Joint Estimation of Additive and Dominance Effects of Markers Using a Genomic Model with a Residual Polygenic Effect

Z. Liu, H. Alkhoder , F. Reinhardt and R. Reents
vit, *Heideweg 1, D-27283 Verden, Germany*

Abstract

Routine genotyping of female animals in genomic selection programme opens up the opportunity to estimate non-additive genetic effects of SNP markers by using direct phenotypes of the genotyped cows. Non-additive effects of markers, e.g. dominance effects, provide useful information for genomic mating and herd management. We extended our current BLUP SNP genomic model, which includes a residual polygenic effect, to additionally estimate dominance effects of SNP markers. A new estimation algorithm was implemented to improve the rate of convergence for all estimated effects of the new genomic model. Parallel computing technique was applied to reduce the total clock time of the full estimation process. To study the convergence behaviour of the genomic dominance model, lactation yield deviations of cows and associated effective data contribution were obtained from German routine genetic evaluation for milk production and somatic cell scores based on a random regression test-day model. Genotypic and phenotypic data of a total of 17,635 cows were used to estimate dominance effects and breeding values of the SNP markers together with residual polygenic effects. Dominance variance of 5% and 10% of the total genetic variance were assumed to investigate the influence of dominance variance. Interim solutions of the model effects were compared to the final estimates from a long iteration process with 10,000 rounds. Based on the correlations between the dominance effect and breeding value estimates, it seems that the dominance effects of a large number of SNP markers can be accurately estimated and properly separated from the breeding value effects within a reasonable time frame. Further results on the convergence behaviour of the new genomic model are discussed.

Key words: genomic model, substitution effect, dominance effect, SNP marker

Introduction

Genomic evaluation for German Holstein (Liu *et al*., 2011; Alkhoder *et al.,* 2014) is based on a bull reference population. As large-scale genotyping of female animals becomes a routine practice, genomic reference populations can now include cows that are preferably non-selectively genotyped or phenotyped. Using a cow reference population, we can avoid the negative impact of genomic pre-selection on genetic evaluation of male animals (Schaeffer, 2014) and provide an alternative reference population for novel traits, such as health traits, for which a bull reference population would require a much longer time to be set up. In addition, we can explore non-additive genetic effects for breeding and mating by using direct phenotypes of genotyped cows (Toro and Varona, 2010). Including dominance effects of SNP markers in genomic evaluation can improve the accuracy of genomic prediction (Toro and Varona, 2010). A genomic mating programme using dominance effects can obtain extra gain in selection response (Sun *et al.,* 2013; Toro and Varona, 2010). A genomic model containing additive and dominance effects of SNP markers were initially developed by Su *et al*. (2012) and others (Wittenburg *et al.,* 2011) and applied in some studies, e.g. Da *et al*. (2014). This genomic model estimates both additive and dominance effects of markers jointly and has been improved by Vitezica *et al*. (2013) for consistency with the conventional polygenic model. Instead of including the additive genetic effects of markers in the model by Su *et al*. (2012), Vitezica *et al*. (2013) proposed estimating substitution or breeding value effects in addition to the dominance effects of the markers. Fitting a residual polygenic (RPG) effect was shown to be important for reducing the inflation of genomic prediction

(Liu *et al.*, 2011). The objectives of this study were to develop a BLUP SNP genomic model with dominance as well as a RPG effect and test a new software for this model, and to investigate the convergence behaviour of the new genomic dominance model.

Materials and Methods

The current genomic evaluation for German Holsteins (Liu *et al.*, 2011; Alkhoder *et al.*, 2014) uses indirect phenotypes of reference bulls, deregressed proofs (DRP). Because DRP are derived from conventional genetic evaluation based on an additive polygenic model, no dominance effects are expected to be contained in the DRP of the bulls. However, adjusted phenotypes of cows, e.g. lactation yield deviations (LYD, Liu *et al.*, 2004), may contain non-additive genetic effects, if they do exist for the analysed traits.

