Single-step genomic model improved reliability and reduced the bias of genomic predictions in Danish Jersey

P. Ma,^{*1} M. S. Lund,^{*} U. S. Nielsen,[†] G. P. Aamand,[‡] and G. Su^{*}

^{*}Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, DK-8830 Tjele, Denmark

[†]Seges, DK-8200 Aarhus, Denmark

[‡]NAV Nordic Cattle Genetic Evaluation, DK-8200 Aarhus, Denmark

ABSTRACT

A bias in the trend of genomic estimated breeding values (GEBV) was observed in the Danish Jersey population where the trend of GEBV was smaller than the deregressed proofs for individuals in the validation population. This study attempted to improve the prediction reliability and reduce the bias of predicted genetic trend in Danish Jersey. The data consisted of 1,238 Danish Jersey bulls and 611,695 cows. All bulls were genotyped with the 54K chip, and 1,744 cows were genotyped with either 7K chips (1,157 individuals) or 54K chips (587 individuals). The trait used in the analysis was protein yield. All cows with EBV were used in a single-step approach. Deregressed proofs were used as the response variable. Four alternative approaches were compared with genomic best linear unbiased prediction (GBLUP) model with bulls in the reference data (GBLUP_{Bull}): (1) GBLUP with both bulls and genotyped cows in the reference data; (2) GBLUP including a year of birth effect; (3) GEBV from a GBLUP model that accounted for the difference of EBV between dams and maternal grandsires; and (4) using a single-step approach. The results indicated all 4 alternatives could reduce the bias of predicted genetic trend and that the single-step approach performed best. However, not all these approaches improved reliability or reduced inflation of GEBV. The reliability was 0.30 and regression coefficients of deregressed proofs on GEBV were 0.69 in the scenario GBLUP_{Bull}. When genotyped cows were included in the reference population, the regression coefficients decreased to 0.59 but the reliability increased to 0.35. If a year effect was included in the model, the prediction reliability decreased to 0.29 and the regres-

sion coefficient improved to 0.75. The method in which GEBV were adjusted for the difference between dam EBV and maternal grandsire EBV led to much lower regression coefficients though the reliability increased to 0.4. The single-step approach improved both the reliability, to 0.38 and regression coefficient to 0.78. Therefore, the bias in genetic trend was reduced. The results suggest that implementing the single-step approach is an effective way to improve genomic prediction in Danish Jersey cattle.

Key words: bias of predicted genetic trend, reliability of genomic prediction, single-step method

INTRODUCTION

Genomic prediction has been widely used in dairy cattle since genome-wide dense marker chips became available. To obtain accurate prediction, a large reference population is needed (Goddard and Hayes, 2009; Hayes et al., 2009a). In dairy cattle, usually progeny-tested bulls are used to form the reference population. In some large populations, such as Holsteins, accurate prediction using genomic information has been obtained (VanRaden et al., 2009; Lund et al., 2011). For Danish Jerseys it is quite challenging to obtain a large reference population because a limited number of progeny-tested bulls are available (Thomasen et al., 2012). One way to overcome this limitation is to add genotyped cows to the reference population. However, previous studies have reported an inflation of the genomic estimated breeding values (GEBV) when cows were included into the training set (Wiggans et al., 2011; Calus et al., 2013), because the genotyped cows are usually elite and possible get preferential treatment. Another strategy is to make use of the phenotypic information from nongenotyped animals. A popular approach is to apply a single-step model which estimates genomic breeding values using the information of genotyped and nongenotyped individuals simultaneously by integrating marker- and pedigree-based relationship matrix

Received April 13, 2015.

Accepted August 14, 2015.

¹Corresponding author: peipei.ma@mbg.au.dk

into a joint relationship matrix (Miształ et al., 2009; Christensen and Lund, 2010; Aguilar et al., 2010).

Nordic routine genomic genetic evaluation has observed a bias of predicted genetic trends in Danish Jerseys. Bias of predicted genetic trends was defined as the annual deviation of GEBV from the deregressed proofs (**DRP**) of the animals in the test population. Bias of predicted genetic trends may lead to an unfair comparison of animals across birth years. The bias could be caused by a discrepancy between assumptions of the genomic prediction models and the selection histories of the practical populations (Vitezica et al., 2011). The genomic prediction models assume there is no selection in the population, which is used for implementing genomic prediction (Hayes et al., 2009b). However, in practice, the genotyped populations usually consist of selected animals such as progeny-tested bulls and elite cows. The single-step approach accounts for the selection by including all records in the model. Therefore, this approach is expected to minimize the bias. Another possible solution to reduce the bias is to add a year of birth effect in the model, which may lead to a robust estimation of genetic trend (Ducrocq, 2010). Therefore, the genetic progress on the maternal side could be taken into account by the year trend. Similarly, adjusting GEBV for the difference between EBV of dam and maternal grandsire (**MGS**) may reduce bias of predicted genetic trend.

