

Meet the Author



Dawid Słomian

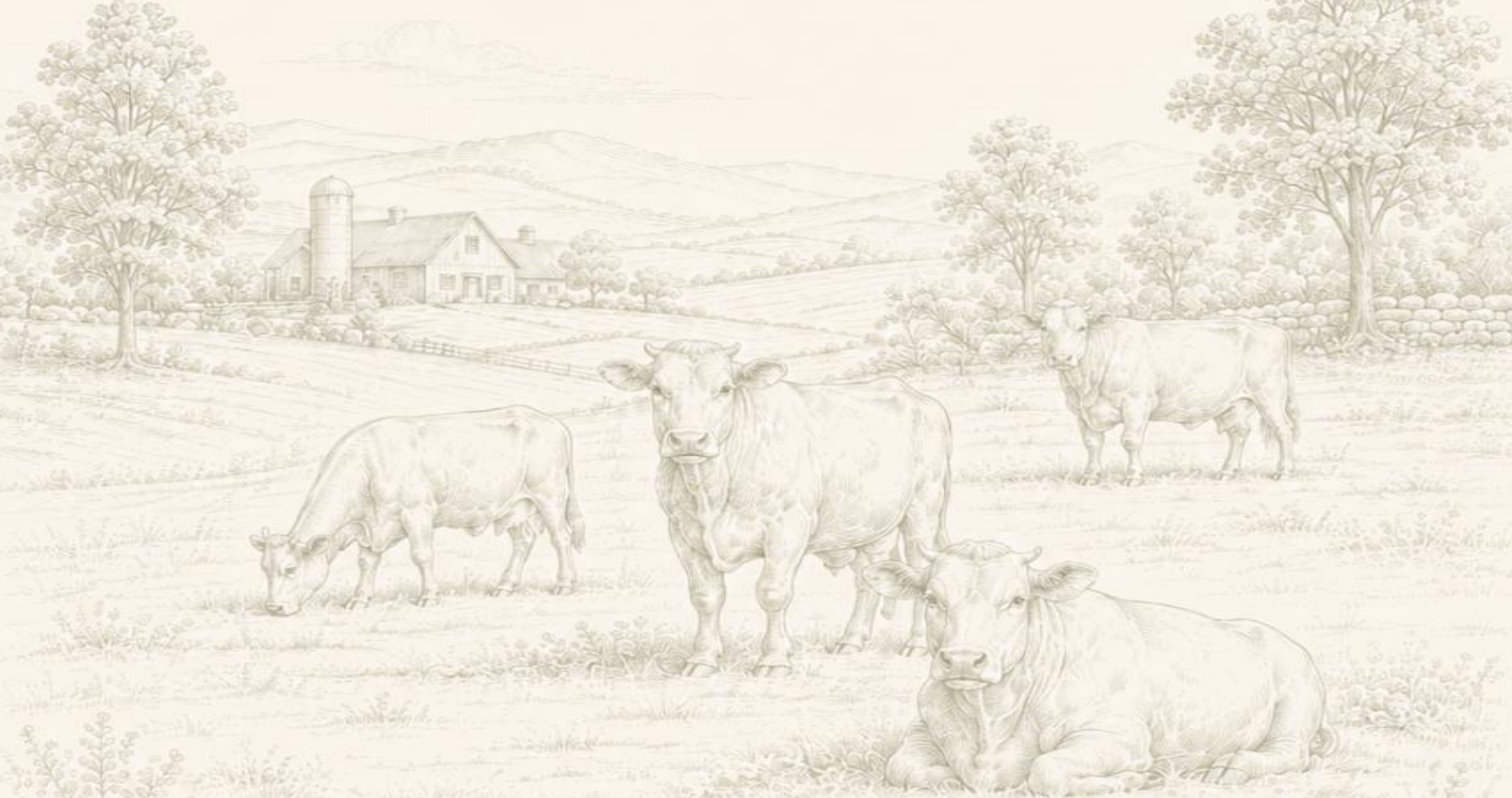
I am a PhD student at the National Research Institute of Animal Production – National Research Institute in Poland. In my research, I focus mainly on routine genetic evaluation, with particular emphasis on single-step genomic models in dairy cattle. My scientific interests also include bioinformatics, genetics, and the use of genomic data in animal breeding.



