total genetic variance and *n* is the number of marker loci).

This assumption (equal variance over all loci) is clearly unrealistic. Genetic variance may not be equal across markers, for example, major genes may exist on some chromosomes. However, BLUP is quick, easy to program and as Meuwissen et al. [21] demonstrated, BLUP performs almost as well as the much more advanced and time consuming Bayesian methods.

The model to estimate the marker effects was

$$y = Xb + Zm + e \qquad (1)$$

where, y is the vector of observations, b is the vector of means, m is the vector of random marker effects, e is the vector of random residual effects, X and Z are coefficient matrices. Row elements of Z consist of 0, 1 and 2 for marker genotype. Then, the expected value of y is 1μ and the variance of y is

$$V(y) = ZIZ'\sigma^2 m + I\sigma^2 e \qquad (2)$$

(assuming equal variance for each marker).
The mixed model equation (MME) for BLUP is

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + I\alpha \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix} \qquad (3)$$

We considered $\alpha = \sigma^2$e/ σ^2m as Meuwissen et al. [21]. After obtaining solution for vector m, GEBV was estimated as

$$GEBV_i = Z_i m_i \qquad (4)$$

The genetic variance was determined as variance of true breeding values among individuals in generation 51 through 55. As haplotyping would increase computation time with little or no gain in accuracy at high marker density [16], we used genotypes rather than haplotypes.

Different scenarios were compared by the accuracy of the estimated genomic breeding values for individuals without a phenotypic record (generation 56). Accuracies were calculated as the correlation between simulated and estimated breeding values.

Each simulated data set was replicated 10 times and results were averaged across replicates. All of the above calculations was done once for each type of both traits(continuous and threshold), but for threshold traits we did one more stage. Before estimation of accuracy of breeding value, at first we changed the structure of data to present a discrete(categorical) distribution of phenotypes.

Table 1. Population structure and parameters used in the simulation.

Parameter	Value
Number of chromosome	3
Number of SNP markers per chromosome	100, 200, 400,800
Genome Length (cM)	300
Marker distance (cM)	1
Number of QTL	30, 60, 120, 240
QTL effects	Normal distribution
Recombination	Haldane map function
Number of generation	56
Generation 1 to 50, create LD	50 male, 50 female
Generation 51 to 56	500 male, 500 female
Training set	Females of generation 51 to 55
Validation set	Females of generation 56
Heritability	0.05, 0.30, 0.80

Table 2. Mean (±SE) of homozygosity and linkage disequilibrium (D′ and r^2) between markers in generation 50.

Parameter	Mean ± SE
D′	0.61 ± 0.003
r^2	0.18 ± 0.002
Homozygosity	0.58 ± 0.002

RESULTS AND DISCUSSION

Results

In this study, six generations with three different levels of heritability (0.05, 0.30, 0.80) and four various number of markers (100, 200, 400, 800) were simulated simultaneously to investigate accuracy of breeding value in two types of traits (continuous and threshold).

The all accuracies of selection are given in tables 3 to 6 and 7 to 10 for threshold and continuous traits respectively. Accuracies showed the correlations between GEBVs and simulated true breeding values in generation 51 to 56.

The highest accuracy for threshold and continuous traits was 0.73 and 0.74 respectively. The lowest accuracy for threshold and continuous traits was 0.21 and 0.24 respectively. The results showed a relatively clear relationship between the number of markers used in the prediction model and the accuracies that were obtained.

The highest accuracy calculated with using 400 and 200 markers for threshold and continuous traits respectively. The accuracy of breeding value for threshold traits varied from 0.21 to 0.64 using 100 markers, for 200 markers from 0.24 to 0.70, with 400 markers from 0.26 to 0.73 and using 800 markers from 0.28 to 0.68 respectively.

The accuracy of breeding value varied from 0.24 to 0.71 using 100 markers, for 200 markers from 0.30 to 0.74, with 400 markers from 0.30 to 0.72 and using 800 markers from 0.28 to 0.69 respectively for continuous traits. Hence, the density of the markers increased the accuracy of breeding values as expected.
In our research for threshold traits, with a heritability of (0.05) the accuracy in generation 51 was 0.45 decreasing to 0.22 in generation 56. With a heritability of (0.30) the accuracy in generation 51 was 0.61 decreasing to 0.27 in generation 56. With a heritability of (0.80) the accuracy in generation 51 was 0.73 decreasing to 0.21. in generation 56.

The accuracy for continuous traits, with a heritability of (0.05) in generation 51 was 0.50 decreasing to 0.26 in generation 56. With a heritability of (0.30) the accuracy in generation 51 was 0.64 decreasing to 0.26 in generation 56. With a heritability of (0.80) the accuracy in generation 51 was 0.74 decreasing to 0.24. in generation 56.

