Reliabilities of estimated breeding values in models with metafounders

Matias Bermann, Ignacio Aguilar, Daniela Lourenco, Ignacy Misztal, Andres Legarra

RESEARCH ARTICLE

Reliabilities of estimated breeding values in models with metafounders

Matias Bermann[1]* , Ignacio Aguilar[2], Daniela Lourenco[1], Ignacy Misztal[1] and Andres Legarra[3,4]

Abstract

Background Reliabilities of best linear unbiased predictions (BLUP) of breeding values are defined as the squared correlation between true and estimated breeding values and are helpful in assessing risk and genetic gain. Reliabilities can be computed from the prediction error variances for models with a single base population but are undefined for models that include several base populations and when unknown parent groups are modeled as fixed effects. In such a case, the use of metafounders in principle enables reliabilities to be derived.

Methods We propose to compute the reliability of the contrast of an individual's estimated breeding value with that of a metafounder based on the prediction error variances of the individual and the metafounder, their prediction error covariance, and their genetic relationship. Computation of the required terms demands only little extra work once the sparse inverse of the mixed model equations is obtained, or they can be approximated. This also allows the reliabilities of the metafounders to be obtained. We studied the reliabilities for both BLUP and single-step genomic BLUP (ssGB-LUP), using several definitions of reliability in a large dataset with 1,961,687 dairy sheep and rams, most of which had phenotypes and among which 27,000 rams were genotyped with a 50K single nucleotide polymorphism (SNP) chip. There were 23 metafounders with progeny sizes between 100,000 and 2000 individuals.

Results In models with metafounders, directly using the prediction error variance instead of the contrast with a metafounder leads to artificially low reliabilities because they refer to a population with maximum heterozygosity. When only one metafounder is fitted in the model, the reliability of the contrast is shown to be equivalent to the reliability of the individual in a model without metafounders. When there are several metafounders in the model, using a contrast with the oldest metafounder yields reliabilities that are on a meaningful scale and very close to reliabilities obtained from models without metafounders. The reliabilities using contrasts with ssGBLUP also resulted in meaningful values.

Conclusions This work provides a general method to obtain reliabilities for both BLUP and ssGBLUP when several base populations are included through metafounders.

*Correspondence:
Matias Bermann
mbermann@uga.edu
[1] Department of Animal and Dairy Science, University of Georgia, Athens, GA, USA
[2] Instituto Nacional de Investigación Agropecuaria (INIA), Montevideo, Uruguay
[3] GenPhySE, INRAE, ENVT, Université de Toulouse, 31326 Castanet-Tolosan, France
[4] Present Address: Council on Dairy Cattle Breeding, Bowie, MD 20716, USA

Background

In the traditional animal model, animals with unknown parents are assumed to be unrelated and to belong to the same base population. This base population is assumed to be of infinite size and with an average breeding value equal to zero. Under this assumption, using the correct model specification and with a complete pedigree, estimation of the breeding values (**u**) and variance components is unbiased [1, 2]. However, pedigrees are incomplete in most livestock populations, in which

case genetic groups or unknown parent groups (UPG) are typically used to handle incomplete pedigrees and unrecorded selection [3]. From the early stages of their implementation, UPG were fitted as fixed effects [3–5], resulting in the final breeding value $\mathbf{u}^* = \mathbf{Qg} + \mathbf{u}$ to be a weighted sum of the fixed group effects $(\mathbf{Qg})$ and a random deviation ($\mathbf{u}$) [3]. As a result, the final breeding values (and their estimates) do not have an explicit base population, therefore, estimated breeding values (EBV) are not estimable, although their contrasts are [3]. Moreover, using fixed UPG is not justified from a quantitative genetics perspective because it assumes that the genetic variance is not changed by drift or selection [6], in other words, it assumes that UPG are of infinite size and not related to each other.

The reliability of an individual EBV can be defined as the squared correlation between the true and estimated breeding value over repeated conceptual sampling: $rel_i = r^2(\widehat{u}, u)$. Reliabilities can be obtained as $rel_i = 1 - \frac{\text{Var}(\widehat{u}_i - u_i)}{\text{Var}(u_i)}$ [7], where $\text{Var}(\widehat{u}_i - u_i)$ is the prediction error variance of $\widehat{u}_i$, which can be obtained from the inverse of the coefficient matrix of the mixed model equations (MME) [8]. However, the expression $rel_i = 1 - \frac{\text{Var}(\widehat{u}_i - u_i)}{\text{Var}(u_i)}$ does not hold when UPG are fitted as fixed effects because then the covariance between true and estimated effects is not equal to the variance of the estimated effects [7]. Instead, when UPG are fitted as fixed effects, the reliability can be defined as $r^2(\widehat{u}^*, u^*) = r^2(\mathbf{q}\widehat{g} + \widehat{u}, \mathbf{q}g + u)$, where $\mathbf{q}$ is a row of $\mathbf{Q}$ (see p. 44 in [8]). However, the value of this correlation depends on the particular generalized inverse that is used to solve the MME (often implicitly) and is, thus, not uniquely defined (see p. 44 in [8]). A practical solution is to calculate the reliabilities from a model without UPG, but results in reliabilities to be overestimated because it ignores uncertainty due to the estimation of UPG [9].

To account for an increase in inbreeding and relationships, VanRaden [10] considered UPG as random effects by assigning them an inbreeding coefficient equal to the average inbreeding of known parents from the same generation. In crosses of populations, Lo et al. [11] modeled the covariances across animals by considering the heterogeneity of the base populations (i.e., different allele frequencies). Their model requires knowledge of the genetic variance of each base population and the segregation variance for each cross, which can differ for each trait. Garcia-Cortés and Toro [12] presented an equivalent model.

