Integration of Interbull’s multiple across-country evaluation approach breeding values into the multiple-trait single-step random regression test-day genetic evaluation for yield traits of Australian red breeds

Vinzent Boerner,1,2*, Thuy T. T. Nguyen,3 and Gert J. Nieuwhof3

1GHPC Consulting and Services Pty. Ltd., Armidale 2351, Australia
2Centre for Quantitative Genetics and Genomics, Aarhus University, Aarhus 8000, Denmark
3DataGene, AgriBio, 5 Ring Road, Bundoora VIC 3083, Australia

ABSTRACT

Interbull’s multiple across-country evaluation approach provides national breeding organizations with breeding values for internationally used bulls, which integrate performance data obtained in different breeding populations, environments, and production systems. However, breeding value-based selection decisions on domestic individuals born to foreign sires can only benefit from Interbull breeding values if they are integrated such that their information can contribute to the breeding values of all related domestic animals. For that purpose, several methods have been proposed which either model Interbull breeding values as prior information in a Bayesian approach, as additional pseudo data points, or as correlated traits, where these methods also differ in their software and parameterization requirements. Further, the complexity of integration also depends on the traits and genetic evaluation models. Especially random regression models require attention because of the dimensionality discrepancy between the number of Interbull breeding values and the number of modeled genetic effects. This paper presents the results from integrating 16,063 Interbull breeding values into the domestic single-step random regression test-day model for milk, fat, and protein yield for Australian Red dairy cattle breeds. Interbull breeding values were modeled as pseudo data points with data point-specific residual variances derived within animal across traits, ignoring relationships between integrated animals. Results suggest that the integration was successful with regard to alignment of Interbull breeding values with their domestic equivalent as well as with regard to the individual and population-wide increase in reliabilities. Depending on the relationship structure between integration candidates, further work is required to account for those relationships in a computationally feasible manner. Other traits with separate parity effects nationally could use a similar approach, even if not modeled with a test-day model.

Key words: cattle, dairy, breeding value, estimation, foreign information

INTRODUCTION

The worldwide dairy cattle population is dominated by only a few breeds, and the regular global exchange of genetic material between local populations has been common for decades. Consequentially, a bilateral and multi-lateral evaluation system has been developed which combine raw records across countries (Edel et al., 2011; Lidauer et al., 2015). However, combined across-country genetic evaluation is still an exception. Consequentially, Interbull (Upsalla, Sweden) is conducting international genetic evaluation for 7 dairy cattle breeds using de-regressed domestic breeding values as pseudo phenotypes in a multiple across-country evaluation approach (MACE; Schaeffer 2001). This allows the integration of all available information about an animal across countries, environments, and production systems, and to compare breeding values at the international level. However, genomic information is not used in MACE, but it is used in genomic MACE for young Holstein bulls and intergenomics for Brown Swiss bulls.

Breeding value-based selection decision on domestic animals born to foreign sires can only benefit from Interbull’s evaluation if the breeding values reported by Interbull are integrated into the national evaluation system. This is even more relevant for single-step genetic evaluation as the number of breeding values affected by the integration is expected to increase compared with genetic evaluation models based on pedigree information due to more nonzero relationships between animals. Several methods have been suggested for in-
integrating external breeding values into national genetic evaluation systems, which differ in their approach and consequentially in their requirement with regard to the used evaluation software and parameterization. Integrating external breeding values as prior information in a Bayesian approach (Quaas and Zhang 2001, 2006; Legarra et al., 2007; Vandenplas and Gengler 2012; Vandenplas et al., 2014; Guarini et al., 2019) requires manipulation of the right-hand side and a special co-variance matrix between animals of the mixed model equation system. Integration as additional pseudo data (Bonaiti and Boichard 1995; Pedersen et al., 1999; VanRaden and Tooker 2012; VanRaden et al., 2014; Pitkänen et al., 2020) requires derivation of 1 or several additional data points and manipulation of the residual co-variance matrix. Treating external breeding values as extra traits (Vandenplas et al., 2015) requires additional co-variances and increases the dimension of the evaluation equations system. Further, the integration of production traits like milk, fat, and protein yield is complicated due to dimensionality discrepancies between the Interbull model and the domestic multiple-trait random regression test-day models. To our knowledge, no entire description of a method for multiple-trait random regression test-day models. To our knowledge, no entire description of a method for integrating external breeding values into a single-step random regression test-day model of a domestic evaluation is available in the scientific literature. Therefore, in this paper, we present results from the integration of Interbull breeding values into the domestic single-step random regression test-day model genetic evaluation for Australian Red Dairy breeds.

