J. Dairy Sci. 101:11143–11149
https://doi.org/10.3168/jds.2018-15090

Short communication: Single-step genomic evaluation of milk production traits using multiple-trait random regression model in Chinese Holsteins

H. Kang,*† C. Ning,† L. Zhou,† S. Zhang,† Q. Yan,‡ and J.-F. Liu†[1]
*School of Life Science and Engineering, Foshan University, Foshan 528231, Guangdong, China
†National Engineering Laboratory for Animal Breeding; Key Laboratory of Animal Genetics, Breeding and Reproduction, Ministry of Agriculture; College of Animal Science and Technology, China Agricultural University, Beijing 100193, China
‡Dairy Association of China, Beijing 100193, China

ABSTRACT

The objectives of this study were to evaluate the prediction performance of the single-step genomic BLUP method using a multi-trait random regression model in genomic evaluation for milk production traits of Chinese Holsteins, and investigate how parameters w, τ, and ω used in the construction of the combined relationship matrix (**H**) affected prediction accuracy and bias. A total of 2.8 million test-day records from 0.2 million cows were available for milk, protein, and fat yields. Pedigree information included 0.3 million animals and 7,577 of them were genotyped with medium-density single nucleotide polymorphism marker panels. Genotypes were imputed into Geneseek Genomic Profiler HD (GeneSeek, Lincoln, NE) including 77K markers. A reduced data set for evaluating models was extracted from the full data set by removing test-day records from the last 4 yr. Bull and cow validation populations were constructed for each trait. We evaluated the prediction performance of the multiple-trait multiple-lactation random regression single-step genomic BLUP (RR-ssGBLUP) models with different values of parameters w, τ, and ω in the **H** matrix, taking consideration of inbreeding. We compared RR-ssGBLUP with the multiple-trait multiple-lactation random regression model based on pedigree and genomic BLUP. De-regressed proofs for 305-d milk, protein, and fat yields averaged over 3 lactations, which were calculated from the full data set, were used for posteriori validations. The results showed that RR-ssGBLUP was feasible for implementation in breeding practice, and its prediction performance was superior to the other 2 methods in the comparison, including prediction accuracy and unbiasedness. For bulls, RR-ssGBLUP models with $w_{0.05}\tau_{2.0}\omega_{1.0}$, $w_{0.05}\tau_{2.5}\omega_{1.0}$, and $w_{0.1}\tau_{1.6}\omega_{0.4}$ achieved the best performance for milk, protein, and fat yields, respectively. For cows, the RR-ssGBLUP with $w_{0.2}\tau_{1.6}\omega_{0.4}$ performed the best for all 3 traits. The **H** matrix constructed with larger τ and smaller ω gave better convergence in solving mixed model equations. Among different RR-ssGBLUP models, the differences in validation accuracy were small. However, the regression coefficient indicating prediction bias varied substantially. The increase of w and τ, and decrease of ω, led to an increase in the regression coefficient. The results demonstrated RR-ssGBLUP is a good alternative to the multi-step approach, but the optimal choice of parameters should be found via preliminary validation study to achieve the best performance.

Key words: genomic evaluation, single-step GBLUP, random regression model, Chinese Holstein

Short Communication

National genomic selection of Chinese Holsteins has been conducted since 2012 using a multi-step procedure, which includes genetic evaluation based on pedigree, genomic evaluation based on genotypic information, and a combination of the results from these 2 steps. Recently, a single-step genomic BLUP (**ssGBLUP**) has been widely used in genomic evaluation. The ssGBLUP is a unified approach to calculate genomic enhanced breeding value (**GEBV**; Legarra et al., 2009; Aguilar et al., 2010; Christensen and Lund, 2010). The main advantages of ssGBLUP over multi-step methods are practical simplicity and resistance to preselection bias (Vitezica et al., 2011; Legarra et al., 2014).

To make predictions utilizing phenotypic, genomic, and pedigree data simultaneously, the single-step method replaces the pedigree-based numerator relationship matrix **A** in mixed model equations (**MME**) with matrix **H**, which integrates matrix **A** and genomic relationship matrix **G**. The construction of the **H** matrix involves several parameters and some of them have been demonstrated to have effects on prediction performance (Vitezica et al., 2011; Christensen et al.,

Received May 21, 2018.
Accepted August 14, 2018.
[1] Corresponding author: liujf@cau.edu.cn

2012). In practical applications, we should determine and apply the optimal values of parameters.