In a genomic context and with access to genotypes, Legarra et al. [13] introduced the concept of metafounders, which are a generalization of genetic groups that account for inbreeding and segregation variances (assumed the same for all traits) by introducing relatedness between base populations—the more different the base populations are, the larger is the genetic segregation variance. However, a base population from which all relationships are computed must be defined when considering different populations and crosses. For the case of two populations, Lo et al. [11] proposed using the F2 cross as a "reference group". Mostly for convenience reasons (compatibility with genomic relationships), metafounders have a conceptual base population that has maximum heterozygosity. This lacks a clear genetic interpretation, as it differs from the base population of a pedigree without metafounders.

Using random UPG or metafounders leads to a model where all components of $\mathbf{u}^*$ are random and, thus, $rel_i = 1 - \frac{\text{Var}(\widehat{u}_i - u_i)}{\text{Var}(u_i)}$ holds. However, in preliminary tests we observed that the computed reliabilities for a model with metafounders were biased (too low), for instance, for proven dairy sires. Thus, we hypothesize that the reliabilities obtained from the inverse of the coefficient matrix when using metafounders do not refer to a meaningful breeding value in the usual quantitative genetics sense because they do not refer to a single unrelated base population. In fact, it is not obvious what the genetic base is when using random UPG or metafounders. Instead, we argue that $rel_i = 1 - \frac{\text{Var}(\widehat{u}_i - u_i)}{\text{Var}(u_i)}$ does represent the squared correlation between the true and estimated breeding values, but across two sampling processes. In the case of metafounders, these two sampling processes are: (i) from the conceptual base population of metafounders (one with maximum heterozygosity) to the metafounders, and (ii) from each metafounder to the individual. In the case of random UPG, there is a sampling from 0 (average UPG effect value a priori) to UPG values and then individual Mendelian samplings for each individual, which sum to UPG fractions, yield the final breeding value. Thus, in order to compute meaningful reliabilities of EBV we must define a proper base (or reference, as in Lo et al. [11]) population, relative to which the breeding values are expressed [14]. Thus, the objectives of this study were to develop theory and computational methods to obtain meaningful reliabilities for EBV for models with metafounders. The method is illustrated with data from the Lacaune dairy breed under pedigree-based and genomic-based models.

Methods
Theory
Consider the following linear mixed model:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Zu} + \mathbf{e}, \tag{1}$$

where $\mathbf{y}$ is the vector of phenotypes, $\boldsymbol{\beta}$ is the vector of fixed effects, $\mathbf{u}$ is the vector of breeding values, $\mathbf{e}$ is the vector of errors, and $\mathbf{X}$ and $\mathbf{Z}$ are incidence matrices. As usual, it is assumed that:

$$E[\mathbf{y}] = \mathbf{X}\boldsymbol{\beta},$$

$$\text{and Var}\begin{pmatrix} \mathbf{u} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} \mathbf{K}\sigma_u^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{pmatrix}. \tag{2}$$

The coefficient matrix of the MME for the model in Eq. (1) is:

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \sigma_u^{-2}\mathbf{K}^{-1} \end{pmatrix}. \tag{3}$$

If the inverse of $\mathbf{C}$ is $\mathbf{C}^{-1} = \begin{pmatrix} \mathbf{C}^{11} & \mathbf{C}^{12} \\ \mathbf{C}^{21} & \mathbf{C}^{22} \end{pmatrix}$, the reliability for the ith animal is calculated as [7]:

$$rel_i = 1 - \frac{\mathbf{C}_{ii}^{22}}{\mathbf{K}_{ii}\sigma_u^2} = 1 - \frac{\text{Var}(\widehat{u}_i - u_i)}{\text{Var}(u_i)}. \tag{4}$$

In "regular" animal models, $\mathbf{K} = \mathbf{A}$, where $\mathbf{A}$ is the numerator relationship matrix. Now, we introduce a distinction by using metafounders. We keep u_i for the breeding value of individual i estimated in the "regular" animal model, and u_i^* for the breeding value estimated with metafounders. When using metafounders [13], $\mathbf{K}^{-1}$ in Eq. (3) is replaced by either $\mathbf{H}_\Gamma^{-1}$ or $\mathbf{A}_\Gamma^{-1}$ if single-step genomic best linear unbiased prediction (ssGBLUP) [13, 15, 16] or pedigree-based BLUP is used, respectively. For the structure of $\mathbf{H}_\Gamma^{-1}$ and $\mathbf{A}_\Gamma^{-1}$, see [13]. In addition, σ_u^2 in Eqs. (3) and (4) is substituted by $\sigma_{u\text{-}related}^2$, which is the genetic variance accounting for relatedness in the base population [13, 17].

Using metafounder models, the reliability calculated following Eq. (4) (i.e., $1 - \frac{\text{Var}(\widehat{u}_i^* - u_i^*)}{\text{Var}(u_i^*)}$) will be referred to as rel_i^*. Initial empirical evidence (examples will be shown later) and analytical proofs (see Appendix) showed that calculating reliabilities with metafounders based on Eq. (4) leads to reliability values that are lower than expected—for instance, 0.60 for a proven bull with 100 progenies with records in a pedigree-BLUP evaluation with $h^2 = 0.25$. This is because Eq. (4) refers to a base population of maximum heterozygosity [13] and with an expected value equal to zero, which has no meaning for breeding purposes. Metafounders are conceptually drawn from this base population with covariance matrix $\Gamma\sigma_{u\text{-}related}^2$ [13], and breeding values of individual animals are subsequently drawn through Mendelian sampling. Although this conceptual base

population has no meaningful genetic interpretation, metafounders do. For example, they represent founders of the pedigree [18], founders of pure breeds [19], or unknown parents of animals born in given time periods [20]. In order to refer reliabilities to a base population [14], we propose to define the reliabilities as a contrast to one of these populations, i.e., contrasts to a reference metafounder.

