Single-step genomic prediction models for metabolic body weight in Nordic Holstein, Red dairy cattle, and Jersey

***R. Kempe*[1], *M. Koivula*[1], *T.J. Pitkänen*[1], *R.B. Stephansen*[2], *J. Pösö*[3], *U.S. Nielsen*[4], *W.F. Fikse*[5], *G.P. Aamand*[6] *and M.H. Lidauer*[1]**
[1] *Natural Resources Institute Finland (Luke), Tietotie 4, FI-31600 Jokioinen, Finland*
[2]*Aarhus University, QGG, C.F. Møllers Allé 3, 8000-Aarhus, Denmark*
[3] *Faba co-op, P.O. BOX 40, FI-01301 Vantaa, Finland*
[4]*SEGES Cattle, Agro Food Park 15, DK-8200 Aarhus N, Denmark*
[5]*Växa, Box 288, SE-75105 Uppsala, Sweden*
[6] *Nordic Cattle Genetic Evaluation, Agro Food Park 15, DK-8200 Aarhus N, Denmark*
Corresponding author: riitta.kempe@luke.fi

Abstract

Nordic Cattle Genetic Evaluation (NAV) introduced a breed-specific index for Saved Feed in 2020, focusing on the maintenance and metabolic efficiency of cows. Maintenance efficiency is based on genomic breeding values for metabolic body weight (MBW), for which a multi-step (genomic) evaluation was implemented in 2019. The model utilizes body weight and conformation observations from Finland and Denmark, but only conformation observations from Sweden. This study aimed to enhance the MBW evaluation by including carcass weight (CARW) from all three countries and by developing a single-step genomic prediction model. The new model includes three MBW traits and two correlated traits: CARW and stature (STA). The data were collected from Danish, Finnish, and Swedish Red Dairy Cattle (RDC), Holstein (HOL), Jersey (JER) cows born between 1990 to 2020. After data editing, the RDC, HOL, and JER datasets comprised of 2.3 million, 4.3 million, and 0.4 million records, including 0.9 million, 0.5 million, and 11 thousand MBW observations, respectively. The pedigree of RDC, HOL, and JER included 3.9, 7.2 and 0.6 million animals, respectively. Among these, 84 232 RDC, 117 845 HOL, and 39 650 JER animals were genotyped since 2009 onwards. To develop single-step genomic best linear unbiased prediction (ssGBLUP) models, we applied VanRaden method I to construct the genomic relationship matrix, with a residual polygenic proportion of 30%. We utilized the ssGTaBLUP method to solve the models. Separate ssGBLUP models were developed for each breed, and these models were validated through forward prediction cross-validation, linear regression of full data breeding values on reduced data breeding values, and comparison of pedigree-based and ssGBLUP breeding values. The inclusion of carcass weight data substantially increased phenotypic information in all three breeds, resulting in enhanced reliability of MBW breeding values. The new ssGBLUP models showed higher validation reliability and better predictive ability than the pedigree-based BLUP models. Furthermore, the new models corrected the genetic trend of MBW, addressing a previous underestimation in all breeds. Including CARW records as correlated observations and applying ssGBLUP models offers a significant improvement for the Nordic metabolic body weight evaluations, thereby enhancing the Saved Feed index.

Key words: animal breeding, genomic predictions, carcass weight, saved feed index

Introduction

The Saved Feed index of the Nordic Cattle Genetic Evaluation (NAV) was integrated into the Nordic Total Merit index in 2020. It comprises two components: maintenance and metabolic efficiency. Metabolic body weight (MBW, kg$^{0.75}$) is the core trait for maintenance feed requirement, while residual feed intake is the main trait for metabolic efficiency (Lidauer et al. 2019, Stephansen et al. 2021). Each breed, including Red Dairy Cattle (RDC), Holstein (HOL), Jersey (JER), has its own evaluation. The current multiple-trait model for maintenance efficiency includes six traits: MBW in the first, second, and third parity, and the conformation traits stature (STA), chest width, and body depth, as correlated indicator traits.

A current challenge is the decreasing number of body weight recordings in Denmark and Finland, and no body weight (BW) data available from Sweden. However, there is a substantial amount of slaughter information available across the Nordic countries. The correlations between carcass weight (CARW) and MBW are high, ranging from 0.77 to 0.85 in RDC (Mehtiö et al. 2021). Additionally, CARW has high heritability with estimates of 0.52 for RDC and 0.37 for Jersey (Mehtiö et al. 2021, 2023). These characteristics make incorporating CARW information highly valuable in the genetic evaluation of MBW.

The aims of this study were to incorporate CARW data into the evaluation of MBW, upgrade the current multiple-step genomic prediction model to a single-step genomic prediction model, and assess the prediction ability of the models through validation tests.

