Animal (2018), **12:5**, pp 898–905 © The Animal Consortium 2017
doi:10.1017/S1751731117002324

animal

Bayesian single-step genomic evaluations combining local and foreign information in Walloon Holsteins

F. G. Colinet[1a], J. Vandenplas[1,2ab], S. Vanderick[1], H. Hammami[1], R. R. Mota[1], A. Gillon[3], X. Hubin[3], C. Bertozzi[3] and N. Gengler[1†]

[1]*TERRA Teaching and Research Centre, Gembloux Agro-Bio Tech, University of Liège, B-5030 Gembloux, Belgium;* [2]*National Fund for Scientific Research, 1000 Brussels, Belgium;* [3]*Research and Development Department, Walloon Breeding Association, 5590 Ciney, Belgium*

(Received 12 January 2017; Accepted 11 August 2017; First published online 16 October 2017)

Most dairy cattle populations found in different countries around the world are small to medium sized and use many artificial insemination bulls imported from different foreign countries. The Walloon population in the southern part of Belgium is a good example for such a small-scale population. Wallonia has also a very active community of Holstein breeders requesting high level genetic evaluation services. Single-step Genomic BLUP (ssGBLUP) methods allow the simultaneous use of genomic, pedigree and phenotypic information and could reduce potential biases in the estimation of genomically enhanced breeding values (GEBV). Therefore, in the context of implementing a Walloon genomic evaluation system for Holsteins, it was considered as the best option. However, in contrast to multi-step genomic predictions, natively ssGBLUP will only use local phenotypic information and is unable to use directly important other sources of information coming from abroad, for example Multiple Across Country Evaluation (MACE) results as provided by the Interbull Center (Uppsala, Sweden). Therefore, we developed and implemented single-step Genomic Bayesian Prediction (ssGBayes), as an alternative method for the Walloon genomic evaluations. The ssGBayes method approximated the correct system of equations directly using estimated breeding values (EBV) and associated reliabilities (REL) without any explicit deregression step. In the Walloon genomic evaluation, local information refers to Walloon EBV and REL and foreign information refers to MACE EBV and associated REL. Combining simultaneously all available genotypes, pedigree, local and foreign information in an evaluation can be achieved but adding contributions to left-hand and right-hand sides subtracting double-counted contributions. Correct propagation of external information avoiding double counting of contributions due to relationships and due to records can be achieved. This ssGBayes method computed more accurate predictions for all types of animals. For example, for genotyped animals with low Walloon REL (<0.25) without MACE results but sired by genotyped bulls with MACE results, the average increase of REL for the studied traits was 0.38 points of which 0.08 points could be traced to the inclusion of MACE information. For other categories of genotyped animals, the contribution by MACE information was also high. The Walloon genomic evaluation system passed for the first time the Interbull GEBV tests for several traits in July 2013. Recent experiences reported here refer to its use in April 2016 for the routine genomic evaluations of milk production, udder health and type traits. Results showed that the proposed methodology should also be of interest for other, similar, populations.

Keywords: dairy cattle, Bayesian integration, genomic evaluation, reliabilities

Implications

All genomic evaluations have to rely on the association of phenotypes with genotypes. In dairy cattle, phenotypic information for genotyped sires often comes from other countries where daughters of these sires are present. This is especially the case for small populations that often rely on importation of bulls from larger populations. Appropriate genomic evaluation systems combining optimally all information are therefore internationally warranted to maximize their positive social impact on the whole breeding industry (i.e. breeders, artificial insemination companies). Only when based on such advances, genomic selection can contribute broadly on an economic and, using the appropriate traits, environmental level. Currently there are few contributions. Therefore this paper shows, using the example of the

[a] These authors contributed equally to this work.
[b] Present address: Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, 6700 AH Wageningen, The Netherlands.
[†] E-mail: nicolas.gengler@ulg.ac.be

https://doi.org/10.1017/S1751731117002324 Published online by Cambridge University Press

Walloon region of Belgium, how such an advanced approach can be implemented in the setting up of a single-step genomic evaluation.

Introduction

Simultaneous use of all relevant data by BLUP is a requirement to predict unbiased breeding values (Henderson, 1984). However, this condition is not always fully met. Especially in dairy cattle, but also in other species where external animals are introduced, foreign artificial insemination (AI) sires are strongly preselected, but foreign raw data used to do this selection is unavailable leading to potential biases in local evaluations. Local estimated breeding values (EBV) might be also less accurate because only incomplete data are available (i.e. foreign raw data is not included). Genomic selection could increase these problems for local genomic evaluations. Currently, most genomic evaluation systems are multi-step (VanRaden *et al*., 2009). The multi-step methods involve a deregression step of, mainly highly reliable, EBV to obtain pseudo-records, followed by regressing these pseudo-records on genomic relationships with a blending of the traditional pedigree-based information (VanRaden, 2008). Usually, the primary sources of foreign phenotypic information are the results from the Multiple Across Country Evaluation (MACE) computations, performed by the Interbull Centre (Uppsala, Sweden). On the other hand, more and more genomic evaluation systems (e.g. Harris *et al*., 2012; Koivula *et al*., 2012; Přibyl *et al*., 2013) show interest switching to approximate, using pseudo-observations, or full single-step genomic BLUP (ssGBLUP), that simultaneously combines genomic, pedigree and all local phenotypic information (Legarra *et al*., 2014). The ssGBLUP, mainly through the simultaneous combination, should reduce potential biases introduced by the multi-step methods in the estimation of genomically enhanced breeding values (GEBV; VanRaden, 2012). However, ssGBLUP normally uses local information and does not natively fulfil the condition by using all important sources of information, as multi-step methods do; that is, by relying on MACE results.