In Chinese Holsteins, the genetic evaluation of milk production traits based on pedigree employs a multiple-trait multiple-lactation random regression test-day model. The benefit of the random regression test-day model is that the partition of variation in phenotype is assumed to be time-dependent, and this consequently increases the accuracy of genetic evaluation (Schaeffer et al., 2000). Integrating the random regression test-day model into ssGBLUP has been demonstrated to be a feasible approach to achieve more accurate and less biased predictions (Koivula et al., 2015; Kang et al., 2017).

The objectives of this study were to evaluate the prediction performance of the multiple-trait multiple-lactation random regression ssGBLUP (**RR-ssGBLUP**) method in the Chinese Holstein population, compared with the random regression test-day model based on pedigree and GBLUP (VanRaden, 2008), and investigate how parameters involved in the construction of the **H** matrix affected prediction accuracy and bias of the RR-ssGBLUP method.

The data analyzed were from those used in the official Chinese Holstein milk production evaluations. The phenotypic data included test-day records for milk, protein, and fat yields from the first 3 lactations. In this study, test-day data included 0.2 million cows with a total of 2.8 million observations recorded from 1995 to 2017, and 0.3 million animals born from 1990 to 2014 were involved in the pedigree. Test-day records were preadjusted for heterogeneous herd-test date-parity variances on a trait by trait basis according to the procedure proposed by Schaeffer et al. (2000). A total of 7,577 animals were genotyped with a medium-density SNP marker panel [i.e., Illumina BovineSNP50 v1 BeadChip (54K SNP), Illumina, San Diego, CA; Illumina BovineSNP50 v2 BeadChip (55K SNP); or Geneseek Genomic Profiler HD (77 K SNP), GeneSeek, Lincoln, NE]. The genotypes were imputed to Geneseek Genomic Profiler HD using FImpute v2.2 (Sargolzaei et al., 2014). The SNP were filtered with the criteria that the minor allele frequency should be larger than 0.01, and the *P*-value for the Hardy-Weinberg equilibrium should be larger than 1×10^{-5}. After quality control, 72,507 autosomal SNP were retained for the following analyses.

A reduced data set was extracted from the full data set to be a reference data set for evaluating models, by removing the last 4 yr of observations. The reduced data set included 0.1 million cows with 1.7 million records. Validation bulls were selected according to the criteria that each bull should have more than 20 daughters with observations in the full data set, and not have daughters in the reduced data set. For validation cows, their reliabilities of breeding values estimated from the full data set should be larger than 0.4, and they should not have test-day records in the reduced data set. For the 3 traits, the bull validation population consisted of 79 genotyped bulls. The cow validation population consisted of 1,073, 1,030, and 430 genotyped cows for milk, protein, and fat yields, respectively.

The reduced data set was analyzed using the multiple-trait multiple-lactation RR-ssGBLUP model as follows:

$$\mathbf{y} = \mathbf{X}_1\mathbf{b}_1 + \mathbf{X}_2\mathbf{b}_2 + \mathbf{Z}\mathbf{a} + \mathbf{W}\mathbf{p} + \mathbf{e},$$

where $\mathbf{y}$ is the vector of test-day observations, and performances of a trait in different parities are considered as different traits; $\mathbf{b}_1$ is the vector of fixed effects of the herd-test date; $\mathbf{b}_2$ is the vector of fixed regression coefficients on Legendre polynomials nested within calving age-calving season effect; $\mathbf{a}$ and $\mathbf{p}$ are the vectors of random regression coefficients on Legendre polynomials nested within the additive genetic effect and permanent environmental effect, respectively; $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{Z}$, and $\mathbf{W}$ are incidence matrices for $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{a}$, and $\mathbf{p}$, respectively; and $\mathbf{e}$ is the vector of the residuals. Legendre polynomials were calculated based on DIM. We determined the orders of Legendre polynomials for different effects according to Akaike information criterion (Akaike, 1974) and Bayesian information criterion (Schwarz, 1978). As a result, Legendre polynomials of order 4 were fitted to the calving age-calving season effect and additive genetic effect. Legendre polynomials of order 3 were fitted to the permanent environmental effect.

It was assumed that

$$\text{var}\begin{bmatrix}\mathbf{a}\\ \mathbf{p}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{H}\otimes\mathbf{C} & 0 & 0\\ 0 & \mathbf{I}\otimes\mathbf{P} & 0\\ 0 & 0 & \mathbf{R}\end{bmatrix},$$

where $\mathbf{I}$ is an identity matrix; $\otimes$ is the Kronecker product; $\mathbf{C}$ and $\mathbf{P}$ are the variance-covariance matrices for additive genetic and permanent environmental effects, respectively; and $\mathbf{R}$ is the residual variance-covariance matrix. The residual covariance matrix depends on DIM (d: 5–45, 46–155, 156–205, and 206–305 d) and lactation number (n: 1, 2, 3). This leads to 12 residual covariance matrices with the following form:

$$\mathbf{R}_{dn} = \begin{bmatrix}r_{11} & r_{12} & r_{13}\\ r_{12} & r_{22} & r_{23}\\ r_{13} & r_{23} & r_{33}\end{bmatrix},$$

where $r_{hh'}$ is the covariance between traits h and h'.