Reliability as a contrast
Thus, we define the reliability rel_{c_i} of the contrast of "the EBV of animal i minus the estimated effect of the metafounder mf", in the usual manner as $rel_{c_i} = r^2\left(\widehat{u}_i^* - \widehat{u}_{mf}^*, u_i^* - u_{mf}^*\right)$, with the following analytical expression:

$$rel_{c_i} = 1 - \frac{\text{Var}\left(\left(\widehat{u}_i^* - u_i^*\right) - \left(\widehat{u}_{mf}^* - u_{mf}^*\right)\right)}{\text{Var}\left(u_i^* - u_{mf}^*\right)}$$

$$= 1 - \frac{\mathbf{C}_{\Gamma_{ii}}^{22} + \mathbf{C}_{\Gamma_{mf,mf}}^{22} - 2\mathbf{C}_{\Gamma_{i,mf}}^{22}}{\left(\mathbf{V}_{\Gamma_{ii}} + \mathbf{V}_{\Gamma_{mf,mf}} - 2\mathbf{V}_{\Gamma_{i,mf}}\right)\sigma_{u\text{-}related}^2}, \tag{5}$$

where $\left(\widehat{u}_{mf}^*\right)u_{mf}^*$ is the (estimated) breeding value of the chosen reference metafounder. In Eq. (5), $\mathbf{V}_{\Gamma_{ii}}$ refers to either $\mathbf{A}_{\Gamma_{ii}}$ or $\mathbf{H}_{\Gamma_{ii}}$, whereas $\mathbf{C}_{\Gamma_{j,k}}^{22}$ refers to the jk element of the corresponding block of the inverse of $\mathbf{C}$ calculated by replacing $\sigma_u^{-2}\mathbf{V}^{-1}$ by $\sigma_{u\text{-}related}^{-2}\mathbf{A}_\Gamma^{-1}$ or $\sigma_{u\text{-}related}^{-2}\mathbf{H}_\Gamma^{-1}$. Note that Eq. (5) can be also used to calculate the reliability of the contrast of the estimated effect of two metafounders. Hence, it is possible to obtain a statistic to evaluate the precision of the estimate of a metafounder's effect, something which is not easy to do for fixed-effect UPG.

Comparison to "regular" animal model reliabilities
Let $rel_i = 1 - \frac{\text{Var}(\widehat{u}_i - u_i)}{A_{ii}\sigma_u^2}$ be the "regular" animal model reliability without UPG or metafounders. In Appendix (Eq. 20), we show that for the case of a single metafounder and under mild assumptions, $rel_i = rel_{c_i}$. This shows that our proposal for rel_{c_i} is a generalization of standard animal breeding theory [7]. The assumption for the proof in the Appendix to hold is that the product of the incidence matrix for the random effects ($\mathbf{Z}$) times a vector of 1s ($\mathbf{1}$) must belong to the column space of the incidence matrix $\mathbf{X}$ for the fixed effects ($\mathbf{Z1} \in \mathcal{C}(\mathbf{X})$, the column space of $\mathbf{X}$). From an interpretational point of view, this assumption implies that an overall mean is fitted implicitly or explicitly in the model ($\mathbf{1} \in \mathcal{C}(\mathbf{X})$) and

Bermann *et al. Genetics Selection Evolution* (2023) 55:6

Page 4 of 13

that the sum of each row of the incidence matrix $\mathbf{Z}$ is constant ($\mathbf{Z1} \propto \mathbf{1}$). This assumption holds for a wide variety of models, including sire, animal (with and without maternal effects), and multi-trait models. For more complex models, such as competition models [21] and models with indirect genetic effects with non-constant group sizes [22], the assumption seems to hold if the models are sensibly specified.

We prove in the Appendix that the expression in Eq. (4) $rel_i^* = 1 - \frac{Var(\hat{u}_i^* - u_i^*)}{Var(u_i^*)}$ gives a systematically lower value of reliability than the regular rel_i; in fact, $rel_i^* = \frac{(2-\gamma)\mathbf{A}_{ii}}{(2-\gamma)\mathbf{A}_{ii}+2\gamma}rel_i$ (see Eq. (19) in Appendix). This is because γ is twice the average inbreeding coefficient of a population; hence, it ranges from 0 to 2 and, therefore, the scalar $\frac{(2-\gamma)\mathbf{A}_{ii}}{(2-\gamma)\mathbf{A}_{ii}+2\gamma}$ always ranges from 0 to 1. For instance, when $\mathbf{A}_{ii} = 1$ (no inbreeding) this gives $rel_i^* = \frac{2-\gamma}{2+\gamma}rel_i$, and for a typical value of γ, say 0.6, this gives $rel_i^* = 0.54rel_i$.

Reliabilities of metafounders
A point that is overlooked in genetic evaluations with UPG is the precision of the estimation of UPG values [9]. In principle, it is possible to compute the standard error of the contrasts of the estimates of two UPG from the elements of the inverse of the MME. However, to our knowledge, this is usually not done. In our study, the reliability of a contrast (Eq. 5) can also be applied to metafounders, which are treated like any other animals. Thus, it is possible to obtain reliabilities of contrasts of metafounders on the same 0 to 1 scale as individuals. So, the reliability of metafounder 2 compared to metafounder 1 also uses Eq. (5). Contrary to contrasts of UPG, the reliabilities of contrasts of metafounders do not need special computational treatment.

Computing strategies
For small datasets, the MME can be solved by inversion of $\mathbf{C}$. Therefore, $\mathbf{C}^{22}_{\Gamma_{ii}}$, $\mathbf{C}^{22}_{\Gamma_{mf,mf}}$, and $\mathbf{C}^{22}_{\Gamma_{i,mf}}$ of Eq. (5) can be obtained from $\mathbf{C}^{-1}$. Also, $\mathbf{K}_\Gamma$ can be explicitly created and used to obtain the elements for the denominator of Eq. (5).

