

Data truncation as a tool for increasing computing efficiency in genomic predictions

Fernando Bussiman¹, Ching-Yi Chen², Justin Holl², Andres Legarra³, Ignacy Misztal¹, Daniela Lourenco¹



¹Department of Animal and Dairy Science, University of Georgia, Athens, GA 30602 USA

²Genus PIC, Hendersonville, TN 37075 USA

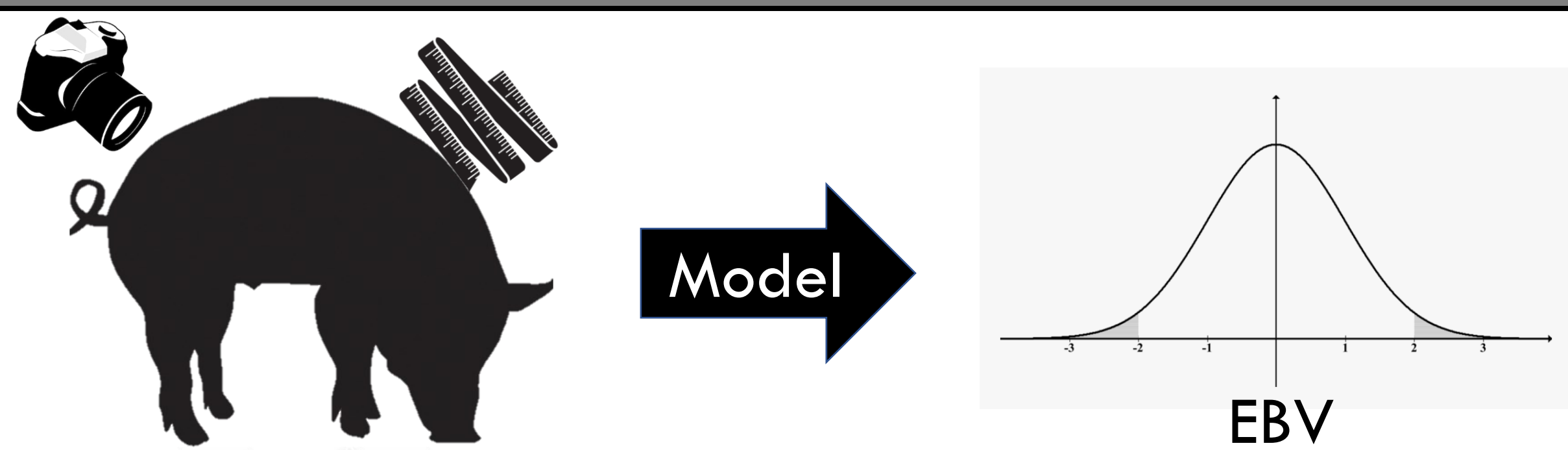
³INRA, UMR1388 GenPhySE, Castanet-Tolosan 31326 France