According to Aguilar et al. (2010), Christensen and Lund (2010), and Tsuruta et al. (2011), the inverse of the **H** matrix is constructed as follows:

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \tau\left[(1-w)\mathbf{G}^{*} + w\mathbf{A}_{22}\right]^{-1} - \omega\mathbf{A}_{22}^{-1} \end{bmatrix},$$

where **A** is the numerator relationship matrix based on pedigree, subscript 2 indicates the genotyped individuals, inbreeding was considered in the constructions of **A** and $\mathbf{A}_{22}$ matrices, and

$$\mathbf{G}^{*} = \mathbf{G}\beta + \alpha,$$

where **G** is the genomic relationship matrix constructed by the first method of VanRaden (2008), and α and β are used to adjust the **G** matrix to avoid the potential incompatibility in scale between the genomic and numerator relationship matrices (Christensen et al., 2012). We used different combinations of parameters, (i.e., $w_{0.05}\tau_{1.0}\omega_{1.0}$, $w_{0.05}\tau_{1.6}\omega_{1.0}$, $w_{0.05}\tau_{2.0}\omega_{1.0}$, $w_{0.05}\tau_{2.5}\omega_{1.0}$, $w_{0.1}\tau_{1.0}\omega_{1.0}$, $w_{0.1}\tau_{1.6}\omega_{0.4}$, $w_{0.1}\tau_{1.6}\omega_{0.5}$, $w_{0.1}\tau_{1.6}\omega_{1.0}$, $w_{0.1}\tau_{2.0}\omega_{1.0}$, $w_{0.1}\tau_{2.5}\omega_{1.0}$, $w_{0.2}\tau_{1.0}\omega_{0.5}$, $w_{0.2}\tau_{1.0}\omega_{0.6}$, $w_{0.2}\tau_{1.0}\omega_{0.7}$, $w_{0.2}\tau_{1.0}\omega_{0.8}$, $w_{0.2}\tau_{1.0}\omega_{0.9}$, $w_{0.2}\tau_{1.0}\omega_{1.0}$, $w_{0.2}\tau_{1.6}\omega_{0.4}$, $w_{0.2}\tau_{1.6}\omega_{0.5}$, $w_{0.2}\tau_{1.6}\omega_{0.8}$, $w_{0.2}\tau_{1.6}\omega_{1.0}$, and $w_{0.25}\tau_{1.6}\omega_{1.0}$) to investigate the parameters' effects on prediction performance of RR-ssGBLUP.

To exploit the potential advantages of RR-ssGBLUP over its conventional counterparts, we additionally performed genetic prediction using a random regression model based on pedigree (**PA**) and subsequent genomic prediction using the method of GBLUP. The model of PA was the same as that of RR-ssGBLUP, except that the **H** matrix in the single-step method was replaced by the **A** matrix in the PA model. Variance components in RR-ssGBLUP and PA were all estimated based on the model of PA using Gibbs sampling implemented in BLUPF90 (Misztal et al., 2002). As Gibbs sampling is time consuming and the execution time increases with the size of data set, we used a small sample of the whole data set to estimate variance components, including approximately 0.3 million test-day records of 20,053 individuals and 44,580 individuals with pedigrees. The sample size was similar to those in other studies (Strabel et al., 2005; Muir et al., 2007). The MME for RR-ssGBLUP and PA with the estimated variance components were solved using a preconditioned conjugate gradient by iteration on data implemented in PIBLUP (Kang et al., 2018).

The GBLUP model (VanRaden, 2008) used was a 3-trait model as follows:

$$\mathbf{y} = \mathbf{u} + \mathbf{Zg} + \mathbf{e},$$

where **y** is the vector of de-regressed proofs (**DRP**) for milk, protein, and fat yields, which are derived from EBV for 305-d yields averaged over the first 3 lactations (Schaeffer et al., 2000); **u** is the vector of overall means for the 3 traits; **g** is the vector of additive genomic effects with a distribution of $N(\mathbf{0}, \mathbf{G} \otimes \mathbf{G}_0)$; **Z** is the corresponding incidence matrix; and **e** is the vector of random residuals with distribution of $N(\mathbf{0}, \mathbf{D} \otimes \mathbf{R})$. **G** is the aforementioned genomic relationship matrix with 0.02 added to its diagonal elements to avoid singularity problems; $\mathbf{G}_0$ is the additive genomic variance-covariance matrix; **R** is the residual variance-covariance matrix; and **D** is a diagonal matrix (see details below). All of the elements in $\mathbf{G}_0$ and **R** are not zero.

