Modeling genetic differences of combined broiler chicken populations in single-step GBLUP

M. Bermann*, D. Lourenco*, V. Breen†, R. Hawken†, F. Brito Lopes†, I. Misztal*

* Department of Animal and Dairy Science, University of Georgia, Athens, GA, USA.
† Cobb-Vantress Inc., Siloam Springs, AR 72761, USA.
‡ UMR GenPhySE, INRA Toulouse, BP52626, 31326 Castanet Tolosan, France.

*This study was supported by Cobb-Vantress Inc. (Siloam Springs, AR)
†Corresponding author: mbermann@uga.edu
Abstract

The introduction of animals from a different environment or population is a common practice in commercial livestock populations. In this study we modeled the inclusion of a group of external birds into a local broiler chicken population for the purpose of genomic evaluations. The pedigree was composed by 242,413 birds and genotypes were available for 107,216 birds. A five-trait model that included one growth, two yield, and two efficiency traits was used for the analyses. The strategies to model the introduction of external birds were to include a fixed effect representing the of origin of parents and to use UPG or metafounders. Genomic estimated breeding values (GEBV) were obtained with single-step GBLUP (ssGBLUP) using the Algorithm for Proven and Young (APY). Bias, dispersion, and accuracy of GEBV for the validation birds, i.e., from the most recent generation, were computed. The bias and dispersion were estimated with the LR-method, whereas accuracy was estimated by the LR-method and predictive ability. When fixed UPG were fit without estimated inbreeding, the model did not converge. In contrast, models with fixed UPG and estimated inbreeding or random UPG converged and resulted in similar GEBV. The inclusion of an extra fixed effect in the model made the GEBV unbiased and reduced the inflation. Genomic predictions with metafounders were slightly biased and inflated due to the unbalanced number of observations assigned to each metafounder. When combining local and external populations, the greatest accuracy can be obtained by adding an extra fixed effect to account for origin of parents plus UPG with estimated inbreeding or random UPG. To estimate the accuracy, the LR-method is more consistent among scenarios, whereas predictive ability greatly depends on the model specification.

Key Words: Introduction of foreign animals, unknown parent groups, metafounders
List of abbreviations

APY: algorithm for proven and young

BLUP: best linear unbiased prediction

ssGBLUP: single-step GBLUP

EBV: estimated breeding value

GEBV: estimated breeding value using genomic information
Introduction

The introduction of animals from a different environment or population is a common practice in some livestock breeding populations (Lo, 1994). This practice is done to increase the genetic performance or to reduce the inbreeding in the current population. The inclusion of external animals should be appropriately accounted for in the genetic evaluation model, otherwise the estimated breeding values (EBV) or the genomic EBV (i.e., GEBV if genomic information is used) may be not accurately assessed and comparison of birds in the same selection group coming from distinct populations can be compromised. Two issues arise from this topic: first, the methods to include external animals in the evaluation, and second, the comparison between those methods and how to choose the best one.

Nowadays, the most used methods to account for the difference in the origin of the animals in the evaluation are to: (i) include an extra fixed effect in the model, (ii) use genetic groups (Quaas, 1988), and (iii) use metafounders (Legarra et al., 2015). The addition of an extra fixed effect in the model can be done by fitting a cross classified effect. This can be interpreted as the inclusion of the mean of each group of animals, for example, one mean per line or origin.

Regarding genetic groups, they are widely used in animal breeding to model the lack of pedigree information. In this setting, animals are assigned to groups according to a specific criterion. Hence, this approach allows to represent the average genetic merit of a group of individuals with missing parents and this is known as Unknown Parent Groups (UPG). The UPG can be easily introduced in genetic evaluations by modifying the inverse of the pedigree relationship matrix ($A^{-1}$), as proposed by Quaas (1988). To obtain a more accurate
estimation of actual relationships, inbreeding could be considered for UPG (VanRaden, 1992). Additionally, UPG can be considered as random effects in the model (Sullivan & Schaeffer, 1994; Legarra et al., 2007; Misztal et al., 2013).

With the advent of genomic information, genetic groups can be used for two purposes. First, to model missing pedigrees as in a pedigree-based evaluation (Misztal et al., 2013), and second to overcome the incompatibility between the pedigree relationship matrix (A) and the genomic relationship matrix (G) (Vitezica et al., 2011). Both issues can be solved with metafounders theory (Christensen, 2012; Legarra et al., 2015). The basic difference between metafounders and UPG is that relationships can exist among metafounders, but not among UPG. Metafounders under single-trait models provided the best results among other methods to model missing pedigrees in simulated datasets (Bradford et al., 2019).