There is a tendency for the average loss to be a little higher from generation 51 to 52 than from generation 52 to 56. Our findings indicate that the decrease in accuracies or predictive power over generations is relatively similar for different heritabilities and higher heritability results in almost higher reliability for all in the generations 51 to 56.

Table 3. Accuracies of breeding value in generation 51 to 56 (Number of Markers= 100, Number of QTL = 30) for threshold traits.

Generation	51	52	53	54	55	56
$h^2 = 0.05$	0.43	0.27	0.27	0.26	0.24	0.22
$h^2 = 0.30$	0.60	0.36	0.34	0.31	0.29	0.27
$h^2 = 0.80$	0.64	0.35	0.33	0.26	0.24	0.21

Table 4. Accuracies of breeding value in generation 51 to 56 (Number of Markers = 200, Number of QTL = 60) for threshold traits.

Generation	51	52	53	54	55	56
$h^2 = 0.05$	0.42	0.28	0.32	0.28	0.25	0.24
$h^2 = 0.30$	0.61	0.41	0.42	0.38	0.35	0.34
$h^2 = 0.80$	0.70	0.42	0.40	0.34	0.33	0.29

Table 5. Accuracies of breeding value in generation 51 to 56 (Number of Markers= 400, Number of QTL = 120) for threshold traits.

Generation	51	52	53	54	55	56
$h^2 = 0.05$	0.45	0.32	0.31	0.30	0.27	0.26
$h^2 = 0.30$	0.59	0.42	0.42	0.41	0.37	0.34
$h^2 = 0.80$	0.73	0.46	0.44	0.41	0.38	0.36

Table 6. Accuracies of breeding value in generation 51 to 56 (Number of Markers = 800, Number of QTL = 240) for threshold traits.

Generation	51	52	53	54	55	56
$h^2 = 0.05$	0.44	0.29	0.31	0.31	0.30	0.28
$h^2 = 0.3$	0.59	0.44	0.42	0.41	0.41	0.40
$h^2 = 0.80$	0.68	0.42	0.41	0.38	0.37	0.36

Table 7. Accuracies of breeding value in generation 51 to 56 (Number of Markers = 100, Number of QTL = 30) for continuous traits.

Generation	51	52	53	54	55	56
$h^2 = 0.05$	0.50	0.35	0.33	0.31	0.29	0.26
$h^2 = 0.30$	0.64	0.37	0.39	0.32	0.28	0.26
$h^2 = 0.80$	0.71	0.40	0.37	0.26	0.25	0.24

Table 8. Accuracies of breeding value in generation 51 to 56 (Number of Markers = 200, Number of QTL = 60) for continuous traits.

Generation	51	52	53	54	55	56
$h^2 = 0.05$	0.49	0.37	0.36	0.34	0.32	0.30
$h^2 = 0.30$	0.63	0.43	0.45	0.42	0.39	0.35
$h^2 = 0.80$	0.74	0.46	0.42	0.36	0.34	0.32

Table 9. Accuracies of breeding value in generation 51 to 56 (Number of Markers = 400, Number of QTL = 120) for continuous traits.

Generation	51	52	53	54	55	56
$h^2 = 0.05$	0.47	0.32	0.33	0.31	0.31	0.30
$h^2 = 0.30$	0.61	0.40	0.39	0.36	0.35	0.33
$h^2 = 0.80$	0.72	0.46	0.44	0.41	0.37	0.35

Table 10. Accuracies of breeding value in generation 51 to 56 (Number of Markers = 800, Number of QTL = 240) for continuous traits.

Generation	51	52	53	54	55	56
$h^2 = 0.05$	0.48	0.35	0.34	0.31	0.29	0.28
$h^2 = 0.30$	0.60	0.38	0.44	0.42	0.40	0.38
$h^2 = 0.80$	0.69	0.42	0.43	0.40	0.39	0.37

DISCUSSION

The estimation model assumes that there is no dominance (i.e., only the additive effects are fitted), and the average effects of the genes are estimated, which is probably satisfactory for the prediction of breeding values in most cases. In theory, this is not a problem, but more research is needed to verify that such estimates are accurate in realistic scenarios.

A relationship between heritability and accuracies was observed; as heritability increased so did the accuracy, this was also observed by Kolbehdari et al. [10].

Meuwissen et al. [21] showed that the accuracy of GEBVs decreased to 0.804, 0.768, 0.758, 0.734 and 0.718 in 5 subsequent generations, respectively. Advantages of Bayesian method to BLUP evaluation has been shown in some studies [3, 4, 21].

