

# Integration of Interbull MACE breeding values into the multiple-trait single-step random regression test-day genetic evaluation for yield traits of Australian red breeds

V. Boerner<sup>1,2 \*</sup>, T. Nguyen<sup>3</sup> and G. Nieuwhof<sup>3</sup>

1: GHPC CONSULTING AND SERVICES PTY. LTD., Armidale 2351, Australia

2: Centre for Quantitative Genetics and Genomics, Aarhus University, Denmark

3: DataGene, AgriBio, 5 Ring Road, Bundoora VIC 3083, Australia

\*vinzent.boerner@gmx.de

---

## Abstract

Interbull's multiple across country evaluation provides national breeding organisations with breeding values for internationally used bulls which must be integrated into the domestic evaluation. For that purpose several methods have been proposed which either model Interbull breeding values as prior information in a Bayesian approach, or as additional pseudo data points. Single-step random regression test-day models add complexity to the integration because of the dimensionality discrepancy between the number of Interbull breeding values and the number of modelled 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. 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.

---

## Introduction

National breeding programs can only benefit from Interbull's multiple across country 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. External breeding values can be integrated as prior information in a Bayesian approach (Vandenplas et al., 2014) which requires manipulation of the mixed model equation system's right-hand side and a special covariance matrix between animals, or as additional pseudo data (Pitkänen et al., 2020) which requires derivation of one or several additional data points and manipulation of the residual co-variances matrix. The integration

for production traits like milk, fat and protein yield is complicated due dimensionality discrepancies between the Interbull model and the domestic multiple-trait random regression test-day models. In this paper we present results from the integration of Interbull breeding values for milk fat and protein yield into the domestic single-step random regression test-day model genetic evaluation for Australian Red Dairy breeds.

## Material and Methods

The data set for genetic evaluation of production traits of Australian red dairy breeds consisted of test-day observation on litre milk, kilogram fat and kilogram protein in the first three lactations with a total of 10,000,614 observations across all nine traits. 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. The total number of bulls to be integrated was 16,063 where 466 bulls had Australian information provided to Interbull (**Abulls**), and 15,597 bulls may have had Australian information but it was not provided to Interbull or did not have Australian information at all (**Bbulls**). Breeding values of **Abulls** 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 an **Abulls**'s breeding values, e.g. for milk yield, but not for fat and protein yield, only the subset was integrated. All breeding values of **Bbulls** were integrated regardless of the Interbull reliabilities. The genetic evaluation model into which the Interbull breeding values were integrated modelled phenotypic observations as linear functions of fixed herd-year, season, age and lactation curve effects, where the latter used Legendre polynomials of order 1 to 4, of correlated direct random genetic effects using Legendre polynomials of order 0 to 2, of correlated direct random permanent environment effects of the animal using Legendre polynomials of order 0 to 2, and correlated random residuals. Vectors of direct genetic effects

$$\sim N(\mathbf{0}, \mathbf{\Gamma} \otimes \Sigma_a)$$

and vectors of permanent environmental effects

$$\sim N(\mathbf{0}, \mathbf{I} \otimes \Sigma_q),$$

where  $\Sigma_q$  is co-variance matrix of the permanent environmental effects of dimension 27×27 and  $\Sigma_a$  is the genetic co-variance matrix of dimension 27×27, and

$$\text{vec}(\mathbf{\Gamma}) = [\mathbf{0}, \mathbf{Q}\mathbf{0}, \mathbf{0}\mathbf{Q}', \mathbf{Q}\mathbf{0}\mathbf{Q}' + \mathbf{H}]$$

where  $\mathbf{0}$  is a diagonal matrix with elements equal to 0.62,  $\mathbf{Q}$  is the pedigree derived genetic group regression matrix, and  $\mathbf{H}$  is the single-step relationship matrix (Christensen and Lund, 2010) constructed using a polygenic weight of 0.2. Interbull breeding values were modelled as additional data points with the data point specific residual variance as the single tuning parameter. For aligning within animal dimensions Interbull breeding values  $u_{MACE}$  and reliabilities  $r_{MACE}$ , and domestic breeding values  $u_{DOM}$  and reliabilities  $r_{DOM}$  provided to Interbull were mapped from  $\mathfrak{R}_3$  to  $\mathfrak{R}_9$  using a block-diagonal matrix  $K$  of dimension 9×3 containing 3×1 column blocks of ones. Further  $\Sigma_a$  was mapped from  $\mathfrak{R}_{27}$  to  $\mathfrak{R}_9$  using a 9×27 block diagonal matrix  $L = \mathbf{t} \otimes \mathbf{I}_9$  with

$$\mathbf{t} = [\mathbf{1}'\mathbf{f}_0(\mathbf{d}), \mathbf{1}'\mathbf{f}_1(\mathbf{d}), \mathbf{1}'\mathbf{f}_2(\mathbf{d})]$$

where  $\mathbf{I}_9$  is an identity matrix of dimensions 9×9,  $\mathbf{t}$  is a vector of length 3,  $\mathbf{d}$  is a vector containing a consecutive sequence from 6 to 305,  $\mathbf{f}_i$  is a vector-valued function applying the Legendre polynomial order  $i$  to the argument, and  $\mathbf{1}$  is a vector of ones.