Genomic Models with Dominance Effects

A BLUP SNP model (Liu *et al.*, 2011) was extended to additionally estimate dominance effects of SNP markers:

$$y_i = \mu + u_i + \sum_{j=1}^{m} z_{ij}\alpha_j + \sum_{j=1}^{m} w_{ij} d_j + e_i \qquad [1]$$

where y_i is LYD of cow *i* of milk production traits or somatic cell score based on a random regression test-day model (Liu *et al.*, 2004), μ is a general mean, u_i is residual polygenic effect of the *i*-th cow, α_j is substitution or breeding value (Vitezica *et al.*, 2013) of SNP marker *j* (*j* =1, …, *m*), *m* is the number of SNP markers, d_j is dominance effect of marker *j* (Vitezica *et al.*, 2013), and e_i is residual effect of cow *i*. For the breeding value α_j of marker *j*, regression coefficients for cow *i* are defined as:

$$z_{ij} = (n_A - 2p_j) \Big/ \sqrt{2p_j(1-p_j)} \qquad [2]$$

where p_j is frequency of allele A of marker *j*, and n_A is the number of copies of allele A in genotype: n_A =2,1,0 for genotype AA, AB or BB respectively. Regression coefficients for dominance effect d_j are (Vitezica *et al.*, 2013):

$$w_{ij} = -2(1-p_j)^2 \big/ 2p_j(1-p_j) = -(1-p_j)\big/ p_j$$

for genotype AA;

$$w_{ij} = 2p_j(1-p_j)\big/ 2p_j(1-p_j) = 1$$

for genotype AB; and

$$w_{ij} = -2(1-p_j)^2 \big/ 2p_j(1-p_j) = -p_j \big/ (1-p_j)$$

for genotype BB. [3]

The regression coefficients on z_{ij} and w_{ij} are both centred and scaled.

Variance of the residual effects is assumed as: $\text{var}(e_i) = \sigma_e^2 / \phi_i$ [4]

with ϕ_i denoting effective data contribution (EDC) of the cow *i*.

The proportion of variance contributed by the residual polygenic effect in the total genetic variance is defined as: $0 < k < 1$. Furthermore, we assume a proportion of genetic variance on genome level contributed by the dominance effect, k_d. Thus, the proportion of genetic variance contributed by breeding values of all the markers is $k_\alpha = 1 - k - k_d$.

An alternative dominance model was developed earlier by Su *et al*. (2012):

$$y_i = \mu + u_i + \sum_{j=1}^{m} z_{ij} a_j + \sum_{j=1}^{m} h_{ij} d_j + e_i \qquad [5]$$

where a_j is additive genetic effect of SNP marker *j*, and h_{ij} represents regression coefficient of marker *j* for the dominance effect of cow *i*.

The differences between the statistical breeding value model [1] and the biological genotypic value model [5] were given by Vitezica *et al*. (2013). Although both models are equivalent and can be converted to each

other, the genomic model [1] of breeding values of SNP markers was preferred for the reason of consistency with conventional genetic evaluation, because we included the RPG effect u_i in our routine genomic prediction (Liu *et al.*, 2011).

Data Materials

Genotype data were obtained from April 2015 official genomic evaluation for German Holsteins. Original genotypes of cows from lower density chips, mainly EuroG10K chip, were imputed to Illumina 50K version 2 using the software findhap. The analysed phenotype data of the genotyped cows were first lactation LYD (Liu *et al.*, 2004) of three milk production yield traits or somatic cell scores. To reduce the impact of short-lactation problem, a minimum of six test-day records was imposed for selecting the cows. A total of 17,635 genotyped and phenotyped cows were finally used as reference animals and the pedigree of the reference cows contained 160,450 animals. Figure 1 shows the numbers of the reference cows and pedigree animals by year of birth. It can be seen that most of the reference cows were born in 2011 or 2012. The majority of the cows born before 2011 were selectively genotyped as dams of bulls or elite cows. There were many generations of ancestors in the deep pedigree of the reference cows. Although the selectively genotyped cows may not be regarded as optimal reference animals, their data can be used, nonetheless, for developing our genomic dominance model.

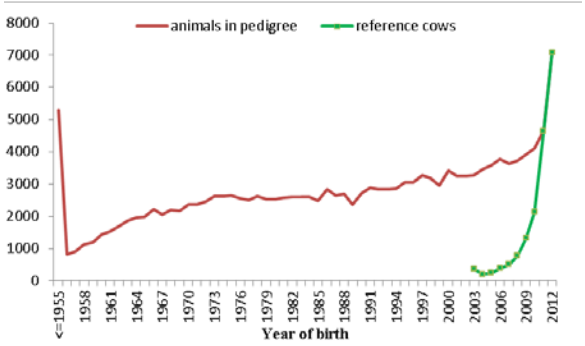

Figure 1. Number of the reference cows and related animals in pedigree by year of birth.