The Walloon dairy cattle population under performance recording is a small-scale population of ~55 000 cows, mainly sired by foreign bulls. For example, 81% of the cows in 1st to 3rd parity in 2015 were sired by AI bulls born outside of the Walloon Region of Belgium. In this context, properties of both multi-step and single-step methods mentioned before, were attractive in the development of Walloon genomic evaluations systems. Due to the widespread use of imported AI bulls, the recovery of sources of information, such as MACE, was required for the Walloon genomic evaluation system. However, due to the type and evolution of the genotyped population (i.e. more genotyped cows) and expected future developments (e.g. novel traits recorded on genotyped cows, multi-trait models) ssGBLUP was considered the best option for the Walloon genomic evaluation, but under the condition to combine simultaneously all available genotype, pedigree, local and foreign information in a single-step genomic evaluation.

Towards this orientation, in order to support local breeding, the objective of this study was to report the development of a full genomic evaluation for Walloon Holsteins by using single-step genomic prediction combining local and foreign information (MACE results).

Material and methods

Hereafter, records will refer to EBV and associated reliabilities (REL), local information will refer to local EBV and associated REL estimated from available local data and foreign records or information will refer to MACE EBV and associated REL.

Genotyped and phenotyped populations

During the April 2016 run, 4847 genotyped animals were included in the Walloon genomic evaluations. Among them, 4324 were born after the year 2001, the remaining genotyped animals being ancestors. A large majority of genotyped animals were bulls (4038 bulls *v*. 809 cows). Among these genotyped bulls were a large proportion of the sires and maternal grandsires of recent Walloon Holstein cows. Approximately half of the cows were genotyped during research projects representing the Walloon variability in Holstein cattle, and the other cows were genotyped upon request from breeders, limiting the risk of preselection. A total of 38 374 single nucleotide polymorphisms (SNP) markers were selected after editing for call rates and minor allele frequencies. The pedigree extracted for the genotyped animals and from the database used for the official Walloon genetic evaluation included 29 383 animals.

Genetic parameters, number of animals and information available from single trait analyses of six traits (milk yield, fat yield, protein yield, somatic cell score (SCS), stature and udder support) are presented in Table 1. The production traits (milk, fat and protein) were defined as 305 days yields, SCS was defined as average value across lactations. All genotyped cows were associated with records in the routine Walloon genetic evaluation, but not all genotyped sires. Numbers of available local and foreign records were slightly different among the traits. Local information included EBV and associated REL for bulls and cows estimated from data provided by the Walloon Breeding Association (subscript W; EBV_W; REL_W) for the official Walloon evaluation of December 2015. Foreign information included MACE EBV and REL for sires provided by the December 2015 MACE evaluation (subscript M; EBV_M, REL_M).

For every trait, contributions of Walloon information into MACE were determined based on the domestic effective daughter equivalents (EDC) associated with EBV_M and REL_M as reported by Interbull. Multiple Across Country Evaluation information free of Walloon information had, therefore, a reported domestic EDC equal to 0. For all animals and traits with a domestic EDC different from 0, Walloon EBV and

https://doi.org/10.1017/S1751731117002324 Published online by Cambridge University Press

associated REL contributing to the December 2015 MACE routine-run (subscript Wc; EBV_{Wc}; REL_{Wc}) were considered to avoid double counting of contributions due to records (Table 1). Information was harmonized between the local and MACE values by adjusting scale and mean difference towards the original expression of the trait reported in the Walloon genetic evaluation computations. Local phenotypic information was, therefore, also summarized using EBV and REL.

Bayesian integration procedure

A Bayesian procedure to integrate multiple sources of external information into local genetic evaluations was outlined by Vandenplas *et al.* (2014). The term 'Bayesian' refers to the alteration of the prior mean and (co)variances of the random additive genetic effect of the local genetic evaluation. The alteration is based on the knowledge of EBV and REL obtained from a genetic evaluation other than the local genetic evaluation, for allowing the consideration of this information by the local genetic evaluation. While it is, at least theoretically, equivalent to an approach based on deregression (Calus *et al.*, 2016), the proposed Bayesian approach has the advantage that it does not rely explicitly on pseudo-observations, because it directly generates an approximate system of equations. Also, Vandenplas *et al.* (2014) outlined that these proposed systems of equations could be extended to integrate multiple sources of external information into ssGBLUP. An advantage of Bayesian procedures is that foreign original phenotypes are not required. The equations blend local and several sources of external information and avoid double counting of contributions due to records and due to relationships. It was adapted to blend Walloon and MACE information into a ssGBLUP for each trait separately. This method, hereafter called single-step genomic Bayesian procedure (ssGBayes), was used as the computational method for the Walloon genomic evaluation system. The equation associated with ssGBayes that blends genomic, Walloon and MACE information and that considers Walloon information contributing to MACE (hereafter called $ssGBayes_{W+M-Wc}$) can be written as follows:

$$\left(\mathbf{G}^{*-1}+\boldsymbol{\Lambda}_{\mathbf{W}}+\boldsymbol{\Lambda}_{\mathbf{M}}-\boldsymbol{\Lambda}_{\mathbf{Wc}}\right)\hat{\mathbf{a}}_{\mathbf{W+M-Wc}} = \mathbf{D}_{\mathbf{W}}^{-1}\hat{\mathbf{u}}_{\mathbf{W}}+\mathbf{D}_{\mathbf{M}}^{-1}\hat{\mathbf{u}}_{\mathbf{M}}-\mathbf{D}_{\mathbf{Wc}}^{-1}\hat{\mathbf{u}}_{\mathbf{Wc}} \quad (1)$$

where $\mathbf{G}^{*}=\mathbf{H}\sigma_a^2$ is the combined genomic-pedigree-based (co)variance matrix, $\mathbf{H}$ being the combined genomic-pedigree-based relationship matrix (e.g. Aguilar *et al.*, 2010; Christensen and Lund, 2010), σ_a^2 being the additive genetic variance; $\hat{\mathbf{a}}_{\mathbf{W+M-Wc}}$ is the unknown vector of Walloon GEBV based on Walloon and MACE information; $\hat{\mathbf{u}}_{\mathbf{W}}$ is the vector of EBV_W; $\hat{\mathbf{u}}_{\mathbf{M}}$ is the vector of EBV_M; $\hat{\mathbf{u}}_{\mathbf{Wc}}$ is the vector of EBV_{Wc}; Λ_i (i = W, M and Wc) is a matrix mimicking the least squares contributions to the left-hand side (LHS) of the hypothetical BLUP equations based on Walloon and MACE information and $\mathbf{D}_{\mathbf{i}}^{-1}$ is the inverse of the prediction error (co)variance matrix of $\hat{\mathbf{u}}_{\mathbf{i}}$. More details how $\boldsymbol{\Lambda}_{\mathbf{i}}$ and $\mathbf{D}_{\mathbf{i}}^{-1}$ were approximated are given later.