Therefore  $\mathbf{u}_{:,m} = \mathbf{K}\mathbf{u}_{:,}$  and  $\mathbf{r}_{:,m} = \mathbf{K}\mathbf{r}_{:,}$ ,

where “:” was either <sub>MACE</sub> or <sub>DOM</sub>

and  $\Sigma_{a,m} = \mathbf{L}\Sigma_a\mathbf{L}'$ .

In case only a subset of breeding values within animal met the integration requirement the above dimension, were adjusted accordingly. Within animal a vector of prediction error variances was calculated by

$$\mathbf{p}_{:,} = \text{diag}(\Sigma_{a,m})(\mathbf{1} - \mathbf{r}_{:,m}),$$

where  $\text{diag}()$  denotes a vector of diagonal elements of a squared matrix. The system for deriving a diagonal matrix  $D_{:,}$  of inverse residual variances across traits within animal was  $\text{diag}((D_{:,} + (\Sigma_{a,m})^{-1})^{-1}) \equiv \mathbf{p}_{:,}$

where  $D_{:,}$  was be obtained directly only if the system's dimension was 1. Otherwise  $D_{:,}$  was established iteratively using a procedure similar to Vandenplas and Gengler (2012).

Within animal vectors of pseudo-phenotypes  $y^*$  were calculated for *ABulls* as

$$D_r = D_{MACE} - D_{DOM}$$

and

$$y^* = D_r^{-1} ((DMACE + (\Sigma a, m)^{-1})^u MACE, m - (DDOM + (\Sigma a, m)^{-1})^u DOM, m)$$

and for *BBulls* as

$$y^* = D_{MACE}^{-1} (DMACE + (\Sigma a, m)^{-1})^u MACE, m.$$

A maximum of nine pseudo-phenotypes  $y^*$  of animal  $i$  were added to the evaluation.

Data preparation and derivation of  $D_i$  was done using the R computing environment (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).

