Transition of the UK dairy national evaluation to across-breed and single-step genomic evaluation: somatic cell counts as a case trait

S. Id-Lahoucine[1], R. Mrode[1,2], M. Winters[3] and M. Coffey[1]

[1] Scotland Rural College, Edinburgh EH9 3JG, United Kingdom
[2] International Livestock Institute, Box 30709, Nairobi, Kenya
[3] Agriculture and Horticulture Development Board, Stoneleigh Park, Kenilworth, Warwickshire, CV8 2TL, United Kingdom
Corresponding author: Samir.IdLahoucine@sruc.ac.uk

Abstract

Current UK genomic evolutions follow a two-step approach; initially, a genetic evaluation based solely on pedigree information, followed by a Single Nucleotide Polymorphism Best Linear Unbiased Prediction (SNPBLUP) analysis for genotyped animals using de-regressed proofs, including MACE proofs from Interbull. Nowadays, with recent advances in computational feasibility and growing interest in across-breed genomic evaluations for dairy herds, there is a compelling need to adopt a single-step across-breed genomic evaluation approach within the UK dairy industry. The single-step method offers notable advantages by simultaneously incorporating genotypes and both recent and historical pedigree and phenotypic data into a single analysis. This integration enhances the accuracy of genetic predictions across diverse breeds, accelerates genetic progress, and improves selection efficiency. This study aims to evaluate the impact of using genomic information and compare the prediction ability of single-step genomic evaluation (using ssSNPBLUP method) and pedigree-based genetic evaluation (PedBLUP) employing cross-validation techniques (Linear Regression method). The trait analysed was somatic cell count (SCC), using data from the UK national evaluations as of the December 2024 official run. The dataset included 11,271,959 animals in the pedigree and 19,056,954 SCC records from 7,527,712 cows. Foreign information was incorporated for 182,844 bulls, with adjustments made to avoid double counting of domestic data. Genotypic data was available for 891,480 animals, imputed at 79,051 SNPs using findhap.f90 V3. Analyses were performed using the MiX99 V23.1026 software, applying an ssSNPBLUP model with 10% polygenic effects. The validation group comprises bulls born after 2016 and cows born after 2018, whose records are set to missing. Results showed a genomic accuracy improvement of up to 54% in cows when comparing ssSNPBLUP to PedBLUP. Among bulls, the greatest gain was observed in Holsteins (+33%), followed by Guernsey and Ayrshire (+30%), and Jersey (+20%). Level bias and dispersion bias was slightly reduced in ssSNPBLUP relative to PedBLUP. Overall, the findings demonstrate that single-step genomic evaluation is a promising and efficient approach for enhancing prediction accuracy in UK dairy cattle.

Key words: single-step genomic evaluation, accuracy, cross-validation, UK dairy industry

Introduction

The genetic evaluation in the UK dairy sector is undergoing a significant transition. Currently, genetic evaluation has relied on a two-step method. This includes conventional pedigree-based evaluations followed by a genomic prediction step using Single Nucleotide Polymorphism Best Linear Unbiased Prediction (SNPBLUP), incorporating de-regressed proofs from Interbull for foreign bulls. Recent advances in computational capacity, as well as the growing interest in across-breed genomic evaluations for dairy herds, there is a compelling need to adopt a single-step across-breed genomic evaluation approach (ssGBLUP) within the UK dairy industry. This approach simultaneously integrates pedigree, phenotypic, and genomic information into a single evaluation model, offering numerous advantages: inclusion of all available data, improving prediction accuracy, mitigate the effects of preselection bias associated with multi-step approach, and, consequently, accelerating the genetic progress.

This report focuses on somatic cell count (SCC) as a trait study to evaluate the prediction ability of ssGBLUP in comparison to traditional pedigree-based BLUP (PedBLUP) employing cross-validation techniques (Linear Regression method).

Materials and Methods

Data

Somatic cell count data was obtained from the UK national evaluations as of the December 2024 official run. The dataset included 11,271,959 animals in the pedigree and 19,056,954 SCC lactation records from 7,527,712 cows. Foreign information was incorporated for 182,844 bulls, with adjustments made to avoid double counting of domestic data. Genotypic data was available for 891,480 animals, imputed at 79,051 SNPs using findhap.f90 V3. This reference imputation panel was derived by USDA (Al-Khudhair et al., 2021).

Model and analyses

SCC was analyzed as log$_e$ SCC and the model included 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 and UPGs were used for missing parents based on year of birth, breed, country of origin and the four paths of dam of cows and bulls, and sire of bulls and cows.

MACE proofs were included following Bonifazi et al. (2023). First, effective record contribution (ERC) was derived from the reliability of the animal ($ERC=\lambda\cdot($ reliability/(1-reliability)); where λ=(1-heritability)/heritability). In second, the de-regressed proofs (DRP) were calculated as follow: DRP=PA+(EBV-PA)/(dERC/(dERC+λ)), where PA is the parent average and dERC=ERC-ERC$_{PA}$. To avoid double counting of the national data (NAT) of bulls with UK daughters from interbull (international, INT) proofs, DRP were blended as follow: DRP*=((dERC$_{INT}\cdot$DRP$_{INT}$)-(dERC$_{NAT}\cdot$DRP$_{NAT}$))/dERC* where dERC*=dERC$_{INT}$-dERC$_{NAT}$. These de-regressed proofs were used as pseudo-records with the corresponding ERC as weights in the model.