The objectives of our study were to investigate the prediction reliability and bias of predicted genetic trend in Danish Jersey. A second objective was to increase prediction reliability and reduce bias of predicted genetic trend using various strategies such as adding genotyped cows to the reference population, including year effect into the prediction model, accounting for the difference of EBV between dam and MGS, and applying a single-step approach.

MATERIALS AND METHODS

Data

Danish Jersey data were used in our study. There were 2,982 genotyped individuals comprising 1,238 bulls born between 1981 and 2009 and 1,744 cows born between 2000 and 2011, with most of them (1,733) born after 2004. Most cows (1,157) were randomly selected from a few herds, whereas the others (587) were selected as potential bull dams by individual farms according to their own breeding schemes. The **DRP** of protein used in different scenarios were calculated from EBV of genetic evaluation in November 2013. When using the single-step approach, all cows with EBV for protein were used in the analysis. After tracing the pedigree

to as many generations as possible for the cows with EBV and bulls with genotypes, the pedigree used for single-step prediction included 819,988 individuals. The **DRP** for all cows were calculated using Mix99 (Lidauer and Strandén, 1999; Strandén and Mäntysaari, 2010); it required that the cows had an effective record contribution (**ERC**) larger than 0.1. This reduced the number of cows with **DRP** to be 611,695. Cows which are daughters of the test bulls (described later) were excluded. After filtering, the number of cows with **DRP** used in the single-step approach was 577,405.

The bulls were genotyped with Illumina BovineSNP50 BeadChip (**54K**; Illumina, San Diego, CA), which includes 54,001 SNP. Bull dams (587) were genotyped with 54K chips. Randomly selected cows (1,157) were genotyped with Illumina BovineLD BeadChip (**LD**) which includes 6,909 SNP. The LD data were imputed to 54K with Beagle (Browning and Browning, 2009) using the 54K genotyped animals as imputation reference population. The markers used for prediction were from 29 autosomes. The genotypes for genomic prediction were edited by deleting the markers with minor allele frequency less than 0.01 and the markers in complete linkage disequilibrium ($r^2 = 1$) with the previous marker. After editing, 38,967 markers were used for genomic prediction.

Methods

To validate the prediction accuracy and unbiasedness, the Jersey bulls were divided into reference and test sets using a cut-off date of birth of January 1, 2005. The bulls born after this date were used as validation animals (208 bulls). Thus, in the scenario using only bull reference data, 1,030 bulls were used as reference population.

Besides the genomic BLUP model (**GBLUP**) with bulls in the reference data (**GBLUP_{Bull}**), 5 alternative approaches were used in our study. The first was including pedigree relationships to weight the genomic relationship (**GBLUPW_{Bull}**). Approach 2 was the **GBLUP** model with both bulls and genotyped cows in reference set (**GBLUP_{Cow}**), in which, 25 cows were dams of test bulls. Approach 3 included a year of birth effect in the **GBLUP** model (**GBLUP_{Year}**) to account for the part of genetic trend that is not accounted for by SNP markers. Approach 4 was to adjust GEBV using the difference of EBV between dams and maternal grandsires (**GBLUP_{Dam_mgs}**). Approach 5 was a single-step method to integrate the information of genotyped and nongenotyped animals for genomic prediction. Two scenarios of this approach were investigated, which were the predictions either using cow genotypes (**SS_{PG}**) or without using cow genotypes (**SS_P**). The numbers of

individuals used in the reference population and test population in different scenarios are shown in Table 1.

Statistical Models

The statistical models used in different scenarios are described below.

GBLUP. The GBLUP model was

$$\mathbf{y} = 1\mu + \mathbf{Z}g + \mathbf{e},$$

where $\mathbf{y}$ is a vector of DRP of animals in reference population; μ is the overall mean; g is the direct genomic value; $\mathbf{Z}$ is the design matrix for linking g to $\mathbf{y}$; and $\mathbf{e}$ is a vector of the random residuals. Random effects were assumed to be distributed as $g \sim N(0, \mathbf{G}_\omega \sigma_g^2)$ and $\mathbf{e} \sim N(0, \mathbf{D} \sigma_e^2)$, where σ_g^2 is the additive genetic variance, $\mathbf{G}_\omega$ is the genomic relationship matrix, σ_e^2 is the residual variance, and $\mathbf{D}$ is a diagonal matrix with elements $d_{ii} = (1 - r_{\text{DRP}}^2) / r_{\text{DRP}}^2$, in which r_{DRP}^2 is the reliability of DRP. The GEBV was calculated as $\text{GEBV} = \hat{\mu} + \hat{g}$. The genomic relationship matrix, $\mathbf{G}_\omega$, is defined as