The performance of genetic evaluation models is usually evaluated by cross-validation (Gianola, 2016). A statistic widely used for this purpose is the predictive ability, which is defined as the correlation between EBV and the phenotypes of the individual adjusted for the fixed effects in the model (Legarra et al., 2008). When divided by the square root of the heritability, the predictive ability gives an estimator of the accuracy of the model. Legarra & Reverter (2018) derived a method called LR-method to obtain estimates of bias, dispersion, accuracy, and other parameters related to the quality of EBV. The statistics from the LR-method are easy to obtain and, theoretically, work with any type of model (Legarra & Reverter, 2018). However, since it is a new method, it has not been tested with complex models and large data sets.

The primary objective of this study was to model the inclusion of a group of external birds into a local population using either a fixed effect representing the of origin of parents, UPG,
or metafounders. As a secondary goal, we determined whether predictive ability or LR-method gives a more reliable estimation of the accuracy of GEBV.

Materials and Methods

Data

The dataset used in this study was provided by Cobb-Vantress Inc. (Siloam Springs, AR). The traits used in the evaluation were weight (WGT), Carcass Yield Component 1 (CYC1), Feed Efficiency 1 (FE1), Carcass Yield Component 2 (CYC2), and Feed Efficiency 2 (FE2). The number of records and genotyped birds per trait and the number of birds used to validate the models (i.e., focal individuals) are shown in Table 1. Genotypes from the 60k SNP panel were available for 107,216 birds and pedigree information was available for 242,413 birds. About 1% of the birds in the whole pedigree had missing parents. Quality control removed SNP with call rates lower than 0.9, minor allele frequencies lower than 0.05, and heterozygosity deviation greater than 0.15 from the Hardy-Weinberg equilibrium expectation (Wiggans et al., 2009). Markers with unknown position or located on sex chromosomes were also excluded from the analyses. After quality control, 47,952 SNP were kept for analysis. All birds were classified into two groups according to their origin, i.e., local and external individuals. Local individuals belonged to the original Cobb-Vantress Inc. (Siloam Springs, AR) breeding program, and the external individuals belonged to another breeding company and were recently acquired. The number of individuals in each group was 240,929 and 1484, respectively. The number of animals with missing parents was 2161 in the first group and 656 in the second. This classification will be referred to as C1. The birds
from the second group of C1 were used as parents and those whose parents were known had records. Based on C1, the birds with both parents known were classified into three groups: (1) with both parents belonging to the original breeding program, (2) with both parents coming from an external population, and (3) with one parent coming from an external population. The number of individuals in each group was 203,759, 18,599, and 16,407, respectively. This second classification will be referred as C2.

The focal individuals used to validate the models were phenotyped individuals from the most recent generation in the pedigree. Table 1 shows the number of focal individuals used for each trait. Hereafter, the subscript \( w \) will denote the scenarios when the phenotypes of the focal individuals were used in the evaluation, whereas the subscript \( p \) will denote when the phenotypes were not used.

**Models**

Two five-trait models that included growth and efficiency traits in broilers were used in the analyses. The first model, denoted as MREG, was:

\[
y = Xb + Zu + Wp + e,
\]

Where \( y \) was the vector of phenotypes, \( b \) was the vector of fixed effects including generation and sex of the bird, \( u \) was the vector of breeding values, \( p \) was the vector of permanent environmental effects for WGT, \( e \) was the vector of errors, and \( X, Z, \) and \( W \) were incidence matrices.
The second model, denoted as MFIX, was equal to MREG plus an extra cross-classified fixed effect representing the origin of parents. Therefore, MFIX was:

\[ y = U\alpha + Xb + Zu + Wp + e, \]

Where \( \alpha \) was a vector representing the C2 classification, and \( U \) was an incidence matrix. For this model, records from animals with missing parents were removed from the analyses.

For both models it was assumed that \( \text{Var}(e) = I \otimes V_e, \text{Var}(u) = H \otimes V_u, \) and \( \text{cov}(e, u) = 0, \) where \( V_e \) and \( V_u \) are the covariance matrices among traits, and \( H \) is defined as in Legarra et al. (2009).

