

Scotland's Rural College

## 292. Within and across breed single step genomic prediction for somatic cell count including foreign information

Mrode, R.; Winters, M.W.; Coffey, M.

*Published in:*

Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP)

*DOI:*

[10.3920/978-90-8686-940-4\\_292](https://doi.org/10.3920/978-90-8686-940-4_292)

First published: 09/02/2023

*Document Version*

Publisher's PDF, also known as Version of record

[Link to publication](#)

*Citation for published version (APA):*

Mrode, R., Winters, M. W., & Coffey, M. (2023). 292. Within and across breed single step genomic prediction for somatic cell count including foreign information. In *Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP): Technical and species orientated innovations in animal breeding, and contribution of genetics to solving societal challenges* (pp. 1229-1232). Wageningen Academic Publishers. Advance online publication. [https://doi.org/10.3920/978-90-8686-940-4\\_292](https://doi.org/10.3920/978-90-8686-940-4_292)

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

## 292. Within and across breed single step genomic prediction for somatic cell count including foreign information

R. Mrode<sup>1,2</sup>, M.W. Winters<sup>3</sup> and M. Coffey<sup>1</sup>

<sup>1</sup>Scotland Rural College, Edinburgh EH9 3JG, United Kingdom; <sup>2</sup>International Livestock Institute, Box 30709, Nairobi, Kenya; <sup>3</sup>Agriculture and Horticulture Development Board, Stoneleigh Park, Kenilworth, Warwickshire, CV8 2TL, United Kingdom; [raphael.mrode@sruc.ac.uk](mailto:raphael.mrode@sruc.ac.uk)

### Abstract

Current UK genetic evaluations for somatic cell count (SCC) are based on an across breed repeatability animal model. Genomic prediction for SCC was investigated using ssGBLUP within the Holstein (HOL) breed only and on an across breed basis. Also, impact of incorporating de-regressed MACE proofs for foreign bulls on accuracy of predictions was investigated. In the within HOL breed analysis, accuracy of bull predictions increased by about 50% to 0.66 with ssGBLUP and by 70% to 0.76 with including foreign information. Results from the across breed evaluations for HOL were similar to results from the within breed analysis. Increases in cow accuracies were low at about 13% with including foreign information. However, there were no increases in accuracies for the other smaller breeds in the across breed evaluations, due may be to the limited number of their bulls in the validation set with genotypes. Evaluations from the various models were highly correlated.

### Introduction

Currently in the UK, conventional genetic evaluations for somatic cell count (SCC) are based on an across breed repeatability animal model using up to 5 lactations with unknown parent groups (UPGs) handled by the QP transformations. The de-regressed proofs from this model are subsequently fed into the multi-step SNPBLUP model that incorporates de-regressed MACE proofs from INTERBULL for genotyped foreign bulls. Genotyped cows have not yet been incorporated into the training population for the multi-step genomic SNPBLUP. However, the simultaneous use of pedigree, phenotypic and genomic information for genomic prediction utilizing a single-step approach (ssGBLUP) would allow genotyped cows to be part of the training population and should, in theory, overcome some of the problems of preselection bias associated with the multi-step SNPBLUP. Therefore, this study examines the use of ssGBLUP for genomic prediction of SCC, initially within the Holstein (HOL) breed and then in an across breed setting involving the Ayrshire (AYR), Jersey (JER) and Guernsey (GUE). In addition, it examines the impact of incorporating de-regressed MACE for foreign bulls on the accuracy of genomic prediction.

### Materials & methods

The SCC data for the official UK conventional genetic evaluations and MACE from INTERBULL for the August 2021 run and corresponding genotypes for the genomic prediction, were used for these analyses. The summary of the data showing number of cows, records, pedigree and genotypes are in Table 1. SCC was analysed as  $\log_e$  SCC and the conventional model (AM1) fitted herd-year-season, lactation, age nested within parity, month of calving as fixed effects and random effects of herd-sire interaction, permanent environment and animal. A pedigree of five generations was used in the analysis with a total of 209 UPGs defined by year of birth, breed, country of origin and the four paths of dam of cows and bulls, and sire of bulls and cows. In the ssGBLUP, 79,051 SNPs used were derived by USDA (Al-Khudhair, *et al.*, 2021). Initially, only HOL was analysed (ssGBLUP-HOL1) with the G matrix computed from genotypes of 9,214 and 45,547 UK bulls and cows, with UPG fitted for both A and the pedigree relationship matrix for genotyped animals ( $A_{22}$ ) (Cesarani, *et al.*, 2021). Thus, the inverse of the H matrix was computed as:

$$\mathbf{H}^{-1} = \mathbf{A} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} & -\mathbf{A}_{22}^{-1}\mathbf{Q}_2 \\ 0 & -\mathbf{Q}_2'\mathbf{A}_{22}^{-1} & \mathbf{Q}_2'\mathbf{A}_{22}^{-1}\mathbf{Q}_2 \end{bmatrix}$$