$$\mathbf{G}_\omega = \omega \mathbf{A} + (1 - \omega) \mathbf{G},$$

where $\mathbf{G}$ is genomic relationship matrix described in VanRaden (2008), and $\mathbf{A}$ is pedigree relationship matrix. In scenarios GBLUP_{Bull} and GBLUP_{Cow}, $\omega = 0$. In scenario GBLUPW_{Bull}, $\omega = 0.2$.

GBLUP with Year Effect. When the year effect is included in a GBLUP model, the model was

$$\mathbf{y} = 1\mu + b\mathbf{X} + \mathbf{Z}g + \mathbf{e},$$

where b is a regression coefficient of $\mathbf{y}$ on birth years, and $\mathbf{X}$ is a vector of birth years, treated as continuous

covariates in this model. The GEBV from GBLUP_{Year} was calculated as $\text{GEBV} = \hat{\mu} + \hat{b} \times \text{year} + \hat{g}$.

Adjusting GEBV for the Difference of EBV Between Dams and MGS. In traditional genetic evaluations for an individual without own or offspring records, when both sire EBV and dam EBV are available, the EBV for the individual is

$$\text{EBV}_o = \frac{1}{2}\text{EBV}_{\text{sire}} + \frac{1}{2}\text{EBV}_{\text{dam}}.$$

When only the bulls' EBV (sire EBV and maternal grandsire EBV) are available, the EBV for the individual is

$$\text{EBV}_o = \frac{1}{2}\text{EBV}_{\text{sire}} + \frac{1}{4}\text{EBV}_{\text{mgs}},$$

where the dam EBV is supposed as the average of EBV from all the daughters of the maternal grandsire, which is not the case because bull dam has high EBV due to selection. The difference between EBV_{mgs} and EBV_{dam} may cause an underestimation of GEBV of candidates when dams are absent in reference population. To reduce the influence by this difference, the GEBV for the validation animals were corrected by adding a value of $\frac{1}{2} \left(\text{EBV}_{\text{dam}} - \frac{1}{2} \text{EBV}_{\text{mgs}} \right)$.

Single-Step Model. The single-step model was as follows

$$\mathbf{y} = 1\mu + \mathbf{Z}\mathbf{a} + \mathbf{e},$$

where $\mathbf{y}$ is the vector of DRP of all the cows with EBV in the whole population, $\mathbf{a}$ is a vector of additive genetic effects, and $\mathbf{Z}$ is the design matrix for additive genetics effects. Random effects were assumed to be normally distributed $\mathbf{a} \sim N(0, \mathbf{H} \sigma_a^2)$ and $\mathbf{e} \sim N(\mathbf{D} \sigma_e^2)$, where σ_a^2 is the additive genetic variance and $\mathbf{H}$ is the

Table 1. The number of individuals in each scenario¹

Item	GBLUP _{Bull} / GBLUPW _{Bull} / GBLUP _{Year} / GBLUP _{Dam_mgs}	GBLUP _{Cow}	SS _P		SS _{PG}	
			No. of genotyped animals	No. of phenotyped animals	No. of genotyped animals	No. of phenotyped animals
Reference set	1,030	2,774	1,030	577,405	2,774	577,405
Test set	208	208	208	208	208	208

¹GBLUP_{Bull} = genomic BLUP model with bulls as reference population; GBLUPW_{Bull} = same as GBLUP_{Bull} but with a genomic relationship matrix $\mathbf{G}_\omega = 0.8\mathbf{G} + 0.2\mathbf{A}$, where $\mathbf{G}$ is a genomic relationship matrix and $\mathbf{A}$ is pedigree relationship matrix; GBLUP_{Cow} = GBLUP model with both genotyped bulls and cows as reference population; GBLUP_{Year} = year effects were included in the model as genetic trend; GBLUP_{Dam_mgs} = genomic EBV from GBLUP model using bull reference data were adjusted for the difference between dam EBV and maternal grandsire (mgs) EBV. SS_P = the single-step approach using phenotypes of all cows and genotypes of genotyped bulls. SS_{PG} = the single-step approach using phenotypes of all cows and genotypes of genotyped bulls and cows.