METHODS

No live human or animal subjects were used, so this analysis did not require approval by an Institutional Animal Care and Use Committee or Institutional Review Board.

Data

The data set for genetic evaluation of production traits of Australian Red dairy breeds consisted of test-day observation on milk yield in liters, fat yield in kilograms, and protein yield in kilograms in the first 3 lactations with a total of 10,000,614 observations across all 9 traits (i.e., each lactation is a separate trait). The pedigree consisted of 73 phantom parents and 975,532 individuals, of which 8,191 were genotyped on various platforms. Genotypes underwent internal quality control and imputation procedures and, after being imputed to a common set of about 73,000 markers, were used to construct a genomic relationship matrix $\mathbf{G}$ using the observed allele frequencies for centering and the average observed allele frequency variance for scaling (VanRaden 2008).

The total number of bulls to be integrated was 16,063, where 466 bulls had Australian information provided to Interbull (A-bulls) and 15,597 bulls had no Australian information provided to Interbull (B-bulls), either because no information was available or the available information was not eligible due to Interbull reporting guidelines. Only 212 A-bulls and 116 B-bulls had genomic information in the Australian population.

Breeding values of A-bulls for integration were identified by a positive difference between the Interbull reliability and the local reliability of at least 0.01. If the difference was only achieved for a subset of A-bulls’s breeding values (e.g., for milk yield, but not for fat and protein yield), for that bull only, the subset was integrated.

All breeding values of B-bulls were integrated regardless of the Interbull reliabilities.

Evaluation Model

The multiple-trait random regression test-day evaluation model, used by DataGene, into which the Interbull breeding values were integrated, was

$$
\begin{align*}
\mathbf{y}_{l,t} & = \mathbf{X}_{l,t} \mathbf{b}_{l,t} + \mathbf{W}_{l,t} \mathbf{c}_{l,t} + \mathbf{e}_{l,t} \\
\mathbf{y}_{3,3} & = 0 \\
\mathbf{y}_{3,1} & = 0
\end{align*}
$$

where subscript $t = [1,2,3]$ indexes traits milk (1), fat (2), and protein (3) within lactation, indexed by subscript $l = [1,2,3]$, $f$ is a matrix valued function applying a normalized Legendre polynomial (Kirkpatrick et al. 1990) order $i$ to the argument, $\mathbf{y}_{l,t}$ is a vector of phenotypic observations of length $n_i$, $\mathbf{X}_{l,t}$, $\mathbf{1}_l$, $\mathbf{Z}_i$, and $\mathbf{W}_l$ are within-lactation trait ($l,t$) or within-lactation across-trait ($l$) design matrices linking respective effects to
observations, where \( \mathbf{1}_l \) is a vector of ones, \( \mathbf{K}_l \) is a diagonal matrix within lactation \( l \), \( \mathbf{K}_l, k \in (3, 9, 15) \) is an identity matrix of subscripted dimension, \( \mathbf{b}_{l,t} \) is a vector of fixed classification effects, \( c_{l,t} \) is a scalar fixed lactation curve effect of normalized Legendre polynomial order \( i \) on trait \( t \) in lactation \( l \); \( \mathbf{u}_{l,t,f} \) and \( \mathbf{q}_{l,t,f} \) are vectors of random direct genetic and random individual permanent environmental effects of normalized Legendre polynomial order \( i \) on trait \( t \) in lactation \( l \), and \( c_{l,t} \) is the respective residual. For the sake of a simplified description, the above notation assumes that all traits within lactation are equally recorded.

Columns in \( \mathbf{X}_l \) were constructed from herd test day, year-season, and age class at calving, and diagonal elements in \( \mathbf{K}_l \) were DIM.