Application of the Dominance Models

Computer programs for our current BLUP SNP model (Liu *et al.*, 2011), which includes a RPG effect, were modified to additionally estimate the dominance effects of SNP markers. The number of estimated marker effects was doubled on top of the RPG effects. The Gauss-Seidel algorithm with a special residual update (Legarra and Misztal, 2008) was kept with more estimation emphasis put on the two sets of marker effects. Because of identical process between markers, the steps of estimating both sets of effects of the markers were parallelised for multiple cores in order to reduce the total clock time.

In literature, studies reported difficulty in accurately estimating dominance effects based on only pedigree information in conventional genetic evaluation (Vitezica *et al.*, 2013). Therefore, emphasis was also put on how well the dominance effects were estimated.

Results and Discussion

Both the statistical genomic dominance model [1] with breeding values of markers by Vitezica *et al*. (2013) and the genomic dominance model [5] with additive effects of markers by Su *et al*. (2012) were applied to the genotype and phenotype data. Two scenarios on dominance variance were investigated: 5% or 10% of the total genetic variance being assumed to be contributed by dominance effects, although the dominance variance could be estimated (Ertl *et al.*, 2013). The number of test runs amounted to 16: 2 genomic dominance models times 4 traits times 2 dominance variances. To investigate the rate of convergence of the genomic dominance models, a long iteration process of 10,000 rounds was performed and solutions from the final round were treated as true values of the model effects.

Estimation of the SNP marker and RPG effects was conducted on a Linux server with Intel Xeon CPU E5-2690 v2 @ 3.00GHz processors. Approximately 6.5 Gb RAM was

required, mainly for storing the regression coefficients of marker effects. The solving program took about 11 hours on 20 cores for the 10,000 rounds of iteration.

Rate of Convergence

Figure 2 shows the rates of convergence for the breeding value and dominance effects of the SNP markers and the residual polygenic effects. Convergence criterion was defined as logarithm of sum of squared solution differences between two consecutive rounds divided by sum of squared solutions of the current round. It can be seen that estimates of the two sets of marker effects were converged very fast and well. In contrast, the residual polygenic effects did not converge as well as the marker effects, probably caused by the very deep pedigree of the reference cows. However, the impact of the convergence rate of the residual polygenic effects is limited due to its small variance. No difference in rates of convergence was observed between the two scenarios of 5% and 10% dominance variances.

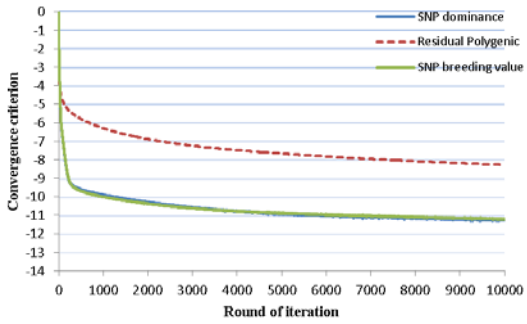

Figure 2. Convergence criteria of the genomic dominance model [1] for first lactation milk yield with 5% dominance variance.

Breeding Value SNP and DGV Effects

Correlations of SNP or DGV interim solutions with the final round increased with the number of rounds and reached 0.99984 at round 1000, indicating both SNP and DGV of breeding value effects were converged to final estimates rather fast. Figure 3 shows changes in estimates of the breeding value DGV effects for milk yield of the scenario of 5% dominance variance. It can be seen that both averages and standard deviations of the DGV changes are small and decrease with longer iteration. For instance, standard deviation of the solution changes is less than 2.5% genetic standard deviations. Maximum changes are not shown here, because these values were associated with cows with extreme LYD or with imputed genotypes.

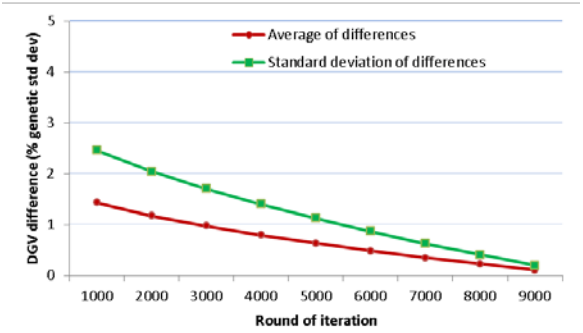

Figure 3. Differences of interim to the final estimates of breeding value DGV for first lactation milk yield with 5% dominance variance.