For large datasets without genomic information, $\mathbf{C}^{22}_{\Gamma_{ii}}$ can be retrieved from a sparse inverse of $\mathbf{C}$, or approximated using a reliability approximation method, e.g. [23, 24], or by a Gibbs sampler. The elements $\mathbf{C}^{22}_{\Gamma_{mf,mf}}$ and $\mathbf{C}^{22}_{\Gamma_{i,mf}}$ are obtained in $\mathbf{x}$ by solving the system $\mathbf{Cx} = \mathbf{e}_{mf}$, where $\mathbf{e}_{mf}$ is a vector of 0s except for a 1 in the position of the reference metafounder. Note that the system needs to be solved only once, for the reference metafounder. The

diagonal elements of $\mathbf{A}_\Gamma$ ($\mathbf{A}_{\Gamma_{ii}}$ and $\mathbf{A}_{\Gamma_{mf,mf}}$) are calculated before setting up the MME, using either a modified version of the Meuwissen and Luo [25] algorithm [13] or recursion. The elements $\mathbf{A}_{\Gamma_{mf,i}}$ are obtained by applying the method of Colleau [22], modified to account for metafounders.

For large datasets with genomic information, $\mathbf{C}^{22}_{\Gamma_{ii}}$ can still be obtained from a sparse inverse of $\mathbf{C}$ if the number of genotyped animals is small. Otherwise, $\mathbf{C}^{22}_{\Gamma_{ii}}$ can be approximated by a Gibbs sampler or by reliability approximation methods that account for genomic information, e.g. [26–28]. The calculation of $\mathbf{C}^{22}_{\Gamma_{mf,mf}}$ and $\mathbf{C}^{22}_{\Gamma_{i,mf}}$ is the same as with pedigree only. The diagonal elements of $\mathbf{H}_\Gamma$ can be calculated using the methods of Legarra et al. [29]. Finally, the elements $\mathbf{H}_{\Gamma_{i,mf}}$ can be obtained in $\mathbf{x}$ by solving the system $\mathbf{H}_\Gamma^{-1}\mathbf{x} = \mathbf{e}_{mf}$. A more efficient way would be to solve separate systems for genotyped and non-genotyped animals. For genotyped animals, it is necessary to calculate $\mathbf{x} = \mathbf{GA}_{\Gamma_{22}}^{-1}\mathbf{A}_{\Gamma_{21}}\mathbf{e}_{mf}$, where the subscripts 1 and 2 refer to the non-genotyped and genotyped animals, respectively. First, the product $\mathbf{w} = \mathbf{A}_{\Gamma_{22}}^{-1}\mathbf{A}_{\Gamma_{21}}\mathbf{e}_{mf}$ can be efficiently calculated using the method of Fernando et al. [30], as the solution to $\mathbf{A}_\Gamma^{22}\mathbf{w} = -\mathbf{A}_\Gamma^{21}\mathbf{e}_{mf}$. Then, values of $\mathbf{x}$ for genotyped animals are obtained from $\mathbf{x} = \mathbf{Gw}$. For non-genotyped animals, it is necessary to calculate $\mathbf{x} = \mathbf{A}_{\Gamma_{11}}\mathbf{e}_{mf} + \mathbf{A}_{\Gamma_{12}}\mathbf{A}_{\Gamma_{22}}^{-1}\left(\mathbf{G} - \mathbf{A}_{\Gamma_{22}}\right)\mathbf{A}_{\Gamma_{22}}^{-1}\mathbf{A}_{\Gamma_{21}}\mathbf{e}_{mf}$. The first term is calculated using the method of Colleau [31], whereas the second uses multiplications from right to left. For more details on these methods, we refer to Colleau et al. [32].

The above methods to implement Eq. (5) have been programmed in the BLUPF90+ [33] software and may be invoked by adding the following options to the parameter file:

```
OPTION store_accuracy n
OPTION store_pec_mf arg
```

where n refers to the number of the animal effect and arg defines which metafounder is used as a contrast (`first`, `last`, `user nmf`). The default uses the first (`first`) metafounder, but any metafounder (`nmf`) can be defined as a contrast through the argument `user nmf`.

Materials
The proposed method was tested on a sheep dataset from the Lacaune breed. The number of animals in the pedigree was 1,961,687 (primarily females), from which 1,791,268 had phenotypes for milk yield. In total, 29,138 rams were genotyped with a 50K single nucleotide polymorphism (SNP) chip. The pedigree records started in

Bermann *et al. Genetics Selection Evolution*   (2023) 55:6

Page 5 of 13

Table 1 Definition of metafounders according to a 2-year interval span and number of animals with records per metafounder

Metafounder code	Metafounder year	Number of animals with records
1	< 1978	102,699
2	1978	17,052
3	1980	15,278
4	1982	14,604
5	1984	14,150
6	1986	13,806
7	1988	12,580
8	1990	11,357
9	1992	10,529
10	1994	10,695
11	1996	10,154
12	1998	12,908
13	2000	10,652
14	2002	8760
15	2004	6808
16	2006	6291
17	2008	7299
18	2010	6159
19	2012	5792
20	2014	5414
21	2016	5431
22	2018	4834
23	2020	1678

1970 and the unknown parents of sheep until 1978 constituted the first metafounder. Then, metafounders were created every 2 years. Although pedigree completeness was above 90% and all males born after 1978 had both parents known, there are females with unknown sire or unknown sire and dam. Table 1 shows the number of animals with records assigned to each metafounder, which ranged from 2000 to 100,000.