Both models were evaluated without the inclusion of UPG (MREG\textsubscript{N} and MFIX\textsubscript{N}), with UPG without estimated inbreeding for UPG (VanRaden, 1992) (MREG\textsubscript{UPG}/MFIX\textsubscript{UPG}), with UPG and estimated inbreeding for UPG (MREG\textsubscript{UPG,INB}/MFIX\textsubscript{UPG,INB}), with random UPG and estimated inbreeding for UPG (MREG\textsubscript{UPG,INB,RAN}/MFIX\textsubscript{UPG,INB,RAN}), and with metafounders (MREG\textsubscript{MF}/MFIX\textsubscript{MF}). Both UPG and metafounders were defined by using the C1 classification, hence, only two UPG or metafounders were used in the analyses. These differences within a model will be referred to as “scenarios”. For example, for MREG\textsubscript{MF}, the model is MREG whereas the scenario is with metafounders.

UPG were defined as in Misztal et al. (2013), where the QP transformation (Quaas and Pollak, 1981) is applied to the pedigree relationship matrix (\( A \)), the genomic relationship matrix (\( G \)), and the pedigree relationship matrix among genotyped animals (\( A_{22} \)). Inbreeding coefficients estimated as in Aguilar et al., (2008) were used to compute both inverses of \( A \) and \( A_{22} \), therefore, inbreeding coefficients were updated when inbreeding for UPG was
Considered. Inbreeding for UPG was calculated as the average inbreeding of known parents of the same period \( t \), that is, \( F_{\text{upg}} = \bar{F}_P \) (VanRaden, 1992).

For metafounders, their covariance matrix \( (\mathbf{\Gamma}) \) was computed by generalized least squares for multiple populations (Legarra et al., 2015; Garcia-Baccino et al., 2017) using Gammaf90 from the BLUPF90 software suite (not publicly released). The resulting matrix was

\[
\mathbf{\Gamma} = \begin{bmatrix} 0.5 & 0.38 \\ 0.38 & 0.57 \end{bmatrix}.
\]

Following the methods in Legarra et al. (2015), the estimated genetic variance \( (\hat{\sigma}_g^2) \) was divided by \( k = 1 + \text{diag}(\mathbf{\Gamma})/2 - \mathbf{\Gamma} \), and the inverse of the numerator relationship matrix was computed using the rules provided in the same study. The variance components used in this study were provided by Cobb-Vantress Inc. (Siloam Springs, AR).

Heritability \( (\hat{h}^2) \), genetic, and phenotypic correlations for all traits are in Table 2.

**Estimation of the model effects**

Estimates for all effects in the models were obtained with ssGBLUP (Legarra et al., 2009; Christensen & Lund, 2010) using BLUP90IOD2 (Misztal et al., 2014) with the Algorithm for Proven and Young (APY) (Misztal, 2016). For BLUP90IOD2, the mixed model equations were solved with the preconditioned conjugate gradient algorithm (PCG), with a convergence statistic equal to the squared norm of the difference between the right hand side and the coefficient matrix times the current vector of solutions, divided by the squared norm of the right hand side. A convergence criterion of \( 10^{-14} \) was adopted. For more details, the reader is referred to Tsuruta et al. (2001). The genomic relationship matrix was computed by the first method of VanRaden (2008) with observed allele frequencies, except for the scenarios with metafounders where the allele frequencies were set to 0.5. For APY, the core was comprised of 15,867 randomly selected birds, which represents more than 99% of the
variance in $G$ but avoids further core changes. As APY requires the inverse of the subset of $G$ for core animals, this matrix was blended with 5% of $A_{22}$.

**Validation**

All scenarios were compared by the estimated bias, dispersion, and accuracy of the predictions for the focal individuals. The bias and dispersion were estimated as in Legarra & Reverter (2018). The accuracy was estimated by two methods: (1) $\hat{d}_{H} = \frac{cov(y_{adj}, \hat{u})}{\hat{r}^2}$, where $y_{adj}$ is the vector of phenotype minus the estimates of all non-genetic effects (Legarra et al., 2008) and $\hat{u}$ is the vector of GEBV, and (2) $\hat{d}_{LR} = \sqrt{\frac{cov(\hat{u}, \hat{u})}{(1-P)\theta^2}}$ (Legarra & Reverter, 2018; Macedo et al., 2020). Both $cov$ and $corr$ correspond to the sample covariance and sample correlation coefficient, respectively. The denominator of $\hat{d}_{LR}$ corresponds to one minus the average inbreeding of the focal individuals ($P$) times the estimate of the additive genetic variance of the trait ($\theta^2$).