The reliabilities of averaged EBV were approximated following the procedure proposed by Jamrozik et al. (2000). The DRP derivation and corresponding reliability were calculated according to Garrick et al. (2009). Diagonal elements in the **D** matrix in the GBLUP model are $d_{ij} = 1/s_{ij}$, where the weighting factor is $d_{ii} = 1/s_i$, with $s_i = r^2_{DRP}/\left(1 - r^2_{DRP}\right)$, being the reliability of DRP for trait j of individual i. We employed the DMU software package (Madsen et al., 2014) to estimate the variance-covariance components and solve MME for the GBLUP model.

For evaluation of different models, DRP r^2_{DRP} and their corresponding reliabilities for individuals in the validation population were estimated from the full data set using the PA model. In model evaluation, EBV and GEBV were predicted based on the reduced data set. Prediction accuracy was evaluated as the Pearson correlation between GEBV (or EBV) and $\left(DRP_{full}\right)$ of individuals in the validation population, corrected with the average of DRP reliability DRP_{full} that is, $\left(R^2_{DRP}\right)$, In addition, the regression coefficient of $r = \dfrac{r_{GEBV,\, DRP_{full}}}{\sqrt{R^2_{DRP}}}$. on GEBV (b) was calculated to assess prediction bias of the model.

In our study, the estimates of parameters α and β were 0.011 and 0.989 (i.e., closer to 1 and 0). This means that the **G** and $\mathbf{A}_{22}$ matrices were relatively compatible and the **G** matrix was little adjusted. Our results showed that the computing time per round of iteration was of the same length for different RR-ssGBLUP models. This is because only small differences existed in MME due to different values of parameters used in **H** matrix construction. We have taken consideration of inbreeding in the construction of **A** and $\mathbf{A}_{22}$ matrix in all RR-ssGBLUP models. We further tested the rate of convergence of RR-ssGBLUP models excluding inbreeding, using DRP_{full} $w_{0.1}\tau_{1.6}\omega_{1.0}$, $w_{0.1}\tau_{2.0}\omega_{1.0}$,

