# Nordic Holstein single-step test day model using metafounders and blending of foreign information

## M. Koivula[1*], I. Strandén[1], G. P. Aamand[2] & E. A. Mäntysaari[1]

[1]Natural Resources Institute Finland (Luke), FI-31600 Jokioinen, Finland; [2]NAV Nordic Cattle Genetic Evaluation, 8200 Aarhus N, Denmark; *Minna.Koivula@luke.fi

## Abstract

The aim of the study was to compare the metafounders approach (MF) to conventional genetic unknown parent groups (UPG) and to estimate the effect of blending the external information. All results were compared also to results from the official EBV model (AM). The results showed that both the traditional UPG and the MF approach can be implemented in ssGTBLUP together with the blending of foreign information. However, the MF approach was slightly more effective in reducing bias in young bulls.

## Introduction

The use of sires from international populations is common in dairy cattle breeding. The exchange of sire genotypes allows national breeding programs to properly integrate multi-trait cross-country assessments (MACE; Schaeffer, 1994) by Interbull (Uppsala, Sweden) into domestic evaluations. The integration of bull information by blending MACE yield indices into the Nordic test-day (TD) model was introduced by Pitkänen et al. (2020).

In single-step models, an unsolved model problem is how to build genomic ($\mathbf{G}$) and pedigree ($\mathbf{A}_{22}$) relationship matrices which meet the theoretical assumptions of the same scale and equal base populations. For example, base population allele frequencies (AF) can be used (VanRaden, 2008), and elements of $\mathbf{G}$ can be scaled and centered to have on average the same diagonal and off-diagonal elements as in $\mathbf{A}_{22}$ (Christensen, 2012). The metafounder (MF) approach proposed by Legarra et al. (2015) attempts to make the $\mathbf{A}$ and $\mathbf{A}_{22}$ matrices compatible with the $\mathbf{G}$ matrix. In the MF approach, allele frequencies (AF) equal to 0.5 for all markers are assumed in the $\mathbf{G}$ matrix (Garcia-Baccino et al. 2017) and the $\mathbf{A}$ matrix has MF or pseudo-individuals with self-relationships. The MF are like unknown parent groups (UPG) but allow a related base population with nonzero inbreeding coefficients (e.g., Legarra et al. 2015; Garcia-Baccino et al. 2017). The relationships within and between the MF are modelled by a $\mathbf{\Gamma}$ matrix, which thereafter is used in forming the pedigree-based relationship matrix ($\mathbf{A}_\Gamma$).

The objectives of this study were to compare the performance of the MF approach to the corresponding model with UPG and to address the effect of integration of external information into Nordic Holstein test day (TD) model evaluations. The MF was applied, and MACE information was included in the single-step (model called $GT_B\_MF$), and in the animal model ($AM_B\_MF$). To compare method differences, we run separately also ssGTBLUP with MACE information and the QP transformation where UPGs were included with the QP transformation ($GT_B$), and ssGTBLUP without MACE information using either the QP transformation (GT) or the metafounders (GT_MF). All the results were compared also to results from the official EBV model (AM).

## Material & methods

The official Nordic HOL milk production evaluation data obtained from the Nordic Cattle Genetic Evaluation (NAV) were used in all analyses of this study. The official evaluation

includes TD records from milk, fat and protein production from the first three lactations as multiple traits. The TD data included 8.5 million cows with a total of 173.7 million records and 10.9 million animals in the pedigree. There were 274 145 genotyped animals. After the same edits used in current NAV HOL genomic evaluations, 46 342 SNP markers on the 29 bovine autosomes were available for further analysis.

*Models.* The Nordic multiple traits reduced rank random regression TD model with heterogeneous variance (HV) correction (Lidauer et al., 2015) was used. The model specifies 27 traits: 3 countries (DFS: Denmark, Finland and Sweden), 3 yield traits, 3 lactations. Each animal receives 15 solutions for the random regression effects. With the data used, the mixed model equations (MME) had 155.8 million animal equations and a total of 444.2 million equations. For the validation of bull (G)EBVs, the last four years of observations were removed from the full data.

All models were solved with MiX99 software (Strandén and Lidauer, 1999). The single-step model used ssGTBLUP with efficient computations (Mäntysaari et al., 2017). In all single-step models, the genomic relationship matrix was based on VanRaden method 1 with a 30% residual polygenic proportion, AF of 0.5 for all the markers, and in UPG models diagonals in $\mathbf{G}$ were scaled by trace($\mathbf{A_{22}}$)/trace($\mathbf{G}$).