Analysis

Two models (BLUP and ssGBLUP) with three versions of reliability scenarios were used to explore the properties of regular reliability and the reliability of the contrast between an individual's EBV and the metafounder's effect. For BLUP, the first scenario did not include UPG or metafounders. The reliabilities calculated from this model $Rel_i = 1 - \frac{\text{Var}(\hat{u}_i - u_i)}{\mathbf{K}_{ii}\sigma_u^2}$ will be referred to as *Rel-NoMF* for BLUP and serve as a reference of properly

defined and accurate reliabilities, as most animals have complete pedigree. For ssGBLUP, this scenario was not considered, as genomic reliabilities depend strongly on the genotype coding [14, 34]. The second scenario (used for BLUP and ssGBLUP) included metafounders but without any correction for the reference population. Thus, the reliability was calculated following Eq. (4): $Rel_i^* = 1 - \frac{\text{Var}(\hat{u}_i^* - u_i^*)}{\mathbf{K}_{\Gamma_{i,i}}\sigma_{u\text{-}related}^2}$. For this method, the reliability calculated from the inverse of the MME coefficient matrix refers to a genetic base population with maximum heterozygosity, which will be referred to as *RelMFnc* for BLUP and *ssRelMFnc* for ssGBLUP. Here, we compute reliabilities with sparse inversion for ssGBLUP because the use of metafounders with estimated $\mathbf{\Gamma}$ puts $\mathbf{G}$ and $\mathbf{A}$ on the "same scale" [13, 17], leading to meaningful reliabilities. The third scenario used the same model with metafounders but reliabilities were defined as a contrast to a reference metafounder, as proposed in this study in Eq. (5): $Rel_{c_i} = 1 - \frac{\text{Var}((\hat{u}_i^* - u_i^*) - (\hat{u}_{mf} - u_{mf}))}{Var(u_i^* - u_{mf})}$ and will be referred to as *RelMFc* and *ssRelMFc* for BLUP and ssGBLUP, respectively. The reference metafounder was the first one, which happens to represent the "oldest" Lacaune population.

For all the scenarios with metafounders, their relationship matrix ($\mathbf{\Gamma}$) was calculated using a modification of the method of Macedo et al. [20] as:

$$\mathbf{\Gamma} = \mathbf{1}\mathbf{1}'\hat{\gamma}_0 + 2k\mathbf{T}\mathbf{T}'\Delta F_\gamma, \tag{6}$$

where $\hat{\gamma}_0$ is the estimate of the self-average relationship of the first metafounder (γ_0), k is the time (in years) between consecutive metafounders, $\mathbf{T}$ is a strictly lower triangular matrix of 1s, and $\Delta F_\gamma = \Delta F(1 - 0.5\gamma_0)$, where ΔF is the change in average inbreeding per year (assumed constant but this can be modified). For the dataset used in this study, $k = 2$ and $\Delta F = 6.442 \times 10^{-4}$ per year as obtained by pedigree analyses. Parameter γ_0 was estimated as $\hat{\gamma}_0 = 2n^{-1}\sum_{i=1}^{n}(2\hat{p}_i - 1)^2$ [18], where n is the number of genotyped markers and $\hat{p}_i$ is the estimate of the minor allele frequency in the base population for the ith marker, which was calculated using the method of Gengler et al. [35], resulting in $\hat{\gamma}_0 = 0.46375$. Note that this technique to obtain $\mathbf{\Gamma}$ is only valid in the context of a single breed and cannot be used when the population includes several breeds and crosses. For simplicity, we assumed a continuous decrease in heterozygosity, but this can be easily relaxed. All computations were done using BLUPF90+ [33].
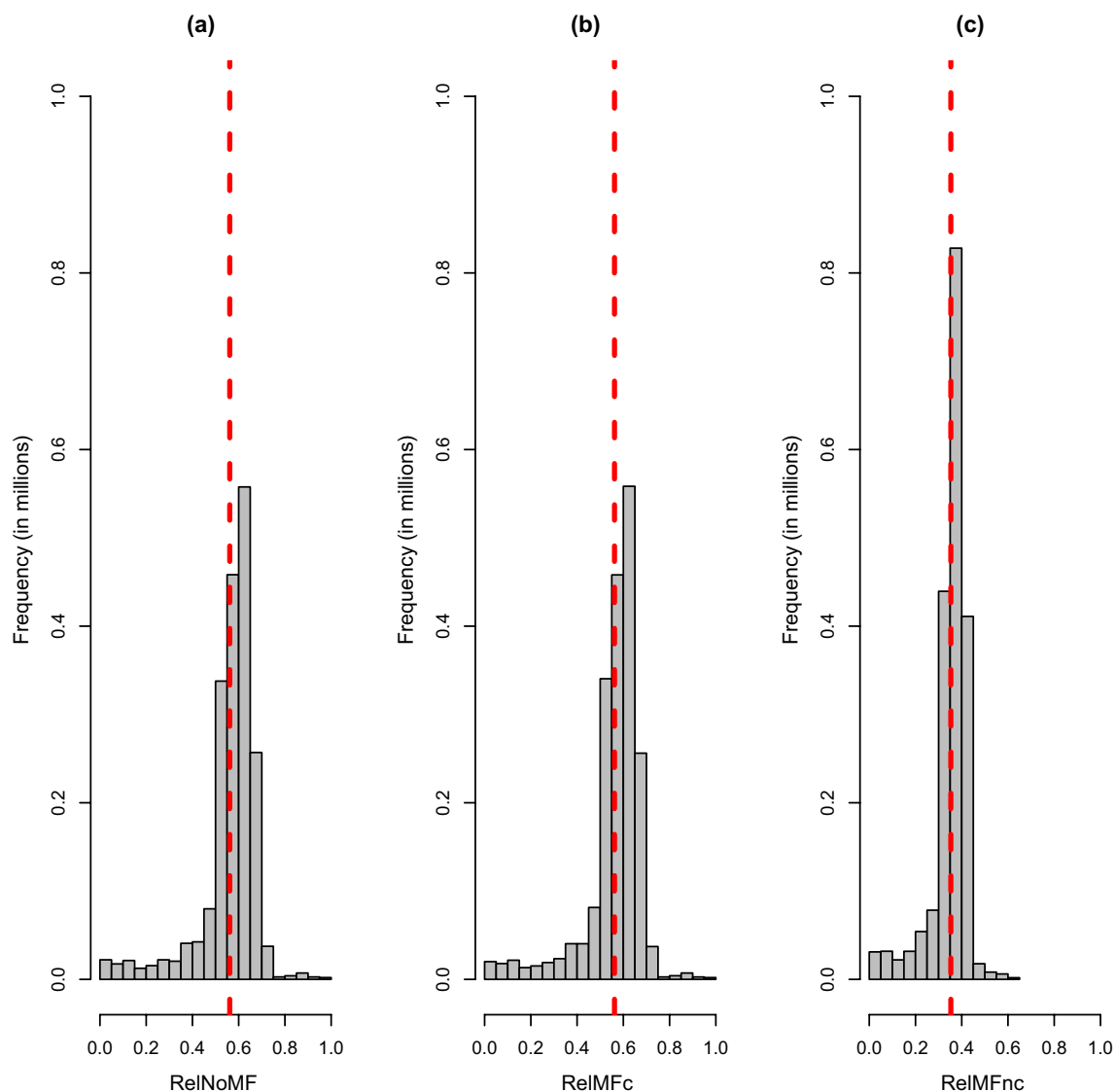
Bermann *et al. Genetics Selection Evolution*   (2023) 55:6