The inverse of the combined genomic-pedigree-based relationship matrix $\mathbf{H}$ was computed using the inverse of the additive pedigree relationship matrix $\mathbf{A}$ and a modified genomic relationship matrix, using the following formula:

$$\mathbf{H}^{-1}=\mathbf{A}^{-1}+\begin{bmatrix}0 & 0\\ 0 & \tau(\alpha\boldsymbol{\Gamma}+(1-\alpha)\mathbf{A}_{22})^{-1}-\omega\mathbf{A}_{22}^{-1}\end{bmatrix} \quad (2)$$

where $\mathbf{A}_{22}$ is the additive pedigree relationship matrix for the genotyped animals; $\boldsymbol{\Gamma}$ is the genomic relationship matrix; τ and ω are empirical weight factors (e.g. Tsuruta *et al.*, 2011), for the inverse of the modified genomic relationship matrix and the inverse of $\mathbf{A}_{22}$, respectively. The matrix $\boldsymbol{\Gamma}$ was computed following the first method of VanRaden (2008), that is:

$$\boldsymbol{\Gamma}=\frac{\mathbf{ZZ}'}{2\sum p_i(1-p_i)}, \quad (3)$$

where $\mathbf{Z}$ is the centred matrix of SNP markers alleles that each individual inherited and p_i is the minor allele frequency at locus i.

For compatibility between $\mathbf{A}_{22}$ and $\boldsymbol{\Gamma}$, diagonal and off-diagonal values of $\boldsymbol{\Gamma}$ were, respectively, centred on the average of diagonal and off-diagonal elements of $\mathbf{A}_{22}$. Based on suggestions from Tsuruta *et al.* (2011), τ and ω were respectively fixed to 1 and 0.70. The value for ω helped to reduce inflation of GEBV and compensated for pedigree incompleteness (Misztal *et al.*, 2013; Masuda *et al.*, 2016). Therefore it also compensated for associated potential incompatibilities between $\mathbf{A}$, and $\mathbf{A}_{22}$, that is due to differences in the consideration of inbreeding, the directly constructed $\mathbf{A}^{-1}$ matrix did not consider inbreeding in order to stay close to the routine polygenic evaluation, but $\mathbf{A}_{22}$, to be compatible with Γ, considered inbreeding.

The values of α reflect the partition of the genetic variance between the SNP and residual polygenic parts and were fixed to 0.60 for milk production traits and to 0.65 for the other traits. These values were based on additional studies (results not shown) verifying the effect of different α in terms of predictability of future (G)EBV from data limited to that collected 4 years earlier.

Regarding the vectors $\hat{\mathbf{u}}_{\mathbf{i}}$ for the three sources of information (i.e. W, M and Wc), it is worth noting that only some animals included in the pedigree were associated with known EBV and REL (hereafter called external animals). Therefore, for each i^{th} source of information, animals not associated with available EBV were called internal animals. The vector of EBV_i (i = W, M and Wc) for all animals included in the pedigree, $\hat{\mathbf{u}}_{\mathbf{i}}$, was estimated as (Vandenplas *et al.*, 2014):

$$\hat{\mathbf{u}}_{\mathbf{i}}=\begin{bmatrix}\mathbf{G}_{\mathbf{i(IE)}}\mathbf{G}_{\mathbf{i(EE)}}^{-1}\hat{\mathbf{u}}_{\mathbf{i(E)}}\\ \hat{\mathbf{u}}_{\mathbf{i(E)}}\end{bmatrix} \quad (4)$$

where $\mathbf{G}=\mathbf{A}\sigma_a^2$ and the subscript i refers to internal animals not associated with the i^{th} source of information, the subscript E refers to external animals associated with the ith

https://doi.org/10.1017/S1751731117002324 Published online by Cambridge University Press

source of information, $\hat{\mathbf{u}}_{\mathbf{i}(\mathbf{E})}$ is the vector of EBV for the i^{th} source of information associated with external animals and:

$$\mathbf{G}^{-1} = \begin{bmatrix} \mathbf{G}_{\mathbf{i}(\mathbf{EE})} & \mathbf{G}_{\mathbf{i}(\mathbf{EI})} \\ \mathbf{G}_{\mathbf{i}(\mathbf{IE})} & \mathbf{G}_{\mathbf{i}(\mathbf{II})} \end{bmatrix}^{-1} \quad (5)$$

is the inverse of the pedigree-based (co)variance matrix.

For the three sources of information (i.e. W, M and Wc), the matrix $\mathbf{D}_{\mathbf{i}}^{-1}$

was approximated as:

$$\mathbf{D}_{\mathbf{i}}^{-1} = \mathbf{G}^{-1} + \mathbf{\Lambda}_{\mathbf{i}}, \quad (6)$$

where $\mathbf{\Lambda}_{\mathbf{i}}$ is a diagonal variance matrix with one diagonal element per animal equal to RE_{ij}/σ_e^2 for $j = 1, 2, \ldots, J$ animals (Vandenplas *et al.*, 2014). The element σ_e^2 is the residual variance and the element RE_{ij} is the effective number of records, so-called record equivalents, for the jth animal. Record equivalents expressed the amount of contributions for an animal (Misztal and Wiggans, 1988). For both Walloon and MACE information, it is worth noting that double counting of contributions due to relationships among related animals could exist. Indeed, contributions for a given animal were initially associated to own information and to those coming from related animals. Therefore, the effective number of records RE_{ij} for the external animals requires to be cleaned for these latter contributions, and was estimated using a two-step algorithm (Vandenplas and Gengler, 2012). The two-step algorithm is a process that reverses the approach used by Misztal and Wiggans (1988) who concentrated the contribution from own records and relatives. The two-step also algorithm takes into account all relationships between animals associated with information and their ancestors. For the internal animals, RE_{ij} is equal to 0, because they have no own records in the considered external genetic evaluation.