Table 2. Mean and SD of deregressed proofs (DRP) and reliability (R^2_{DRP}) of DRP for protein in different data sets

Trait	Genotyped bulls		Genotyped cows		Cows used in single-step approach	
	Mean	SD	Mean	SD	Mean	SD
Protein	89.29	13.95	106.62	18.70	78.89	29.09
R^2_{DRP}	0.92	0.04	0.44	0.07	0.36	0.06

relationship matrix of all the individuals as defined below. Here the reliability of DRP (r^2_{DRP}) was $\text{ERC}/(\text{ERC} + \lambda)$, where $\lambda = (1 - h^2)/h^2$.

Following Legarra et al. (2009), Aguilar et al. (2010), and Christensen and Lund (2010),

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

where $\mathbf{A}$ is pedigree relationship matrix and can be partitioned as $\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$ with subscript 1 for non genotyped individuals and 2 for genotyped individuals, and $\mathbf{G}_\omega = (1 - \omega)\mathbf{G} + \omega\mathbf{A}_{22}$. In our study the $\mathbf{G}$ matrix was adjusted for the differences in location and scale of pedigree-based relationship matrix ($\mathbf{A}_{22}$) using the method proposed by Christensen et al. (2012). Furthermore, ω was set as 0.2 according to the study by Gao et al. (2012).

The inverse of $\mathbf{H}$ (Aguilar et al., 2010; Christensen and Lund, 2010) was

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}_\omega^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}.$$

GEBV was calculated as $\text{GEBV} = \hat{\mu} + \hat{a}$. In all the models, the DMU package (Madsen et al., 2010) was used to estimate variance components and predict breeding values.

Validation of Predictions

The reliability of predictions was calculated as the squared correlation between GEBV and DRP divided

by the average reliability of the DRP in the test set. The bias was investigated by the regression coefficient and intercept of DRP corrected with model mean on estimated genetic effects (the year trend was added to the direct genomic values in scenario $\text{GBLUP}_{\text{Year}}$) and predicted genetic trend. Bias of predicted genetic trends was assessed by comparing year mean of GEBV with year mean of DRP for test individuals.

RESULTS

Descriptive statistics of DRP in different data sets are shown in Table 2. The mean DRP differ because individuals in different data sets were born in different periods. The average DRP of genotyped bulls was lower than the average DRP of the genotyped cows, whereas it was higher than the average of all the cows used in the single-step approach. This is caused by genetic progress over years due to selection and that genotyped cows were born in recent years.

The number of test individuals in each year varied from 43 to 55, except for year 2009, in which there were only 16 test individuals (Table 3). The mean of the DRP in each year varied from 103.06 to 109.08, and the standard deviation varied from 6.56 to 8.88.

The reliabilities of GEBV, as well as regression coefficients and intercept of DRP on GEBV for different scenarios are shown in Table 4. The reliabilities ranged from 0.29 to 0.40 in different scenarios. The reliability of GEBV from basic model ($\text{GBLUP}_{\text{Bull}}$) was 0.30. In scenario $\text{GBLUP}_{\text{Cow}}$, the reliability of GEBV increased. The scenarios of the single-step approach gained a large increase of reliability regardless of including cow genotypes or not. The reliability was 0.38 for scenario SS_{PC} , whereas it was 0.36 for Scenario SS_{P} . The highest reliability (0.40) was achieved in the scenario

Table 3. The number of individuals and mean and SD of deregressed proofs in each year in test set

Year	2005	2006	2007	2008	2009
No.	46	48	55	43	16
Mean	103.06	103.06	104.15	105.73	109.08
SD	7.09	7.28	8.42	8.88	6.56

Table 4. Reliabilities (R^2_{GEBV} ; Rel.) of genomic EBV (GEBV), regression coefficient (Reg. coef.), and intercept (Int.) of deregressed proofs on GEBV of test individuals in different scenarios¹

Item	GBLUP _{Bull}	GBLUPW _{Bull}	GBLUP _{Cow}	GBLUP _{Year}	GBLUP _{Dam_mgs}	SS _P	SS _{PG}
Rel.	0.30	0.29	0.35	0.31	0.40	0.36	0.38
Reg. coef.	0.69	0.72	0.60	0.75	0.58	0.78	0.74
Int.	6.00	6.55	7.97	5.90	3.79	5.88	6.92

¹GBLUP_{Bull} = genomic BLUP model with bulls as reference population; GBLUPW_{Bull} = same as GBLUP_{Bull} but with a genomic relationship matrix $\mathbf{G}\omega = 0.8\mathbf{G} + 0.2\mathbf{A}$, where $\mathbf{G}$ is the genomic relationship matrix and $\mathbf{A}$ is the pedigree relationship matrix; GBLUP_{Cow} = GBLUP model with both genotyped bulls and cows as reference population; GBLUP_{Year} = year effects were included in the model as genetic trend; GBLUP_{Dam_mgs} = GEBV from GBLUP model using bull reference data were adjusted for the difference between dam EBV and maternal grand-sire EBV; SS_P = the single-step approach using phenotypes of all cows and genotypes of genotyped bulls; SS_{PG} = the single-step approach using phenotypes of all cows and genotypes of genotyped bulls and cows.