Modelling missing parents in single-step test-day SNP-BLUP evaluation of dairy cattle

D. Słomian^{1}, K. Żukowski¹, M. Skarwecka¹, J. Ten Napel², J. Vandenplas², J.Szyda^{1,3}*

*1 National Research Institute of Animal Production, Balice, Poland;
2 Wageningen University & Research, Wageningen, Netherlands;
3 Wroclaw University of Science, and Technology, Poland;*



Project Question

How does the modelling of missing parents affect genomic breeding value prediction in single-step dairy cattle evaluation?

Abstract

Introduction

Materials & Methods

Results

Conclusion

Modelling missing parents in single-step test-day SNP-BLUP evaluation of dairy cattle

Abstract

In many countries, single-step genomic models have replaced multiple step models for routine evaluation. These models use all available information on animals' phenotypes, genotypes, and pedigrees, yet missing parental information in pedigrees remains a challenge affecting genomic breeding value (GEBV) predictions. Therefore, the choice of method for handling missing parents can affect the prediction of breeding values. Here, we compared three approaches to model missing parental information for three levels of missing pedigree data:

- **P_Real** – pedigree from the routine evaluation,
- **P_2010** – at least 20 percent of dams and 10 percent of sires born before 2019 were set to missing,
- **P_4020** – at least 40 percent of dams and 20 percent of sires born before 2019 were set to missing.

Missing parents' information was expressed through missing codes in the raw pedigree (**RP**), by defining genetic groups (**GG**) that represent missing parents grouped based on year of birth, sex, and country of

origin, or by defining metafounders (**MF**), which represent missing parents grouped by average genetic relationships estimated from the genomic information of their descendants. The genomic breeding values for fat yield were estimated using the single-step test-day SNP-BLUP model, implemented with MiXBLUP software.

For the considered scenarios, results were presented separately for sires and dams, as well as for genotyped and ungenotyped individuals. We observed differences in prediction quality between genotyped and ungenotyped animals. While GEBV predictions for the former were generally stable across scenarios, the predictions for ungenotyped individuals varied. Particularly, the removal of parental information led to less stable results when missing parental information was expressed by **MF**, where insufficient pedigree completeness resulted in an overestimation of the genetic trend. In conclusion, for informative pedigrees with a small percentage of missing parents, the incorporation of **GG** and **MF** results in very similar GEBV predictions, however **GG** appear to be a more robust approach for ungenotyped individuals in highly incomplete pedigrees.

Modelling missing parents in single-step test-day SNP-BLUP evaluation of dairy cattle

Introduction

Incomplete pedigree



Method of dealing with
missing parents



GEBV prediction



Validation

Background

Single-step genomic evaluation combines phenotype, genotype, and pedigree information in one model. In dairy cattle breeding, incomplete pedigree information is common and may affect the accuracy and bias of genomic breeding values.

Problem

Missing parents in pedigree data can be handled in different ways. The traditional solution is to use genetic groups, while a newer approach uses metafounders, which model relationships among unknown ancestors using genomic information.

Aim

The aim of this study was to compare three approaches for handling missing parents in the pedigree in a single-step SNP-BLUP test day model for fat yield in the Polsih Holstein population.

Modelling missing parents in single-step test-day SNP-BLUP evaluation of dairy cattle

Materials and Methods

Data

The study used data from the Polish national evaluation of fat yield from April 2024. The full phenotype dataset included over 63 million test-day records from 3.7 million cows. A truncated dataset, excluding the youngest animals, was used for validation. Genomic information was available for 181,999 animals, and pedigree information for 4.7 million individuals.