Foreign information was included for a total of 31,633 HOL bulls with genotypes and MACE proofs, using two slightly different approaches. For bulls with UK daughters included in their MACE, an adjustment was made to avoid double counting of this domestic information. The procedure described by Guarini *et al.* (2019) was implemented as:  $PTA_{adj} = \frac{(DE_M+k) \times PTA_M - (DE_D+k) \times PTA_D}{(DE_M-DE_D)+k}$  and  $Rel_{adj} = \frac{DE_M-DE_D}{(DE_M-DE_D)+k}$ ; where  $PTA_M$  and  $PTA_D$  are PTAs from MACE and domestic evaluations respectively;  $DE_M$  and  $DE_D$  are corresponding daughter equivalents calculated from their respective reliabilities for  $PTA_M$  and  $PTA_D$ ; and  $k$  is a variance ratio computed as  $(4-h^2)/h^2$ . The de-regressed  $PTA_{adj}$  multiplied by two were used as the response variable in the analysis for these bulls with a weighting of corresponding daughter equivalents (DE) computed from  $Rel_{adj}$ . However, for foreign bulls with no domestic information included in MACE, their de-regressed PTAs using a full pedigree and multiplied by two were used in the analysis, weighted by their corresponding DEs computed from MACE reliability. To handle the increased number of genotypes by including foreign information, the inverse of G inverse was computed using the APY approach, with 35,000 random bulls with genotypes used as core animals. The model with foreign information included was termed as ssGBLUP-HOL2.

The ability to predict the future performance of young candidates was used for validation of accuracy of genomic prediction. The validation cows comprised of 282,728 cows with 357,591 SCC records born in 2017 and onwards and their records were set to missing in the validation run. Yield deviations (YD) were computed for validation cows by adjusting their records for all fixed effects in the model. Validation bulls were sires with all their UK daughters born within the validation years for validation cows. The YD of validation cows were further adjusted for half the breeding value of their dams and averaged by sires, to compute estimated daughter yield deviations (DYD) for the validation bulls. The accuracy of predictions was computed as the correlation between (G)EBV and YD or DYD for validation cows or bulls. Regression coefficients of YD or DYD on (G)EBV were used to measure the inflation of predictions in validation cows and bulls.

**Across breed predictions.** The conventional repeatability animal model used for HOL was implemented including other breeds; AYR, JER and GUE and various crossbreeds (AM2). This involved 69,737,471 cows with 17,572,710 SCC records with a pedigree consisting of 9,063,330 animals (Table 1). The same UPG structure was utilized since animals from some of the other breeds were dams of cows in the Holstein evaluation. Similarly, ssGBLUP was implemented for the across breed model (ssGBLUP-ALL1) resulting in the inclusion of an additional 1,583 bulls and 6,402 cows with genotypes from the AYR, JER and GUE in the construction of  $G_{22}$ . Moreover, foreign information was included for an additional 10,924 foreign bulls for these breeds with genotypes and MACE proofs as described for the HOL only evaluations. The ssGBLUP with foreign information included was termed ssGBLUP-ALL2. The validation bulls and cows were defined as for the within HOL analysis and comprised of 311,633 cows with 393,492 SCC records born in 2017 and onwards and their records were set to missing in the validation run. Similarly, accuracies and measures of inflation of predictions were computed as described for the within HOL analysis. The heritability of 0.11 was used in all analyses within the HOL breed and in the across breed evaluations. All analyses were carried out using the MiX99 software (Strandén *et al.*, 2017).

**Table 1.** Summary statistics of number of animals, records and genotypes.

	Evaluations	
	Holstein breed only	Across breed
Cows	6,453,896	6,973,747
Records	16,245,062	17,572,710
Pedigree size	8,420,844	9,063,330
UK bull genotypes	9,214	10,797
Cow genotypes	45,547	51,949
Foreign bull genotypes	31,633	38,130

## Results

**Accuracies and measures of inflation of predictions.** The estimates of correlations between (G) EBVs and DYDs for bulls or YDs for cows in the validation set are shown in Table 2 for HOL from the various models. In the within HOL run, the accuracy of genomic prediction for bulls increased from 0.44 with the AM1 by about 50% to 0.66 with ssGBLUP. However, the inclusion of foreign information increased accuracies by 72% relative to the AM1 to 0.76. The pattern of results for HOL from the across breed evaluations was similar to results from the within breed analysis. In terms of measures of inflation of predictions, the regression coefficients (Table 2) were close to unity apart from the ssGBLUP in the within HOL analysis where there was some under-prediction. The accuracies of genomic prediction were much lower for cows at 0.23 from the AM1 and increased slightly to 0.26 with ssGBLUP and with foreign information included. In general, the regression coefficients were close to unity for cows from all models.