The model also had vector \( \left( \mathbf{q}_{l,t,f}, \ldots, \mathbf{q}_{3,3,t,f} \right) \sim N \left( \mathbf{0}, \mathbf{I}_3 \otimes \mathbf{G} \right) \) and vector \( \left( \mathbf{u}_{l,t,f}, \ldots, \mathbf{u}_{3,3,t,f} \right) \sim N \left( \mathbf{0}, \mathbf{G} \otimes \mathbf{Q}_3 \right) \), where \( \mathbf{Q}_3 \) is co-variance matrix of the permanent environmental effects of dimension \( 27 \times 27 \) and \( \mathbf{G} \) is the genetic co-variance matrix of dimension \( 27 \times 27 \), and

\[
\mathbf{\Gamma} = \begin{bmatrix}
\mathbf{\Theta} & \mathbf{\Theta} \mathbf{Q}' \\
\mathbf{Q} \mathbf{\Theta} & \mathbf{Q} \mathbf{\Theta} \mathbf{Q}' + \mathbf{H}
\end{bmatrix},
\]

[2]

of which the inverse is

\[
\mathbf{\Gamma}^{-1} = \begin{bmatrix}
\mathbf{\Theta}^{-1} + \mathbf{Q} \mathbf{H}^{-1} \mathbf{Q}' & \mathbf{Q} \mathbf{H}^{-1} \\
\mathbf{H}^{-1} \mathbf{Q} & \mathbf{H}^{-1}
\end{bmatrix},
\]

[3]

where \( \mathbf{\Theta} \) is a diagonal matrix with elements equal to 0.62, implying that genetic groups are fitted as random with a variance of 0.62. \( \mathbf{Q} \) is the pedigree-derived genetic group regression matrix, and \( \mathbf{H} \) is the single-step relationship matrix constructed using a weighted genomic relationship matrix \( \left( \mathbf{G}_w = 0.8 \times \mathbf{G} + 0.2 \times \mathbf{A}_{22} \right) \) (Christensen and Lund 2010).

**Integration Methodology**

Interbull breeding values were modeled as additional data points with the data point specific residual variance as the single tuning parameter.

The integration approach must account for the fact that the model provides 27 factor level effects (random regression coefficients) at the polynomial multi-trait level, whereas the information sent to and received from Interbull provides cumulative breeding values within trait and subsequently averaged across lactations. That is, although the domestic evaluation model provides 27 factor level effects (random regression coefficients) and 9 reliabilities (1 reliability per trait) per animal, only 3 breeding values and reliabilities per animal (lactation averages within milk, fat and protein) are sent to and received from Interbull. Note that the number of reliabilities results from the used software and approximation algorithm.

For aligning within-animal dimensions, Interbull breeding values \( \mathbf{u}_{\text{MACE}} \) and reliabilities \( \mathbf{r}_{\text{MACE}} \) and sent breeding values \( \mathbf{u}_{\text{SENT}} \) and reliabilities \( \mathbf{r}_{\text{SENT}} \) were mapped from \( \mathbf{R}_3 \) to \( \mathbf{R}_9 \) using a block-diagonal matrix \( \mathbf{K} \) of dimension \( 9 \times 3 \) containing \( 3 \times 1 \) column blocks of ones. Further, \( \mathbf{\Sigma}_a \) was mapped from \( \mathbf{R}_{27} \) to \( \mathbf{R}_9 \) using a \( 9 \times 27 \) block-diagonal matrix \( \mathbf{L} = \mathbf{t} \otimes \mathbf{I}_9 \) with \( \mathbf{t} = \left( \mathbf{t}_1 \right) \left( \mathbf{d} \right), \mathbf{1}_f \left( \mathbf{d} \right), \mathbf{1}_f \left( \mathbf{d} \right) \) \( \mathbf{1}_f \left( \mathbf{d} \right) \), where \( \mathbf{I}_9 \) is an identity matrix of dimensions \( 9 \times 9 \), \( \mathbf{t} \) is a vector of length 3, \( \mathbf{d} \) is a vector containing a consecutive sequence from 6 to 305, \( \mathbf{1}_f \) is a vector-valued function applying the Legendre polynomial order \( i \) to the argument, and \( \mathbf{1}_l \) is a vector of ones. Therefore, \( \mathbf{u}_m = \mathbf{Ku} \) and \( r_{m} = \mathbf{Kr} \), where \( \mathbf{1} \) represents either MACE or SENT and \( \mathbf{\Sigma}_{a,m} = \mathbf{L} \mathbf{\Sigma}_a \mathbf{L}' \), where \( \mathbf{\Sigma}_{a,m} \) is of dimension \( 9 \times 9 \). In case only a subset of breeding values within animal met the integration requirement, the above dimensions were adjusted accordingly.