**Results**

Models with fixed UPG did not converge, hence, results for MREG$_{UPG}$ and MFIX$_{UPG}$ were not obtained. The lack of convergence is because the coefficient matrix became non-positive definite. All the results regarding UPG with estimated inbreeding and random UPG were the same due to the small number of UPG in the models and the large number of records for each trait. Thus, referring to one of them will suffice.

Figure 1 shows the rounds until convergence for each model and each scenario. For both models, the use of metafounders reached convergence in 500-1000 rounds earlier than the
rest of the methods. The slowest convergence was seen for fixed and random UPG with estimated inbreeding.

Table 3 shows the estimated accuracy of the predictions for each scenario and each trait. Values near one indicate that the EBV are accurately estimated. The concordance between $\hat{\text{acc}}_{PA}$ and $\hat{\text{acc}}_{LR}$ depends on the model and the trait. For MREG, both $\hat{\text{acc}}_{PA}$ and $\hat{\text{acc}}_{LR}$ were different for CYC1, FE1, and FE2. When metafounders were used, the two measures of prediction accuracy became similar for FE1 and FE2. On the other hand, for MFIX, predictive ability and LR-method performed similarly, except for CYC1. The largest difference between $\hat{\text{acc}}_{PA}$ and $\hat{\text{acc}}_{LR}$ was 0.34 for MREG and 0.15 for MFIX. In general, differences among models were not large. The highest accuracies were obtained using metafounders in MREG, whereas for MFIX the greatest accuracies were obtained with UPG with estimated inbreeding. Within models, no differences were observed among scenarios for some traits (e.g. CYC1).

Table 4 shows the estimated dispersion of the predictions from the LR-method for each scenario and each trait. Values near one indicate that there is no over/under dispersion in the GEBV. For MREG, the GEBV for CYC1, FE1, and FE2 were inflated except when metafounders were implemented in the evaluation. No considerable inflation was observed for any situation in MFIX. Within the range 0.90-1.10, we considered the inflation/deflation as acceptable. Overall, for MREG, the best performance was achieved when metafounders were included in the model, whereas the worst performance was with UPG and estimated inbreeding. For MFIX, the best performance was obtained without modeling the genetic groups.
Table 5 shows the estimated bias of the predictions for each scenario and each trait, expressed proportionally to the genetic standard deviation of each trait. Values near zero indicate that the EBV are unbiased. For all the traits, the bias was reduced when a fixed effect representing the origin of parents was fitted to the data (i.e., MFIX). The absolute value of the bias for MREG ranged from 0.10 to 0.69 standard deviations, whereas for MFIX the absolute bias ranged from 0.01 to 0.18. Except for WGT, the least biased predictions for MREG were obtained when metafounders were included in the model. For MFIX, the least biased predictions were obtained without including genetic groups and with UPG with estimated inbreeding.

Discussion

The best way to model the inclusion of a group of external individuals in the evaluation was via the addition of an extra fixed effect representing the origin of the parents of the individuals, i.e., local and external. Although the impact in accuracy is small, the GEBV with MFIX were almost unbiased and without inflation/deflation. For all traits, the phenotypic means in C2 were different among groups. Thus, dismissing α leads to a model misspecification (Rencher & Schaalje, 2008, Section 7.9).

Although the data and results show that α should be included in the model, this new parameter is hard to interpret in this situation because all focal individuals were born in the local environment. Hence, it is likely that α may be modeling a genetic effect absent in MREG. In fact, fitting α as a fixed effect accounts for non-genetic and genetic differences between parents of local and external individuals, whereas UPG account only for the genetic
differences. If only UPG are considered, the non-genetic differences are not going to be modeled. The consideration of $\alpha$ as a fixed effect instead of random was to avoid the re-estimation of variance components and to keep model simplicity. Although the re-estimation of variance components may not be a limitation for smaller datasets, it can be time-consuming when the number of genotyped animals and traits is large. This can lead to a limitation in the interval between successive genetic evaluations and, therefore, become a problem for companies.