Comparisons with computations without Multiple Across Country Evaluation results

Because a major feature of this optimized single-step genomic evaluation system, as implemented in the Walloon genomic evaluation system, is its ability to use MACE information, in comparison with standard ssGBLUP, the influence of the use of this information was tested. Therefore ssGBayes was run considering only Walloon information ($\text{ssGBayes}_{\text{W}}$) using the following, simplified, system of equations:

$$\left(\mathbf{G}^{*-1} + \mathbf{\Lambda}_{\mathbf{W}}\right) \hat{\mathbf{a}}_{\mathbf{W}} = \mathbf{D}_{\mathbf{W}}^{-1} \hat{\mathbf{u}}_{\mathbf{W}}, \quad (7)$$

where $\hat{\mathbf{a}}_{\mathbf{W}}$ is the vector of Walloon GEBV only based on Walloon information.

By comparing [6] and [7], it is important to highlight that under the hypothesis that no animals were genotyped ($\mathbf{G}^{*-1} = \mathbf{G}^{-1}$), we obtained $\hat{\mathbf{a}}_{\mathbf{W}} = \hat{\mathbf{u}}_{\mathbf{W}}$ which is the original set of Walloon EBV. Because the equations associated with $\text{ssGBayes}_{\text{W+M-Wc}}$ and $\text{ssGBayes}_{\text{W}}$, respectively, represented hypothetical mixed model equations, REL was computed using the standard formula:

$$\text{REL}_{\text{j}} = 1 - \text{PEV}_{\text{j}} / \sigma_{\text{j}}^2, \quad (8)$$

where σ_j^2 is the diagonal element of $\mathbf{G}^*$ for animal j and PEV_j is the prediction error variance obtained from the diagonal element of the inverted left-hand side for this animal j. Computations of (G)EBV and associated PEV by the two ssGBayes runs were executed using BLUPF90 (Misztal, 2013) modified to implement ssGBayes equations.

Enhanced breeding values validation

The GEBV validation was based on the GEBV test (Mäntysaari *et al.*, 2010) established by Interbull Centre (Sweden). The GEBV Interbull test compares the bias and the improvement of accuracy of a forward prediction when genomic information is used (i.e. a genomic evaluation), or is not used (i.e. a conventional genetic evaluation). The forward prediction uses EBV or GEBV predicted by using data available 4 years ago, that is we must exclude the last 4 years of data and re-run the analyses with reduced data, and using the same model that is being tested.

To test bias in national genomic evaluations, a regression model is used to estimate the slope (b_1), to compare with its expectation and test bias on GEBV. The statistical significance is tested by using a t test against H_0 (Confidence interval of 95%). Moreover, improvement in accuracy is performed by comparing the coefficient of determination (R^2) of the reduced genomic model and equivalent reduced conventional model (from 4 years ago) regressed to current performances. This validation accuracy is calculated as follows:

$$R^2_{validation} = \frac{R^2}{\overline{w}}, \quad (9)$$

where $\overline{w}$ are a function of average EDC of all tested bulls (Mäntysaari *et al.*, 2010).

The GEBV test consists basically in two tests, for bias and improvement from polygenic evaluation. To pass the GEBV test, first, for a given country by trait combination, the value of b_1 must be greater than the lower endpoint of the 95% confidence interval or biological equivalent. Moreover, the R^2 from the model including genomic information must be higher than the model including only parent average information (polygenic). Detailed information regarding the Interbull validation test for genomic evaluations (GEBV test) can be found in http://www.interbull.org/ib/cop_appendix8.

In order to check the influence of MACE results, we tested two types of genomic evaluations, with and without foreign information.

Results

Table 1 contains the number of (genotyped) animals with local and foreign information available for the six analysed traits (milk yield, fat yield, protein yield, SCS, stature and udder support). Approximately 3750 bulls evaluated by Interbull were also included in the computations. Walloon information for around 20% of these bulls contributed to the December 2015 MACE routine-run. Table 1 also indicates

https://doi.org/10.1017/S1751731117002324 Published online by Cambridge University Press

that at least 61% of the 4038 genotyped bulls and at least 13% of the 29 383 animals in the pedigree file had foreign information.

Table 2 gives details on the improvement of average REL when only genomic and local information was considered. For the genotyped bulls with low reliable official Walloon EBV ($REL_W < 0.50$), the genomic information allowed a substantial increase of 0.26 points for average REL of these bulls according to the studied traits.

The genomic information also increased average REL for bulls with already more accurate Walloon EBV. Indeed, the average REL was increased by 0.07 to 0.08 points for the bulls with REL_W between 0.50 and 0.75. Even for locally well proven bulls with $REL_W \geqslant 0.75$, the genomic information added 0.01 to the average REL, for example this represents 11 EDC for milk yield.

Considering the simultaneous combination of genomic and foreign information (i.e. $ssGBayes_{W+M-Wc}$), the increases of the averaged REL for each of the three mentioned categories of genotyped bulls (Table 2) were higher than those associated with $ssGBayes_W$. As expected, the highest increase of REL was observed for the bulls with the lowest REL_W. The use of $ssGBayes_{W+M-Wc}$ led to an increase of average REL between 0.14 points for udder support and 0.22 points for SCS compared with $ssGBayes_W$. The increase was lower for genotyped bulls with REL_W included in the range (0.50 to 0.75) with 0.14 to 0.20 additional points of REL, respectively for udder support and SCS. Even for the already locally well proven bulls (i.e. $REL_W \geqslant 0.75$), $ssGBayes_{W+M-Wc}$ still provided more reliable GEBV than $ssGBayes_W$. Additional points of REL ranged from 0.03 for milk production traits to 0.05 for SCS (Table 2).