Materials and Methods

Data

Phenotypic data and pedigree were obtained from the February 2022 NAV Saved Feed evaluation. Breeding organizations Faba, Växa, and Seges extracted country-specific carcass weight data for this study. Observations were from the Danish, Finnish and Swedish RDC, HOL, and JER cows born between 1990 and 2020. The data included all available MBW observations (kg), the first parity STA observation (cm) from the NAV routine conformation evaluation (NAV, 2022) and CARW data from the year 2007 onwards. The CARW data were further restricted to: a) parities 1 to 5, b) 60-550 days after the last calving, c) animals aged 24–110 months at slaughter, and d) herds with more than three CARW records. CARW records deviating more than 3 SD from the mean were removed as outliers. The BW observations were pre-processed as described by Lidauer et al. (2019) to obtain one MBW observation per lactation. After editing, the RDC and HOL data consisted of 0.93 and 0.54 million MBW observations, respectively. The JER data had 11 thousand MBW observations. The number of phenotypic records is presented in Table 1.

Table 1. Number of records for metabolic body weight in the first three parities (MBW1, MBW2, MBW3), first parity stature (STA), and carcass weight (CARW) in Red Dairy Cattle (RDC), Holstein, and Jersey dairy cows.

	RDC	Holstein	Jersey
	N	N	N
MBW1	521 132	293 237	6 064
MBW2	318 764	173 686	3 458
MBW3	93 502	72 766	1 926
STA	349 329	740 521	301 844
CARW	686 946	1 740 589	175 636

Genotype data from February 2022 included 84k RDC, 117k HOL, and 39k JER animals. The genomic data were truncated, retaining only the most resent genotyped animals from the year 2009 onwards. The pedigrees of the RDC, HOL and JER cows with observations were pruned for five generations, including 3.9, 7.2 and 0.6 million animals, respectively. Genetic groups were formed by categorizing unknown parents within country and breed based on 5-year birth year classes, resulting in 182, 202 and 70 unknow parent groups (UPG) for RDC, HOL, and JER, respectively.

Models

The pedigree-based Best Linear Unbiased Prediction (BLUP) models developed for the NAV routine MBW evaluation (Lidauer et al., 2019) served as the foundation for building the single-step genomic prediction (ssGBLUP) models. These multiple-trait models were updated by replacing the traits chest width and body depth with CARW, resulting in multiple-trait BLUP models with five traits: MBW in the 1st, 2nd, and 3rd parity, first parity STA, and CARW. Multiple-trait linear mixed animal models for the 1st, 2nd, and 3rd parity MBW and STA are detailed in Lidauer et al. (2019). The linear model for CARW was as follows:

$$y_{ijkln} = sageP_i + catsP_j + sym_k + shy_l + a_n + e_{ijkln},$$

where y_{ijkln} is a CARW observation, $sageP_i$ is the slaughter age × parity × 5–year period interaction, where year periods are constructed from the birth years; $catsP_j$ is the fixed effect of days from calving to slaughter × parity × 5–year period, again with periods based on birth years; sym_k is the fixed effect of slaughter year × month; shy_l is the fixed effect of slaughter herd × birth year; a_n is the random additive genetic effect of animal, and e_{ijkln} is the random residual.

Single-step models were solved with the ssGTaBLUP approach (Mäntysaari et al. 2017). The VanRaden method I (VanRaden 2008) was used for building the genomic relationship matrix by blending the **G** matrix with a 30% residual polygenic proportion. Pedigree inbreeding coefficients were considered in $\mathbf{A}^{-1}$ and $\mathbf{A}_{22}^{-1}$. Genetic groups were included in the single-step models using the partial QP transformation that omitted $\mathbf{G}^{-1}$ in QP (Koivula et al. 2021).

For each animal, combined MBW breeding values (BV), including estimated breeding value (EBV) and genomic enhanced breeding value (GEBV), were formed using the BVs from the 1st, 2nd, and 3rd parities with weighting coefficients of 0.30, 0.25, and 0.45, respectively.

BLUP and ssGBLUP BVs were validated using forward prediction cross-validation. For the evaluations with reduced data, observations from the most recent four years (2016-2020) were excluded. Candidate bulls for validation were chosen from genotyped bulls born between 2013 and 2018 that had an effective record contribution (ERC) >1 in the full data and ERC=0 (i.e., no daughters) in the reduced data. For the cow validation group, genotyped cows born between 2015 and 2020, with no records in the reduced data (ERC=0) and at least one record in the full data (ERC>0), were considered as candidate cows. In the validation cohort, we had 43 503 RDC cows and 290 RDC bulls, 75 707 HOL cows and 470 HOL bulls and 18 235 JER cows and 150 JER bulls. The same pedigree and genomic information were used in the reduced data as for the full data set evaluations to obtain BVs (either EBV or GEBV) for candidates (BV_c). Cross-validation reliability (r^2_{cv}) was calculated as:

$$r^2_{cv} = \mathrm{corr}\,(DRP, BV_c)^2 / r^2_{DRP},$$

i.e., squared correlation between deregressed proofs (DRP) estimated from the full data and BV_c divided by the average reliability of the DRPs. The second statistic applied was the linear regression of full data breeding values on reduced data breeding values (Legarra and Reverter 2018).