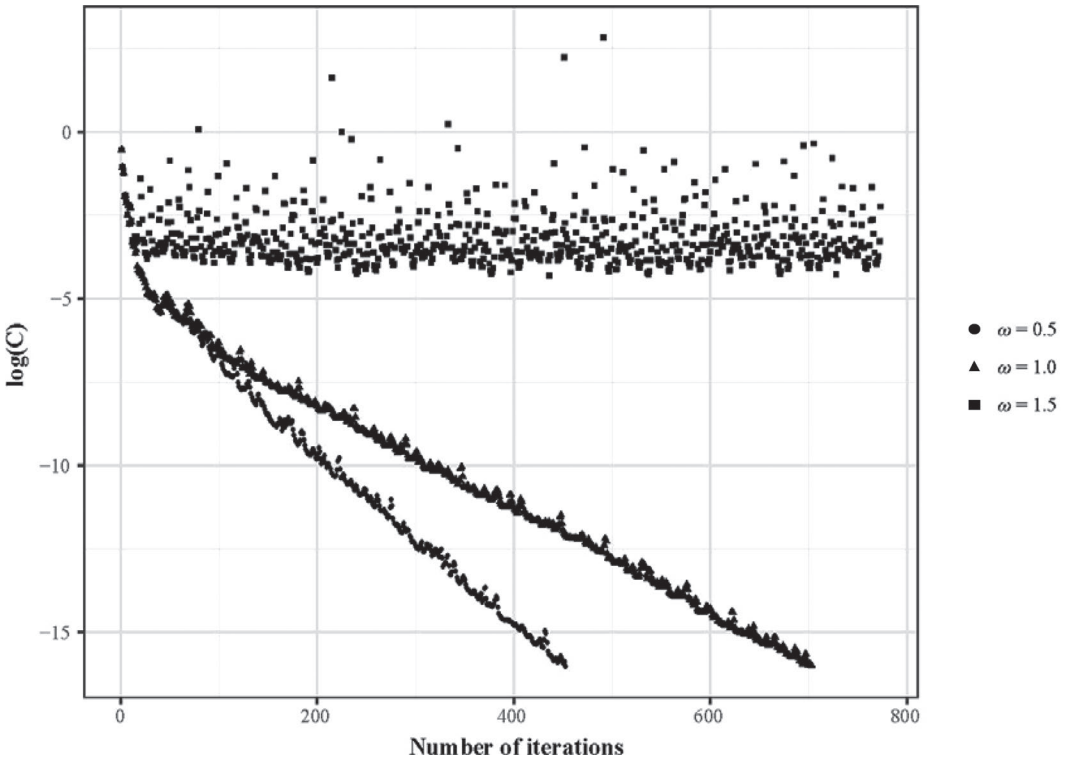

Figure 1. Convergence rates of the multiple-trait multiple-lactation random regression single-step genomic BLUP method with $w = 0.1$, $\tau = 1.6$, and different values of ω (0.5, 1.0, and 1.5) in analyses of the reduced data set. w = proportion of polygenic variance; τ = weight for $\mathbf{G}^{-1}$; ω = weight for $\mathbf{A}_{22}^{-1}$. The log(C) denotes the logarithm to base 10 of the convergence criterion.

$w_{0.2}\tau_{1.6}\omega_{1.0}$, $w_{0.2}\tau_{1.6}\omega_{0.8}$, and $w_{0.1}\tau_{1.6}\omega_{0.5}$, Results showed that considering inbreeding reduced the number of PCG iterations by 440, 81, and 192 for RR-ssGBLUP with $w_{0.2}\tau_{1.6}\omega_{0.4}$. $w_{0.1}\tau_{1.6}\omega_{1.0}$, and $w_{0.1}\tau_{2.0}\omega_{1.0}$, Similar results for ssGBLUP with $\omega = 1.0$ were also observed in other studies (Matilainen et al., 2016; Masuda et al., 2018). But for RR-ssGBLUP with $\omega < 1.0$, that is, $w_{0.2}\tau_{1.6}\omega_{1.0}$. $w_{0.2}\tau_{1.6}\omega_{0.8}$, and $w_{0.1}\tau_{1.6}\omega_{0.5}$, considering inbreeding had little effect on convergence, which only increased or decreased several rounds of iterations. Therefore, including inbreeding can only improve the convergence of RR-ssGBLUP models with $\omega = 1.0$. Moreover, for RR-ssGBLUP models, parameters τ and ω in $\mathbf{H}^{-1}$ had effects on the rate of convergence, no matter whether inbreeding was considered or not. Large values of τ can accelerate the convergence. For example, for RR-ssGBLUP including inbreeding and with $w = 0.1$ and $\omega = 1.0$, the numbers of iterations were 703, 653, and 617 for values of τ of 1.6, 2.0, and 2.5, respectively. Compared with τ, the parameter ω had a larger effect on convergence. Figure 1 shows the rate of convergence of RR-ssGBLUP including inbreeding and with $w = 0.1$, $\tau = 1.6$, and different values of ω. Smaller ω tended to give better convergence, whereas larger ω ($\omega = 1.5$) led to divergence. This is in agreement with the findings of Koivula et al. (2015). This can be explained because the convergence rate is influenced by the property of the $\mathbf{H}$ matrix when MME is solved by iterative methods. Larger values of ω causes the $\mathbf{H}$ matrix to be less positive semi-definite, thus leading to slower convergence or even divergence. This negative effect can be compensated by larger τ to make the $\mathbf{H}$ matrix more positive semi-definite (Martini et al., 2018).

Tables 1 and 2 present prediction performance of different methods for candidate bulls and cows, respectively. The tables show validation accuracies (r) and regression coefficients (b). In general, genomic selection methods had superior prediction performance in comparison with the PA method, and RR-ssGBLUP performed the best among all the models evaluated.

The RR-ssGBLUP models generally achieved the same or slightly higher accuracy compared with the GBLUP method. Our results indicated that the differences in validation accuracy among the RR-ssGBLUP models with different values of parameters were small, which is consistent with previous studies (Koivula et al., 2015). For bulls, validation accuracies of RR-ssGBLUP models with different values of parameters varied between 0.42 and 0.47 for milk yield, between

0.36 and 0.43 for protein yield, and between 0.20 and 0.24 for fat yield. The RR-ssGBLUP model with the parameters providing the best performance improved the accuracy by 0.31 for milk, 0.29 for protein, and 0.15 for fat, compared with the PA method. For cows, validation accuracies of the RR-ssGBLUP models with different **H** matrices varied between 0.34 and 0.35 for milk yield, between 0.28 and 0.30 for protein yield, and between 0.21 and 0.27 for fat yield. The RR-ssGBLUP model with the best choices of parameters improved the accuracy by 0.10 for milk yield and 0.05 for protein yield.