Within animal, a vector of prediction error variances was calculated by \( \mathbf{p} = \text{diag}(1 - r_m) \), where \( \text{diag}(\cdot) \) denotes a vector of diagonal elements of a squared matrix. The system for deriving a diagonal matrix \( \mathbf{D} \) of inverse residual variances across traits within animal was

\[
\text{diag} \left( \mathbf{D} \left( \mathbf{\Sigma}_{a,m} \right)^{-1} \right) = \mathbf{p},
\]

[4]

where \( \mathbf{D} \) was obtained directly only if the dimension of the system was 1. Otherwise, \( \mathbf{D} \) was established iteratively (Vandenplas and Gengler 2012; Barwick et al. 2013). Within-animal vectors of pseudo-phenotypes (“deregressed proofs”) \( \mathbf{y}^* \) were calculated for A-bulls as

\[
\mathbf{D}_r = \mathbf{D}_{\text{MACE}} - \mathbf{D}_{\text{SENT}}
\]

\[
\mathbf{a}_{\text{MACE}} = \left( \mathbf{D}_{\text{MACE}} + \left( \mathbf{\Sigma}_{a,m} \right)^{-1} \right)^{-1} \mathbf{u}_{\text{MACE},m}
\]

\[
\mathbf{a}_{\text{SENT}} = \left( \mathbf{D}_{\text{SENT}} + \left( \mathbf{\Sigma}_{a,m} \right)^{-1} \right)^{-1} \mathbf{u}_{\text{SENT},m}
\]

\[
\mathbf{y}^* = \mathbf{D}_r^{-1} \left( \mathbf{a}_{\text{MACE}} - \mathbf{a}_{\text{SENT}} \right)
\]

[5]

(Pitkänen et al., 2020), and for B-bulls as
Boerner et al.: INTERBULL BREEDING VALUES IN RANDOM REGRESSION TEST MODEL

\[ y_i^* = D_{MACE}^{-1} \mathbf{a}_{MACE} \].

A maximum of 9 pseudo-phenotypes \( y_i^* \) of animal \( i \) were added to the evaluation using the following model:

\[
\begin{pmatrix}
  y_{i,t,t}^* \\
y_{i,3,3}^*
\end{pmatrix}
= \begin{pmatrix}
  b_{i,t,1} \\
  b_{i,3,3}
\end{pmatrix}
+ \begin{pmatrix}
  (1'f_0(d), 1'f_1(d), 1'f_2(d)) \otimes L_0 \\
  u_{i,3,3}
\end{pmatrix}
+ \begin{pmatrix}
  e_{i,t,t} \\
e_{i,3,3}
\end{pmatrix},
\]

with \( e_{i,t,t} \sim N(0, D_i^{-1}) \), where \( d \) is a vector containing a sequence from 6 to 305, \( D_i^{-1} \) for animal \( i \) is either \( D_{A-bulls}^{-1} \) or \( D_{B-bulls}^{-1} \) for B-bulls. In other words, \( D_i^{-1} \) is the residual co-variance matrix of the respective pseudo-phenotypes derived as a function of the prediction error variance of the respective domestic and Interbull breeding values. Note that \( [b_{i,t,t}, b_{i,3,3}] \) are general means modeled within trait \( t \) in lactation \( l \) and bull group \( b \).

**Evaluating Results**

For evaluating results, genetic effects were mapped from \( \mathcal{R}_{27} \) to \( \mathcal{R}_3 \) and subsequently averaged across lactations within trait. The reliability approximation software provided reliabilities only in \( \mathcal{R}_3 \), which were averaged across lactations within trait. Single-step reliabilities were approximated with a combination of the method of Jamrozik et al. (2000) and Misztal et al. (2013; I. Strandén, Natural Resources Institute (LUKE), Jokioinen, Finland; personal communication).