Pedigree Scenarios

P_Real – original pedigree

P_2010 – increased missing parents: ~20% dams and ~10% sires

P_4020 – highly incomplete pedigree: ~40% dams and ~20% sires

Method to deal with missing parents

RP – raw pedigree, missing parents kept unknown

GG – genetic groups based on birth year, sex, and country

MF – metafounders based on genomic relationships of descendants

Model

$$y = Xh + Wf + Vp + Vu + e,$$

y – test-day records for fat yield

h – fixed effects of herd-test_date-parity-milking_frequency

f – fixed lactation curve coefficients

p – permanent environmental random effect

u – random additive genetic effect

X, W, V – design matrices for effects

Validation

GEBVs for fat yield were predicted using a single-step test-day SNP-BLUP model in MiXBLUP 3.0. Validation was performed using the GEBVtest method to compare GEBVs from the full and truncated datasets.

Modelling missing parents in single-step test-day SNP-BLUP evaluation of dairy cattle

Results

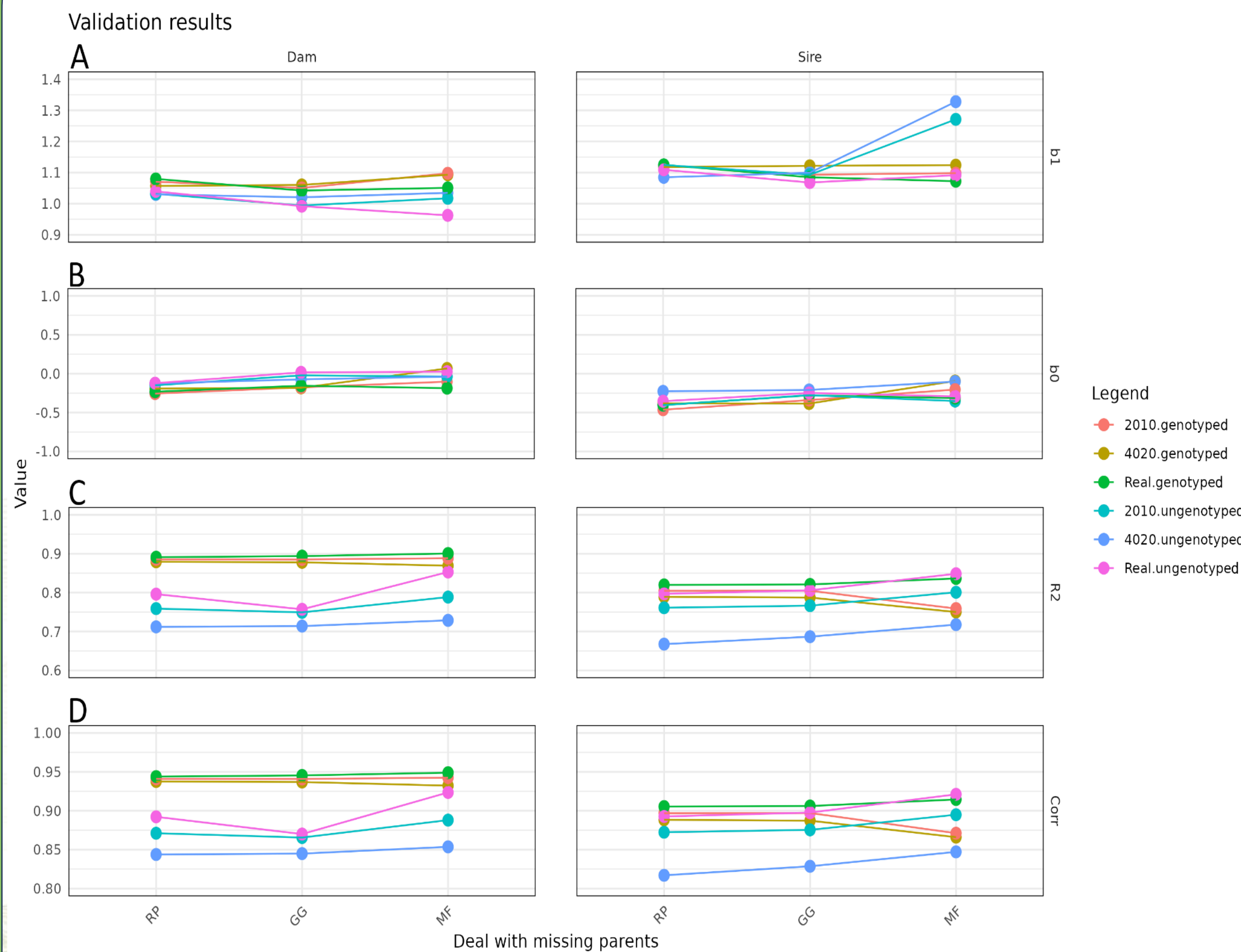


Figure 1. Validation of GEBV predictions across pedigree scenarios and missing-parent methods. Regression slope, intercept, R^2 , and Pearson correlation were compared separately for genotyped and ungenotyped animals.

- Most regression slopes were close to 1, indicating generally acceptable prediction dispersion.
- Metafounders showed overdispersion for ungenotyped bulls when pedigree incompleteness increased.
- Genotyped cows showed more stable prediction accuracy across scenarios.

- Mean GEBV increased after 2010, especially for sires.
- The highest genetic trend was observed for **P_Real** when missing parents were modelled using metafounders.
- However, under highly incomplete pedigrees, metafounders may overestimate genetic trends in ungenotyped animals.