Results and Discussion

Results showed that the genetic trend of combined MBW is increasing in each breed (Figure 1). The current BLUP models underestimate the genetic trend for MBW compared to the new BLUP or ssGBLUP models. The new ssGBLUP models give a slightly higher trend compared to new BLUP-models. This was an expected result because the ssGBLUP models incorporate genomic information directly, which increases the accuracy of estimated breeding values and allows to account for genomic pre-selection.

The cross-validation results are given in Table 2. The correlations between candidates' BV_c and their future DRP were the highest when BV_c were estimated with ssGBLUP for both bulls and cows in all breeds. On average, correlations between candidates' BV_c and their

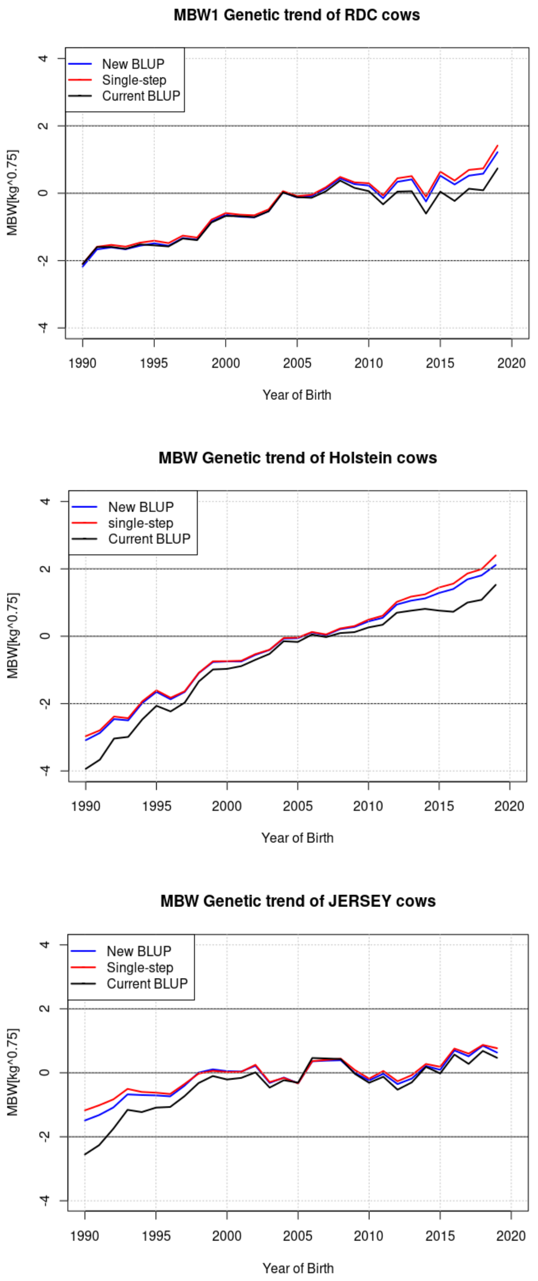

Figure 1. Genetic trends of MBW (G)EBVs by birth year in Red Dairy Cattle, Holstein, and Jersey cows. Trends of MBW from the single-step model (red line), new BLUP model (blue line), and current BLUP-model used by NAV (black line) are expressed as standardized breeding values for cows born between the years 2005-2007.

DRP were 20.0 and 14.0 percentage units higher for the single step models compared with BLUP models in cows and bulls, respectively.

Table 2. Cross-validation and Legarra-Reverter (LR) estimates: Correlation between DRP and BV of canidates ($r_{(DRP,\ BVc)}$), regression coefficient (b_1), validation reliability (r^2_{cv}), and coefficient of determination (R^2) for RDC, HOL and JER bull and cow candidate groups by different models.