For A-bulls, results were evaluated by regressing breeding values and reliabilities received from Interbull on breeding values and reliabilities from pre- and post-integration evaluation, and by Pearson correlations between sets. Because only breeding values with a positive reliability difference were integrated, it was expected that a successful integration produces breeding values and reliabilities which are highly correlated with those supplied by Interbull.

For B-bulls, the same evaluation parameters were calculated, but post-integration breeding values were expected to be a weighted average of domestic and Interbull information and post-integration reliabilities to be greater than the domestic or Interbull reliabilities. Therefore, Pearson correlations between post-integration breeding values and reliabilities and their Interbull equivalents were expected to be lower than that of A-bulls.

For the entire evaluation population, it was expected that post-integration reliabilities are either equal to their pre-integration values or higher, where the magnitude is a function of the relation of the respective animals to the integrated bulls.

**Software**

Data preparation and derivation of \( D \) was done using the R computing environment (version 4; R Development Core Team, 2011). The genetic evaluation model was solved and reliabilities were deterministically calculated using MiX99 and Apax, respectively (Lidauer et al. 2017).

**RESULTS**

Results for protein yield are summarized in Figures 1, 2, 3, and 4. Result trends for fat and milk yield were very similar and are, therefore, not separately shown. For A-bulls, correlations between Interbull breeding values and domestic evaluation breeding values increased from 0.94, 0.93, and 0.94 pre-integration to 0.98, 0.98, and 0.99 post-integration for milk, fat, and protein yield, respectively. The regression intercept changed from \(-328\), \(-9.16\), and \(-8.94\) to \(-334\), \(-9.72\), and \(-9.08\), and the regression slope from \(1.03\), \(1.08\), and \(1.1\) to \(0.974\), \(1.01\), and \(0.998\) for milk, fat, and protein yield, respectively. For B-bulls, correlations between Interbull breeding values and domestic evaluation breeding values increased from \(\leq 0.01\) pre-integration to \(\geq 0.98\) for all the 3 traits.

Results were similar for the reliabilities, but correlations and regression parameters, due to the nature of reliabilities, did not provide the same goodness-of-fit information as for the breeding values. However, post-integration reliabilities of all bulls were at least as high as their Interbull reliabilities. For A-bulls, the post-integration reliabilities aligned very well with the Interbull reliabilities. Further, for B-bulls with breeding values with pre-integration reliabilities of zero, the procedure returned post-integration reliabilities similar to the Interbull reliabilities.

For both A-bulls and B-bulls, the results were not affected by the genotyping status of the integration animals.

For the entire evaluation population, excluding the integrated bulls, about 40% of the reliabilities of all 3 traits were not affected by the integration, and only 5% of the reliabilities increased by more than 0.1 (see Fig-
Contrarily, every single breeding value changed as a result of the integration with 27, 29, and 29% of the breeding values for milk, fat, and protein yield, respectively, having changed by more than 10% when compared with the pre-integration breeding values. However, for all 3 traits, only about 15% of the breeding values changed by more than 25% when compared with their pre-integration value. The relatively small number of animals with substantial changes in breeding values is also reflected by the re-ranking of animals.

Figure 1. Breeding values (upper row) and reliabilities (lower row) for kilograms of protein per 305-d lactation pre-integration (left) and post-integration (right) for bulls with Australian information (A-bulls) in the Interbull evaluation. The numbers of genotyped and non-genotyped animals were 212 and 254, respectively. MACE = Multiple Across-Country Evaluation.
(see Figure 4). Although the pre- and post-integration rank was equal for only about 100 animals, only for 4% of all animals the difference between the pre- and post-integration rank was greater than 100,000, resulting in a Spearman’s rank correlation between pre- and post-integration ranks of 0.99.

**DISCUSSION**

Results shown in this study align very well with the formulated expectations and, therefore, the integration of Interbull breeding values into the domestic single-step random regression test-day evaluation is regarded
as successful. The nonzero intercept when regressing Interbull breeding values on domestic breeding values, pre- as well as post-integration, is a result of the average superiority of the integrated bulls within the Australian population as all base population differences were accounted for by fitting bull group specific fixed effects. The finding that all breeding values were affected by the integration, but more than 50% of all reliabilities remained unaffected, is a feature of the reliability approximation (Jamrozik et al., 2000). Although the added data points changed the information content for all modeled factor levels, the accuracy approximation can only account for information changes within groups of closely related animals (e.g., full-sib and half-sib families.)