Table 3 shows the enhancements of reliabilities for genotyped animals only associated with EBV_W, that is without foreign information, but sired by a genotyped bull with MACE results. These genotyped animals were Walloon cows as well as foreign or Walloon bulls to be tested. Again, similarly to Table 2, even if $ssGBayes_W$ allowed an increase of average REL from 0.05 to 0.31 additional points, $ssGBayes_{W+M-Wc}$ led to higher REL. For the six traits, $ssGBayes_{W+M-Wc}$ provided an average REL higher than 0.50 for the genotyped animals with a REL_W lower than 0.25.

Obtained REL from the Walloon genomic evaluations for young genotyped AI bulls, but having not yet daughters with phenotypic data in Wallonia or own MACE results, are reported in Table 4. If their sires were genotyped and had MACE results, the REL of GEBV of these animals were higher than the publishable threshold in more than 89% of the cases.

The values of bias and accuracy are presented in Table 5. All reported traits passed the GEBV test established by Interbull for bias and improvement from conventional evaluations by including MACE information. The b_1 values were greater than the lower endpoint of the 95% confidence interval, whereas R^2 values from genomic were higher than R^2 values using only local information. GEBV validation test results from April 2016 showed improvements in R^2 of 9.0% points for milk yield, 8.6% points for fat yield, 15.5% points for protein yield, 22.7% points for SCS, 24.5% points for stature and 4.4% points for udder support.

In addition, by excluding MACE information, all traits passed the test as well (Table 5). However, the R^2 values were lower compared with those by including MACE. For bias, the expected regression coefficients for the GEBV test were within confidence intervals based on two times

Table 1 *Used genetic parameters and amount of local and foreign information available for the genomic evaluation in April 2016*

		Number of animals with			Number of genotyped animals with		
Traits	Heritability	EBV_W	EBV_M	EBV_{Wc}	EBV_W	EBV_M	EBV_{Wc}
Milk yield	0.38	23 197	3792	771	4709	2505	433
Fat yield	0.43	23 197	3792	771	4709	2505	433
Protein yield	0.41	23 197	3792	771	4709	2505	433
SCS	0.13	23 198	3804	727	4709	2516	402
Stature	0.52	23 131	3660	760	4586	2429	435
Udder support	0.19	22 462	3646	734	4571	2430	435

EBV = estimated breeding value with subscripts: W = Walloon genetic evaluation, M = Multiple Across Country Evaluation (MACE), Wc = Walloon contribution to MACE evaluation; SCS = somatic cell score.

Table 2 *Average reliabilities (REL; SD in parentheses) associated to estimated (genomically enhanced) breeding values (EBV_W, $GEBV_W$ and $GEBV_{W+M-Wc}$) when genomic and local information is considered*

	REL_W[1] (0.00 to 0.50)				REL_W[1] (0.50 to 0.75)				REL_W[1] (0.75 to 0.99)			
Traits	*n*	RELW[1]	$GREL_W$[2]	$GREL_{W+M-Wc}$[3]	*n*	RELW[1]	$GREL_W$[2]	$GREL_{W+M-Wc}$[3]	*n*	REL_W[1]	$GREL_W$[2]	$GREL_{W+M-Wc}$[3]
Milk yield	3032	0.23 (0.10)	0.49 (0.08)	0.68 (0.13)	238	0.63 (0.07)	0.71 (0.05)	0.86 (0.05)	631	0.90 (0.07)	0.91 (0.06)	0.94 (0.04)
Fat yield	3051	0.23 (0.10)	0.49 (0.08)	0.67 (0.13)	249	0.63 (0.07)	0.71 (0.05)	0.86 (0.05)	601	0.90 (0.07)	0.91 (0.06)	0.94 (0.04)
Protein yield	3040	0.23 (0.10)	0.49 (0.07)	0.67 (0.12)	239	0.63 (0.07)	0.71 (0.05)	0.85 (0.05)	622	0.90 (0.07)	0.91 (0.06)	0.94 (0.04)
SCS	3107	0.23 (0.10)	0.49 (0.07)	0.71 (0.15)	273	0.64 (0.07)	0.71 (0.04)	0.91 (0.07)	521	0.90 (0.07)	0.91 (0.06)	0.96 (0.03)
Stature	2927	0.26 (0.09)	0.52 (0.06)	0.72 (0.14)	226	0.63 (0.07)	0.71 (0.05)	0.89 (0.07)	625	0.91 (0.07)	0.92 (0.06)	0.96 (0.03)
Udder support	3020	0.25 (0.10)	0.51 (0.06)	0.65 (0.11)	283	0.62 (0.07)	0.70 (0.05)	0.84 (0.07)	460	0.90 (0.08)	0.91 (0.07)	0.94 (0.04)

W = Walloon genetic evaluation; M = Multiple Across Country Evaluation (MACE); Wc = Walloon contribution to MACE evaluation; SCS = somatic cell score.
[1]REL obtained from Walloon polygenic evaluation.
[2]Genomic REL obtained from Walloon genomic evaluation using only EBV_W.
[3]Genomic REL obtained from Walloon genomic evaluation using EBV_W, EBV_M and EBV_{Wc}.

https://doi.org/10.1017/S1751731117002324 Published online by Cambridge University Press

Table 3 *Average reliabilities (REL; SD in parentheses) associated to estimated (genomically enhanced) breeding values (EBV_W, $GEBV_W$ and $GEBV_{W+M-Wc}$) for genotyped animals without Multiple Across Country Evaluation (MACE) (M) results and sired by genotyped bulls with MACE*

