J. Dairy Sci. 100:453–464
https://doi.org/10.3168/jds.2016-11496

Assessing genomic prediction accuracy for Holstein sires using bootstrap aggregation sampling and leave-one-out cross validation

Ashley A. Mikshowsky,*[1] Daniel Gianola,*†‡ and Kent A. Weigel*
*Department of Dairy Science,
†Department of Animal Sciences, and
‡Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison 53706

ABSTRACT

Since the introduction of genome-enabled prediction for dairy cattle in 2009, genomic selection has markedly changed many aspects of the dairy genetics industry and enhanced the rate of response to selection for most economically important traits. Young dairy bulls are genotyped to obtain their genomic predicted transmitting ability (GPTA) and reliability (REL) values. These GPTA are a main factor in most purchasing, marketing, and culling decisions until bulls reach 5 yr of age and their milk-recorded offspring become available. At that time, daughter yield deviations (DYD) can be compared with the GPTA computed several years earlier. For most bulls, the DYD align well with the initial predictions. However, for some bulls, the difference between DYD and corresponding GPTA is quite large, and published REL are of limited value in identifying such bulls. A method of bootstrap aggregation sampling (bagging) using genomic BLUP (GBLUP) was applied to predict the GPTA of 2,963, 2,963, and 2,803 young Holstein bulls for protein yield, somatic cell score, and daughter pregnancy rate (DPR), respectively. For each trait, 50 bootstrap samples from a reference population comprising 2011 DYD of 8,610, 8,405, and 7,945 older Holstein bulls were used. Leave-one-out cross validation was also performed to assess prediction accuracy when removing specific bulls from the reference population. The main objectives of this study were (1) to assess the extent to which current REL values and alternative measures of variability, such as the bootstrap standard deviation (SD) of predictions, could detect bulls whose daughter performance deviates significantly from early genomic predictions, and (2) to identify factors associated with the reference population that inform about inaccurate genomic predictions. The SD of bootstrap predictions was a mildly useful metric for identifying bulls whose future daughter performance may deviate significantly from early GPTA for protein and DPR. Leave-one-out cross validation allowed us to identify groups of reference population bulls that were influential on other reference population bulls for protein yield and observe their effects on predictions of testing set bulls, as a whole and individually.
Key words: genomic prediction, leave-one-out cross validation, bootstrap sampling, dairy cattle

INTRODUCTION

Since the release of genomic evaluations for Holsteins and Jerseys in January 2009 (Wiggans et al., 2011), the transmitting abilities of nearly all young animals perceived to be genetically elite have been predicted using molecular markers spanning the entire genome (Meuwissen et al., 2001). The resulting genomic data are integrated into the national genetic improvement program, which is managed by the Council on Dairy Cattle Breeding (**CDCB**; Bowie, MD). The genomic PTA (**GPTA**) of young genome-tested animals are predicted using data from a reference population of older animals with genotypic and phenotypic data, and breeding companies and dairy farmers use this information for selection decisions on a weekly basis.

Genomic BLUP (**GBLUP**), a common method used for genome-enabled prediction, is used to compute GPTA. The resulting GPTA take into account information from all genotyped relatives and nongenotyped offspring of genotyped sires. The official GPTA use selection index blending to include information from animals' nongenotyped ancestors as well. The reliability (**REL**) values corresponding to the GPTA reflect the approximate amount of information contributed by an animal's parents, progeny, own performance, and molecular markers. They are estimated with an approximation that uses a weighted sum of the genomic relationships between an animal and the reference population (Wiggans and VanRaden, 2010).

Genomic PTA are used by dairy producers to identify groups of young bulls to be used as service sires and groups of young heifers that should be retained as

Received May 20, 2016.
Accepted October 5, 2016.
[1]Corresponding author: amikshowsky@gmail.com

future herd replacements (Weigel et al., 2012). However, REL values are lower than those usually reached by progeny testing (VanRaden et al., 2009), and the GPTA for some bulls deviate from their eventually realized daughter performance, as measured by subsequent daughter yield deviations (**DYD**) for production traits or daughter deviations (**DD**) for health traits of these bulls.