Analyses were performed using the MiX99 V23.1026 software (Strandén et al., 2017), applying a ssSNPBLUP (Liu et al., 2014) model with 10% polygenic effects.

Validation

A cross-validation technique, linear regression method, was used to compare ssGBLUP and PedBLUP following Legarra & Reverter (2018). The different estimators' statistics of bias (difference of means), dispersion (slope of the regression) and accuracy were calculated based on partial and full runs. The validation group included bulls born after 2016 and cows after 2018. The phenotypes of these cows together with the de-regressed proofs of these bulls were excluded from the "partial" runs.

Results & Discussion

The implementation of the single-step genomic BLUP model led to substantial improvements in prediction accuracy over the traditional pedigree-based BLUP (PedBLUP), with gains observed across all breeds: Guernsey, Holstein, Jersey, and Ayrshire. In cows, prediction accuracy improved dramatically, with Holstein showing the largest relative gain of +102% (accuracy increased from 0.32 to 0.648). Guernsey and Ayrshire cows followed with increases of +65% and +56%, respectively, while Jersey cows exhibited a +43% improvement, but notably had the 2nd highest ssGBLUP accuracy. Bulls also benefitted from the use of ssGBLUP, with Holstein bulls achieving the highest accuracy gain of +33%, followed by Guernsey and Ayrshire (+30%), and Jersey (+20%). The advantage of ssGBLUP is coming from leverages genomic information to capture Mendelian sampling more effectively than PedBLUP. The improved prediction accuracies of ssGBLUP in both sexes are summarized in Tables 1 (bulls) and 2 (cows).

Level bias, as illustrated in Figure 1, was marginally reduced under ssGBLUP relative to PedBLUP across all breeds and both sexes. For example, the level bias for Jersey bulls was -0.001 and -0.056 for both ssGBLUP and PedBLUP, respectively, reducing the level bias to mostly null when using ssGBLUP. This reduction indicates that genetic evaluations using ssGBLUP are more centered on the true genetic values, thus improving the reliability of selection. Although the magnitude of the reduction was modest generally across breeds and sexes, it consistently favored ssGBLUP and supports its use for less biased evaluations.

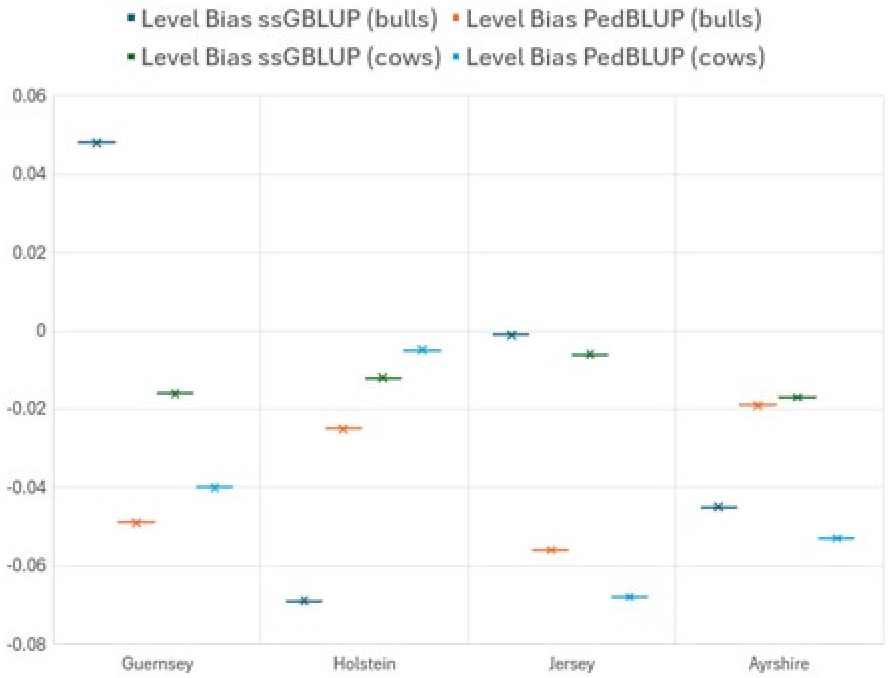

Figure 1. Level Bias for single-step genomic evaluation approach (ssGBLUP) and pedigree-based BLUP (PedBLUP) accords breeds.