	REL_W^1 (0.00 to 0.25)				$RELw^1$ (0.25 to 0.50)				REL_w^1 (0.50 to 0.75)			
Traits	*n*	REL_W^1	$GREL_W^2$	$GREL_{W+M-Wc}^3$	*n*	REL_W^1	$GREL_W^2$	$GREL_{W+M-Wc}^3$	*n*	REL_W^1	$GREL_W^2$	$GREL_{W+M-Wc}^3$
Milk yield	623	0.17 (0.07)	0.47 (0.07)	0.54 (0.04)	533	0.35 (0.07)	0.57 (0.03)	0.59 (0.03)	287	0.55 (0.04)	0.65 (0.03)	0.66 (0.02)
Fat yield	637	0.17 (0.07)	0.47 (0.07)	0.54 (0.04)	611	0.37 (0.08)	0.56 (0.04)	0.59 (0.03)	197	0.54 (0.04)	0.64 (0.03)	0.66 (0.03)
Protein yield	628	0.17 (0.07)	0.47 (0.07)	0.54 (0.04)	577	0.36 (0.08)	0.56 (0.04)	0.59 (0.03)	238	0.54 (0.03)	0.65 (0.03)	0.66 (0.02)
SCS	672	0.16 (0.07)	0.47 (0.07)	0.57 (0.03)	703	0.38 (0.08)	0.57 (0.04)	0.62 (0.02)	73	0.56 (0.08)	0.67 (0.05)	0.69 (0.05)
Stature	472	0.18 (0.06)	0.49 (0.06)	0.57 (0.03)	674	0.33 (0.06)	0.57 (0.03)	0.61 (0.03)	265	0.70 (0.04)	0.75 (0.03)	0.76 (0.02)
Udder support	557	0.17 (0.06)	0.47 (0.05)	0.54 (0.04)	694	0.34 (0.07)	0.56 (0.03)	0.59 (0.03)	210	0.54 (0.04)	0.65 (0.03)	0.67 (0.03)

W = Walloon genetic evaluation; Wc = Walloon contribution to MACE evaluation; SCS = somatic cell score.
[1]REL obtained from Walloon polygenic evaluation.
[2]Genomic REL obtained from Walloon genomic evaluation using only EBV_W.
[3]Genomic REL obtained from Walloon genomic evaluation using EBV_W, EBV_M and EBV_{Wc}.

Table 4 *Average genomic reliabilities ($GREL_{W+M-Wc}^1$; SD in parentheses) associated with genomically enhanced estimated breeding values ($GEBV_{W+M-Wc}^1$) for genotyped bulls without own Multiple Across Country Evaluation (MACE) (M) results, without own daughters with phenotypic data in Wallonia, and sired by genotyped bulls with MACE*

Traits	Publication rule: GREL⩾	Number of bulls	Averaged $GREL_{W+M-Wc}$ (SD)	Number of publishable $GEBV_{W+M-Wc}$
Milk yield	0.50	923	0.56 (0.04)	842
Fat yield	0.50	923	0.55 (0.04)	839
Protein yield	0.50	923	0.55 (0.04)	829
SCS	0.45	916	0.58 (0.03)	914
Stature	0.50	917	0.59 (0.04)	874
Udder support	0.50	917	0.56 (0.05)	817

W = Walloon genetic evaluation; Wc = Walloon contribution to MACE evaluation; SCS = somatic cell score.
[1]$GEBV_{W+M-Wc}$ and $GREL_{W+M-Wc}$ are computed using results from Walloon genetic evaluation (W) and MACE (M) corrected for contribution from contribution of Walloon genetic evaluation to MACE (Wc).

Table 5 *Bias (SD), expected bias (Exp. b_1) and accuracy (R^2) of genomically enhanced estimated breeding value (GEBV) test by comparing current performances (genomic) with and without Multiple Across Country Evaluation (MACE) information and data available from 4 years ago (conventional evaluation)*

	Bias						R^2		
	Conventional		Genomic without MACE		Genomic with MACE				
Traits	b_1	Exp. b_1	b_1	Exp. b_1	b_1	Exp. b_1	Conventional (%)	Genomic without MACE (%)	Genomic with MACE (%)
Milk yield	1.31 (0.17)	1	1.23 (0.17)	0.91	1.21 (0.14)	0.92	46.8	47.2	55.8
Fat yield	1.50 (0.19)	1	1.41 (0.17)	0.95	1.07 (0.11)	0.96	51.5	53.8	60.1
Protein yield	1.40 (0.21)	1	1.34 (0.18)	0.96	1.14 (0.12)	0.97	44.6	48.7	60.1
SCS	0.77 (0.28)	1	0.83 (0.28)	0.93	1.14 (0.23)	0.95	18.0	20.5	40.7
Stature	0.89 (0.22)	1	0.89 (0.21)	0.91	0.88 (0.13)	0.94	25.7	44.6	50.2
Udder support	1.33 (0.24)	1	1.02 (0.17)	0.89	1.02 (0.17)	0.89	40.2	28.3	44.6

SCS = somatic cell score.

standard errors (please see http://www.interbull.org/ib/gebvtest for further information).

Discussion

The computational strategies based on the ssGBayes method used in the Walloon genomic evaluations allowed the incorporation of large amounts of additional information and increased the overall REL of the produced GEBV. Given the limited size of the genotyped population, results were in line with expectations or exceeded them (e.g. Wiggans *et al.*, 2011a). The method should therefore be internationally appealing for small and medium sized populations. Computed REL were based on the inverse of the coefficient

https://doi.org/10.1017/S1751731117002324 Published online by Cambridge University Press

matrices and are therefore not directly comparable with results reported in other studies which were based on validation R^2. We are aware that EDC are computed through approximation, but this is a general feature of most REL computations. What is really relevant is that computed REL are compatible inside and across different types of inputs. In our situation, all evaluations (Walloon, MACE) use compatible REL computation methods. It is only at output level where we compute exacted REL by using diagonal elements of inverted LHS.

