Inclusion of MACE Proofs in Single-Step Genomic Analysis

G.J. Nieuwhof[1], T.T.T. Nguyen[1] and M. Abdelsayed[1]
[1] *DataGene, AgriBio, 5 Ring Road, Bundoora Vic 3083, Australia*

Abstract

The integration of multiple across country evaluation (MACE) proofs in single-step genomic analysis is important to provide the dairy industry with the best estimated breeding values (EBVs), especially in countries that import a major part of their genetics. The method developed earlier that uses Deregressed Proofs (DRP) that account for correlations between traits, but not relationships among MACE bulls was largely successful, but as we show here leads to large overestimations if MACE bulls are related. We developed an alternative approach that uses DRPs which take into account relationships among bulls, but still uses the old weights based on the assumption of no relationship. This proofed to be a better predictor of performance in Australia for bulls that had both a genotype and MACE proof partially based on their Australian daughters. An additional adjustment to account for that daughter information sent to Interbull proved ineffective, with the regression coefficient 0.82 in both cases. Bulls that were not expected to be affected by the singe-step procedure as they had no Australian daughters and no genotype, did in fact show large changes (regression coefficient 0.66), showing that the weights need to be in-line with the DRP estimation procedure.

Keywords: MACE, genomics, single-step

Introduction

DataGene delivers a single-step genomic evaluation for milk, fat and protein yield and Somatic Cell Count (SCC) for Red breeds, as described in Boerner et al (2022). This procedure uses deregressed MACE proofs as input data alongside test day observations for cows. The deregression method takes into account that MACE milk, fat and protein proofs are correlated and come from a multi-trait analysis in Australia and other countries. It assumes that MACE bulls are unrelated.

We observed an overestimation of breeding values for MACE bulls, especially for protein yield, and attributed this to the assumption of unrelatedness among them. This was initially resolved by setting parents of MACE bulls to missing, although this could not be done for bulls that had both local and overseas daughters. This approach no longer worked when we obtained genotypes on many MACE bulls – confirming relationships among them.

This paper describes how we have succeeded in replacing the deregression procedure with an alternative that takes into account relations among MACE bulls, though not correlations among different traits. It shows how this markedly improves EBVs for some animals but not for others.

Materials and Methods

Current Method

Our current method described by Boerner et al (2022) includes the following steps to create pseudo records and adjust the pedigree

1. Calculation of within animal residual variance
2. Adjustment of residual variance for bulls who had their EBV included in MACE (this will be referred to as 'sent')
3. Deregression of MACE proofs
4. Pedigree adjustment
5. MiX99 run

In the calculation of within animal residual variance a data point specific residual variance is modelled such that a within-animal multi-

trait mixed model equation system would yield reliabilities equal to those derived from Interbull reliabilities.

The pedigree adjustment consisted of replacing the sire and dam for a bull that had a MACE proof but no Australian daughters with a phantom group. Different phantom groups were used for sires and dams.

Alternative Method

A suit of programs tailored to the deregression of MACE proofs based on the deregression method of Jairath et al (1998) was kindly provided by Zenting Liu (VIT Germany). In here, deregress.f90 is the main program. It estimates deregressed proofs for all bulls with daughters in a MACE proof file using iteration on data and full sire-dam pedigree. A Gauss-Seidel algorithm is used to solve the equation system with pre-defined convergence criteria.

Deregressed proofs from this calculation were used to replace the DRPs calculated in step 3 above. Step 4. Pedigree adjustment was omitted.

The new DRP calculation gives one DRP per trait per animal, unlike the current procedure which calculates a DRP for each of the first 3 lactations, although they tended to be similar. We therefore tested two scenarios; one where a MACE bull only had an observation for the first lactations (the observation being the new DRP), or where it had the same observation for all three lactations.

Note that in this approach we do not make an adjustment of DRPs for bulls who had their EBV based on Australian daughters sent to Interbull (equivalent of step 2). As an alternative, we therefore further adjusted the new DRPs calculated above, by weighing them and DRPs calculated from the sent EBVs according to their Effective Daughter Contribution (EDC) as described by Pitkanen (2021).

Data

The impact of the alternative deregression method was investigated using the December 2022 Red Dairy Cattle (RDC) MACE proofs for milk, fat and protein yield. The EBVs that Australia contributed to this MACE run were based on data from the 25 October 2022, but from a special Australia-only conventional analysis (i.e. excluding MACE proofs and genomics). Breeding values from the 'current' method are those published on 6 December 2022.

The production file for RDC in December 2022 consists of 17081 bulls, of which 16721 are of breed RDC or Milking Shorthorn (MSH). Of these 836 had an Australian EBV included.

We identified 57 bulls with genomics and at least 100 test day observations per trait on their daughters and whose EBV was sent to Interbull. These 57 serve as the main validation group, and in various scenarios we remove their daughter observations, their MACE proofs or both from the data to ascertain how well the analysis predicts the Australian performance.