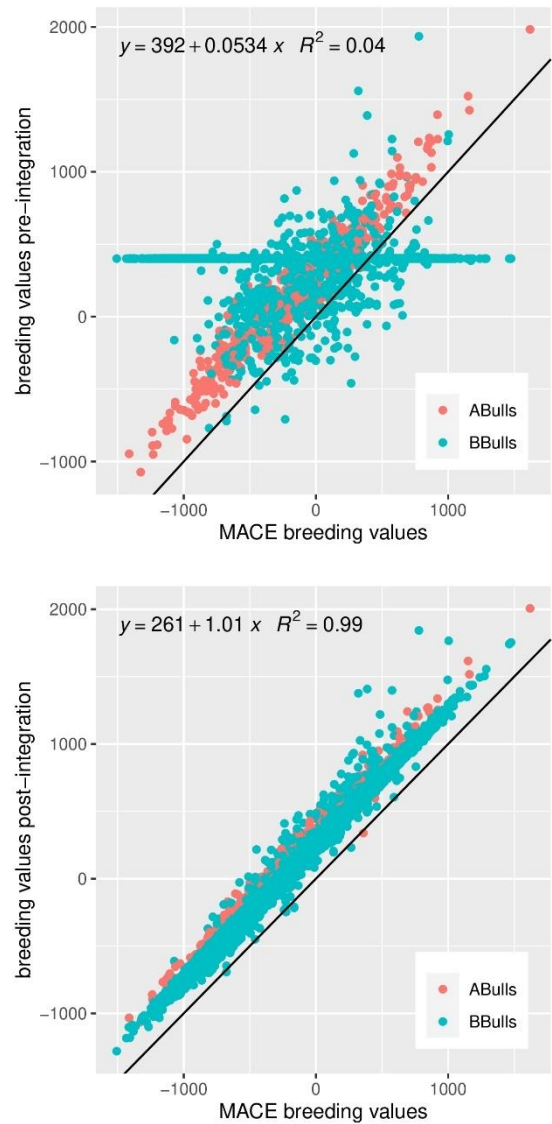
## Results

Generally, results for all three traits were very similar. Therefore, only results for milk yield are summarised in Figures 1, 2, and 3. Correlations between Interbull breeding values and domestic evaluation breeding values increased from about 0.05, 0.05 and 0.04 pre-integration to 0.99, 0.98 and 0.99 post-integration for milk, fat and protein yield respectively. The regression intercept changed from 392, 12.4 and 11.3 to 261, 14.3 and 7.11, and the regression slope from 0.05, 0.06 and 0.06 to 1.01, 1.01 and 1.01 for milk fat and protein yield, respectively. Therefore, the non-zero intercept when regressing pre-integration breeding values on Interbull breeding values was also found when post-integration breeding values were used as response variables.

Further, pre-integration breeding values of *BBulls* which did not have domestic information constituted a horizontal line, the genetic group solution, in the first plot of Figure 1 which was found in plots using post-integration breeding values.

Results were similar for the reliabilities, but correlations and regression parameters are weaker and, due to the nature of reliabilities, do not provide the same goodness-of-fit information as for the breeding values. However, post-integration reliabilities for all

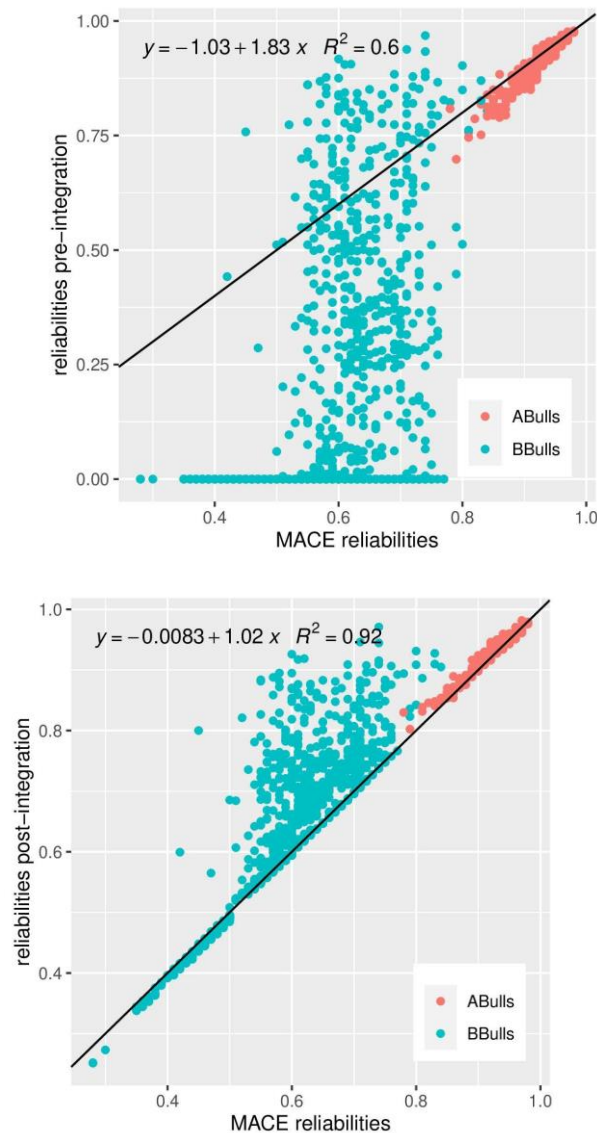
bulls were at least as high as the Interbull reliabilities.



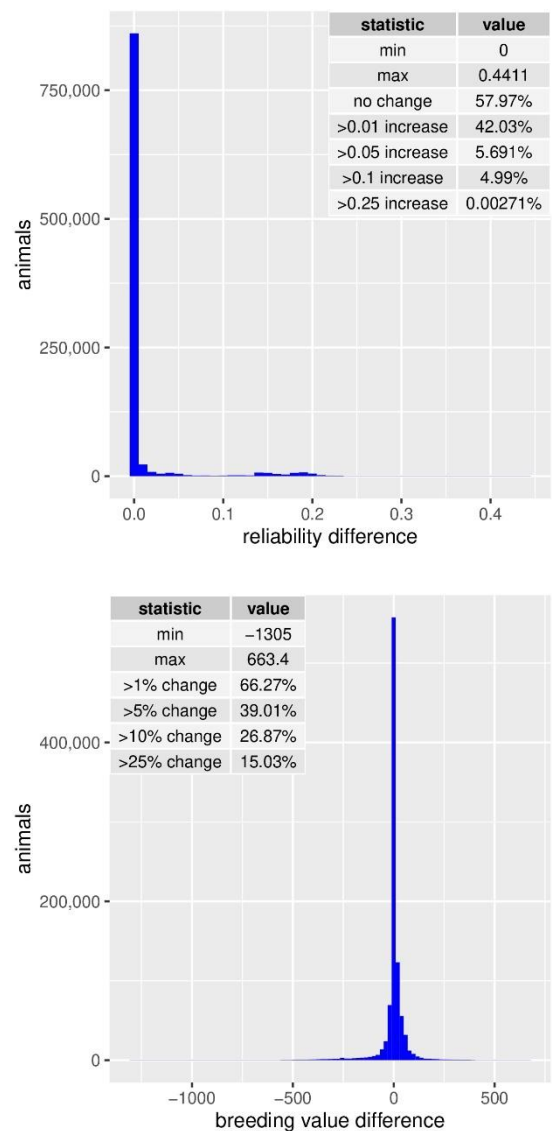
**Figure 1.** Breeding values for litre milk per 305 day lactation pre-integration (first plot) and post-integration (second plot) for bulls with (*ABulls*) and without (*BBulls*) Australian information in the Interbull evaluation, respectively.