In this study, two estimators for the accuracy were used. Predictive ability is widely used but depends on model specification and will tend to be higher with over-parametrized models. That is, when increasing the number of fixed effects, predictive ability will tend to overestimate the accuracy of GEBV. As an extreme example, with a saturated model (one fixed effect per animal) $\hat{\alpha}_{cPA}$ will be equal to 1, whereas $\hat{\alpha}_{cLR}$ will be equal to zero. It can be observed in Table 3 that $\hat{\alpha}_{cPA}$ greatly varied when the extra fixed effect was added into the model, whereas $\hat{\alpha}_{cLR}$ showed smaller changes. The average difference between $\hat{\alpha}_{cPA}$ and $\hat{\alpha}_{cLR}$ for MREG was 0.10, whereas for MFIX it was 0.06. It can be observed that for some traits $\hat{\alpha}_{cPA}$ greatly varied from MREG to MFIX. For example, it varied from 0.39 to 0.63 for FE1 when no genetic groups were modeled. On the other hand, $\hat{\alpha}_{cLR}$ varied from 0.63 to 0.65 for the same scenario. The addition of a simple fixed effect may not justify the increase in accuracy when using $\hat{\alpha}_{cPA}$. Additionally, as an example, the correlation between GEBV for FE1 estimated with MREG and $N_{MFIX}$ was 0.92. Clearly, this shows that the accuracy is wrongly estimated by $\hat{\alpha}_{cPA}$ in MREG. For the same scenario in Table 5, it can be observed that the bias greatly decreased from -0.69 to 0.01 standard deviations. Thus, it can be deduced that $\hat{\alpha}_{cPA}$ is more influenced by the model specification than $\hat{\alpha}_{cLR}$. In other
words, when a fixed effect like $\alpha$ is omitted from the model, the estimates of the accuracy with predictive ability greatly vary. Therefore, the use of $\hat{a}\hat{c}_{CLR}$ for further evaluations is recommended since it depends less on the modeling of fixed effects. Different ways to compute $\hat{a}\hat{c}_{CLR}$ exist depending how the denominator is constructed (Legarra & Reverter, 2018; Macedo et al. 2020). However, the differences may be small and all the forms are proportional.

The accuracy of predictions with metafounders estimated with both $\hat{a}\hat{c}_{PA}$ and $\hat{a}\hat{c}_{CLR}$ decreased significantly from MREG to MFIX. Also, predictions with metafounders in MFIX were more biased than with other methods. These issues may be related with the scaling of the genetic variance due to metafounders. As mentioned in Materials and Methods, when using metafounders in ssGBLUP, the genetic variance, and consequently the inverse of the relationship matrix, are divided by $k$. The way of computing $k$ implicitly assumes that all metafounders have the same amount of information in the population. When groups to define metafounders are very unbalanced, as in the case of C1, this assumption may be incorrect. If the group of external birds was larger, there would possibly be no issues of reduced accuracy and biased predictions with the inclusion of metafounders. However, further research is needed to solve this issue when the groups used to define metafounders are unbalanced and no other group definition can be used to determine the number of metafounders. In agreement to previous studies (van Grevenhof et al., 2020), the use of metafounders resulted in better convergence properties than other methods because of a reduction in the condition number of the system.
Conclusion

The inclusion of animals from an external group into local genetic evaluations should be correctly modeled to obtain accurate, unbiased, and not under/overdispersed GEBV. This can be done by including an extra fixed effect in the model, by using UPG or metafounders. For most of the traits in this study, the best scenario includes an extra fixed effect and with an extra adjustment as UPG with estimated inbreeding or metafounders. When the groups to define the metafounders greatly differ in the number of animals, predictions can be biased. In such a situation, it may be better to use UPG. Further research on this topic is needed. In this study, UPG without estimated inbreeding did not converge. Thus, taking into account estimated inbreeding for UPG is recommended. The use of metafounders reduced the condition number of the system, resulting in faster convergence for both models, however, this result cannot be generalized to other models and datasets. The estimation of accuracy using the LR-method is less sensitive to the fixed effects of the model than predictive ability, therefore, being a proper method to the estimate accuracy of GEBV, independently of model specification.
Acknowledgments

This study was supported by Cobb-Vantress Inc.

Conflict of interest

The authors declare that they do not have any conflict of interest.