The Walloon genomic evaluation system is constantly undergoing optimizations and enhancements. First, the proportion of genotyped cows available is expected to increase in the next years through scientific projects and by genotyping cows at the request of breeders and producers whose interest could potentially be boosted by increased alternative uses of genotypes (e.g. Yin and König, 2016). Second, genotyping of new key bulls will be optimized. For example, bulls with phenotypes through their daughters (e.g. maternal grandsires, brothers, half-brothers), related to selection candidates, will be genotyped, in order to increase the links between genotyped animals with phenotypes and the selection candidates. Therefore, an increase in REL, and of the proportion of publishable GEBV, is expected. As expected by definition, in general, average REL_W were about the same within classes by trait. Similar increases from REL_W to $GREL_W$ were observed for all traits. This can happen by chance as animals are not exactly the same between traits. Another reason is that the only additional information from REL_W to $GREL_W$ is the quite similar genomic relationship matrix between animals, even if the animals are not 100% the same. On the other hand, increases in REL were observed from $GREL_W$ to $GREL_{W+M-Wc}$. Here, additional information comes not only from genomic relationships, but also from MACE phenotypes (Misztal *et al.*, 2012). Although VanRaden *et al.* (2012) reported similar REL gains for traits such as milk, fat and protein yield, Wiggans *et al.* (2011b) observed gains between 25.6 and 35.0 points for the same traits. Lower gains in REL for udder support compared with other traits may happen due to the lower heritability of this trait. Similar smaller gains in udder support were reported by VanRaden *et al.* (2009) and Schenkel *et al.* (2009), with values of 0.14 and 0.10, respectively.

Third, another important methodological element is the optimization of the weight α used in the combination of the raw genomic relationships and the pedigree relationships. As explained earlier this coefficient was recognized as describing the partition of variance (Christensen and Lund, 2010). Several tests with our data but also research done by others (e.g. Liu *et al.*, 2011) showed that the used weight has a large influence and that the optimal proportion between raw genomic and pedigree relationships is critical. Exploiting this, now better understood, partitioning of total genetic variance into variances explained by SNP effects and by the residual polygenic effect should allow better estimates of α. Fourth, as was shown by Gengler *et al.* (2012) and Liu *et al.* (2014, 2016) future equivalent single-step genomic models could avoid completely the setting up of explicit genomic relationship matrices. This theory can be developed further to become a very attractive computing strategy inside our ssGBayes method. Fifth, the ssGBayes method used in the Walloon genomic evaluation system can also be set up as a multi-trait model allowing the genomic evaluation of only locally available traits (e.g. novel traits based on fine milk composition (Gengler *et al.*, 2016), methane emissions (Vanlierde *et al.*, 2015)) using external information from correlated traits (e.g. traits evaluated by Interbull). A multi-trait model would also allow avoiding harmonization of external information (e.g. expressed as indices or standardized estimates) and internal information (Vandenplas *et al.* 2015). This is especially desirable in the context of dairy cattle evaluations in which EBV are often not on the original phenotypic scale (e.g. EBV defined as average 305 day milk yield over first three lactations when the original records are test-day yields). Our current implementation avoids this issue by using EBV based on the reported and Interbull trait definitions, but, thus, approximating only local phenotypic information. However, the ssGBayes methodology is very flexible and could be set up to use directly local phenotypic data (e.g. Vandenplas *et al.*, 2017).

The GEBV test can be considered as quality assurance assessment for national genomic evaluations. In this study, GEBV test results showed the importance of the combination of local and foreign information. The fact that all traits passed the test, by using MACE or not, but associated to strong increases in R^2 when adding MACE, demonstrated the importance of both sources of information, genomic relationships and MACE. Finally, high differences in R^2 to conventional evaluations, for most of the traits, validated our national genomic model for the studied traits. Based on all results, the computed GEBV by using ssGBayes method should, therefore, be considered accurate and using also correctly local and MACE results.

Conclusions

Results presented in this study showed that the ssGBayes method is a good choice for the Walloon genomic evaluation system and an interesting option for other countries and evaluations. It was validated through the GEBV Test. This system in its use for the Walloon Holstein population appropriately integrates MACE results into ssGBLUP and recovers indirectly large amounts of external phenotypic information. Generated GEBV were associated to satisfactory REL given the size of the population. The whole genomic evaluation system will allow breeders to improve their selection decisions. The computed REL are taking into account genetic distances from selection candidates to genotyped animals with phenotypes. Extensions to other breeds (i.e. Dual Purpose Belgian Blue) are under development. Finally, the Walloon genomic evaluation system and the ssGBayes method in general have a large potential for future developments and enhancements. The flexibility of ssGBayes based systems can be used to adapt it to many other

https://doi.org/10.1017/S1751731117002324 Published online by Cambridge University Press

situations and implementations both at a Walloon level but also in other countries (e.g. Vandenplas *et al.*, 2017) or even international settings.

Acknowledgements

The authors acknowledge the Ministry of Agriculture of the Walloon Region of Belgium (Service Public de Wallonie, DGO3) for its financial support. Research leading to the implementation of this system was supported through several grants, that is research projects D31-1233 and D31-1274. Jérémie Vandenplas, as former research fellow, and Nicolas Gengler, as an honorary senior research associate, acknowledge the former and ongoing support of the National Fund for Scientific Research (F.R.S.-FNRS, Brussels, Belgium) for their positions and through several grants. The help of S. Tsuruta (University of Georgia, Athens, USA) and I. Aguilar (Instituto Nacional de Investigación Agropecuaria, Las Brujas, Urugay) concerning the BLUPF90 family of programs made available to us by I. Misztal (University of Georgia, Athens, USA), on which the developed computing programs are based, is acknowledged. Additional computational resources have been provided by the Plateau de Calcul et Modélisation Informatique of the University of Liège Gembloux Agro-Bio Tech and the Consortium des Équipements de Calcul Intensif funded by the National Fund for Scientific Research (F.R.S.-FNRS, Brussels, Belgium) under grant no. 2.5020.11.

References

Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S and Lawlor TJ 2010. Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. Journal of Dairy Science 93, 743–752.

Calus MPL, Vandenplas J, ten Napel J and Veerkamp RF 2016. Validation of simultaneous deregression of cow and bull breeding values and derivation of appropriate weights. Journal of Dairy Science 99, 6403–6419.

Christensen OF and Lund MS 2010. Genomic prediction when some animals are not genotyped. Genetics Selection Evolution 42, 2.

Gengler N, Nieuwhof G, Konstantinov K and Goddard M 2012. Alternative single-step type genomic prediction equations. Presented at the 63rd Annual Meeting European Federation of Animal Science, 27–31 August 2012, Bratislava, Slovakia. Retrieved on 11 January 2017, from http://hdl.handle.net/2268/138366.