The regression coefficient (b) of DRP_{full} on GEBV (or EBV) reflects the unbiasedness of prediction. Optimal prediction for candidate individuals should have a regression coefficient of 1. In most cases, EBV or GEBV were inflated (b <1), especially for the PA and GBLUP models. The deflation of EBV or GEBV (b >1) were observed mainly in genomic prediction of milk yield for bulls, and some in genomic prediction of protein yield for bulls. Similar to validation accuracy, GEBV by GBLUP and RR-ssGBLUP methods were generally less biased than the PA method. When GEBV obtained using the GBLUP method were inflated or deflated, RR-ssGBLUP with a specific combination of parameters could obtain more unbiased predictions. For both bulls and cows, increase of w and τ, and decrease of ω led to an increase of the regression coefficient. Moreover, our results indicated that changes in w and τ had smaller effects on regression coefficients compared with that of ω. The effects of parameters τ and ω on regression coefficients can be explained because both an increase in τ and decrease in ω reduced the variance of the predicted genetic value, thus resulting in larger regression coefficients (Martini et al., 2018).

Considering both prediction accuracy and bias, the best method for prediction was the RR-ssGBLUP method. However, the best choice of parameters in the construction of the **H** matrix depended on traits and candidate populations. For bulls, the optimal parame-

Table 1. Bull validation results from different methods, showing validation accuracies (r) and regression coefficients (b) from the parent average (PA), genomic BLUP (GBLUP), and random regression single-step GBLUP (RR-ssGBLUP) methods with different combinations of parameters

	Milk yield		Protein yield		Fat yield	
Method	r	b	r	b	r	b
PA	0.16	0.47	0.14	0.38	0.09	0.27
GBLUP	0.45	1.11	0.40	0.97	0.19	0.52
RR-ssGBLUP[1]						
$w_{0.2}\tau_{1.6}\omega_{0.4}$.	0.44	0.77	0.37	0.68	0.20	0.38
$w_{0.05}\tau_{1.0}\omega_{1.0}$	0.46	0.94	0.40	0.86	0.22	0.51
$w_{0.05}\tau_{1.6}\omega_{1.0}$	0.47	1.03	0.42	0.95	0.23	0.58
$w_{0.05}\tau_{2.0}\omega_{1.0}$	0.47	1.12	0.43	1.04	0.24	0.65
$w_{0.05}\tau_{2.5}\omega_{1.0}$	0.44	0.80	0.37	0.70	0.20	0.39
$w_{0.1}\tau_{1.0}\omega_{1.0}$	0.44	1.56	0.38	1.39	0.23	0.99
$w_{0.1}\tau_{1.6}\omega_{0.4}$	0.45	1.50	0.39	1.35	0.23	0.94
$w_{0.1}\tau_{1.6}\omega_{0.5}$	0.46	0.97	0.40	0.89	0.22	0.52
$w_{0.1}\tau_{1.6}\omega_{1.0}$	0.47	1.06	0.42	0.98	0.23	0.59
$w_{0.1}\tau_{2.0}\omega_{1.0}$	0.47	1.15	0.43	1.07	0.24	0.67
$w_{0.1}\tau_{2.5}\omega_{1.0}$	0.42	1.37	0.36	1.20	0.21	0.82
$w_{0.2}\tau_{1.0}\omega_{0.5}$	0.43	1.31	0.36	1.15	0.21	0.77
$w_{0.2}\tau_{1.0}\omega_{0.6}$	0.44	1.23	0.37	1.09	0.21	0.71
$w_{0.2}\tau_{1.0}\omega_{0.7}$	0.44	1.12	0.37	1.00	0.21	0.63
$w_{0.2}\tau_{1.0}\omega_{0.8}$	0.45	0.99	0.37	0.88	0.21	0.53
$w_{0.2}\tau_{1.0}\omega_{0.9}$	0.44	0.84	0.37	0.74	0.20	0.42
$w_{0.2}\tau_{1.0}\omega_{1.0}$	0.44	1.58	0.38	1.40	0.22	0.99
$w_{0.2}\tau_{1.6}\omega_{0.4}$	0.45	1.53	0.39	1.37	0.23	0.96
$w_{0.2}\tau_{1.6}\omega_{0.5}$	0.46	1.29	0.40	1.17	0.23	0.76
$w_{0.2}\tau_{1.6}\omega_{0.8}$	0.46	1.02	0.40	0.94	0.22	0.56
$w_{0.2}\tau_{1.6}\omega_{1.0}$	0.46	1.05	0.40	0.96	0.22	0.58

[1] w = proportion of polygenic variance; τ = weight for $\mathbf{G}^{-1}$; ω = weight for $w_{0.25}\tau_{1.6}\omega_{1.0}$

Table 2. Cow validation results from different methods, showing validation accuracies (r) and regression coefficients (b) from the parent average (PA), genomic BLUP (GBLUP), and random regression single-step GBLUP (RR-ssGBLUP) methods with different combinations of parameters