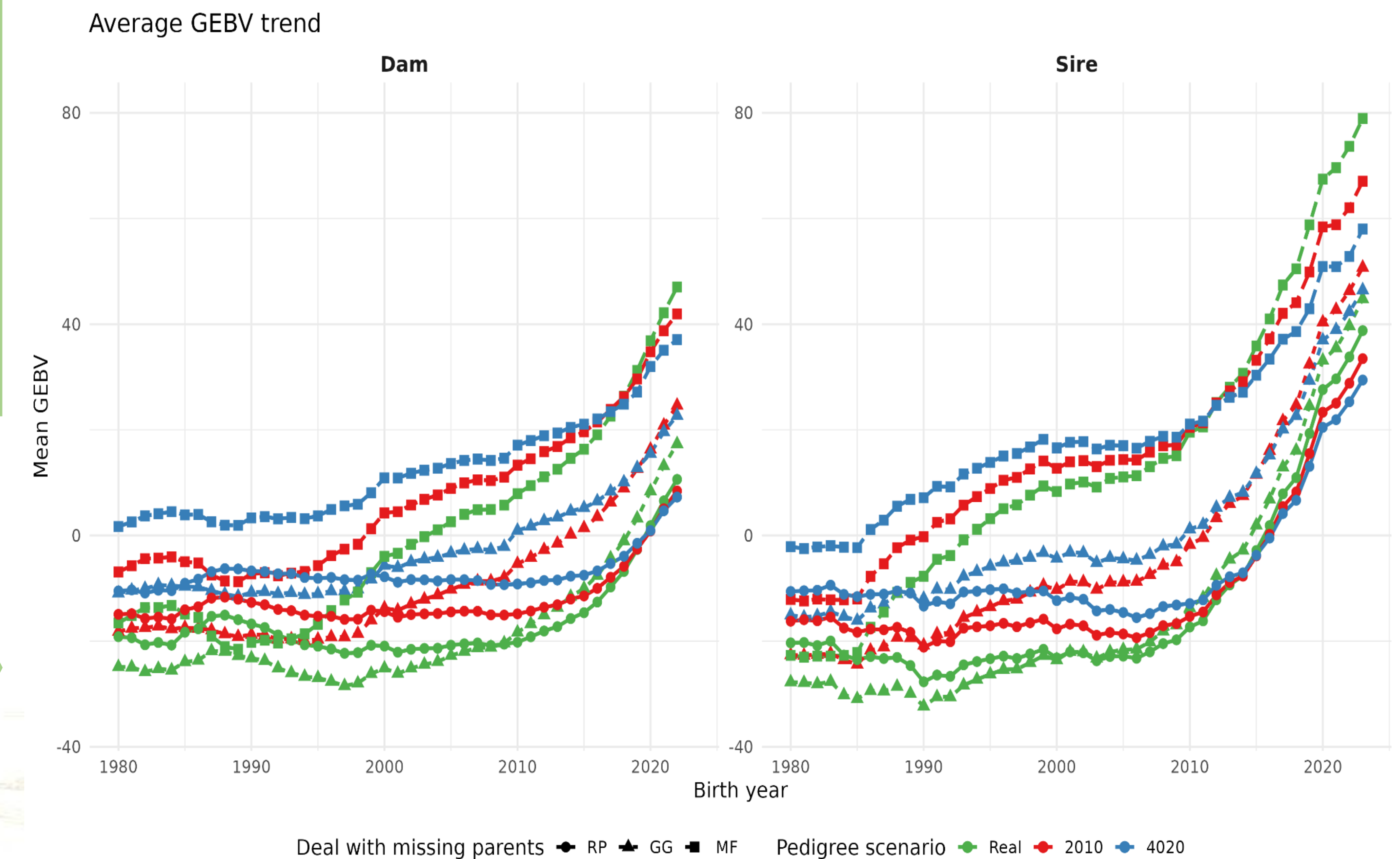


Figure 2. Average GEBV trends by sex, pedigree scenario, and missing-parent method.

Modelling missing parents in single-step test-day SNP-BLUP evaluation of dairy cattle

Conclusion

- The method used to handle missing parents in the pedigree affects GEBV prediction, especially when pedigree incompleteness increases.
- Metafounders can improve prediction in some cases, particularly for animals with genomic information.
- However, when the proportion of missing parents increases, metafounders may lead to biased predictions for ungenotyped animals.
- For genotyped individuals, differences between missing-parent approaches were small.
- Correct modelling of missing parents becomes increasingly important as incomplete pedigree records become more frequent.

Take-home message

Metafounders are a useful tool for modelling missing pedigree information, but their application should be carefully evaluated, especially for ungenotyped animals and highly incomplete pedigrees.

References

- Legarra, A., Bertrand, J., Strabel, T., Sapp, R., Sánchez, J., & Misztal, I. (2007). Multi-breed genetic evaluation in a Gelbvieh population. *J. Anim. Breed. Genet*, 124(5), 286–295. <https://doi.org/10.1111/j.1439-0388.2007.00671.x>
- Legarra, A., Christensen, O. F., Aguilar, I., & Misztal, I. (2014). Single Step, a general approach for genomic selection. *Livest. Sci.*, 166, 54–65. <https://doi.org/10.1016/j.livsci.2014.04.029>