fob@uga.edu

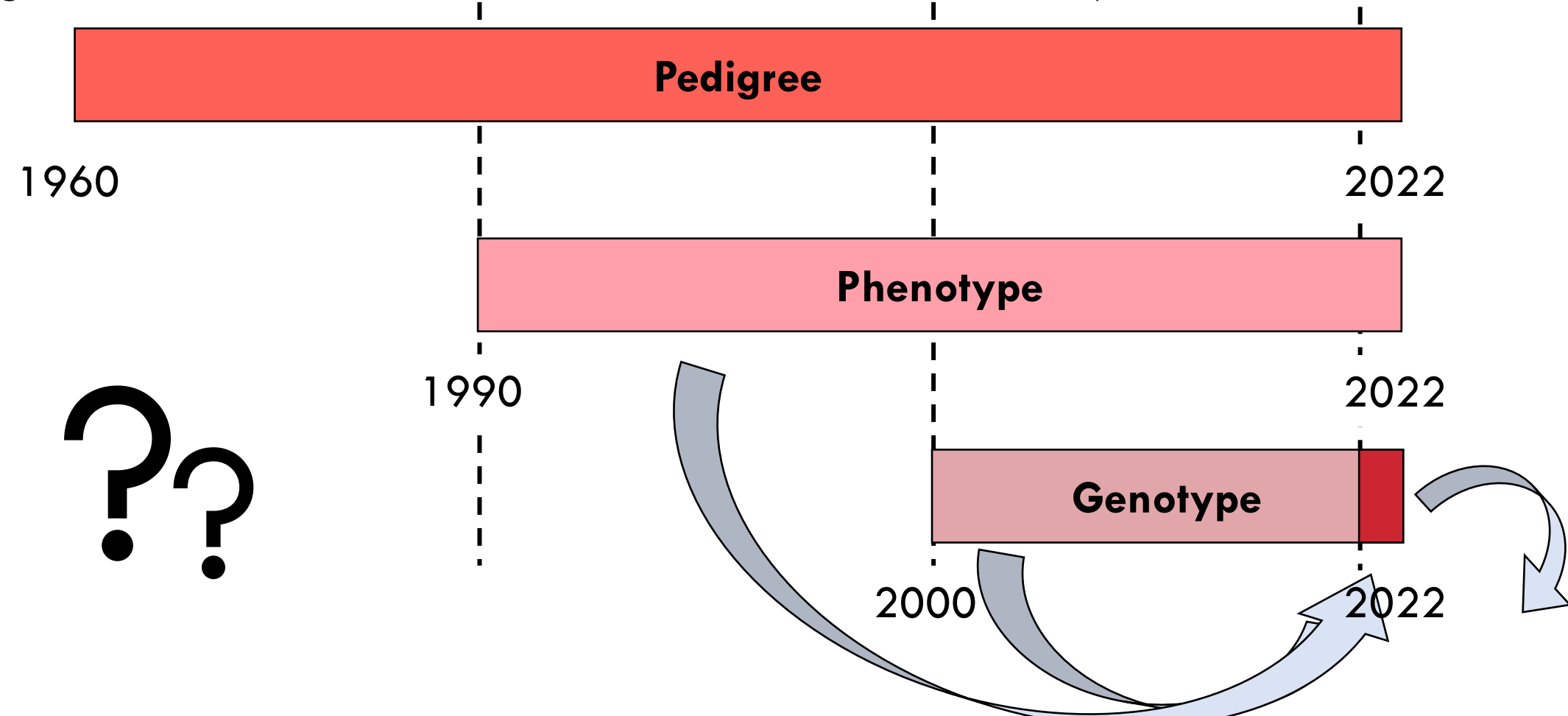


Animal Breeding and Genetics Group
College of Agricultural & Environmental Sciences
UNIVERSITY OF GEORGIA

