J. Dairy Sci. 98:4999–5003
http://dx.doi.org/10.3168/jds.2015-9371

Short communication: Reliability of single-step genomic BLUP breeding values by multi-trait test-day model analysis

J. Bauer,[1] J. Přibyl, and L. Vostrý
Institute of Animal Science, Přátelství 815, 10401 Praha-Uhříněves, Czech Republic

ABSTRACT

The purpose of our study was to develop an approximation procedure to estimate reliabilities of single-step genomic BLUP breeding values in a test-day model for routine evaluation of milk yield in a dairy cattle population. Input data consisted of 20,220,047 first-, second-, and third-lactation test-day milk yield records of 1,126,102 Czech Holstein cows (each lactation being considered a separate trait), with 1,844,679 animals in the pedigree file and with genomic data from 2,236 bulls. Evaluation was according to a multi-lactation model. The procedure was based on the effective number of records per animal from milk recording as well as from genomic and pedigree relationships. Traits were analyzed individually, and genetic covariances among traits were subsequently taken into account. The use of genomic information increased average reliability in young bulls from 0.276 to 0.505, but increased reliability in proven bulls only from 0.828 to 0.855. The reliabilities of genomic breeding values in multi-trait evaluation for first, second and third lactations, respectively, averaged 0.652, 0.673, and 0.633 for young bulls and 0.907, 0.894, and 0.852 for proven bulls. For an index combining all 3 lactations, the average reliability of a single-step genomic BLUP prediction was 0.712 and 0.925 for younger and proven bulls, respectively. Increased reliability due to genotyping in the population of all genotyped and nongenotyped animals was very small (<0.01) because of the small proportion of genotyped animals in the population.
Key words: reliability, genomic prediction, single-step genomic BLUP, test-day model

INTRODUCTION

Utilizing information from genomic breeding values is a promising procedure to increase accuracy of genetic evaluations of farm animals, especially from young animals without performance data. Two types of genomic prediction of breeding values are currently used: the multi-step method of Meuwissen et al. (2001) and VanRaden (2008) and the single-step method of Misztal et al. (2009) and Christensen and Lund (2010).

Reliabilities of genomic breeding values are required to achieve efficient selection of parents for future generations, especially for the international Interbull multiple across-country evaluation (MACE), in which reliabilities are used as weighting factors for various information sources (Schaeffer, 1994; Sullivan and Jakobsen, 2012). They can be calculated by inversion of the left hand side of a BLUP system of equations (VanRaden, 2008) but this is usually not feasible due to large population sizes and massive computational requirements. For this reason, methods to approximate reliabilities of genomic EBV (**GEBV**) were developed by Szyda et al. (2011) for the multi-step method and by Misztal et al. (2013) for the single-step procedure.

Přibyl et al. (2012, 2013) and Zavadilová et al. (2014) developed a workable procedure for genomic evaluation of dairy cattle in the Czech Republic. However, for routine use, it is necessary to develop a method to calculate reliabilities of genomic breeding values. The aim of our current investigation, therefore, was to develop a method for routine approximation of the reliability of single-step genomic breeding values for test-day model evaluation.

We used 20,220,047 test-day milk yield (kg) records of 1,126,102 Czech Holstein cows. Observations were from the first 3 lactations, with each considered as a separate trait (9,480,924, 6,681,967, and 4,057,156 first-, second-, and third-lactation records, respectively). With inclusion of 6 generations of ancestors, the total number of animals in the evaluation was 1,844,679. There were 2,236 genotyped bulls in the population, of which 445 were considered unproven "young" bulls (fewer than 3 offspring each) and the rest were considered proven. The proven bulls had 240 daughters on average.

The Illumina BovineSNP50 Beadchip V2 (Illumina Inc., San Diego, CA) was used for genotyping. To eliminate possible input errors, data were edited for minor allele frequency (MAF) <0.05, number of loci per bull <90% of all possible loci, number of bulls per locus <90% of all possible bulls and large discrepancy

Received January 22, 2015.
Accepted March 20, 2015.
[1]Corresponding author: bauer.jiri@vuzv.cz

of pedigree relationship $\mathbf{A_{22}}$ with genomic relationship **G** (absolute difference in relationship to others >3 animals >0.30), and proportion of Holstein genes <85%. After filtering, 40,653 SNP loci were used in the analysis.