	Milk yield		Protein yield		Fat yield	
Method	r	b	r	b	r	b
PA	0.25	0.61	0.25	0.59	0.27	0.64
GBLUP	0.33	0.85	0.27	0.71	0.18	0.47
RR-ssGBLUP[1]						
$\mathbf{A}_{22}^{-1}$.	0.34	0.64	0.28	0.52	0.21	0.39
$w_{0.05}\tau_{1.0}\omega_{1.0}$	0.34	0.73	0.28	0.60	0.23	0.48
$w_{0.05}\tau_{1.6}\omega_{1.0}$	0.34	0.77	0.29	0.64	0.23	0.53
$w_{0.05}\tau_{2.0}\omega_{1.0}$	0.34	0.81	0.29	0.67	0.24	0.58
$w_{0.05}\tau_{2.5}\omega_{1.0}$	0.35	0.66	0.28	0.54	0.22	0.41
$w_{0.1}\tau_{1.0}\omega_{1.0}$	0.34	0.98	0.30	0.84	0.26	0.79
$w_{0.1}\tau_{1.6}\omega_{0.4}$	0.34	0.96	0.30	0.82	0.26	0.76
$w_{0.1}\tau_{1.6}\omega_{0.5}$	0.35	0.75	0.29	0.62	0.23	0.50
$w_{0.1}\tau_{1.6}\omega_{1.0}$	0.35	0.80	0.29	0.66	0.24	0.55
$w_{0.1}\tau_{2.0}\omega_{1.0}$	0.34	0.83	0.29	0.69	0.24	0.60
$w_{0.1}\tau_{2.5}\omega_{1.0}$	0.34	0.92	0.30	0.79	0.26	0.73
$w_{0.2}\tau_{1.0}\omega_{0.5}$	0.34	0.90	0.30	0.77	0.26	0.70
$w_{0.2}\tau_{1.0}\omega_{0.6}$	0.35	0.87	0.30	0.74	0.26	0.65
$w_{0.2}\tau_{1.0}\omega_{0.7}$	0.35	0.82	0.30	0.70	0.25	0.59
$w_{0.2}\tau_{1.0}\omega_{0.8}$	0.35	0.77	0.30	0.64	0.24	0.52
$w_{0.2}\tau_{1.0}\omega_{0.9}$	0.35	0.69	0.29	0.57	0.23	0.44
$w_{0.2}\tau_{1.0}\omega_{1.0}$	0.34	0.99	0.30	0.86	0.27	0.82
$w_{0.2}\tau_{1.6}\omega_{0.4}$	0.34	0.98	0.30	0.84	0.27	0.79
$w_{0.2}\tau_{1.6}\omega_{0.5}$	0.35	0.90	0.30	0.76	0.26	0.67
$w_{0.2}\tau_{1.6}\omega_{0.8}$	0.35	0.79	0.30	0.65	0.24	0.53
$w_{0.2}\tau_{1.6}\omega_{1.0}$	0.35	0.80	0.30	0.67	0.24	0.55

[1]w = proportion of polygenic variance; τ = weight for $\mathbf{G}^{-1}$; ω = weight for $w_{0.25}\tau_{1.6}\omega_{1.0}$

ters for milk yield were $w_{0.2}\tau_{1.6}\omega_{0.4}$, for protein yield were $w_{0.05}\tau_{2.0}\omega_{1.0}$, and for fat yield were $w_{0.05}\tau_{2.5}\omega_{1.0}$, For cows, the best choice of parameters for all the 3 milk production traits was $w_{0.1}\tau_{1.6}\omega_{0.4}$.

Compared with the multi-step method employed by the national Chinese Holstein milk production evaluation, the RR-ssGBLUP method could simplify the procedure and obtain the combined index in a single step. The advantage of RR-ssGBLUP was shown to exist in prediction accuracy. Furthermore, the more obvious and attractive benefit of RR-ssGBLUP is that it can achieve almost unbiased prediction with specific parameter combinations. In accordance with previous studies (Koivula et al., 2015; Baba et al., 2017), we demonstrated the superiority of RR-ssGBLUP in prediction performance and the feasibility of its implementation in breeding practice. Our results show that the best choice of parameters involved in RR-ssGBLUP depended on traits and candidate populations. We should figure out the optimal values of parameters via preliminary validation study and further apply them in practical applications.

ACKNOWLEDGMENTS

This work was supported by the National High Technology Research and Development Program of China (863 Program 2013AA102503, Beijing), the National Natural Science Foundations of China (31661143013, 31272419; Beijing), the Program for Changjiang Scholar and Innovation Research Team in University (IRT_15R62, Beijing), and Scientific Research Start-up Fund for High-level Talents of Foshan University (gg07079, Foshan).