Integration of Interbull breeding values as pseudo data points was considered the only options as, to our knowledge, the used software did not allow for equation system manipulation required for a Bayesian approach. Further, integration of Interbull breeding values as additional traits was not considered given the anticipated effects on solver runtime and equation system dimensions, and requirements for pipeline manipulation and variance component estimation.

![Figure 3](image1.png)

Figure 3. Histograms of differences between pre- and post-integration reliabilities (left) and breeding values (right) for protein yield in kilograms of the entire evaluation population without the integrated bulls. Min = minimum change in reliability or breeding value; max = maximum change in reliability or breeding value; no change = proportion of animals without any change in reliabilities; >1% change, >5% change, >10% change, and >25% change = proportion of animals with more than 1, 5, 10, and 25% change in reliability or breeding value, respectively. Proportional changes were calculated using the pre-integration results as a reference.

![Figure 4](image2.png)

Figure 4. Plot of pre-integration ranks versus post-integration ranks of breeding values for protein yield in kilograms of the entire evaluation population without the integrated bulls. The diagram shows a 2-dimensional heatmap of a 3-dimensional histogram with 50 bins.
When deriving matrix $D$, the fixed effect structure associated with the pseudo-data points was not accounted for. This is not problematic as long as the number of bulls within a fixed effect level is sufficiently large to ensure that the detrimental effect of fixed effects on reliabilities is limited. The current integration approach ignores the relationships between integrated bulls when deriving individual-specific residual variances. Therefore, the methodology used is the same regardless of the target equation system using genomic information or not. Empirically, ignoring the relationships among integration animals does not pose a problem if the reliabilities are high and integration candidates are not closely related, where the latter condition may be unrealistic given the level of global genetic exchange between local dairy populations.

Accounting for relationships among integration animals poses a computational problem. For the approach presented, the procedure for deriving $D$ of system

$$\text{diag} \left( D + \Sigma_a^{-1} \otimes \Gamma_{\pi}^{-1} \right)^{-1} \equiv p \text{ required up to tens of thousands of iterations (inversions), which is in sharp contrast to Calus et al. (2016), who used only 5 iterations. However, Vandenplas and Gengler (2012) and Calus et al. (2016) derived } D \text{ within traits across animals (i.e., } \Sigma_a^{-1} \text{ was a scalar and } \Gamma_{\pi}^{-1} \text{ was a matrix), whereas we derived } D \text{ across traits within animals (i.e., } \Sigma_a^{-1} \text{ was a matrix and } \Gamma_{\pi}^{-1} \text{ was 1). This is an important difference, as actual number of necessary inversions for convergence depends on the sensitivity of the } \text{diag} \left( D + \Sigma_a^{-1} \otimes \Gamma_{\pi}^{-1} \right)^{-1} \text{ to changes of } D, \text{ where the sensitivity is affected by starting values, reliability differences, and the spectral radius of } \Sigma_a^{-1} \otimes \Gamma_{\pi}^{-1}. \text{ The latter is entirely dependent on the choice of } \Gamma_{\pi}^{-1} (\text{e.g., } I, A^{-1}, H^{-1}) \text{ and the properties of } \Sigma_a. \text{ Therefore, any results regarding convergence are data set and implementation specific. For our data set and approach, where } \Sigma_a = \Sigma_{a,m} \text{ was a derivative of the genetic co-variance matrix of a random regression test-day model and } \Gamma_{\pi}^{-1} \text{ was 1, we found that updating of } D, \text{ as proposed by Vandenplas and Gengler (2012) and Calus et al. (2016), was too coarse and the algorithm never converged. Instead, we updated } D \text{ with empirically determined small values (Barwick et al. 2013), which required up to tens of thousands of inversions to achieve convergence. Therefore, extending the approach to a system across traits and animals will certainly affect inversion time and, for a given data set, the required number of iterations as the spectral radius of } \Sigma_a^{-1} \otimes \Gamma_{\pi}^{-1} \text{ will increase if } \Gamma_{\pi}^{-1} \text{ deviates from } I. \text{ As an example, for the data set presented deriving } D \text{ jointly for the entire integration population would have required inverting a system of rank 144,000 probably several hundred thousand times.}