Four single-step models were tested: GT, $GT_B$, GT_MF and $GT_B$_MF. In the GT and $GT_B$ models, 176 genetic groups were included using QP transformation (detailed description in Koivula et al., 2021). The GT_MF and $GT_B$_MF models used the MF approach. The MF approach needs a covariance matrix, denoted $\mathbf{\Gamma}$, for the metafounders. The 176 UPGs were used as metafounders. The UPGs were based on 4 breed groups (HOL, RDC, JER, other) and 5 country of origin groups within Holstein (HOLdnk, HOLswe, HOLfin, HOLother, HOLred). Within each of these 9 sources, UPG were further grouped by birth year decade and by selection path when appropriate. For the $\mathbf{\Gamma}$ matrix of the 9 base metafounders, we defined a covariance function (CF) model (Kirkpatrick et al., 1994) described in Kudinov et al. (2021). After solving the $\mathbf{\Gamma}$ matrix for the 176 MF, we computed the $\mathbf{\Gamma}$ matrix compliant inbreeding coefficients needed for the inverses $\mathbf{A_\Gamma}$ and $\mathbf{A_{22}}^\Gamma$ in the MME.

Foreign information was included in the $GT_B$ and $GT_B$_MF models. The external information from MACE was included as described by Pitkänen et al. (2020). A bull was considered to have additional information in MACE when its reliability for milk, protein and fat indices in MACE were at least 0.01 units higher compared to the Nordic evaluation reliability. The integration process had three steps: 1) effective record contributions (ERC) in MACE and in the national evaluations were calculated for the selected bulls; 2) yield indices and ERC were used to calculate multi-trait deregressed proofs (DRPs) separately using the Nordic and the MACE data. Finally, 3) based on the two DRPs and the two ERCs, pseudo-observations were calculated for the selected 31 779 bulls and were included in the TD data. In addition, the official animal model (AM) and animal model $AM_B$_MF using blending with MF were performed.

## Results and Discussion

Table 1 illustrates the linear regression validation results (LR, Legarra & Reverter, 2018) from the different models for 524 DFS Holstein validation bulls. The level differences in the (G)EBV predictions were corrected by standardizing the mean of cows born 2007 to be the same in all models. The $b_0$ column has the mean difference (kg) between the full and reduced run (G)EBVs. This illustrates the realized bias between the full and reduced evaluation. The results indicate that the use of the MF approach led to a somewhat lower bias and higher coefficient of correlation ($R^2$) in the single-step evaluations. The regression coefficient ($b_1$) was on average

2% units higher for the MF models than for the QP models suggesting somewhat less overdispersion of GEBVs. These results support the earlier findings that the MF approach may improve the single-step evaluations. The inclusion of external information also improved the evaluations and together with the MF approach gave the best validation results. The blending of foreign information had, in general, a positive effect on the $R^2$ values but, on the other hand, does not improve the bias problem.

**Table 1. Bull LR validation (Bulls=524) results. Regression coefficients ($b_1$) and coefficient of correlation ($R^2$) from different models. The $b_0$= mean(Full_(G)EBV – reduced_(G)EBV). The single-step models: unknown parent groups (UPG) using the QP transformation (GT) or the metafounder (MF) approach (GT_MF), UPG using the QP transformation and the blending of foreign information ($GT_B$), and the MF approach with the blending of foreign information ($GT_B$_MF). The animal models: the official model (AM), and the MF approach with the blending of foreign information ($AM_B$_MF).**

|         | Model      | $b_0$    | $b_1$ | $R^2$ |
|---------|------------|----------|-------|-------|
| Milk    | AM         | -101.7   | 0.84  | 0.32  |
|         | $AM_B$_MF  | -109.86  | 0.89  | 0.35  |
|         | GT         | -319.84  | 0.87  | 0.67  |
|         | GT_MF      | -272.31  | 0.89  | 0.68  |
|         | $GT_B$     | -306.32  | 0.86  | 0.69  |
|         | $GT_B$_MF  | -267.07  | 0.88  | 0.71  |
| Protein | AM         | 0.80     | 0.74  | 0.24  |
|         | $AM_B$_MF  | -0.31    | 0.83  | 0.28  |
|         | GT         | -11.10   | 0.81  | 0.63  |
|         | GT_MF      | -9.71    | 0.83  | 0.64  |
|         | $GT_B$     | -11.04   | 0.79  | 0.64  |
|         | $GT_B$_MF  | -9.73    | 0.82  | 0.67  |
| Fat     | AM         | -2.18    | 0.73  | 0.23  |
|         | $AM_B$_MF  | -2.55    | 0.80  | 0.26  |
|         | GT         | -16.16   | 0.82  | 0.64  |
|         | GT_MF      | -14.67   | 0.85  | 0.65  |
|         | $GT_B$     | -14.76   | 0.81  | 0.66  |
|         | $GT_B$_MF  | -13.38   | 0.84  | 0.68  |

The genetic trend of protein for the genotyped Nordic Holstein bulls is presented in Figure 1. The figure shows that with the full data the genetic trends were similar with different single-step models, and the correlations between GEBVs varied from 0.985 to 1.00. The differences between the models can be observed in the reduced data runs. The most obvious difference is seen between the models using the QP transformation and the MF approach. The MF approach model does not increase the genetic trend of young bulls as much as the QP model. Also, the blending of foreign information increases the genetic trend of young bulls, but the increase is not as high when the MF approach is used. The animal model EBVs showed a much lower genetic trend compared to the single-step models, but with the MF approach and the inclusion of MACE information, the $b_1$ and $R^2$ also improved in the animal model.

