J. Dairy Sci. 100:1–14
https://doi.org/10.3168/jds.2016-11733

National single-step genomic method that integrates multi-national genomic information

J. Vandenplas,*†[1,2] M. Spehar,‡§ K. Potocnik,§ N. Gengler,* and G. Gorjanc§#
*Agriculture, Bio-engineering and Chemistry Department, Gembloux Agro-Bio Tech, University of Liege, 5030 Gembloux, Belgium
†National Fund for Scientific Research, 1000 Brussels, Belgium
‡Croatian Agricultural Agency, 10000 Zagreb, Croatia
§Biotechnical Faculty, University of Ljubljana, 1000 Ljubljana, Slovenia
#The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Easter Bush Research Centre, Midlothian EH25 9RG, United Kingdom

ABSTRACT

The aim of this paper was to develop a national single-step genomic BLUP that integrates multi-national genomic estimated breeding values (EBV) and associated reliabilities without double counting dependent data contributions from the different evaluations. Simultaneous use of all data, including phenotypes, pedigree, and genotypes, is a condition to obtain unbiased EBV. However, this condition is not always fully met, mainly due to unavailability of foreign raw data for imported animals. In dairy cattle genetic evaluations, this issue is traditionally tackled through the multiple across-country evaluation (MACE) of sires, performed by Interbull Centre (Uppsala, Sweden). Multiple across-country evaluation regresses all the available national information onto a joint pedigree to obtain country-specific rankings of all sires without sharing the raw data. In the context of genomic selection, the issue is handled by exchanging sire genotypes and by using MACE information (i.e., MACE EBV and reliabilities), as a valuable source of "phenotypic" data. Although all the available data are considered, these "multi-national" genomic evaluations use multi-step methods assuming independence of various sources of information, which is not met in all situations. We developed a method that handles this by single-step genomic evaluation that jointly (1) uses national phenotypic, genomic, and pedigree data; (2) uses multi-national genomic information; and (3) avoids double counting dependent data contributions from an animal's own records and relatives' records. The method was demonstrated by integrating multi-national genomic EBV and reliabilities of Brown Swiss sires, included in the InterGenomics consortium at Interbull Centre, into the national evaluation in Slovenia. The results showed that the method could (1) increase reliability of a national (genomic) evaluation; (2) provide consistent ranking of all animals: bulls, cows, and young animals; and (3) increase the size of a genomic training population. These features provide more efficient and transparent selection throughout a breeding program.

Key words: single-step genomic BLUP, combination, multi-national, genomically enhanced estimated breeding values

INTRODUCTION

This paper presents a national single-step genomic method that integrates multi-national genomic EBV and associated reliabilities. Breeding programs collect phenotypic and genetic data and distill it in the form of EBV and associated reliabilities by using BLUP (Henderson, 1984). Simultaneous use of all data is a condition to obtain unbiased EBV but this condition is not always fully met. For example, most breeding programs make some use of elite parents imported from other populations. When the imported individuals or their relatives are evaluated as a part of the routine national genetic evaluation, only their national data are commonly used. The inability to include foreign data in national evaluations can lead to bias and lower reliability of EBV. This is particularly the case in small breeding programs that often rely more on importation than do large programs.

In dairy cattle, the issue of incomplete across-country data is traditionally tackled through the multiple across-country evaluation of sires (**MACE**; Schaeffer, 1994). A unit of information in such evaluations is a sire's EBV and associated reliability. Conducted by Interbull Centre (Uppsala, Sweden), MACE simultaneously regresses all the available national information

Received July 12, 2016.
Accepted October 5, 2016.
[1]Current address: Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, 6700 AH Wageningen, the Netherlands.
[2]Corresponding author: jeremie.vandenplas@wur.nl

onto a joint pedigree to obtain country-specific rankings of all sires without sharing the "raw" data. Because all sires are evaluated on each country scale, MACE has proved to be a valuable source of "phenotypic" data for genomic evaluations. By using the MACE information and exchanging sire genotypes, the national programs have been able to build large training populations that facilitate accurate genomic predictions (VanRaden et al., 2009; Lund et al., 2010; Jorjani et al., 2012).