Data Availability Statement

The data belongs to Cobb-Vantress and, therefore, cannot be shared.
References

Aguilar, I., & Misztal, I. (2008). Technical note: recursive algorithm for inbreeding coefficients assuming nonzero inbreeding of unknown parents. Journal of dairy science, 91(4), 1669–1672. https://doi.org/10.3168/jds.2007-0575

Bradford, H. L., Masuda, Y., VanRaden, P. M., Legarra, A., & Misztal, I. (2019). Modeling missing pedigree in single-step genomic BLUP. Journal of Dairy Science, 102(3), 2336–2346. https://doi.org/10.3168/jds.2018-15434

Christensen, Ole F, & Lund, M. S. (2010). Genomic prediction when some animals are not genotyped. Genetics Selection Evolution, 1–8. https://doi.org/10.1186/1297-9686-42-2

Christensen O. F. (2012). Compatibility of pedigree-based and marker-based relationship matrices for single-step genetic evaluation. Genetics, selection, evolution : GSE, 44(1), 37. https://doi.org/10.1186/1297-9686-44-37

Gianola, D., & Schön, C. C. (2016). Cross-Validation Without Doing Cross-Validation in Genome-Enabled Prediction. G3 (Bethesda, Md.), 6(10), 3107–3128. https://doi.org/10.1534/g3.116.033381

Legarra, A., Bertrand J.K., Strabel T., Sapp R.L., Sanchez J.P, Misztal I. (2007). Multi-breed genetic evaluation in a Gelbvieh population. J. Anim. Breed. Genet., 124, 286–295.

Legarra, A., Robert-Granié, C., Manfredi, E., & Elsen, J. M. (2008). Performance of genomic selection in mice. Genetics, 180(1), 611–618. https://doi.org/10.1534/genetics.108.088575

Legarra, A., Aguilar, I., & Misztal, I. (2009). A relationship matrix including full pedigree and genomic information. Journal of Dairy Science, 92(9), 4656–4663. https://doi.org/10.3168/jds.2009-2061

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

Legarra, A., & Reverter, A. (2018). Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. Genetics Selection Evolution, 50(1), 1–18. https://doi.org/10.1186/s12711-018-0426-6

Lo, L.L. (1994). Genetic evaluation and selection in multibreed population. Ph.D dissertation, University of Illinois at Urbana-Champaign. http://hdl.handle.net/2142/22306
Macedo, F.L., Reverter, A., Legarra, A. (2020). Behavior of the Linear Regression method to estimate bias and accuracies with correct and incorrect genetic evaluation models, Journal of Dairy Science, Volume 103, Issue 1, 529-544, https://doi.org/10.3168/jds.2019-16603.

Misztal, I., Vitezica, Z., Legarra, A., Aguilar, I. and Swan, A. (2013), Unknown-parent groups in single-step genomic evaluation. J. Anim. Breed. Genet., 130: 252-258. doi:10.1111/jbg.12025

Misztal, I., Tsuruta, S., Lourenco, D. A. L., Aguilar, I., Legarra, A., & Vitezica, Z. (2014). Manual for BLUPF90 family of programs. http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all7.pdf

Misztal I. (2016). Inexpensive computation of the inverse of the genomic relationship matrix in populations with small effective population size. Genetics. 202:401–9.

Quaas, R. L. (1988). Additive Genetic Model with Groups and Relationships. Journal of Dairy Science, 71(5), 1338–1345. https://doi.org/10.3168/jds.S0022-0302(88)79691-5

Quaas R.L., Pollak E.J. (1981) Modified equations for sire models with groups. J. Dairy Sci., 64, 1868–1872.

Rencher, A. and Schaalje, G., 2008. Linear Models In Statistics. Hoboken: Wiley & Sons.

Sullivan P.G., Schaeffer L.R. (1994). Fixed versus random genetic groups. In: Proceedings of the 5th World Congress on Genetics Applied to Livestock Production, Dept. of Animal & Poultry Science, University of Guelph, Guelph, ON, Canada, 7–12 August 1994. Vol. 18, pp. 483–486.

Tsuruta, S., Misztal, I., & Strandén, I. (2001). Use of the preconditioned conjugate gradient algorithm as a generic solver for mixed-model equations in animal breeding applications. Journal of animal science, 79(5), 1166–1172. https://doi.org/10.2527/2001.7951166x

van Grevenhof EM, Vandenplas J, Calus MPL. Genomic prediction for crossbred performance using metafounders. J Anim Sci. 2019 Feb 1;97(2):548-558. doi: 10.1093/jas/sky433. Erratum in: J Anim Sci. 2020 Jan 1;98(1): PMID: 30423111; PMCID: PMC6358227.