In conclusion, it seems that both the traditional genetic groups and the MF approach can be implemented in ssGTBLUP together with the blending of foreign information. However, it seems that the MF approach can be slightly more efficient in reducing bias in young bulls.
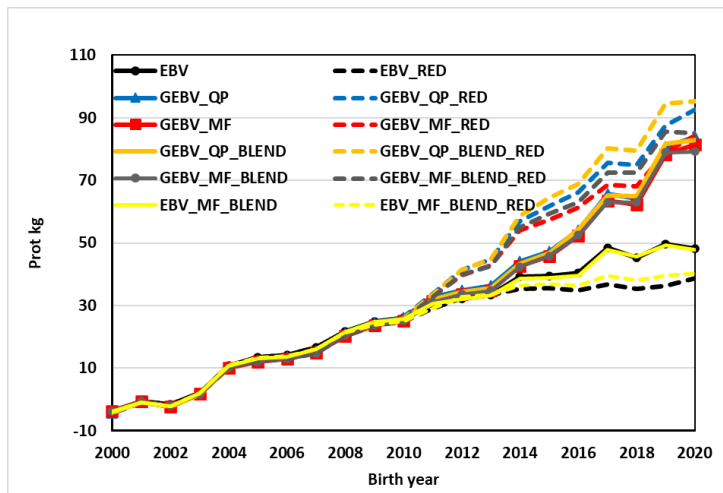
**Figure 1. Genetic trends by birth year for protein from different models. The single-step models: unknown parent groups (UPG) using the QP transformation (GT) or the metafounder (MF) approach (GT_MF), UPG using the QP transformation and the blending of foreign information (GT$_B$) and the MF approach with the blending of foreign information (GT$_B$_MF). The animal models: the official model (AM), and the MF approach with the blending of foreign information (AM$_B$_MF). The *RED indicates the reduced data run.**

## References

Christensen O.F. (2012) Genet. Sel. Evol. 44: 37. https://doi.org/10.1186/1297-9686-44-37.

Garcia-Baccino C. A., Legarra A., Christensen O.F., Misztal I., Pocrnic I., et al. (2017) Genet. Sel. Evol. 49: 34. https://doi.org/10.1186/s12711-017-0309-2.

Kirkpatrick, M., Hill W.G. and Thompson R. (1994) Genet. Res. 64: 57–69. https://doi.org/10.1017/S0016672300032559.

Koivula, M., Strandén, I., Aamand, G.P, and Mäntysaari, E.A. (2021) J. Dairy Sci., 104: 10049-10058. https://doi.org/10.3168/jds.2020-19821.

Kudinov A.A., Koivula M., Strandén I., Aamand G.P. and Mäntysaari E.A. (2021) Interbull Bulletin 56:174-179.

Legarra, A., Christensen, O.F., Vitezica, Z.G. Aguilar, I. and Misztal, I. (2015) Genetics 200: 455-468. https://doi:10.1534/genetics.115.177014.

Legarra A. and Reverter, A. (2018) Genet. Sel. Evol, 50:53. https://doi.org/10.1186/s12711-018-0426-6.

Lidauer, M., J. Pösö, J. Pederson, J. Lassen, P. Madsen, et al. (2015) J. Dairy Sci. 98:1296–1309. https://doi.org/10.3168/jds.2014-8307.

Mäntysaari, E.A., Evans, R.D. and Strandén, I. (2017) J. Anim. Sci. 95:4728-4737. https://doi.org/10.2527/jas2017.1912.

Pitkänen, T.J., Koivula, M.,Strandén, I., Aamand, G.P. and Mäntysaari, E.A. (2020) EAAP Book of Abstracts, Vol. 26 Virtual Meeting.

Schaeffer L. R. (1994) J. Dairy Sci., 77: 2671-2678. https://doi.org/10.3168/jds.S0022-0302(94)77209-X.

Strandén I. and Lidauer M. (1999) J. Dairy Sci. 82:2779-2787. https://doi.org/10.3168/jds.S0022-0302(99)75535-9.

VanRaden P.M. (2008) J. Dairy Sci. 91:4414-4423. https://doi.org/10.3168/jds.2007-0980.