- Legarra, A., Christensen, O. F., Vitezica, Z. G., Aguilar, I., & Misztal, I. (2015). Ancestral relationships using metafounders: finite ancestral populations and across population relationships. *Genetics*, 200(2), 455–468. <https://doi.org/10.1534/genetics.115.177014>
- Liu, Z., Reinhardt, F., Bünger, A., & Reents, R. (2004). Derivation and calculation of approximate reliabilities and Daughter Yield-Deviations of a random Regression Test-Day model for genetic evaluation of dairy cattle. *J. Dairy Sci*, 87(6), 1896–1907. [https://doi.org/10.3168/jds.s0022-0302\(04\)73348-2](https://doi.org/10.3168/jds.s0022-0302(04)73348-2)
- Mäntysaari, E. A., Evans, R. D., & Strandén, I. (2017). Efficient single-step genomic evaluation for a multibreed beef cattle population having many genotyped animals1. *J. Anim. Sci.*, 95(11), 4728–4737. <https://doi.org/10.2527/jas2017.1912>
- Strandén, I. (2014). *RelaX2 program for pedigree Analysis, User's Guide for Version 1.65*.
- Vandenplas, J., Veerkamp, R., Calus, M., Lidauer, M., Strandén, I., Taskinen, M., Schrauf, M., & Napel, J. T. (2022). 358. MiXBLUP 3.0 – software for large genomic evaluations in animal breeding programs. *Proceedings of 12th World Congress on Genetics Applied to Livestock Production*, 1498–1501. https://doi.org/10.3920/978-90-8686-940-4_358