The number of bulls with 50 UK daughters in the validation set for the other breeds were too few to obtain any reliable results. Using only bulls with at least 10 UK daughters resulted in a total of 149 bulls across the three breeds, AYR, JER and GUE in the validation set. The accuracy of predictions using these 149 bulls was 0.43 from AM2 but decreased to 0.42 from ssGBLUP-ALL1 with no foreign information included and further to 0.30 when foreign information was incorporated. The regression coefficients were 1.02 for both AM2, ssGBLUP-ALL1 and lower at 0.84 for ssGBLUP-ALL2. In terms of the accuracies of predictions for cows, the accuracies were in generally similar to those obtained to HOL from the AM2 but ssGBLUP with or without inclusion of foreign information did not result in any increases in accuracies.

**Table 2.** Estimates of correlations between (G)EBVs and DYDs of bulls and YDs of cows in validations data set and corresponding estimates of regression coefficients of DYDs or YDs on (G)EBVs.<sup>1</sup>

Model	Bulls		Cows	
	Correlation	Regression	Correlation	Regression
AM1	0.44	1.00	0.23	0.95
ssGBLUP-HOL1	0.66	1.20	0.24	1.01
ssGLUP-HOL2	0.76	0.98	0.26	0.99
AM2	0.43	0.98	0.23	0.96
ssGBLUP-ALL1	0.66	1.23	0.24	1.06
ssGBLUP-ALL2	0.77	1.00	0.26	1.00

<sup>1</sup> AM1= Within HOL repeatability animal model; AM2= Across breed repeatability animal model; ssGBLUP-HOL1 and ssGBLUP-HOL2= within HOL ssGBLUP without (1) and with (2) foreign information; ssGBLUP-ALL1 and ssGBLUP-ALL2= Across breed ssGBLUP without (1) and with (2) foreign information.

**Genetic trends and correlations among various evaluations.** The correlation between (G)EBVs for bulls from the AM1 and ssGBLUP-HOL1 was 0.99, 0.98 between AM1 and ssGBLUP-HOL2 and 0.99 between ssGBLUP-HOL1 and ssGBLUP-HOL2. Corresponding estimates for cow evaluations were all 0.99. The estimates of correlation for bulls and cows for the other breeds in the across breed evaluations followed the same pattern with correlations generally between 0.97 to 0.99. Estimates of genetic trends for HOL from all the models within and across breed evaluations were very similar.

## Discussion

Increased accuracy of prediction has been one of the major advantages of genomic prediction and an increase of about 50% in accuracy of predictions for the HOL have been observed with ssGBLUP from the within and across breed evaluations. Cesarani *et al.* (2021) reported similar ranges of increases in reliabilities of bulls implementing ssGBLUP for milk, fat and protein yields in the USA. However, Guarini *et al.* (2019) reported lower increases of about 10 to 19% for milking temperament and speed. However, the largest increase in accuracy for HOL of about 70% resulted in the inclusion of foreign information. Guarini *et al.* (2019) reported similar increases in accuracy of about 90% with inclusion of foreign information for milking temperament and speed. However, the levels and increases in accuracies for cow predictions were low and similar to those reported by Cesarani, *et al.* (2021) for productive traits in USA cows. Similar to the results of Guarini, *et al.* (2019), the inclusion of foreign information generally improved the bias of genomic predictions with regression coefficients being closer to unity. However, there were no improvements in accuracy for both bulls and cows for other breeds in the across breed evaluations with ssGBLUP or inclusion of foreign information. The limited number of bulls with genotypes in the validation set may account for this and this will further be investigated.

However, correlations and estimates of genetic trends showed that evaluations from the various models were consistent.

## References

- Al-Khudhair A., VanRaden P.M., Null D.J., and Li B (2021) J. Dairy Sci. 104(4):4478–4485. <https://doi.org/10.3168/jds.2020-19260>
- Cesarani A., Masuda Y., Tsuruta S., Nicolazzi E.L., VanRaden P.M., *et al.* (2021) J. Dairy Sci. 104(9):5843–5853 <https://doi.org/10.3168/jds.2020-19789>
- Guarini A.R., Lourenco D.A.L., Brito L.F., Sargolzaei M., Baes C.F., *et al.* (2019) J. Dairy Sci. 102(9):8175–8183. <https://doi.org/10.3168/jds.2018-15819>
- Strandén, I., Matilainen, K., Aamand, G.P. and Mäntysaari, E.A. (2017). J Anim Breed Genet. 134:264-274. <https://doi.org/10.1111/jbg.12257>