Previous genomic evaluations computed REL by inversion of the mixed model equation, but this approach was stopped when data sets became very large. Single-step GBLUP was used to estimate reliabilities by Misztal et al. (2013). The algorithm they developed, which used inversion of a matrix containing inverses of both the genomic and pedigree relationship matrix for genotyped animals, was found to be reasonably accurate and low cost for data sets containing fewer than 100,000 genotypes.

One way to assess the effect of individual animals in the reference population is through leave-one-out cross validation (**LOOCV**). In LOOCV, a model is continuously refit, each time removing a single reference population individual and computing predictions for all other animals. Although LOOCV has been criticized for having high variance (Breiman and Spector, 1992), it has the advantage of allowing us to observe the influence of each individual reference population animal on the predictions of the remaining animals and provides a conservative measure of prediction error. When each animal in the reference population is removed once, the resulting predictions for the other animals can be compared with the predictions obtained from using the full reference population. Influential animals in the reference population who have many offspring may cause significant variation in the testing set animals' predictions, and these influential ancestors can be identified and further analyzed.

An alternative method of estimating the stability of a bull's GPTA is through the use of bootstrap aggregation sampling, also known as "bagging." Bagging was recently used for Jersey sires by Mikshowsky et al. (2016). It is a resampling method that is simple to implement, and it can increase the accuracy of predictions in situations where sampling from the training set leads to large variance in the predictor (Breiman, 1996). It involves repeated sampling with replacement from the original reference population to create a set of predictors, which are averaged across samples to calculate the bagged predictor. Gianola et al. (2014) first applied this methodology to genome-enabled prediction and computed bagged GBLUP (hereafter **BGBLUP**) predictors.

This paper builds upon the work of Mikshowsky et al. (2016), in which bootstrap aggregation sampling was used for Jersey sires in an attempt to find other measures of REL that might provide a useful alternative to published REL values. The authors concluded that BGBLUP did not improve the accuracy of genomic predictions in Jerseys, but it allowed computation of bootstrap prediction REL across random samples of the reference population. These bootstrap prediction REL could be used to construct useful diagnostic tools for assessing genomic prediction systems or for evaluating the composition of a genomic reference population. The bootstrap SD of GBLUP was found to be a weak indicator of the magnitude of prediction errors for protein, but not for SCS or daughter pregnancy rate (**DPR**).

The present study uses BGBLUP for Holstein sires. The Holstein population is much larger than that of Jerseys, so many more bulls are available for the study (approximately 5 times as many bulls in the reference population). The initial BGBLUP analyses were carried out as in the previous study but, with the larger number of bulls, we could look more closely at the effect of the sire in the reference population. A LOOCV was also conducted to analyze prediction stability for candidate bulls and to identify bulls in the reference population who may cause major changes in testing set predictions. The objectives of this study were (1) to determine if bagging GBLUP for protein yield, SCS, and DPR of young Holstein bulls could provide a measure to aid in the identification of bulls whose daughter performance will deviate significantly from early genomic predictions, and (2) to identify characteristics of the reference population that are associated with inaccurate or highly variable genomic predictions.

MATERIALS AND METHODS

Data

The genotypes of 17,276 Holstein bulls were provided by the Cooperative Dairy DNA Repository (Columbia, MO). The original genomic data included 60,671 SNP markers for each bull. Single nucleotide polymorphisms with >20% missing values and those with a minor allele frequency ≤5% were discarded, and any missing values still in the data set were imputed based on the allele frequency of the marker. A total of 57,169 markers remained for analysis.

Three phenotypic traits were analyzed: protein yield (kg), SCS $\{\log_2[(\text{cells/mL})/100{,}000] + 3\}$, and DPR (%). The PTA values for all 3 traits, as well as DYD for protein yield and DD for SCS and DPR, were obtained from the CDCB for the August 2011 and August 2014 sire summaries. Holstein bulls with at least 50 US daughters for a given trait in August 2011 were used as the reference population, and Holstein bulls with