Dominance SNP and DGV Effects

Similar as for the breeding value effects, correlations of dominance SNP or DGV interim solutions with the final round increased with the number of rounds and reached 0.99986 at round 1000, indicating both SNP and DGV of dominance effects were converged to final estimates rather fast, too. From Figure 4 it can be seen that averages and standard deviations of the differences of dominance DGV are already small at round 1000, c.a. 1% genetic standard deviations, even smaller than the breeding value DGV in Figure 3 due to the smaller assumed value of dominance variance.

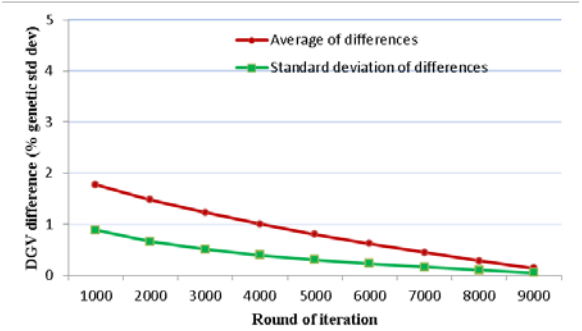

Figure 4. Differences of interim to the final estimates of dominance DGV for first lactation milk yield with 5% dominance variance.

Correlations of breeding value with dominance effects of markers

On both individual SNP markers and DGV levels, correlations between breeding values and dominance effects reached their respective final values already within first 100 rounds of iteration (Figure 5), suggesting that separation of the two sets of effects is straightforward based on genomic information. The correlation of breeding value with dominance effects was null for SNP effects. However, the DGV correlation between the two sets of effects was slightly positive, 0.187. It is unclear if the DGV correlation for the selected reference cows should be expected to be zero.

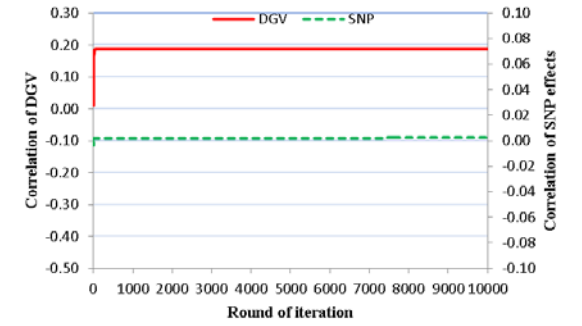

Figure 5. Correlations of breeding value with dominance effects for SNP effects as well as for DGV of first lactation milk yield with 5% dominance variance.

Correlations of RPG effects with DGV

Figure 6 shows the correlations of RPG effects with breeding value or dominance DGV during the iteration process. We can clearly see that the correlations of the RPG effects are stabilised rather slowly, with both breeding value and dominance DGV. This slower convergence in the correlations may possibly be caused by the high depth of pedigree of the reference cows. This may suggest a longer iteration process to achieve better convergence of the effect estimates.

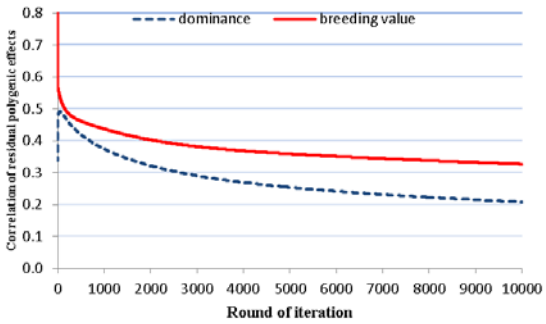

Figure 6. Correlations of residual polygenic effects with breeding value or dominance DGV for first lactation milk yield with 5% dominance variance.

Comparison of the two genomic dominance models

The two genomic dominance models [1] and [5] were applied to the same data sets in order to compare their effect estimates. Little difference was observed in rate of convergence between the two dominance models. Correlations of SNP effect estimates were 0.987 between breeding value effects of Vitezica *et al*'s (2013) model and additive effects of Su *et al*. (2012), and 0.918 between dominance effects of both models. For DGV effect estimates, correlation was 0.849 for dominance effect and 0.980 for breeding value

and additive genetic effects. Despite the less than unity correlations for either dominance or additive effects, sums of DGV dominance and additive effects were practically fully correlated, with a correlation of 0.998, indicating that the two dominance models separate the two sets of effects differently.

Conclusion

Dominance effects can be straightforwardly and accurately estimated based on genomic information, in contrast to conventional evaluation based on only pedigree and phenotype data. Despite of twice as many marker effects to be estimated in the genomic dominance model, the statistical problem of $p>>n$ did not become worse, compared to a genomic model with only additive genetic effects of markers. Estimating both breeding value and dominance effects of SNP markers can be run on multiple computer cores simultaneously to reduce total clock time, thanks to the identical processes between SNP markers.