Gengler N, Soyeurt H, Dehareng F, Bastin C, Colinet F, Hammami H, Vanrobays ML, Lainé A, Vanderick S, Grelet C, Vanlierde A, Froidmont E and Dardenne P 2016. Capitalizing on fine milk composition for breeding and management of dairy cows. Journal of Dairy Science 99, 4071–4079.

Henderson CR 1984. Applications of linear models in animal breeding, 2nd edition. University of Guelph, Guelph, ON, Canada.

Harris BL, Winkelman AM and Johnson DL 2012. Large-scale single-step genomic evaluation for milk production traits. Interbull Bulletin 46, 20–24.

Koivula M, Strandén I, Pösö J, Aamand GP and Mäntysaari EA 2012. Single step genomic evaluations for the Nordic Red Dairy cattle test day data. Interbull Bulletin 46, 115–120.

Legarra A, Christensen OF, Aguilar I and Misztal I 2014. Single step, a general approach for genomic selection. Livestock Science 166, 54–65.

Liu Z, Goddard ME, Reinhardt F and Reents R 2014. A single-step genomic model with direct estimation of marker effects. Journal of Dairy Science 97, 5833–5850.

Liu Z, Goddard ME, Hayes BF, Reinhardt F and Reents R 2016. Technical note: equivalent genomic models with a residual polygenic effect. Journal of Dairy Science 99, 2016–2025.

Liu Z, Seefried FR, Reinhardt F, Rensing S, Thaller G and Reents R 2011. Impacts of both reference population size and inclusion of a residual polygenic effect on the accuracy of genomic prediction. Genetics Selection Evolution 43, 19.

Masuda Y, Misztal I, Tsuruta S, Legarra A, Aguilar I, Lourenco DAL, Fragomeni BO and Lawlor TJ 2016. Implementation of genomic recursions in single-step genomic best linear unbiased predictor for US Holsteins with a large number of genotyped animals. Journal of Dairy Science 99, 1968–1974.

Mäntysaari E, Liu Z and VanRaden P 2010. Interbull validation test for genomic evaluations. Interbull Bulletin 41, 17–21.

Misztal I 2013. BLUPF90 family of programs. Retrieved on 11 January 2017, from http://nce.ads.uga.edu/wiki/doku.php.

Misztal I, Tsuruta S, Aguilar I, Legarra A, VanRaden PM and Lawlor TJ 2012. Methods to approximate reliabilities in single-step genomic evaluation. Journal of Dairy Science 96, 647–654.

Misztal I, Vitezica ZG, Legarra A, Aguilar I and Swan AA 2013. Unknown-parent groups in single-step genomic evaluation. Journal of Animal Breeding and Genetics 130, 252–258.

Misztal I and Wiggans GR 1988. Approximation of prediction error variance in large-scale animal models. Journal of Dairy Science 71 (suppl. 2), 27–32.

Přibyl J, Madsen P, Bauer J, Přibylová J, Šimečková M, Vostrý L and Zavadilová L 2013. Contribution of domestic production records, Interbull estimated breeding values, and single nucleotide polymorphism genetic markers to the single-step genomic evaluation of milk production. Journal of Dairy Science 96, 1865–1873.

Tsuruta S, Misztal I, Aguilar I and Lawlor TJ 2011. Multiple-traits genomic evaluation of linear type traits using genomic and phenotypic data in US Holsteins. Journal of Dairy Science 94, 4198–4204.

Schenkel F, Sargolzaei M, Kistemaker G, Jansen G, Sullivan P, Van Doormaal B, VanRaden P and Wiggans G 2009. Reliability of genomic evaluation of Holstein cattle in Canada. Interbull Bull 39, 51–58.

Vandenplas J, Colinet FG and Gengler N 2014. Unified method to integrate and blend several, potentially related, sources of information for genetic evaluation. Genetics Selection Evolution 46, 59.

Vandenplas J, Colinet FG, Glorieux G, Bertozzi C and Gengler N 2015. Integration of external estimated breeding values and associated reliabilities using correlations among traits and effects. Journal of Dairy Science 98, 9044–9050.

Vandenplas J and Gengler N 2012. Comparison and improvements of different Bayesian procedures to integrate external information into genetic evaluations. Journal of Dairy Science 95, 1513–1526.

Vandenplas J, Spehar M, Potocnik K, Gengler N and Gorjanc G 2017. National single-step genomic method that integrates multi-national genomic information. Journal of Dairy Science 100, 465–478.

Vanlierde A, Vanrobays ML, Dehareng F, Froidmont E, Soyeurt H, McParland SC, Lewis E, Deighton MH, Grandl F, Kreuzer M, Gredler B, Dardenne P and Gengler N 2015. Hot topic: Innovative lactation-stage-dependent prediction of methane emissions from milk mid-infrared spectra. Journal of Dairy Science 98, 5740–5747.

VanRaden PM 2008. Efficient methods to compute genomic predictions. Journal of Dairy Science 91, 4414–4423.

VanRaden PM 2012. Avoiding bias from genomic pre-selection in converting daughter information across countries. Interbull Bulletin 45, 29–33.

VanRaden PM, Olson K, Null D, Sargolzaei M, Winters M and Van Kaam JB 2012. Reliability increases from combining 50,000-and 777,000-marker genotypes from four countries. Interbull Bulletin 46, 75–79.

VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF and Schenkel FS 2009. Invited review: reliability of genomic predictions for North American Holstein bulls. Journal of Dairy Science 92, 16–24.

Wiggans GR, VanRaden PM and Cooper TA 2011a. The genomic evaluation system in the United States: past, present, future. Journal of Dairy Science 94, 3202–3211.

Wiggans G, VanRaden P and Cooper T 2011b. Dairy genomics in application. Paper presented at the Minnesota Dairy Health Conference 2011. May 17th - 19th, Bloomington, Minnesota, USA. Retrieved on 10 August 2017, from http://hdl.handle.net/11299/118925.

Yin T and König S 2016. Genomics for phenotype prediction and management purposes. Animal Frontiers 6, 65–72.

https://doi.org/10.1017/S1751731117002324 Published online by Cambridge University Press