GBLUP_{Dam_mgs}. The reliability increased 1 percentage point in the scenario GBLUP_{Year}. However, in scenario GBLUPW_{Bull}, the reliability decreased 1 percentage point. The regression coefficients varied from 0.58 to 0.78 in different scenarios. The regression coefficient from basic model (GBLUP_{Bull}) was 0.69. The regression coefficients increased to 0.72, 0.75, 0.74, and 0.78 in scenarios GBLUPW_{Bull}, GBLUP_{Year}, SS_{PG}, and SS_P, respectively. The regression coefficient in scenario GBLUP_{Cow} was 0.09 lower than in the scenario GBLUP_{Bull}. The regression coefficient was the lowest in the scenario GBLUP_{Dam_mgs}. The intercept for different scenario were much larger than 0, which indicated that the mean of GEBV was lower than the mean of DRP.

The trends of GEBV and DRP for genotyped bulls are shown in Figure 1. Bias of predicted genetic trend was observed in the scenario GBLUP_{Bull}. The difference between DRP and GEBV from GBLUP_{Bull} was around 5, which was statistically significant. Compared with scenario of GBLUP_{Bull}, all the alternative approaches reduced bias of predicted genetic trend to some extent except GBLUPW_{Bull}. Bias of predicted genetic trend was partly corrected in the scenario GBLUP_{Cow}. The scenario SS_P and SS_{PG} greatly reduced bias of predicted genetic trend. Scenario GBLUP_{Dam_mgs} also reduced bias of predicted genetic trend. Bias of predicted genetic trend was reduced slightly in scenario GBLUP_{Year}. Figure 2 shows the boxplots results for GEBV-DRP for each scenario in each birth year.

DISCUSSION

Our study investigated strategies to improve the prediction reliability and reduce bias of predicted genetic trend observed in the Danish Jersey population. Several strategies were tested; that is, including cows in the reference, including a year of birth effect in the prediction model, adjusting GEBV with the difference between dam EBV and MGS EBV, and using a single-step approach. The results showed that these strategies could

reduce bias of predicted genetic trend to some extent. However, the prediction reliability and regression coefficients did not consistently improve in parallel with the reduction in the bias of predicted genetic trend.

The regression coefficients in different scenarios were smaller than 1 in our study. One possible reason could be that markers were not in complete linkage disequilibrium with causal genes, and thus could not fully account for the total genetic variance. Another reason could be that the data used in the analysis were not a random sample, but selected data.

The reliability of prediction for protein was improved when the genotyped cows were included in the reference, as it clearly enlarged the size of reference population, which is the most important factor affecting prediction reliability (Goddard and Hayes, 2009). However, we observed that GEBV were more inflated when both genotyped cows and genotyped bulls were used as reference population. Inflation may be caused by preferential treatment of cows included in the reference population (Wiggans et al., 2011; Kuhn et al., 1994). The results from our study were consistent with the results reported by Wiggans et al. (2011); in their study, the regression coefficients of DRP on GEBV of protein decreased from 0.86 to 0.83 when the cows were added into the reference population. On the other hand, bias of predicted genetic trend was reduced when genotyped cows were included into the reference population. The reason could be that the cows which were bull dams and sibs of test bulls may account for the contribution of the test bulls' dam to the bulls.

Previous studies reported that including a polygenic effect in a SNP-BLUP model or Bayesian model led to less inflation of GEBV (Solberg et al., 2009; Liu et al., 2011; Su et al., 2014). The regression coefficient was improved from 0.69 to 0.72 in the current study when the polygenic effect was included in the model (GBLUPW_{Bull}). However, the bias of predicted genetic trend was not reduced compared with the model without polygenic effect.