Page 6 of 13

Fig. 1 Histogram of the three reliabilities for all animals: **a** *RelNoMF*, **b** *RelMFc*, and **c** *RelMFnc*. The dashed red line denotes the average reliability for each scenario

Reliabilities obtained from each scenario were compared for all animals and by the following categories of animals: selection candidates and males and females with two, one, or no parents known. Selection candidates were defined as genotyped young males without progeny.

Results

Figure 1 shows histograms for reliabilities for *RelNoMF*, *RelMFc*, and *RelMFnc*. The leftmost and larger peaks contain females, most having own records, whereas the rightmost and smaller peak consists of proven males. The average reliabilities of ~0.55 for females and ~0.90 for proven males in Fig. 1 agree with values from simple selection index theory, as the females have one to three records and proven males have a minimum of 30 daughters each, the heritability of the trait was 0.30 and the repeatability was 0.50. The values for *RelNoMF* and *RelMFc* align extremely well, as also shown in Fig. 2 (for females with both parents known). However, *RelMFnc* resulted in reliabilities that are too low both for males (~0.55) and females (~0.40), as expected based on the
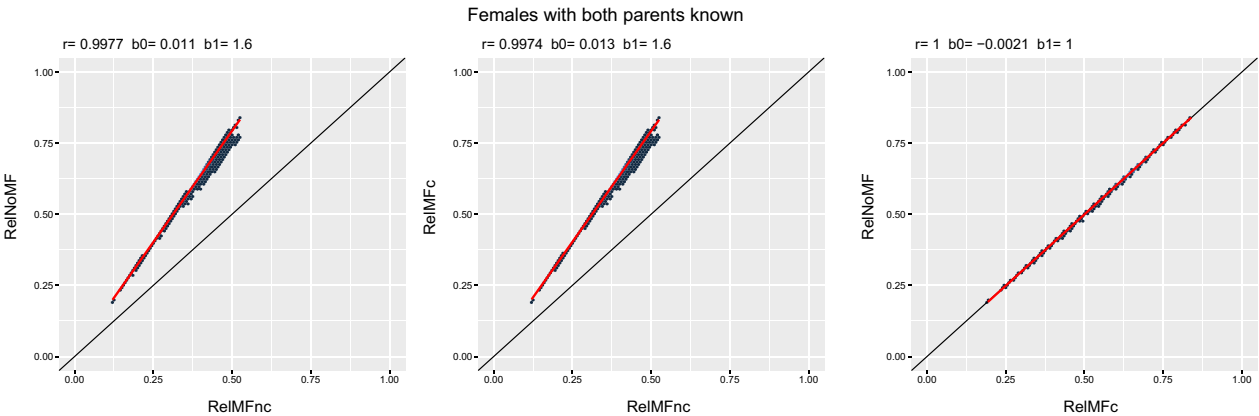
Fig. 2 Scenario comparison for females with both parents known. The red line is the fitted regression
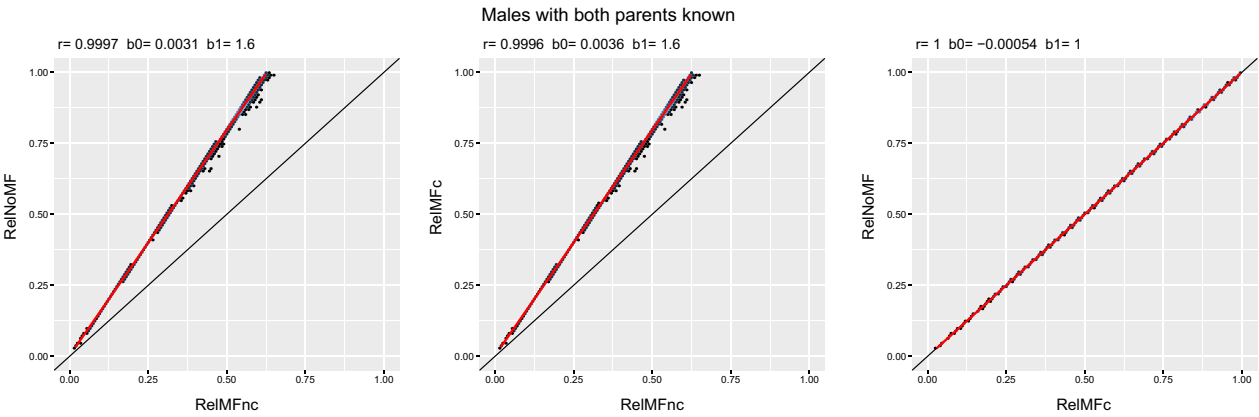