A second validation set consisted of the 15 556 bulls that had neither Australian daughters nor a genotype in Australia. The expectation was that a correct procedure would return EBVs and reliabilities from the single-step genomic analysis that are essentially the same as their MACE proofs and reliabilities.

Results & Discussion

As the original issue mainly showed for protein yield, most results presented below are for protein. Results for milk and fat yield are in line with these.

Figure 1, shows how various datasets predict the Australian-only conventional breeding value for 57 validation bulls for protein using the current procedure for deregression, but with full pedigree. Datasets with only genotypes but no MACE and with both genotypes and MACE perform reasonably well with slopes of 0.75 and

0.64 respectively. A dataset that includes MACE proofs but no genotypes however has a slope of only 0.35, indicating a large overestimation of true performance in Australia. Note the inverse relationship between slope and R^2.

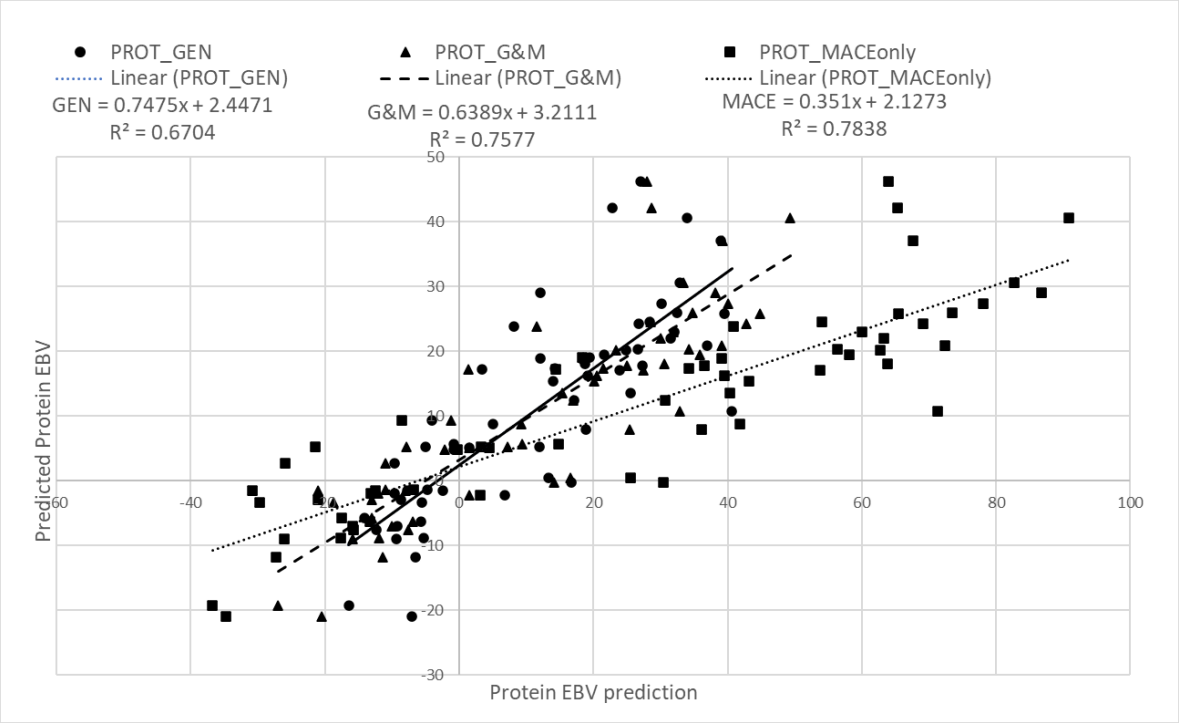

Figure 1. Australian-only Protein EBV predicted for validation bulls using the current procedure for deregression, but with full pedigree (Boerner et al. 2022), comparing predictions based on genomics only (PROT_GEN), MACE only (PROT_MACEonly) and the combination (PROT_G&M).

Figure 2 compares the Genomics & MACE from Figure 1 with the alternatives using the same data but with the new DRP calculation (referred to as ZT) and the additional adjustment for sent EBVs (referred to as ST). The ZT version is the one with the same DRP for each of 3 lactations, rather than the one with only a DRP for lactation 1, which performed slightly less. The ST version is based on 3 lactations as well. Note that ZT is not visible as ZT and ST are virtually identical for this group of bulls.

The alternative DRP calculations clearly give superior results, both in terms of slope (0.82) and R^2.

The effect of the new deregression method on the prediction bias in bulls with MACE proofs was analysed by comparing the MACE proof with the single-step breeding values for the 836 bulls that were sent to Interbull.

The adjustment for 'sent' EBV was specially meant for this group of bulls but it had minimal effect, with slopes being the same with and without the adjustment (0.93) and still showing some bias. R^2 was 0.981 for both DRPs. This may be because the adjustment is designed for a single-trait analysis, not a multi-trait.

The second validation set which had no Australian daughters and no genotype included in the analysis, showed large overestimations of protein EBVs for both the ZT and ST method when the single-step genomic EBV was regressed on the MACE proof; regression coefficients of 0.66 for EBV and 0.77 for reliability. The reason for this is most likely that while the MACE deregression accounted for relationships, the error variance did not and thereby put too much weight on the DRPs.