The routine random regression test-day model of animal evaluation in the Czech Republic is described in studies of Zavadilová et al. (2005a,b) and evaluates cow test-day milk production by fixed effects: herd-test-day, Legendre polynomial regression coefficients for lactation and contemporary group, regressor for lactation day, and random effects of Legendre polynomials for permanent environment and additive genetic effect of animal.

We developed our own programs for prediction of genomic breeding values and approximation of reliabilities parallel in 2 languages: SAS (SAS Institute, 2005) and COBOL (ISO, 2014).

In routine BLUP evaluation (EBV prediction), a pedigree relationship matrix **A** is used to quantify the additive genetic effect, whereas, for single-step genomic (**ssGBLUP**) evaluation (GEBV prediction), a relationship matrix **H** is used, in which is **A** augmented by a genomic relationship matrix **G**, with weights 20 and 80% respectively (Christensen and Lund, 2010). Matrix **G** is constructed according to deviations from the averages of observed allele frequencies and standardized so the average of diagonals equals 1 (Forni et al., 2011) and then shifted so the elements of the pedigree relationship matrix of genotyped animals $\mathbf{A}_{22}$ and elements of **G** will have the same average (Vitezica et al., 2011).

Overall GEBV or EBV were calculated in our current case as an index (**i**) by summing values of genetic polynomials for 300 d of lactation and 3 lactations, and then calculating the average across those 3 lactations:

$$\mathbf{i} = (\Sigma\Sigma \mathrm{a}_{zjm} \times \mathrm{v}_{mt})/3,$$

where $\Sigma\Sigma$ = sum of random Legendre polynomial regression coefficients m and lactations j, a_{zjm} = mth random effect of Legendre polynomial regression coefficient for animal z in lactation number j with covariance matrix (12 × 12) covering random regression coefficients over all 3 lactations and connected to relationship matrix, and v_{mt} = mth regressor for lactation day t ($t = 6, \ldots, 305$).

Reliabilities of predicted genomic breeding values were estimated based on the procedure of Misztal et al. (2013). To take into account the multi-trait nature of the evaluation, we used the method described by Strabel et al. (2001). The entire procedure can be characterized as a sequence of the following consecutive steps:

1. Approximation of the reliability of BLUP breeding values for all animals and for each trait by the iterative approach of Misztal et al. (1993). This method is based on the approximation of reliability using the effective number of records arising from performance observations within individual contemporary groups and from relationships among animals in the pedigree. To take into account a permanent environmental effect common for all test-days of the same individual, the effective number of records **d** for each trait was updated before the iterative procedure by the formula $\mathbf{d^*} = \tau\mathbf{d}/(\tau + \mathbf{d})$ (Misztal et al., 1991), where the asterisk is differentiation mark between $\mathbf{d}^*$ and nonupdated **d** and τ is the ratio of residual variance to permanent environmental variance. For this step of the procedure, genetic parameters were averaged throughout individual lactations, and observations were taken into account as repeated records.
2. Calculation of the reliability of genotyped animals with inclusion of the increase due to the additional genomic information in prediction of ssGBLUP genomic breeding values. The calculation was done as follows: reliability $r_i^2 = 1 - \alpha q_{ii}$, where α is the ratio of error variance to animal genetic variance and q_{ii} is the diagonal elements of the $\mathbf{Q}^{-1}$ matrix:

$$\mathbf{Q}^{-1} = \left[\mathbf{D} + \left(\mathbf{I} + \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}\right)\ \alpha\right]^{-1},$$

   where **D** is the contribution of the records and pedigrees to the reliability from step (1), **I** is the identity matrix, $\mathbf{G}^{-1}$ is the inverse of the genomic relationship matrix, and $\mathbf{A}_{22}^{-1}$ is the inverse of the section of the pedigree-based relationship matrix that contains relationship information from only the genotyped animals (Misztal et al., 2013). We applied this procedure for each trait separately.
3. Addition of the contribution of genotyping to the reliabilities of nongenotyped animals. This step was completed using a procedure similar to that described in the first step. The reliabilities of the genotyped bulls were kept constant to avoid double counting the contribution of the relationships among these bulls and to avoid changing the already fully conveyed values.
4. Re-evaluation of individual reliabilities of breeding values due to the covariances among traits in the multi-trait model by the procedure of Strabel et al. (2001). This method approximates reli-