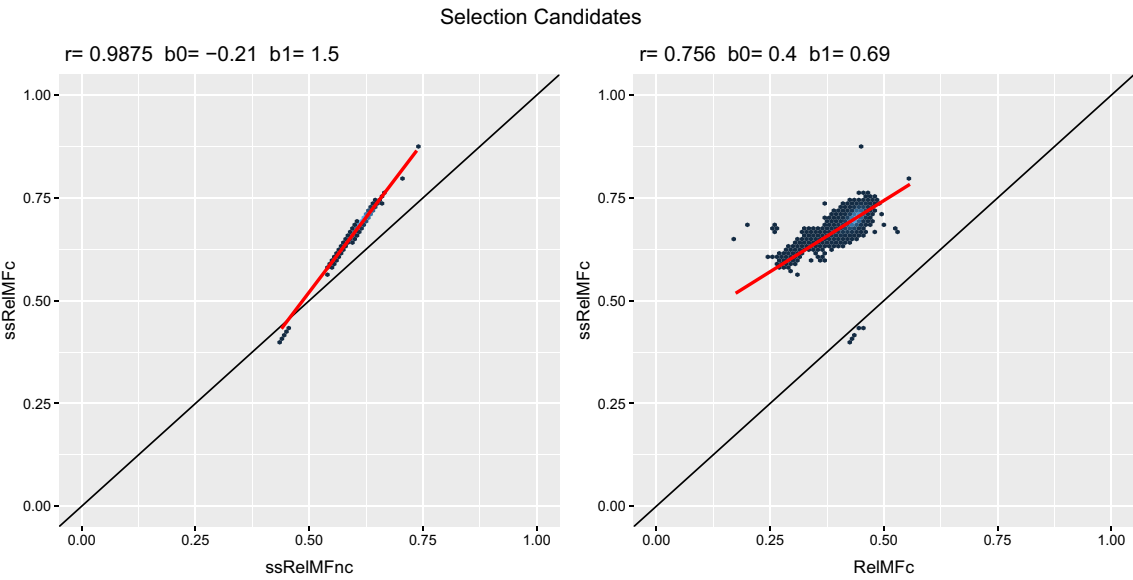**Fig. 3** Scenario comparison for males with both parents known. The red line is the fitted regression

Fig. 4 Comparison between *ssRelMFnc*, *ssRelMFc*, and *RelMFc* for selection candidates

expression $Rel_i^* = \frac{(2-\gamma)\mathbf{A}_{ii}}{(2-\gamma)\mathbf{A}_{ii}+2\gamma}rel_i$. *RelMFnc* also gave reliabilities that are too low because they refer to a conceptual base population with maximum heterozygosity.

Figures 2 and 3 show the comparison between scenarios using BLUP for females and males, respectively. The leftmost plots in Figs. 2 and 3, i.e., the comparison between *RelNoMF* and *RelMFnc* evidence the inadequacy of Eq. (4) for calculating reliabilities with metafounders. The dots over the red dashed line indicate that the reliability is miscalculated because the addition of metafounders in *RelMFnc* should result in very similar reliabilities as *RelNoMF*, given that most individuals have both parents known. The comparison between *RelMFc* against *RelMFnc* in Figs. 2 and 3 demonstrate the underestimation of the reliabilities when using metafounders without the correction that is suggested in Eq. (5). Finally, the rightmost plots show how the calculation of the reliability as a contrast with a reference metafounder corrects the underestimation in *RelMFnc*.

For ssGBLUP, because the vast majority of individuals are not genotyped, results for non-genotyped animals were equal to those from Figs. 2 and 3. For genotyped selection candidates (males with genotype but no offspring yet), Fig. 4 shows the comparison between *ssRelMFnc* and *ssRelMFc*. Again, *ssRelMFnc* was systematically lower than *ssRelMFc*, with the mode for genotyped selection candidates approximately equal to 0.62 for *ssRelMFnc* and 0.70 for *ssRelMFc*. Given that *ssRelMFnc* systematically underestimated reliabilities, *ssRelMFc* provides accurate reliabilities. The comparison of *ssRelMFc*
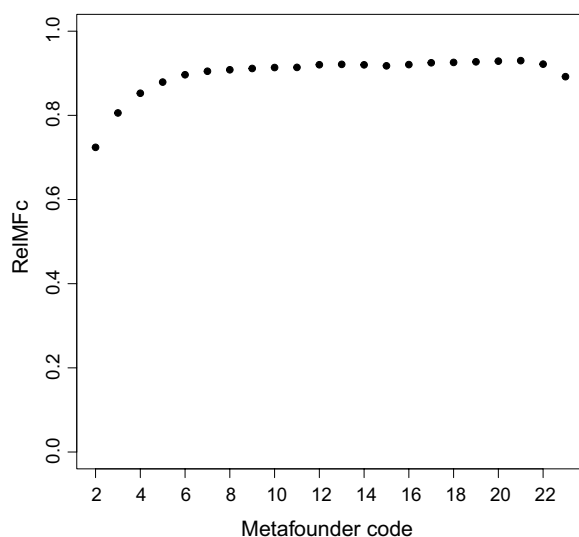
and *RelMFc* in Fig. 4 shows the gain in reliability by adding genomic information.

Finally, Fig. 5 shows the reliabilities of the contrasts of each metafounder with the reference metafounder. Most reliabilities oscillated around 0.90. In particular, the reliabilities declined in the most recent years because there is less information. Given the large number of records per metafounder (Table 1), we expected higher reliabilities. This may illustrate that correct estimation of different base populations (metafounders or UPG) is difficult, probably due to confounding of metafounders with environmental effects (flock-year) and poor genetic connections between metafounders (which, at best, are based through many common offspring of metafounders).