INTRODUCTION



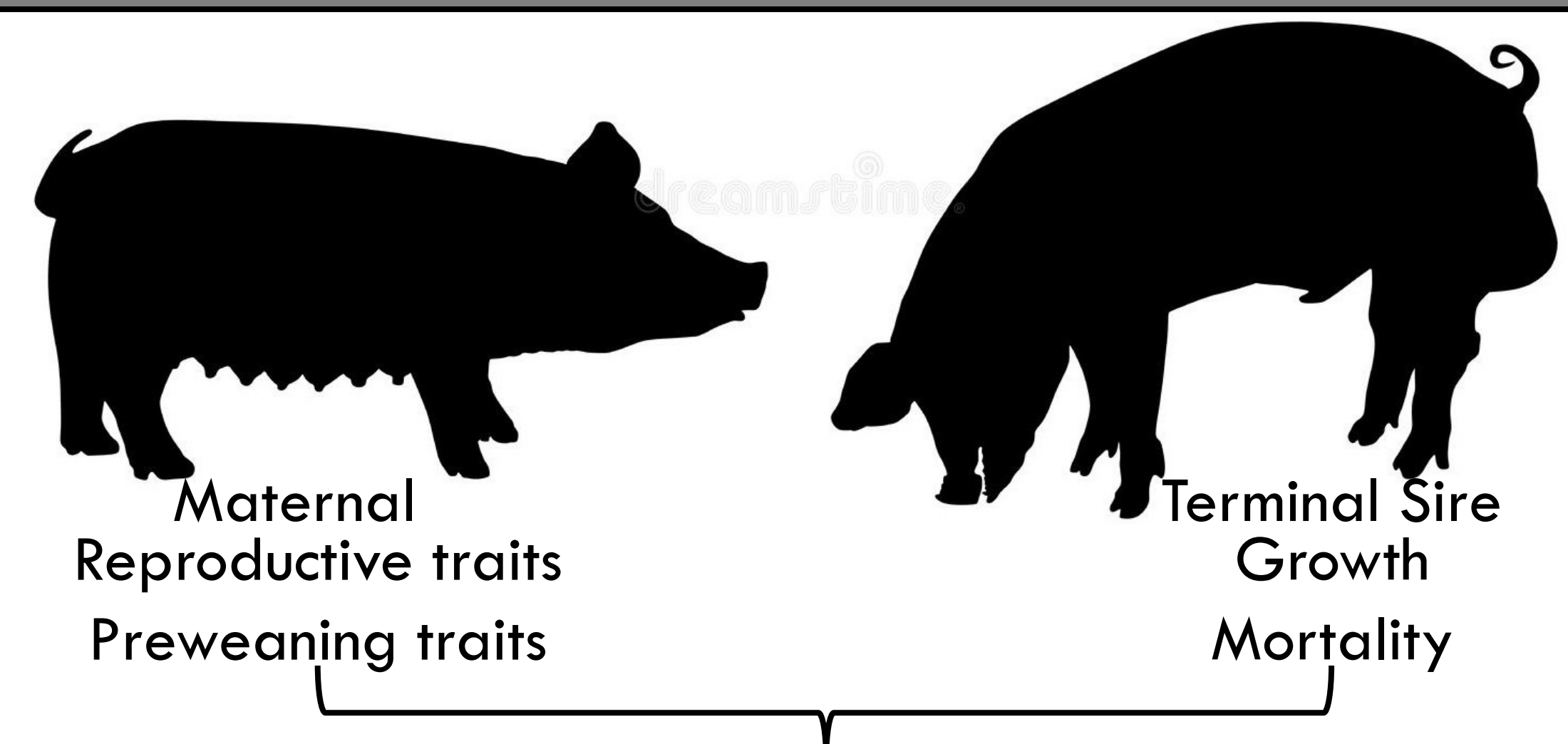
- If the model is right, adding more pedigree generations increases prediction accuracy (Henderson, 1984)
- Models are approximations and contributions from previous generations decay with time (Lourenco et al., 2014)
- More data requires more computing power and time
- Small wrong model decisions accumulate over time (Macedo et al., 2022)



OBJECTIVE

- Investigate how different criteria to remove genotyped animals along with pedigree and phenotype truncation impact the prediction of GEBV of genotyped selection candidates