For *ABulls*, the post-integration reliabilities aligned very well with the Interbull reliabilities. Further, for *BBulls* breeding values with a pre-integration reliability of zero the procedure returned a post-integration reliability similar to the Interbull reliability. For the entire evaluation population, but excluding the integrated bulls, about 40% of the reliabilities of all three traits were not affected by the integration, and only 5% of the reliabilities increased by more than 0.1 when

compared to their pre-integration value. 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 to their pre-integration value. However, for all three traits only about 15% of the breeding values changed by more than 25%.



**Figure 2.** Reliabilities for litre milk per 305 day lactation pre-integration (first plot) and post-integration (second plot) for bulls with (ABulls) and without (BBulls) Australian information in the Interbull evaluation, respectively.



**Figure 3.** Histograms of the integration impact on breeding values and reliabilities of the entire evaluation population

## Discussion

Results align very well with the expectations and therefore the integration of Interbull breeding values into the domestic single-step random regression test-day evaluation is regarded as successful. The non-zero 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). While the added data points change the information content for all modelled factor levels, the reliability approximation can only account for information changes within groups of closely related animals (e.g. full-sib and half-sib families).

The current integration approach ignores the relationships between integrated bulls when deriving individual specific residual variances. Empirically, ignoring the relationships between integration animals does not pose a problem if the reliabilities are high and integration candidates are not closely related. Accounting for relationship poses a computational problem (Vandenplas et al., 2014). For the approach presented the procedure for deriving  $D_i$  required up to tens of thousand of inversions. Extending the approach to system of  $\Sigma_a^{-1} \otimes (\Gamma, :)^{-1}$  will certainly affect inversion time and most likely also the required number of iterations, and will also require block extractions from a complex  $\Gamma$  matrix. An intermediate approach is to approximate some matrix

$$K = C_{a_i, a_i} - C_{a_i, -a_i} (C_{-a_i, -a_i})^{-1} C_{-a_i, a_i},$$

where  $C$  is the mixed model coefficient matrix and  $a_i$  is a vector indexing the block in  $C$  related to animal  $i$ . Subsequently  $K$  is used instead of  $\Sigma a$  in  $\text{diag}((D_i + K)^{-1}) \equiv p_i$  (Pitkänen, 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. However, it might be computationally prohibitive to apply such procedure to a single-step system. In any case computation time maybe 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.

## Conclusion

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 mixed model equation system's right-hand side or a special co-variance structure between individuals. Further work may be required to allow for co-variance between integration candidates.

## References

- Christensen O.F. and Lund M.S. 2010. *Genet. Sel. Evol.* 42:2.  
<http://dx.doi.org/10.1186/1297-9686-42-2>
- Jamrozik J., Schaeffer L., and Jansen G. 2000. *Livest. Prod. Sci.* 66(1):85–92.
- Lidauer M., Matilainen K., Mäntysaari E., Pitkänen T., Taskinen M., and Strandén I. 2017. *Natural Resources Institute Finland* (Luke).
- Pitkänen T., Koivula M., Strandén I., Aamand G., and Mäntysaari E. 2020. In Proc. of the 71th Annual Meeting of the European Association for Animal Production, Porto, Portugal, page 582. Wageningen Academic Publishers.
- R Development Core Team. 2011. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>, ISBN 3-900051-07-0.
- Vandenplas J., Colinet F.G., and Gengler N. 2014. *Genet. Sel. Evol.* 46.
- Vandenplas J. and Gengler N. 2012. *J. Dairy Sci.* 95:1513–1526.