VanRaden, P.M., (1992). Accounting for Inbreeding and Crossbreeding in Genetic Evaluation of Large Populations, Journal of Dairy Science, Volume 75, Issue 11, 3136-3144, https://doi.org/10.3168/jds.S0022-0302(92)78077-1.
VanRaden, P. M. (2008). Efficient Methods to Compute Genomic Predictions. Journal of Dairy Science, 91(11), 4414–4423. https://doi.org/10.3168/jds.2007-0980

VanRaden, P. M., C. P. VanTassel, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and F. S. Schenkel. (2009). Invited review: Reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci. 92:16-24.

Vitezica, Z. G., Aguilar, I., Misztal, I., & Legarra, A. (2011). Bias in genomic predictions for populations under selection. Genetics, (2011), 357–366. https://doi.org/10.1017/S001667231100022X

Wiggans, G. R., Sonstegard, T. S., Vanraden, P. M., Matukumalli, L. K., Schnabel, R. D., & Taylor, J. F. (2009). Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. Journal of Dairy Science, 92(7), 3431–3436. https://doi.org/10.3168/jds.2008-1758
Table 1. Number of records and number of focal individuals of each trait. The number of genotyped animals is in parenthesis.

| Trait                      | Abbreviation | Number of records | Number of focal individuals |
|----------------------------|--------------|-------------------|----------------------------|
| Weight                     | WGT          | 74,484            | 8,689 (3,030)              |
| Carcass Yield Component 1  | CYC1         | 15,293            | 823 (585)                  |
| Feed Efficiency 1          | FE1          | 59,567            | 8,917 (3,012)              |
| Carcass Yield Component 2  | CYC2         | 15,208            | 820 (584)                  |
| Feed Efficiency 2          | FE2          | 60,126            | 8,963 (3,004)              |
Table 2. Heritability (diagonal), genetic correlation (above-diagonal), and phenotypic correlation (below-diagonal) for each trait.

|        | WGT  | CYC1 | FE1  | CYC2 | FE2  |
|--------|------|------|------|------|------|
| WGT    | 0.16 | 0.17 | 0.18 | 0.08 | -0.09|
| CYC1   | 0.46 | 0.59 | -0.10| -0.49| -0.09|
| FE1    | 0.01 | -0.04| 0.36 | 0.07 | 0.87 |
| CYC2   | 0.01 | -0.38| 0.06 | 0.62 | 0.05 |
| FE2    | -0.37| -0.12| 0.73 | 0.01 | 0.24 |
Table 3. Accuracy of predictions for each combination of model-scenario\(^1\) and each trait estimated with both \(\hat{a}_{cc_{PA}}\) and \(\hat{a}_{cc_{LR}}\).

|               | WGT  | CYC1 | FE1  | CYC2 | FE2  |
|---------------|------|------|------|------|------|
|               | \(\hat{a}_{cc_{PA}}\) | \(\hat{a}_{cc_{LR}}\) | \(\hat{a}_{cc_{PA}}\) | \(\hat{a}_{cc_{LR}}\) | \(\hat{a}_{cc_{PA}}\) | \(\hat{a}_{cc_{LR}}\) | \(\hat{a}_{cc_{PA}}\) | \(\hat{a}_{cc_{LR}}\) |
| MREG\(_N\)  | 0.72 | 0.72 | 0.50 | 0.59 | 0.39 | 0.63 | 0.59 | 0.57 | 0.29 | 0.63 |
| MREG\(_{UPG\_INB}\) | 0.73 | 0.71 | 0.50 | 0.58 | 0.40 | 0.59 | 0.58 | 0.61 | 0.32 | 0.52 |
| MREG\(_{UPG\_INB\_RA}\) | 0.73 | 0.71 | 0.50 | 0.58 | 0.40 | 0.59 | 0.58 | 0.61 | 0.32 | 0.52 |
| MREG\(_{MF}\) | 0.73 | 0.70 | 0.62 | 0.83 | 0.69 | 0.69 | 0.67 | 0.80 | 0.58 | 0.61 |
| MFI\(_N\)   | 0.79 | 0.74 | 0.59 | 0.69 | 0.63 | 0.65 | 0.58 | 0.61 | 0.63 | 0.64 |
| MFI\(_{UPG\_INB\_RAN}\) | 0.82 | 0.79 | 0.59 | 0.73 | 0.63 | 0.67 | 0.57 | 0.65 | 0.63 | 0.66 |
| MFI\(_{UPG\_RAN}\)  | 0.82 | 0.79 | 0.59 | 0.73 | 0.63 | 0.67 | 0.57 | 0.65 | 0.63 | 0.66 |
| MFI\(_{MF}\) | 0.71 | 0.69 | 0.58 | 0.74 | 0.59 | 0.62 | 0.61 | 0.71 | 0.52 | 0.57 |