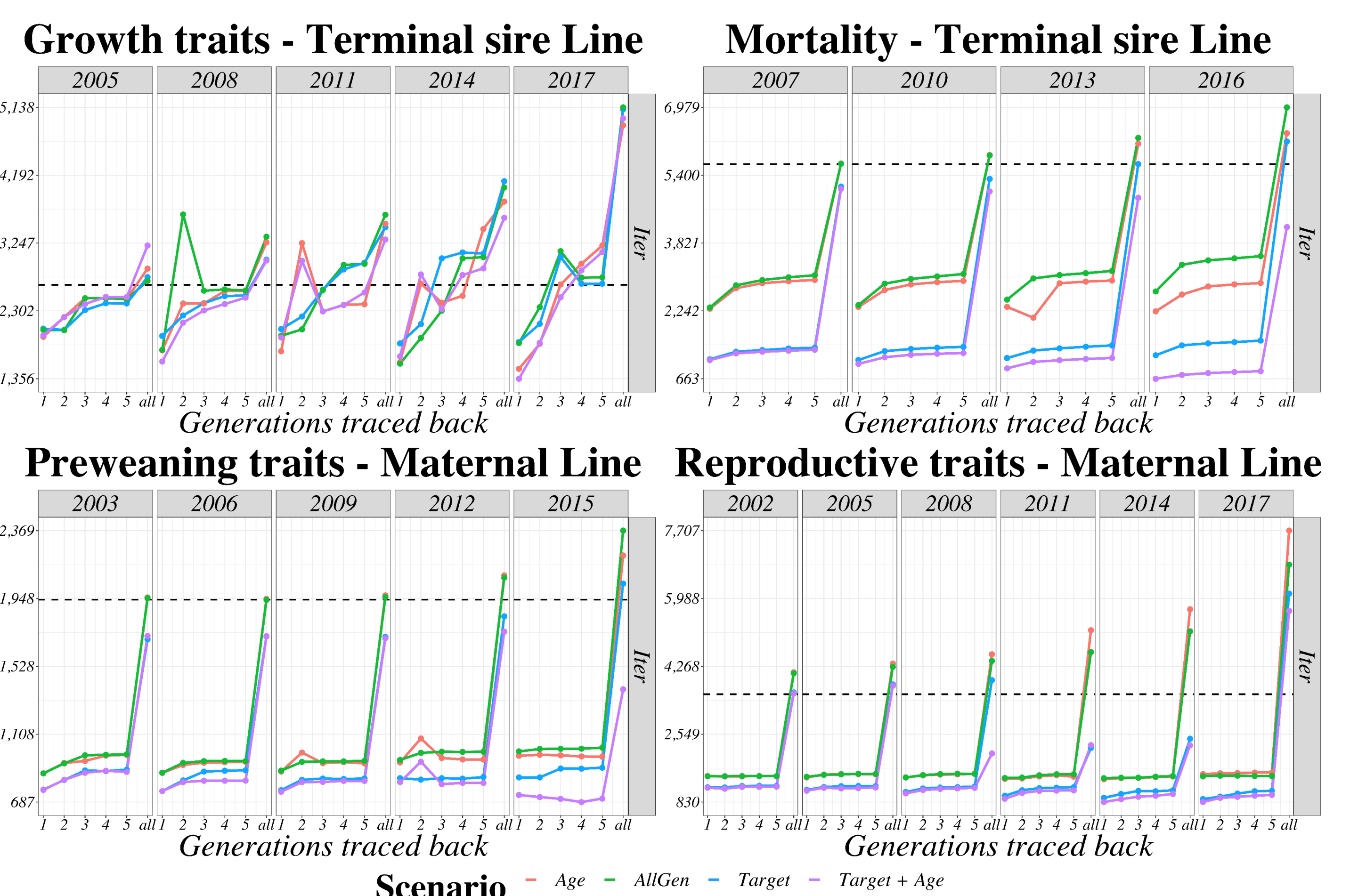
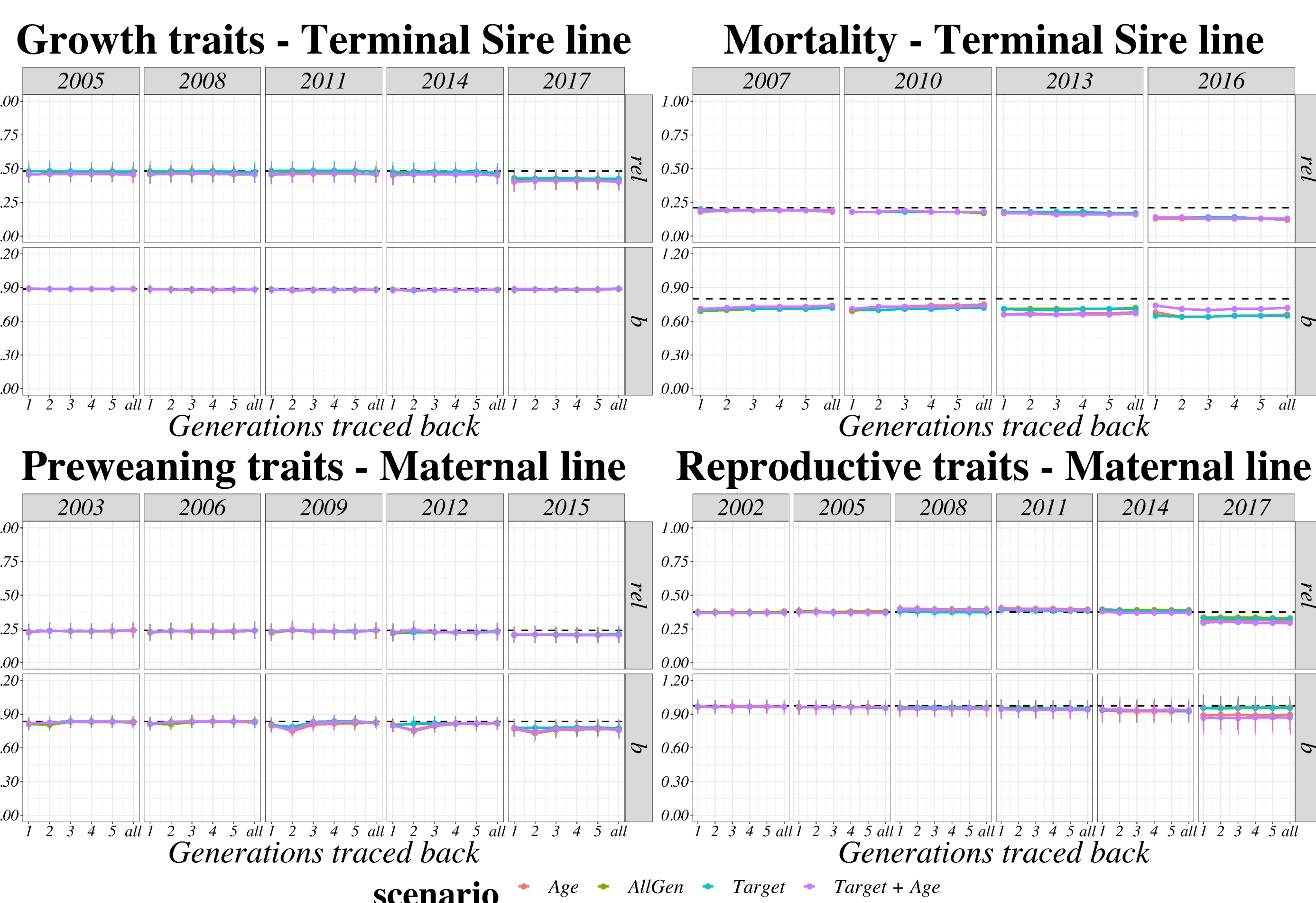
MATERIAL AND METHODS