		Cross-validation			LR	
		$r_{(DRP,\ BVc)}$	b_1	r^2_{cv}	b_1	R^2
	BLUP[1]					
RDC	cows	0.25	1.04	0.19	1.04	0.43
	bulls	0.63	1.03	0.43	1.03	0.39
HOL	cows	0.19	0.87	0.11	0.92	0.33
	bulls	0.60	0.88	0.40	0.88	0.36
JER	cows	0.14	0.97	0.09	0.94	0.28
	bulls	0.46	0.91	0.23	0.98	0.21
	ssGBLUP[2]					
RDC	cows	0.48	1.32	0.73	1.08	0.80
	bulls	0.76	0.92	0.61	1.06	0.71
HOL	cows	0.36	1.08	0.39	1.00	0.70
	bulls	0.74	0.85	0.60	0.98	0.67
JER	cows	0.34	1.39	0.51	1.03	0.59
	bulls	0.61	0.88	0.40	1.08	0.59

[1] BLUP = Best Linear Unbiased Prediction

[2] ssGBLUP = Single-step Genomic BLUP

BVc = Breeding Value for candidates

The optimal prediction of genetic merit of young individuals should have a regression coefficient (*b1*) of one. In our cross-validation of bulls, the *b1* estimates obtained using the BLUP model were slightly better compared to those from the ssGBLUP model (Table 2). This difference is likely because our DRPs were based on the BLUP model. Using the Legarra-Reverter (LR) validation method, all *b1* values were close to one for both bulls and cows in both the BLUP and ssGBLUP models, except in HOL. The BLUP model appeared to slightly overpredict the future breeding values for HOL candidate cows and bulls.

The validation reliabilities (r^2_{cv}) for the BLUP model varied between 0.23 and 0.43 for RDC, JER, and HOL bulls, and between 0.09 and 0.19 for cows (Table 2). In contrast, the r^2_{cv} for the ssGBLUP model varied between 0.40 and 0.61 for bulls, and between 0.39 and 0.73 for cows. This indicates that, across all breeds, the validation reliability was on average 18.3 percentage units higher for the single-step model in bulls and 41.3 percentage units higher in cows. Additionally, using the LR method, the coefficients of determination (R^2) were on average 33.7 percentage units higher for the single-step model in bulls and 35.0 percentage units higher in cows. These results suggest a better predictive ability of the model with genomic data.

Conclusions

In this study we developed models that include carcass weight data as correlated information for predicting genomic breeding values for MBW. The CARW data significantly increased the amount of phenotypic information used for the genomic evaluation in all Nordic breeds. This, along with the development of single-step genomic prediction, contributes positively to the reliability and unbiasedness of predictions of breeding values for maintenance. As a result, animals will receive more accurate breeding values.

Acknowledgments

This work was a part of the A^{++}COW -project: “Breeding dairy cattle for resource efficiency and environmental sustainability” (2019-2023) and was supported by the Finnish Ministry of Agriculture (453/03.01.02/2018), VikingGenetics Ltd, Valio Ltd, Nordic Cattle Genetic Evaluation (NAV), and Faba co-op. We thank the NAV and VikingGenetics for providing the phenotypic and genomic data for this study.

References

Koivula, M., Strandén, I., Aamand, G.P. and Mäntysaari, E.A. 2021. Practical implementation of genetic groups in single-step genomic evaluations with Woodbury matrix identity based genomic relationship inverse. *J. Anim. Sci. (*104), 10049–10058. DOI: 10.3168/jds.2020-19821

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), 1–18. DOI: 10.1186/s12711-018-0426-6

Lidauer, M.H. Leino, A.-M., Stephansen, R.S., Pösö, J., Nielsen, U.S., Fikse, W.F., Aamand, G.P. 2019. Genetic evaluation for Maintenance – Towards breeding values for Saved feed in Nordic dairy cattle. *Interbull Bulletin* (55), 21–25.

Mehtiö, T., Pitkänen, T., Leino, A.-M., Mäntysaari, E.A., Kempe R., Negussie, E. & Lidauer, M.H. 2021. Genetic analyses of metabolic body weight, carcass weight and body conformation traits in Nordic dairy cattle. *Animal* (15), 100398. DOI: 10.1016/j.animal.2021.100398

Mehtiö, T., Mäntysaari, P., Negussie, E., Kempe, R., Sevón-Aimonen, M.L., Chegini, A., Hietala, S., Kostensalo, J. and Lidauer, M. 2023. Breeding Dairy Cattle for Resource Efficiency and Environmental Sustainability: Final report of the A++COW -project (2019-2023). *Natural resources and bioeconomy studies* 54/2023. Natural Resources Institute Finland (Luke). Helsinki. 94 p.

Mäntysaari, E.A., Evans, R.D. and Strandén, I. 2017. Efficient single-step genomic evaluation for a multibreed beef cattle population having many genotyped animals. *J. Anim. Sci.* (95), 4728–4737. DOI: 10.2527/jas2017.1912

NAV 2022. NAV official genetic evaluation of Dairy Cattle – data and genetic models. May 2022, 14th edition. 85 p.

Stephansen, R.B., Lidauer, M.H., Nielsen, U.S., Pösö, J., Fikse, F., Manzanilla-Pech, C.I.V. and Aamand, G.P. Genomic prediction of residual feed intake in the Nordic breeds using data from research herds and 3D cameras in commercial herds. *Interbull Bulletin* (56), 162–166.

VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* (91), 4414–4423. DOI: 10.3168/jds.2007-0980