Wang et al. [5] introduced a threshold model to the framework of GS. They specifically extended three Bayesian method BayesA, BayesB and BayesCπ on the basis of threshold model for estimation genomic breeding value of threshold traits and termed correspondingly extended methods BayesTA, BayesTB and BayesTCπ. They showed new approaches generally performed better than the corresponding normal Bayesian methods, in particular, when the number of phenotypic categories was small. Their calculating accuracies at first generations were similar to our results, but after generation 2, their finding was higher than our study except BayesA method. In conclusion, they suggested that threshold model for predicting GEBVs of threshold traits.

We calculated relatively higher accuracies by increasing the number of markers. Solberg et al. [20] reported that the accuracy of selection increased from 0.63 to 0.83 as the density of markers increased.

The presented simulations assumed a relatively small effective population size of Ne = 100, which generates linkage disequilibrium(LD) between the markers and QTLs for creating the marker effects. The LD is the key factor that is driving the genomic prediction process, and, to further confirm this, some simulations were tested to how the accuracy changed with increasing LD. The simulated population was therefore stopped before it reached a balance between recombination and drift, and before the equilibrium amount of LD was reached, resulting in a lower LD. The expected amount of disequilibrium in a stable population represents a balance between its creation by drift and its decay by recombination.

A reduced heritability will lead to a decrease in accuracy of predicting the breeding value but can be compensated for by using a larger number of observations to estimate the marker effects. A large part of this reduction can probably be avoided by updating the prediction model with the most recent data. Calus & Veerkamp [15] confirmed that accuracy increased as data from the latest generation was included in the genetic evaluation of GEBV. The basis for the high reduction and possible improvements should be investigated more thoroughly.

Results for low heritability trait confirmed that accuracy of GEBV will improve by increasing the number of phenotypic records in the training set even if new records are from more distant generations. Accuracy of estimated breeding values based on marker distance and number of phenotypic records in the training set is similar to other studies like Meuwissen et al. [21].

Muir [23] proved that after several generations following estimation of marker effect, the accuracy of GEBVs reduces and these effects should be re estimated. However, there is a tendency for the rate of decay to be a little higher for the first two to three generations, which is the critical time frame for the use of genomic predictions in dairy cattle. We found similar trend for decay in accuracies in our research. It could be interesting to understand the effective factors of this decay and to see if relative improvements in the genomic prediction models could be made to keep reliabilities high over a longer time span.

Because of the only gradual decay in the reliability of GEBV over time, it can be concluded that GS will be well applicable to dairy breeding setting without progeny testing. As progeny testing is not necessary, costs of breeding may be reduced and in addition genetic gain per year is accelerated by reducing the generation interval. Implementation of GS for dairy breeding requires high accuracies of GEBV for at least two to three generations ahead without having phenotypes.

However, high reliabilities could be obtained for all of heritabilities with appropriate models, showing that GS can be especially useful for low heritable traits. Nejati- Javaremi et al. [1] replaced pedigree based relationship by marker-based total allelic relationship and documented its impact on reducing prediction error variance, hence, increasing accuracy of evaluation. Habier et al. [9] indicated that genomic selection uses genomic relationship among individuals and LD between markers and QTL to improve accuracy of GEBVs. They reported increase in accuracy of evaluation is partly due to using genomic relationship information among individuals.

It is important to notice, that in the presented simulations the population is random mating. Muir [23] reported that the decay of accuracies were faster in situations with directional selection compared with random mating, which means that change in allele frequencies, and generation of LD also may be effective. For a trait with a heritability of 0.10, he found that random selection resulted in a decay in accuracies from about 0.6 down to 0.35 over six generations if population started in Hardy–Weinberg equilibrium (HWE) and from 0.65 to 0.55 if population started in mutation-drift equilibrium (MDE). The faster reduction in accuracies with selection is explained by changes in allele frequencies, genetic variance and generation LD [14].

GS will be useful in dairy cattle breeding because the reliabilities of GEBV are high, and the decay in reliability over generations is slow. This is also useful for traits with low heritability, and therefore GS can lead to more balanced selection schemes. Models with effects of haplotypes perform better than single SNP effects, but the optimal length of haplotypes will depend on different factors such as LD, marker distances and the population and must therefore be optimized for the specific data.

With traditional selection schemes, low heritable traits are only improved slowly because reliability of the breeding value depends strongly on the heritability. With GS the reliability related to different factors like size of dataset used to make the marker associations, recombinations, LD and marker density which is similar for both low and high heritability traits. Therefore, use of GS may lead to a more balanced selection index than obtained in the classic selection methods. With GS the necessity to have phenotypes on close relatives of breeding candidates is relaxed, as the prediction ability persists over several generations. This gives the opportunity to perform difficult recordings in designated herds, which are then used to construct prediction models for the total population. Thus, the number of observations needed to estimate accuracy increases with decreased heritability. This is most likely why we see a decline in accuracies with lower heritabilities.