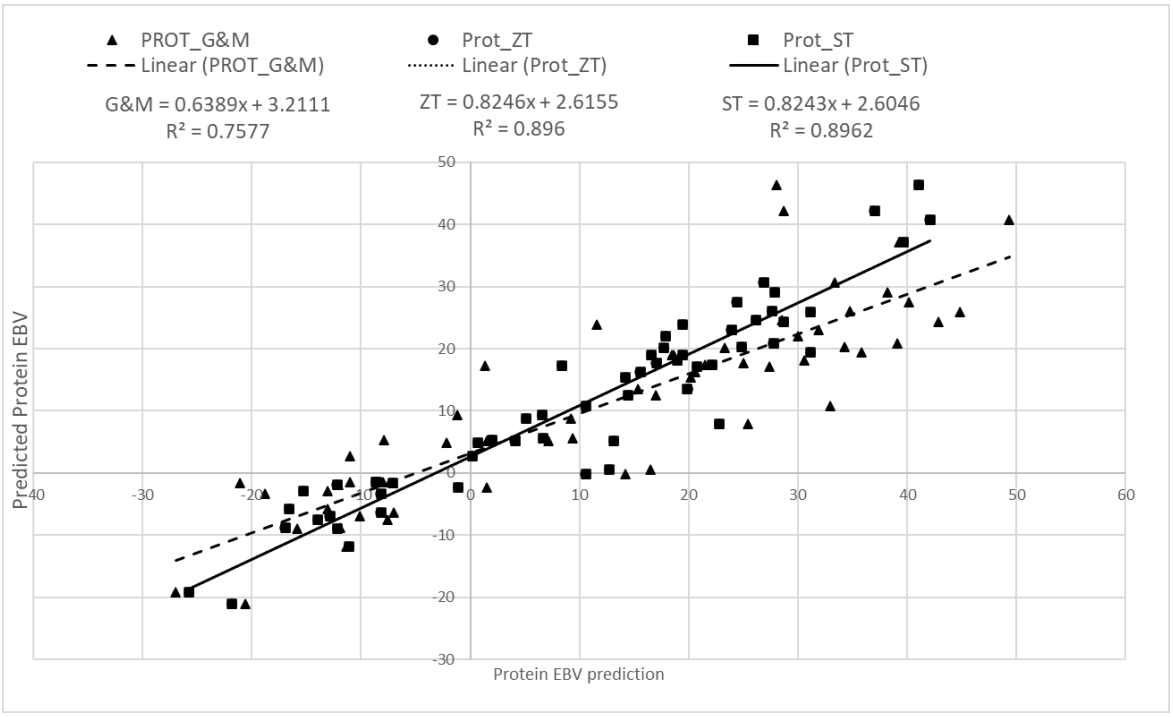

Figure 2. Australian-only Protein EBV predicted for validation bulls using different methods for deregression. a) Boerner et al 2022 (Prot_G&M); b) Zengting Liu's method (Prot_ZT), c) Deregressed using Zengting Liu's method followed by an adjustment using Pitkanen (2021, Prot_ST). Note that the last two overlap completely in the figure.

Conclusions

We have taken a pragmatic approach to try and remove a bias from single-step genomic breeding values that include MACE. We replaced the DRPs from a procedure that ignores relationships among MACE bulls with one that does, but in the process ignored lactation specific DRPs and maintained the weightings as calculated for the old procedure.

The bias in prediction of breeding values was considerably reduced, with the regression coefficient increasing from 0.64 to 0.82

An adjustment was made to the DRPs for animals who had their EBVs included in MACE. This proved to have minimal if any effect on breeding values from the genomic analysis and no effects on reliabilities at all. It appears the 'old' weights took care of this. For bulls that had no Australian daughters and no genotype included in the analysis, EBVs from the single-step genomic analysis grossly overestimated MACE proofs, while they were expected to be similar.

Acknowledgments

The authors wish to thank Zengting Liu from VIT (Germany) for providing the deregression software and assistance with implementing it, as well as Mekonnen Haile-Mariam from Agriculture Victoria (Australia) for assisting with the latter. Support from Timo Pitkanen (LUKE Finland) for providing the correction procedure for the sent EBV is also acknowledged.

References

Pitkanen, T. 2021. Blending with mix99. Blending process, first draft. Memo 22 November 2021.

Boerner, V., Nguyen, T. and Nieuwhof, G. 2022. Integrating MACE breeding values into multiple-trait single step random regression test day evaluations of Aussie Reds. Proceedings of 12th World Congress on Genetics Applied to Livestock

Production (WCGALP). DOI: 10.3920/978-90-8686-940-4 _670.

Jairath, L., Dekkers, J.C.M., Schaeffer, L.R., Liu, Z., Burnside, E.B., and Kolstad, B. 1998. Genetic Evaluation for Herd Life in Canada J. Dairy Sci. 81:550-562.