The most common approach to use MACE information in genomic evaluation is via the so-called multi-step method. The method involves deregressing the MACE EBV to obtain pseudo-records and regressing these on genomic relationships with implicit or explicit blending of the traditional pedigree-based information (VanRaden, 2008). Although breeding programs perform such evaluations nationally, the evaluations involve a multi-national training population with domestic and foreign information. We refer to such an evaluation as a "multi-national genomic evaluation."

An example of the multi-step method is the genomic evaluation in the Brown Swiss breed operated through the InterGenomics consortium at Interbull Centre (Jorjani et al., 2012). Individual Brown Swiss breeding programs in several countries (Austria, France, Germany, Italy, Slovenia, Switzerland, and the United States) have limited training populations for genomic selection because of the limited size of their respective national populations. To enable accurate genomic evaluation, the programs agreed to combine national genotype data sets, and the phenotypic data (the MACE information) are readily available for all sires on each country scale.

An alternative to the multi-step method is the single-step method that jointly uses phenotypic, genomic, and pedigree data in one analysis (Legarra et al., 2014). Although the multi-step method is practical, it rests on several assumptions that are not met in all situations (Legarra et al., 2014). For example, the multi-step method assumes that pseudo-records are independent. This can be assumed for pseudo-records with high reliability, as is the case with sires tested on a large number of progeny in large breeding programs. However, independence cannot be assumed for pseudo-records with low reliability, as is the case with sires tested on a smaller number of progeny in small breeding programs. When records are dependent, the analysis should take this into account to avoid double counting of information and consequently overestimating reliability (e.g., Calus et al., 2016). The single-step method, referred here to as the single-step genomic BLUP (**ssGBLUP**), avoids this because it simultaneously utilizes any phenotypic, genomic, and pedigree data in a single evaluation (Legarra et al., 2014).

However, standard ssGBLUP cannot integrate multi-national information. A breeding program with national evaluation based on ssGBLUP needs a way to integrate multi-national information to use all available sources of information optimally. Such a method would have to consider that the multi-national information might be partially based on national information, which needs to be accounted for to avoid double counting. Specifically, double counting can occur due to double use of data that pertains to an individual animal, as well as to double use of data that is correlated among relatives (Vandenplas and Gengler, 2012; Vandenplas et al., 2014). We will refer to these 2 as double counting contributions from records and double counting contributions from relatives, respectively.

The aim of this study was to develop and demonstrate the potential of a national single-step genomic method that integrates multi-national genomic information and avoids the double counting. The method delivers a national genomic evaluation that uses all the available data in an appropriate manner and presents results on the same scale for bulls, cows, and young animals. These 2 properties enable efficient and transparent selection among all animals.

MATERIALS AND METHODS

The first part of this section describes theory of (1) the standard ssGBLUP, (2) the integration of multi-national ssGBLUP information into a national ssGBLUP, and (3) the required corrections to avoid double counting. The second part describes a demonstration.

Theory

The Standard ssGBLUP. The standard ssGBLUP can be applied to a univariate linear mixed model:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e}, \qquad [1]$$

where $\mathbf{y}$ is the vector of phenotypic records, $\mathbf{b}$ is the vector of fixed effects, $\mathbf{a}$ is the vector of random additive genetic effects, and $\mathbf{e}$ is the vector of residuals. The matrices $\mathbf{X}$ and $\mathbf{Z}$ are incidence matrices linking $\mathbf{y}$ with, respectively, $\mathbf{b}$ and $\mathbf{a}$.

In the context of ssGBLUP, it is assumed that $\mathbf{a} \sim MVN\left(0, \mathbf{H}\sigma_a^2\right)$, where MVN = multivariate normal, $\mathbf{H}$ is a combined genomic and pedigree relationship matrix, σ_a^2 is the additive genetic variance of the trait; and that $\mathbf{e} \sim MVN\left(0, \mathbf{I}\sigma_e^2\right)$, where $\mathbf{I}$ is an identity matrix and σ_e^2 is the residual variance (Legarra et al., 2014).

The inverse of $\mathbf{H}$ is