- **Four scenarios of Genotype truncation:**
 - **Target:** removing genotyped animals without phenotype and progeny
 - **Age:** removing old genotyped animals
 - **Target + Age:** combining Target and Age in this order
 - **AllGen:** no genotype truncation
- **Phenotype truncation:**
 - Old phenotypic information was removed based on generation interval (GI) of three years
- **Pedigree truncation:**
 - Keeping all pedigrees or tracing back from one to five generations in each phenotype truncation step
- **Validation:**
 - Genotyped animals born in 2019 and subsequent years
 - LR validation (Legarra and Reverter, 2018)
- **Software:**
 - BLUPF90IOD2OMP1 (Misztal et al., 2014)

RESULTS AND DISCUSSION

- Rank and correlations between expected progeny differences (EPD) in truncated and full datasets varied from 0.87 to 0.99 → Truncation does not affect the prediction of young, genotyped animals



- More generations in the pedigree ≠ higher reliability
- Keeping all pedigree and genotypes along with truncating the phenotypic information creates convergence problems
- Data truncation may be affected by the GI, changes in the trait definition, and changes in variance components over time

CONCLUSIONS

- Data truncation reduces computing costs
- Keeping 2 or 3 generations of data and 3 generations of pedigree is enough
- h^2 and number of traits in the model should also be considered
- Old phenotyped and/or genotyped animals can be removed
- Removing genotyped, non-phenotyped with no progeny is a better strategy

ACKNOWLEDGEMENTS

- This work was supported by the Pig Improvement Company

REFERENCES

Henderson, C. R. 1984. Applications of Linear Models in Animal Breeding Models. University of Guelph, Guelph, Canada.
Legarra, A., O. F. Christensen, I. Aguilari, and I. Misztal. 2014. Single Step, a general approach for genomic selection. *Livest Sci.* 166:54-65. doi:10.1016/j.livsci.2014.04.029.
Lourenco, D. A., L. I. Misztal, S. Tsuruta, I. Aguilari, T. J. Lawlor, S. Forri, and J. L. Welton. 2014. Are evaluations on young genotyped animals benefiting from the past generations? *J Dairy Sci.* 97:3930-3942. doi:10.3168/jds.2013-7769.
Macedo, F. L., J. A. Astor, T. H. E. Meuwissen, and A. Legarra. 2022. Removing data and using metafounders alleviates biases for all traits in Laccane dairy sheep predictions. *J Dairy Sci.* 105:2439-2452. doi:10.3168/jds.2021-20860.
Misztal, I., S. Tsuruta, D. A. L. Lourenco, Y. Masuda, I. Aguilari, A. Legarra, and Z. G. Vitezica. 2014. Manual for BLUPF90 family of programs. Available from http://nce.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_allB.pdf. Accessed in November 2022.