Results also showed the accuracies of the traits with higher heritability is more accurate than accuracies for the traits of lower heritability. Similar results have been reported from other studies [13, 22]. However, if the cost of genotyping is an issue it may be recommended to use genotypes (and phenotypic information) of individuals from more recent generations.

Therefore, GS is possible and a very interesting approach to replace or supplement progeny testing. Results showed that there is no barrier to achieving such accuracies using genomic evaluations in practice. GS has improved dairy cattle breeding by greatly increasing the accuracies of genetic merit estimation and the rate of genetic progress by shortening the generation interval. GS so far has focused on continuous traits. However, many of traits that significantly affect profitability and are difficult to be selected belong to threshold traits.

Among different existing approaches for estimating genomic breeding values of quantitative traits, the three normal Bayesian methods (BayesA, BayesB and BayesCπ) are commonly used, although, they are not suitable for threshold traits, because they are based on linear models.

CONCLUSION

Genomic selection for threshold traits can be used as well as continuous traits in practical. However, method for estimating genomic breeding values of threshold traits is scarce yet and maybe it needs to be done more research because of having different data structure compare with continuous traits.

REFERENCES

[1] A. Nejati-Javaremi, C. Smith, J. P. Gibson, *J. Anim. Sci*, **1997**, 75, 1738-1745.
[2] A. J. Wilson, J. A. Hutchings, M. M. Ferguson, *J. Evol. Biol*, **2003**, 16, 584-594.

[3] B. J. Hayes, P. J. Bowman, A. J. Chamberlin, M. E. Goddard, *J. Dairy Sci*, **2009**, 92, 433-443.
[4] B. J. Hayes, P. J. Bowman, A. C. Chamberlain, K. Verbyla, M. E. Goddard, *Genet. Sel. Evol*, **2009**, 41, 41-51.
[5] C. L. Wang, X. D. Ding, J. Y. Wang, J. F. Liu, W. X. Fu, Z. J. Yin , Q. Zhang, *Heredity*, **2013**, 110, 213-219.
[6] D. Gianola, R. L. Fernando, A. Stella. *Genetics*, **2006**, 173, 1761-1776.
[7] D. Habier, J. Tetens, F. R. Seefried, P. Lichtner, G. Thaller, *Genet. Sel. Evol*, **2010**, 42, 5-17.
[8] D. Habier, R. L. Fernando, K. Kizilkaya, D. J. Garrick, *BMB Bioinformatics*, **2011**, 186-198.
[9] D. Habier, R. L. Fernando, J. C. M. Dekkers, *Genetics*, **2007**, 177, 2389-2397.
[10] D. Kolbehdari, L. R. Schaeffer, J. A. B. Robinson, *J. Anim. Breed. Genet*, **2007**, 124, 356-361.
[11] J. Y. Wang, Y. R. Luo, W. X. Fu, X. Lu, J. P. Zhou, X. D. Ding, J. F. Liu, Q. Zhang, *Anim. Genet*, **2012**, 44, 34-43.
[12] L. R. Schaeffer, *J. Anim. Breed. Genet*, **2006**, 123, 218-223.
[13] M. E. Goddard, *Genetica*, **2009**, 136, 245-257.
[14] M. G. Bulmer, *Am. Nat*, **1971**, 105, 201-211.
[15] M. P. L. Calus, R. F. Veerkamp. **2007**, *J. Anim. Breed. Genet*, 124, 362-368.
[16] M. P. L. Calus, T. H. E. Meuwissen, A. P. W. Deroos, R. F. Veerkamp, *Genetics*, **2008**, 178, 553-561.
[17] P. M. VanRadan, *J. Dairy Sci*, **2008**, 91, 4414-4423.
[18] S, Xu, *Genetics*, **2003**, 163, 789-801.
[19] S, Xu, Z. Jia, *Genetics*, **2007**, 175, 1955-1963.
[20] T. R. Solberg, A. K. Sonesson, J. A. Woolliams, T. H. E. Meuwissen, *J. Anim. Sci*, **2008**, 86, 2447-2454.
[21] T. H. E. Meuwissen, B. J. Hayes, M. E. Goddard, *Genetics*, **2001**, 157, 1819-1829.
[22] T. M. Villumsen, L. Janss, M. S. Lund, *J. Anim. Breed. Genet*, **2009**, 126, 3-13.
[23] W. M. Muir, *J Anim. Breed. Genet*, **2007**, 124, 342-355.