Large-scale genotyping and phenotyping female animals enable set up a cow reference population, particularly for novel traits. Direct phenotypes of the genotyped cows allow joint estimation of non-additive genetic effects, e.g. dominance effects, together with additive genetic effects. For testing our genomic dominance model with a residual polygenic effect, first lactation yield deviations of milk production yield traits and somatic cell scores were used as test traits. A total of 17,635 reference cows were analysed with a pedigree containing 160,450 related animals. The statistical genomic dominance model [1] by Vitezica *et al*. (2013) was compared to the biological genomic dominance model [5] by Su *et al*. (2012). These two genomic dominance models differ in modelling additive effects as well as dominance effects. Breeding value and dominance effects of SNP markers were converged much better than the RPG effects that had a very deep pedigree for the reference cows. Already at round 1000, estimates of both SNP effects or DGV of the two sets of effects no longer changed much. The breeding value and dominance effects were able to be separated accurately already within the first 100 rounds of iteration, however, separating residual polygenic effects from the two sets of SNP marker effects seemed to require many rounds of iteration, indicated by the slower convergence in correlations of residual polygenic effects with either set of effects. Although the two genomic dominance models led to identical sums of DGV estimates of the two sets of effects, Vitezica *et al*'s model is preferred, because this dominance model is consistent with classical definitions of breeding value and dominance effects.

Our genomic dominance model needs to be extended to novel traits, in particular to those with low heritability. Optimal variance of dominance needs to be determined either via genomic validation or by direct estimation. Predictive ability of the genomic dominance model needs to be investigated and compared to the current genomic model with additive genetic effects of SNP markers only.

Acknowledgements

The German research project Kuh-L is acknowledged for financial support. Dr. Guosheng Su, Denmark, is thanked for discussion about the two genomic dominance models.

References

Alkhoder, H., Liu, Z., Bach, Th., Pasman, E. & Reinhardt, F. 2014. A Continuous Genomic Evaluation System for German Holsteins. *Interbull Bulletin 48,* 100-104.

Da, Y., Wang, C., Wang, S. & Hu, G. Mixed Model Methods for Genomic Prediction and Variance Component Estimation of Additive and Dominance Effects Using SNP Markers. 2014. *PLos ONE* 9(1), e87666.

Ertl, J., Legarra, A., Vitazica, Z.G., Varona, L., Edel, C., Emmerling, R. & Götz, K.-U. 2013. Genomic Analysis of Dominance Effects in Milk Production and Conformation Traits of Fleckvieh Cattle. *Interbull Bulletin 47,* 28-31.

Legarra, A. & Misztal, I. 2008. *Technical Note:* Computing Strategies in Genome-Wide Selection. *J. Dairy Sci. 91:1,* 360-366.

Liu, Z., Reinhardt, F., Bünger, A. & Reents, R. 2004. Derivation and Calculation of Approximate Reliabilities and Daughter Yield-Deviations of a Random Regression Test-Day Model for Genetic Evaluation of Dairy Cattle. *J. Dairy Sci. 87:6,* 1896-1907.

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

Schaeffer, L.R. 2014. Is the animal model obsolete? *Document Online*.

Su, G., Christensen, O.F., Ostersen, T., Henryon, M. & Lund, M.S. 2012. Estimating Additive and Non-Additive Genetic Variances and Predicting Genetic Merits Using Genome-Wide Dense Single Nucleotide Polymorphism Markers. *PLoS ONE* 7(9):e45293.

Sun, C., VanRaden, P.M., O'Connell, J.R., Weigel, K.A. & Gianola, D. 2013. Mating programs including genomic relationships and dominance effects. *J. Dairy Sci. 96:12,* 8014-8023.

Toro, M. & Varona, L. 2010. A note on mate allocation for dominance handling in genomic selection. *Genet. Sel. Evol. 42:33.*

Vitezica, Z.G., Varona, L. & Legarra, A. 2013. On the Additive and Dominant Variance and Covariance of Individuals Within the Genomic Selection Scope. *Genetics 195:4,* 1223-1230.

Wittenburg, D., Melzer, N. & Reinsch, N. 2011. Including non-additive genetic effects in Bayesian methods for the prediction of genetic values based on genome-wide markers. *BMC Genetics 12:74.*