An intermediate approach is to approximate some matrix $J = C_{a_i} - C_{a_i} \{C_{a_i} = a_i\} C_{a_i}$, where $C$ is the mixed model coefficient matrix and $a_i$ is a vector indexing the block in $C$ related to animal $i$, which is similar to approximating prediction error variances when calculating reliabilities. Subsequently, $J$ is used instead of $\Sigma_a^{-1} \otimes \Gamma_{\pi}^{-1}$ in $\text{diag} \left( D + J \right)^{-1} \equiv p$. [T. Pitkänen, Natural Resources Institute (LUKE), Jokioinen, Finland; personal communication]. Depending on the actual implementation, the approach either ignores that all other bulls have fractional observations when deriving $J$ for bull $i$, or it must be run iteratively until convergence. Further, it might be computationally prohibitive to apply such procedure to a single-step system. In any case, computation time may be crucial as imputing data points is done within the genetic evaluation pipeline before the actual solving step, and the frequency of evaluation and available hardware resources may set a tight time frame.

The presented approach assumes that genetic correlations between Interbull and domestic breeding values is 1.0, where for B-bulls a deviation from 1.0 can be accounted for by using conditional expectations and associated reliabilities instead of Interbull breeding values and reliabilities. However, for A-bulls, the approach may not be straightforward as an imperfect correlation is a function of perfect and imperfect correlations between the domestic and international parts of the Interbull breeding value and its domestic equivalent, respectively.

**CONCLUSIONS**

The outlined methodology and presented results demonstrate a possible approach for integrating Interbull breeding values into the domestic single-step random regression test-day model where the pseudo data approach provides good alignment between Interbull and post-integration breeding values and is applicable to software which does not allow for manipulating the right-hand side or a special co-variance structure between individuals of the mixed model equation system. Further work may be required to allow for co-variance between integration candidates.

**ACKNOWLEDGMENTS**

This work was funded by DataGene (AgriBio, Bundoora, Australia). We thank Ismo Strandén and Timo
Pitkänen for discussions regarding methodology and MiX90 implementation. The initial research, which enabled single-step genetic evaluation in the Australian Red dairy breeds, was done by Irene van den Berg, Iona MacLeod, Mekonnen Haile-Mariam, and Tuan Nguyen at DairyBio (AgriBio, La Trobe University, Bundoora, Australia). Their continued support throughout the project is highly appreciated. We thank Christy Vander Jagt (Agriculture Victoria Research, AgriBio, Bundoora, Australia) for proofreading the manuscript. The authors have not stated any conflicts of interest.

REFERENCES

Barwick, S. A., B. T. Tier, A. A. Swan, and A. L. Henzell. 2013. Estimation of accuracies and expected genetic change from selection for selection indexes that use multiple-trait predictions of breeding values. J. Anim. Breed. Genet. 130:341–348. https://doi.org/10.1111/j.12011.

Bonaiti, B., and D. Boîchtard. 1995. Accounting for foreign information in genetic evaluation. Interbull Bull. 11.

Calus, M. P. L., J. Vandenplas, J. ten Napel, and R. F. Veerkamp. 2016. Validation of simultaneous deregression of cow and bull breeding values and derivation of appropriate weights. J. Dairy Sci. 99:6403–6419. https://doi.org/10.3168/jds.2016-11028.

Christensen, O. F., and M. S. Lund. 2010. Genomic prediction when some animals are not genotyped. Genet. Sel. Evol. 42:2. https://doi.org/10.1186/1297-9686-42-2.

Edel, C., H. Hamann, S. Neuner, R. Emmelting, and K. Götz. 2011. The German-Austrian genomic evaluation system for Fleckvieh (Siemmental) cattle. Interbull Bull. 44:152–156.