Figure 1. The deregressed proofs (DRP) and genomic EBV (GEBV) trends of protein in different scenarios. $GBLUP_{Bull}$ = genomic BLUP (GBLUP) model with bulls as reference population; $GBLUPW_{Bull}$ = same as $GBLUP_{Bull}$ but with a genomic relationship matrix $G\omega = 0.8G + 0.2A$, where G is the genomic relationship matrix and A is the pedigree relationship matrix; $GBLUP_{Cow}$ = GBLUP model with both genotyped bulls and cows as reference population; $GBLUP_{Year}$ = year of birth effects were included in the model as genetic trend; $GBLUP_{Dam\_mgs}$ = GEBV from GBLUP model using bull reference data were adjusted for the difference between dam EBV and maternal grandsire EBV; SS_P = the single-step approach using phenotypes of all cows and genotypes of genotyped bulls; SS_{PG} = the single-step approach using phenotypes of all cows and genotypes of genotyped bulls and cows. Color version available online.

The single-step model used cows' deregressed EBV as a response variable rather than the raw phenotypic data. However, the effect of genomic preselection is minor in the current Jersey data and the deregressed cow EBV should not be biased at all. Therefore, the results could be considered single-step prediction using raw data. However, if genomic preselection is used in breeding schemes, the EBV estimated using pedigree will be biased. In this case, it is better to use raw data as a response variable in single-step approach. The single-step prediction, which used all the females' DRP and pedigree as well as genotypes from genotyped bulls and cows, increased the reliability and reduced inflation of GEBV and bias of predicted genetic trend. These results were consistent with previous reports (Vitezica

et al., 2011; Koivula et al., 2012; Su et al., 2012b). As DRP of nongenotyped animals also contributes to the prediction through a combined matrix, the prediction reliability was improved. Moreover, single-step models could reduce bias of predicted genetic trend by including all the records to trace selection (Vitezica et al., 2011). Similar to a GBLUP model including genotyped cows in the reference data, the regression coefficient decreased when the cow genotypes were included into the single-step approach. The selection index blending (VanRaden et al., 2009; Su et al., 2012b) with the same information used in the single-step approach without cow genotype data was compared with single-step approach in our study (data not shown). The prediction reliability was 0.32, which was higher than reliability

Figure 2. Boxplots for difference between genomic EBV (GEBV) and deregressed proofs (DRP; i.e., GEBV-DRP). The box shows the first to third quartile of GEBV-DRP. The horizontal bar inside the box shows the median value. The upper and lower bars show the upper and lower values within 1.5 times the interquartile range (IQR) from the upper and lower quartile. The IQR is the interquartile range that is defined as the distance between the 1st and 3rd quartile. Less than 3 IQR from either end of the box are labeled as outliers (o). $GBLUP_{Bull}$ = genomic BLUP (GBLUP) model with bulls as reference populations; $GBLUPW_{Bull}$ = same as $GBLUP_{Bull}$ but with a genomic relationship matrix $G\omega = 0.8G + 0.2A$, where G is the genomic relationship matrix and A is the pedigree relationship matrix; $GBLUP_{Cow}$ = GBLUP model with both genotyped bulls and cows as reference population; $GBLUP_{Year}$ = year effects were included in the model to account for part of genetic trend; $GBLUP_{Dam\_mgs}$ = GEBV from GBLUP model using bull reference data were adjusted for the difference between dam EBV and maternal grandsire EBV; SS_P = the single-step approach using phenotypes of all cows and genotypes of genotyped bulls; SS_{PG} = the single-step approach using phenotypes of all cows and genotypes of genotyped bulls and cows.

of GEBV directly from $GBLUP_{Bull}$ but lower than scenario SS_P even though the information used in these 2 methods was the same. The bias of predicted genetic trend was corrected for individuals born in 2005 and 2006, but not for the individuals born after 2006 when the blending index was used. Genomic relationship matrix was modified with pedigree relationship matrix in single-step approach. As the pedigree relationship has influence on the regression coefficient, to be consistent, the scenario $GBLUPW_{Bull}$ was investigated. The results from $GBLUPW_{Bull}$ showed GBLUP model with 20% of the pedigree relationship matrix did not increase

the prediction reliability and reduce bias of predicted genetic trend. However, the regression coefficients were improved by the weighted G matrix (from 0.69 of $GBLUP_{Bull}$ to 0.72 of $GBLUPW_{Bull}$) and by the single-step approach (from 0.72 of $GBLUPW_{Bull}$ to 0.78 of SS_P). These results suggest that using a single-step method is an effective approach to increase the prediction reliability and reduce the bias of predicted genetic trend.

Including the year of birth effect reduced the bias of predicted genetic trend and improved the regression coefficients. The reason could be that the year effect partly accounted for the trend of selection among the

dams. The GEBV together with the year effect captured the genetic progress across years, which led to a robust estimation of genetic trend (Ducrocq, 2010).