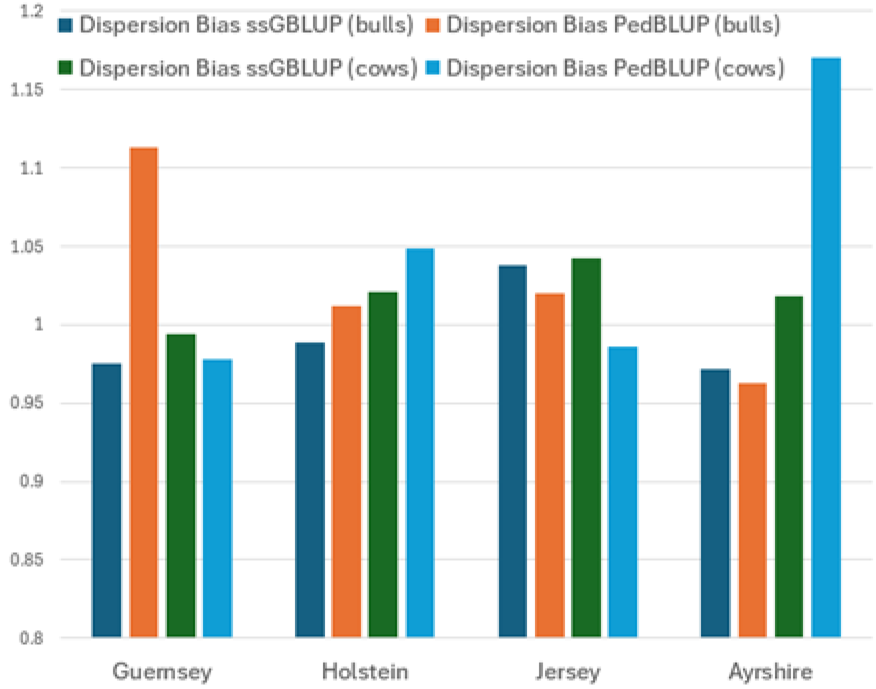

Figure 2. Dispersion Bias for single-step genomic evaluation approach (ssGBLUP) and pedigree-based BLUP (PedBLUP) accords breeds.

On the other hand, as shown in Figure 2, ssGBLUP consistently produced dispersion bias values closer 1.00. In contrast, PedBLUP tended to over-dispersion in several cases, for example, Ayrshire cows displayed a dispersion bias of 1.17 versus 1.02 for ssGBLUP. Also, Guernsey bulls presented a dispersion bias of 1.11 in comparison to 0.975 for ssGBLUP. Overall, ssGBLUP showed more balanced dispersion across all breeds and sexes, particularly improving prediction spread in cows.

Additionally, within breed ssGBLUP were performed which in all cases yielded lower accuracies compared to the across breed ssGBLUP (results not shown in this paper).

These results highlight ssGBLUP's ability to provide more reliable genetic evaluations.

Table 1: Comparison of prediction ability in bulls.

Breed	Number of animals	Accuracy ssGBLUP	Accuracy PedBLUP	Increase % ACC	Correl Full PedBLUP # Partial ssGBLUP
Gue	28	0.50	0.38	29.2	0.50
Hol	6,087	0.79	0.69	13.9	0.90
Jer	453	0.61	0.57	8.3	0.87
Ayr	245	0.51	0.44	14.9	0.79

*Gue: Guernsey; Hol: Holstein; Jer: Jersey; Ayr: Ayrshire.

Table 2: Comparison of prediction ability in cows.

Breed	Number of animals	Accuracy ssGBLUP	Accuracy PedBLUP	Increase % ACC	Correl Full PedBLUP # Partial ssGBLUP
Gue	855	0.45	0.27	65.1	0.40
Hol	95,965	0.65	0.32	102.2	0.60
Jer	3,512	0.50	0.35	43.1	0.60
Ayr	382	0.42	0.27	56.1	0.42

*Gue: Guernsey; Hol: Holstein; Jer: Jersey; Ayr: Ayrshire.

Conclusions

The cross-validation results clearly demonstrate that single-step genomic evaluation is a more effective evaluation method for predicting genetic merit, particularly in genotyped animals. The results support the adoption of a unified Single-Step Genomic BLUP framework for genetic evaluation in UK dairy cattle. By combining genomic, pedigree, and phenotypic information in a single model, ssGBLUP improves prediction accuracy, increases selection efficiency, and accelerates genetic gain across breeds. A national implementation would simplify evaluations, allow the evaluation of crossbreed animals, and help ensure a more competitive and sustainable breeding program.

References

Al-Khudhair, A, VanRaden, P.M, Null, D.J. and Li B. 2021. Marker selection and genomic prediction of economically important traits using imputed high-density genotypes for 5 breeds of dairy cattle. *J. Dairy Sci.* 104(4):4478–4485. doi.org/10.3168/jds.2020-19260

Legarra, A. and Reverter A. 2018. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. *Genet. Sel. Evol.* 50, 53. doi.org/10.1186/s12711-018-0426-6

Liu, Z, Goddard, M.E, Reinhardt, F. and Reents, R. 2014. A single-step genomic model with direct estimation of marker effects. *J. Dairy Sci.* 97.9, pp. 5833– 5850. doi:10.3168/jds.2014-7924.

Strandén, I, Matilainen, K, Aamand, G.P. and Mäntysaari, E.A. 2017. Solving efficiently large single-step genomic best linear unbiased prediction models. *J Anim Breed Genet.* 134:264-274. doi.org/10.1111/jbg.12257