Guarini, A. R., D. A. L. Lourenco, L. F. Brito, M. Sargolzaei, C. F. Baes, F. Miglior, S. Tsuruta, I. Mislal, and F. S. Schenkel. 2019. Use of a single-step approach for integrating foreign information into national genomic evaluation in Holstein cattle. J. Dairy Sci. 102:8175–8183. https://doi.org/10.3168/jds.2018-15819.

Jamrozik, J., L. R. Schaeffer, and G. B. Jansen. 2000. Approximate accuracies of prediction from random regression models. Livest. Prod. Sci. 66:85–92. https://doi.org/10.1016/S0301-6226(00)00158

Kirkpatrick, M., D. Lofsvold, and M. Bulmer. 1990. Analysis of the inheritance, selection and evolution of growth trajectories. Genetics 124:979–993. https://doi.org/10.1093/genetics/124.4.979.

Legarra, A., J. K. Bertrand, T. Strabel, R. L. Sapp, J. P. Sanchez, and I. Mislal. 2007. Multi-breed genomic evaluation in a Gelbvieh population. J. Anim. Breed. Genet. 124:286–295. https://doi.org/10.1111/j.1439-0388.2007.00671.x.

Lidauer, M., K. Mattilainen, E. Mäntysaari, T. Pitkänen, M. Tassinen, and I. Strandén. 2017. MiX99: Technical Reference Guide for MiX99 Solver. Natural Resources Institute Finland.

Lidauer, M. H., J. Pöös, J. Pedersen, J. Lassen, P. Madsen, E. A. Mäntysaari, U. S. Nielsen, J.-A. Eriksson, K. Johansson, T. Pitkänen, I. Strandén, and G. P. Aaamnd. 2015. Across-country test-day model evaluations for Holstein, Nordic Red Cattle, and Jersey. J. Dairy Sci. 98:1296–1309. https://doi.org/10.3168/jds.2014-8307.

Mislal, I., S. Tsuruta, I. Aguilar, A. Legarra, P. VanRaden, and T. Lawlor. 2013. Methods to approximate reliabilities in single-step genomie evaluation. J. Dairy Sci. 96:647–654. https://doi.org/10.3168/jds.2012-5656.

Pedersen, E. G., J. Pedersen, U. S. Nielsen, and P. Madsen. 1999. Experiences of blending foreign information in the national genetic evaluation. Interbull Bull. 22:61–63.

Pitkänen, T. J., M. Koivula, I. Strandén, G. P. Aamnd, and E. A. Mäntysaari. 2020. Integration of MACE breeding values into domestic multi-trait test-day model evaluations. Page 582 in Book of Abstracts of the 71th Annual Meeting of the European Association for Animal Production, Porto, Portugal. Wageningen Academic Publishers.

Quaas, R. L., and Z. Zhang. 2006. Multiple-breed genomic evaluation in the US beef cattle context: Methodology. Pages 12–24 in Proc. 8th World Congress Applied to Livestock Production, Belo Horizonte, Brazil.

Quaas, R. L., and Z. W. Zhang. 2001. Incorporating external information in multibreed genomic evaluation. J. Anim. Sci. 79:S342.

R Development Core Team. 2011. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. http://www.R-project.org/.

Schaeffer, L. R. 2001. Multiple trait international bull comparisons. Livest. Prod. Sci. 69:145–153. https://doi.org/10.1016/S0301-6226(00)00255-4.

Vandenplas, J., F. G. Colinet, and N. Gengler. 2014. Unified method to integrate and blend several, potentially related, sources of information for genetic evaluation. Genet. Sel. Evol. 46:59. https://doi.org/10.1186/s12711-014-0059-3.

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

Vandenplas, J., and N. Gengler. 2012. Comparison and improvements of different Bayesian procedures to integrate external information into genetic evaluations. J. Dairy Sci. 95:1513–1526. https://doi.org/10.3168/jds.2011-4322.

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414–4423. https://doi.org/10.3168/jds.2007-0980.

VanRaden, P. M., and M. E. Tooker. 2012. Methods to include foreign information in national evaluations. J. Dairy Sci. 95:S446.

Vincent Boerner. https://orcid.org/0000-0001-8005-9253

Thuy T. Nguyen. https://orcid.org/0000-0002-9880-6859

ORCIDS

Journal of Dairy Science Vol. 106 No. 2, 2023