The mean of GEBV adjusted for the difference between dam EBV and MGS EBV were much closer to the mean of DRP in the test population compared with the GEBV without adjustment. The reliability was improved greatly, which may have been caused by a possible autocorrelation between dam EBV and the progeny DRP. However, the regression coefficients deviated more from unity, which may have been caused by the preferential treatment of selected cows. Bias of prediction trend was corrected in a form of large inflation of the GEBV. Therefore, it is not a good approach to correct for bias of predicted genetic trend.

The results from the current study indicate that the regression coefficient, which has mainly been used in previous studies (Verbyla et al., 2009; Su et al., 2012a), should not be the only criterion to measure the unbiasedness of predictions. The regression coefficient is not always consistent with the bias of predicted genetic trends. As the prediction trend is important when the individuals across generations are compared, it should also be included in the evaluation criteria. The year mean of DRP could be expressed as the year mean of GEBV times the regression coefficients plus the intercept. Therefore, the bias of predicted genetic trend could be predicted using the regression coefficient and intercept. Therefore the intercept together with the regression coefficients should be given attention in genomic prediction.

CONCLUSIONS

The main reason for the bias of predicted genetic trend could be that the reference animals did not have all the information required to trace selection, especially the information of dams. Consequently, methods using more information related to selection can reduce the bias. The most efficient way is to implement a single-step approach for genomic prediction, as the single-step approach increased the prediction reliability, improved the regression coefficients, and led to an unbiased prediction trend. As bias of predicted genetic trends can be measured by the intercept and regression coefficient of observations on GEBV, both intercept and regression coefficients should be taken into consideration in validation of genomic predictions.

ACKNOWLEDGMENTS

This work was performed within the project “Genomics in herds,” funded by VikingGenetics (Randers,

Denmark) and Nordic Cattle Genetic Evaluation (Aarhus, Denmark).

REFERENCES

- Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* 93:743–752. <http://dx.doi.org/10.3168/jds.2009-2730>.
- Browning, B. L., and S. R. Browning. 2009. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am. J. Hum. Genet.* 84:210–223. <http://dx.doi.org/10.1016/j.ajhg.2009.01.005>.
- Calus, M. P. L., Y. de Haas, and R. F. Veerkamp. 2013. Combining cow and bull reference populations to increase accuracy of genomic prediction and genome-wide association studies. *J. Dairy Sci.* 96:6703–6715. <http://dx.doi.org/10.3168/jds.2012-6013>.
- Christensen, O. F., and M. S. Lund. 2010. Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol.* 42:2. <http://dx.doi.org/10.1186/1297-9686-42-2>.
- Christensen, O. F., P. Madsen, B. Nielsen, T. Ostersen, and G. Su. 2012. Single-step methods for genomic evaluation in pigs. *Animal* 6:1565–1571. <http://dx.doi.org/http://dx.doi.org/10.1017/S1751731112000742>.
- Ducrocq, V. 2010. Sustainable dairy cattle breeding: Illusion or reality. In *Proc. 9th World Congr. Genet. Appl. Livest. Prod.*
- Gao, H., O. F. Christensen, P. Madsen, U. S. Nielsen, Y. Zhang, M. S. Lund, and G. Su. 2012. Comparison on genomic predictions using three GBLUP methods and two single-step blending methods in the Nordic Holstein population. *Genet. Sel. Evol.* 44:8. <http://dx.doi.org/10.1186/1297-9686-44-8>.
- Goddard, M. E., and B. J. Hayes. 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat. Rev. Genet.* 10:381–391. <http://dx.doi.org/10.1038/nrg2575>.
- Hayes, B. J., P. J. Bowman, A. J. Chamberlain, and M. E. Goddard. 2009a. Invited review: Genomic selection in dairy cattle: progress and challenges. *J. Dairy Sci.* 92:433–443. <http://dx.doi.org/10.3168/jds.2008-1646>.
- Hayes, B. J., P. M. Visscher, and M. E. Goddard. 2009b. Increased accuracy of artificial selection by using the realized relationship matrix. *Genet. Res. (Camb.)* 91:47–60. <http://dx.doi.org/10.1017/S0016672308009981>.
- Koivula, M., I. Strandén, J. Pösö, G. P. Aamand, and E. A. Mäntysaari. 2012. Single step genomic evaluations for the Nordic Red Dairy cattle test day data. *Interbull Bull.* 46:115–120.
- Kuhn, M. T., P. J. Boettcher, and E. Freeman. 1994. Potential biases in predicted transmitting abilities of females from preferential treatment. *J. Dairy Sci.* 77:2428–2437. [http://dx.doi.org/10.3168/jds.S0022-0302\(94\)77185-X](http://dx.doi.org/10.3168/jds.S0022-0302(94)77185-X).
- Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. *J. Dairy Sci.* 92:4656–4663. <http://dx.doi.org/10.3168/jds.2009-2061>.
- Lidauer, M., and I. Strandén. 1999. Fast and flexible program for genetic evaluation in dairy cattle. *Interbull Bull.* 20:19–24.
- Liu, Z., F. R. Seefried, F. Reinhardt, S. Rensing, G. Thaller, and R. Reents. 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. <http://dx.doi.org/10.1186/1297-9686-43-19>.
- Lund, M. S., A. P. W. De Roos, A. G. De Vries, T. Druet, V. Ducrocq, S. Fritz, F. Guillaume, B. Guldbrandtsen, Z. Liu, R. Reents, C. Schrooten, F. Seefried, and G. Su. 2011. A common reference population from four European Holstein populations increases reliability of genomic predictions. *Genet. Sel. Evol.* 43:43. <http://dx.doi.org/10.1186/1297-9686-43-43>.
- Madsen, P., G. Su, R. Labouriau, and O. F. Christensen. 2010. DMU—A package for analyzing multivariate mixed models. Page 732 in