\(^1\) Models are defined as MREG and MFIX, while scenarios are denoted with the subscript in each line (N, UPG, UPG\_INB, UPG\_INB\_RAN, and MF).
Table 4. Dispersion of predictions for each scenario\(^1\) and each trait.

|       | WGT | CYC1 | FE1 | CYC2 | FE2 |
|-------|-----|------|-----|------|-----|
| MREG\(_N\) | 1.01 | 0.87 | 0.71 | 1.05 | 0.62 |
| MREG\(_{UPG\_INB}\) | 0.98 | 0.85 | 0.69 | 1.00 | 0.62 |
| MREG\(_{UPG\_INB\_RAN}\) | 0.98 | 0.85 | 0.69 | 1.00 | 0.62 |
| MREG\(_{MF}\) | 1.05 | 0.90 | 1.11 | 0.96 | 1.01 |
| MFIX\(_N\) | 0.99 | 0.99 | 1.01 | 1.00 | 0.98 |
| MFIX\(_{UPG\_INB}\) | 0.98 | 0.96 | 0.98 | 0.95 | 0.95 |
| MFIX\(_{UPG\_INB\_RAN}\) | 0.98 | 0.96 | 0.98 | 0.95 | 0.95 |
| MFIX\(_{MF}\) | 0.99 | 0.93 | 0.97 | 0.92 | 0.91 |

\(^1\) Models are defined as MREG and MFIX, while scenarios are denoted with the subscript in each line (N, UPG, UPG\_INB, UPG\_INB\_RAN, and MF).
Table 5. Bias of predictions for each scenario\(^1\) expressed in terms of the genetic standard deviation of each trait.

| Scenario          | WGT | CYC1 | FE1 | CYC2 | FE2 |
|-------------------|-----|------|-----|------|-----|
| MREG\(_N\)        | -0.10 | 0.30 | -0.69 | -0.24 | -0.59 |
| MREG\(_{UPG\_INB}\) | -0.07 | 0.27 | -0.61 | -0.21 | -0.51 |
| MREG\(_{UPG\_INB\_RAN}\) | -0.07 | 0.27 | -0.61 | -0.21 | -0.51 |
| MREG\(_{MF}\)     | 0.19 | 0.17 | 0.52 | -0.02 | 0.40 |
| MFIX\(_N\)        | -0.01 | 0.03 | 0.01 | -0.02 | 0.02 |
| MFIX\(_{UPG\_INB}\) | -0.01 | 0.03 | 0.01 | -0.02 | 0.02 |
| MFIX\(_{UPG\_INB\_RAN}\) | -0.01 | 0.03 | 0.01 | -0.02 | 0.02 |
| MFIX\(_{MF}\)     | 0.03 | 0.09 | 0.18 | -0.09 | 0.18 |

\(^1\) Models are defined as MREG and MFIX, while scenarios are denoted with the subscript in each line (N, UPG, UPG\_INB, UPG\_INB\_RAN, and MF).
Figure 1. Rounds until convergence for each scenario. The first box corresponds to MREG, while the second corresponds to MFIX. The dashed line denotes the logarithm with base 10 of the convergence criterion.\textsuperscript{1,2}

\textsuperscript{1} MREG refers to a model without parents’ origin fitted as fixed effect, whereas MFIX to a model with parents’ origin fitted as fixed effect.

\textsuperscript{2} The subscript N denotes no inclusion of unknown parent groups (UPG) or metafounders (MF), UPG\_INB\_FIX denotes fixed UPG with inbreeding, and UPG\_INB\_RAN denotes random UPG with inbreeding.
Figure 1

MREG

MFIX

log C convergence criterion

round