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.

Akaike, H. 1974. A new look at the statistical model identification. IEEE Trans. Automatic Control 19:716–723.

Baba, T., Y. Gotoh, S. Yamaguchi, S. Nakagawa, H. Abe, Y. Masuda, and T. Kawahara. 2017. Application of single-step genomic best linear unbiased prediction with a multiple-lactation random regression test-day model for Japanese Holsteins. Anim. Sci. J. 88:1226–1231. (Nihon Chikusan Gakkaiho).

Christensen, O. F., and M. S. Lund. 2010. Genomic prediction when some animals are not genotyped. Genet. Sel. Evol. 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.

Garrick, D. J., J. F. Taylor, and R. L. Fernando. 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. Genet. Sel. Evol. 41:55.

Jamrozik, J., L. R. Schaeffer, and G. B. Jansen. 2000. Approximate accuracies of prediction from random regression models. Livest. Prod. Sci. 66:85–92.

Kang, H., C. Ning, L. Zhou, S. Zhang, N. Yang, and J. F. Liu. 2018. PIBLUP: High-performance software for large-scale genetic evaluation of animals and plants. Frontiers Genet. 9:226.

Kang, H., L. Zhou, R. Mrode, Q. Zhang, and J. F. Liu. 2017. Incorporating the single-step strategy into a random regression model to enhance genomic prediction of longitudinal traits. Heredity 119:459.

Koivula, M., I. Strandén, J. Pösö, G. P. Aamand, and E. A. Mäntysaari. 2015. Single-step genomic evaluation using multitrait random regression model and test-day data. J. Dairy Sci. 98:2775–2784.

Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. J. Dairy Sci. 92:4656–4663.

Legarra, A., O. F. Christensen, I. Aguilar, and I. Misztal. 2014. Single Step, a general approach for genomic selection. Livest. Sci. 166:54–65.

Madsen, P., P. Sørensen, G. Su, L. H. Damgaard, H. Thomsen, and R. Labouriau. 2014. DMU–A package for analyzing multivariate mixed models. Pages 27–11 in Proc. 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Minas Gerais, Brazil, 13–18 August, 2006.

Martini, J. W. R., M. F. Schrauf, C. A. Garcia-Baccino, E. C. G. Pimentel, S. Munilla, A. Rogberg-Muñoz, R. J. C. Cantet, C. Reimer, N. Gao, V. Wimmer, and H. Simianer. 2018. The effect of the H^{-1} scaling factors τ and ω on the structure of H in the single-step procedure. Genet. Sel. Evol. 50:16.

Masuda, Y., P. M. Vanraden, I. Misztal, and T. J. Lawlor. 2018. Differing genetic trend estimates from traditional and genomic evaluations of genotyped animals as evidence of preselection bias in US Holsteins. J. Dairy Sci. 101:5194–5206.

Matilainen, K., M. Koivula, I. Strandén, G. P. Aamand, and E. A. Mäntysaari. 2016. Managing genetic groups in single-step genomic evaluations applied on female fertility traits in Nordic Red dairy cattle. Pages 71–75 in Interbull Bulletin. Vol. 50, Puerto Varas, Chile.

Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet, and D. H. Lee. 2002. BLUPF90 and related programs (BGF90). Page 743 in Proc. World Congress on Genetics Applied To Livestock Production, Montpellier, France.

Muir, B. L., G. Kistemaker, J. Jamrozik, and F. Canavesi. 2007. Genetic parameters for a multiple-trait multiple-lactation random regression test-day model in Italian Holsteins. J. Dairy Sci. 90:1564–1574.

Sargolzaei, M., J. P. Chesnais, and F. S. Schenkel. 2014. A new approach for efficient genotype imputation using information from relatives. BMC Genomics 15:478.

Schaeffer, L. R., J. Jamrozik, G. J. Kistemaker, and B. J. Van Doormaal. 2000. Experience with a test-day model. J. Dairy Sci. 83:1135–1144.

Schwarz, G. 1978. Estimating the dimension of a model. Ann. Stat. 6:461–464.

Strabel, T., J. Szyda, E. Ptak, and J. Jamrozik. 2005. Comparison of random regression test-day models for Polish Black and White cattle. J. Dairy Sci. 88:3688–3699.

Tsuruta, S., I. Misztal, I. Aguilar, and T. J. Lawlor. 2011. Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in US Holsteins. J. Dairy Sci. 94:4198–4204.

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414–4423.

Vitezica, Z. G., I. Aguilar, I. Misztal, and A. Legarra. 2011. Bias in genomic predictions for populations under selection. Genet. Res. (Camb.) 93:357–366.