- Proc. 9th World Congr. Genet. Appl. Livest. Prod., Leipzig, Germany. Gesellschaft für Tierzuchtwissenschaft e.V., Bonn, Germany.
- Misztal, I., A. Legarra, and I. Aguilar. 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *J. Dairy Sci.* 92:4648–4655. <http://dx.doi.org/10.3168/jds.2009-2064>.
- Solberg, T. R., A. K. Sonesson, J. A. Woolliams, J. Odegard, and T. H. E. Meuwissen. 2009. Persistence of accuracy of genome-wide breeding values over generations when including a polygenic effect. *Genet. Sel. Evol.* 41:53. <http://dx.doi.org/10.1186/1297-9686-41-53>.
- Strandén, I., and E. A. Mäntysaari. 2010. A recipe for multiple trait deregression. *Interbull Bull.* 42:21–24.
- Su, G., R. F. Brøndum, P. Ma, B. Guldbrandtsen, G. P. Aamand, and M. S. Lund. 2012a. Comparison of genomic predictions using medium-density (~54,000) and high-density (~777,000) single nucleotide polymorphism marker panels in Nordic Holstein and Red Dairy Cattle populations. *J. Dairy Sci.* 95:4657–4665. <http://dx.doi.org/10.3168/jds.2012-5379>.
- Su, G., O. F. Christensen, L. Janss, and M. S. Lund. 2014. Comparison of genomic predictions using genomic relationship matrices built with different weighting factors to account for locus-specific variances. *J. Dairy Sci.* 97:6547–6559. <http://dx.doi.org/10.3168/jds.2014-8210>.
- Su, G., P. Madsen, U. S. Nielsen, E. A. Mäntysaari, G. P. Aamand, O. F. Christensen, and M. S. Lund. 2012b. Genomic prediction for Nordic Red Cattle using one-step and selection index blending. *J. Dairy Sci.* 95:909–917. <http://dx.doi.org/10.3168/jds.2011-4804>.
- Thomasen, J. R., B. Guldbrandtsen, G. Su, R. F. Brøndum, and M. S. Lund. 2012. Reliabilities of genomic estimated breeding values in Danish Jersey. *Animal* 6:789–796. <http://dx.doi.org/10.1017/S1751731111002035>.
- VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91:4414–4423. <http://dx.doi.org/http://dx.doi.org/10.3168/jds.2007-0980>.
- VanRaden, P. M., C. P. Van Tassell, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and F. S. Schenkel. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci.* 92:16–24. <http://dx.doi.org/10.3168/jds.2008-1514>.
- Verbyla, K. L., B. J. Hayes, P. J. Bowman, and M. E. Goddard. 2009. Accuracy of genomic selection using stochastic search variable selection in Australian Holstein Friesian dairy cattle. *Genet. Res. (Camb.)* 91:307–311. <http://dx.doi.org/10.1017/S0016672309990243>.
- Vitezica, Z. G., I. Aguilar, and I. Misztal, and A. Legarra. 2011. Bias in genomic predictions for populations under selection. *Genet. Res. (Camb.)* 93:357–366. <http://dx.doi.org/10.1017/S001667231100022X>.
- Wiggans, G. R., T. A. Cooper, P. M. VanRaden, and J. B. Cole. 2011. Technical note: Adjustment of traditional cow evaluations to improve accuracy of genomic predictions. *J. Dairy Sci.* 94:6188–6193. <http://dx.doi.org/10.3168/jds.2011-4481>.