Discussion

We have not been able to find a discussion in the scientific literature on the definition of reliabilities for models with fixed UPG. Presumably, this is mainly because UPG are used in large datasets and reliabilities for EBV are approximated while ignoring UPG [23, 24, 36]. Another reason might be that reliabilities are not uniquely defined because the inverse of the additive relationship matrix is non-full rank [8]. Da et al. [37] presented methods to calculate prediction error variances for models with fixed UPG. Although these prediction error variances can be used in REML procedures, they cannot be used to obtain reliabilities because these prediction error variances and their derived reliabilities depend on the choice of the generalized inverse that is used to solve the mixed model equations. Hickey et al. [38] proposed a sampling-based method to calculate the reliability of EBV by simulating the true breeding value using random UPG. However, in this method, UPG effects are either drawn as random and estimated as random (in which case the associated variance component is unclear and the reliabilities refer to an undefined population), or UPG effects are drawn as fixed and estimated as fixed (in which case they suffer the same problem of lack of uniqueness). In our opinion, metafounders is a structure that is better defined from a quantitative genetics point of view, as relationships between metafounders are functions of heterozygosity at markers [18].

Another point is how to define a base point in genetic evaluations [14]. The metafounder approach uses a conceptual, non-existing population that has maximum heterozygosity, which leads to reliabilities that are too low, i.e. in some way, the maximum heterozygosity population is too distant from the actual animals—for the difference to be correctly estimated. In animal breeding practice, EBV are often referred to the average of a group of recent animals, such as "purebred females born in a given year". This suggests that reliabilities should be defined as

Fig. 5 Reliability for selection candidates using the contrast between all the metafounders and the reference metafounder

contrasts from the same average, which is conceptually feasible but not practical as it would imply manipulating thousands of prediction error variances and covariances. Thus, the breeder must choose a metafounder to provide a base for the contrast. In purebreds, a meaningful choice is an "old" population (representing the oldest founders of the breed), but it could also be the metafounder that has the most unknown parents assigned to it. In crossbreds, a natural choice is one of the parental breeds (or the oldest population within the breed), or even different breeds, which would result in different reliabilities for contrasts from different breeds. Our proposal of using metafounders within breeds assumes that differences in variances per breed and segregation variances are correctly defined by the matrix $\mathbf{\Gamma}$.

Using genomic information added another layer of complexity to the interpretation of reliabilities calculated from the inverse of the MME. Stranden and Christensen [34] showed that different allele coding results in different reliabilities, although the EBV remain the same. Tier et al. [14] proposed a method to obtain prediction error variances that are independent of allele coding for GBLUP. Stranden et al. [39] proposed a method to fit the so-called J-factors and genetic groups, similar to metafounders, but as fixed instead of random effects. One of the drawbacks of their method is that the reliability is undefined. Our proposal of using metafounders together with contrasts overcomes these issues since the arbitrariness of allele frequencies is overcome by fixing them to 0.5 and the base population to which the reliabilities refer is explicitly addressed. Thus, our method adequately reports reliabilities of EBV from ssGBLUP models with different base populations.

An issue with the use of metafounders is the variance components that should be used for the covariance structure because they have no clear genetic interpretation without the associated $\mathbf{\Gamma}$. For a single metafounder, the variance components associated with the metafounder and the genetic (co)variance of pedigree founders are proportional by $\left(1 - \frac{\gamma}{2}\right)$ [13]. For several metafounders, the proportionality constant is $\left(1 + \frac{\overline{diag(\Gamma)}}{2} - \overline{\mathbf{\Gamma}}\right)$ [13]. However, this assumes a mixture in equal proportions of all individuals' origins, which is not necessarily valid. Thus, future research could focus on which covariance matrix should be used when many metafounders are fitted in a model. In our Lacaune example, the scaling factor $1/\left(1 + \frac{\overline{diag(\Gamma)}}{2} - \overline{\mathbf{\Gamma}}\right)$ resulted in a value of 0.7654,

which is very close to $1/\left(1 - \frac{\Gamma_{(1,1)}}{2}\right)$ (where $\Gamma_{(1,1)} = 0.4638$) because the increase in coancestry over time is very small compared to the initial heterozygosity of the breed. This may be true when using metafounders to model missing pedigrees within a breed.

However, in breed crosses, it is not obvious how to obtain variance components with metafounders from "routine" evaluations. The current modeling of breed crosses does not consider segregation variances, so a model with metafounders is just different. However, Poulsen et al. [40] showed by simulation that there is very good agreement between simulated segregation variances and their expected values based on $\mathbf{\Gamma}$ multiplied by a single variance component $\sigma^2_{u\text{-}related}$. Nevertheless, more work is needed in this regard. Alternatively, one could simply estimate $\sigma^2_{u\text{-}related}$ and, if needed, express genetic variances on the usual "unrelated" scale, combining it with $\mathbf{\Gamma}$ [41].

As mentioned before, Eq. (5) allows the reliability of the estimates of genetic groups or metafounders with respect to a reference group to be computed. Reliabilities of the effects of genetic groups are usually not calculated in genetic evaluations. The proposed Eq. (5) could help identify non-reliable groups, which can be a sign of non-optimal assignment of genetic groups. For example, if 2 consecutive metafounders defined by year of birth have a low reliability, combining them would allow more accurate modeling of genetic groups.

Conclusions

To date, there was no expression for the reliabilities with different base populations. Reliabilities that are calculated directly from prediction error variances obtained from the inverse of MME in models with metafounders underestimate the reliability of EBV because they refer to a conceptual base population that has maximum heterozygosity. We propose to calculate reliability as a contrast to a chosen metafounder. This leads to expressions for reliability that are mathematically identical to those for a single population and using pedigree-BLUP, and our empirical results show that they are adequate when there are several base populations. Given that computations are not difficult, we propose that reliabilities as a contrast to a reference metafounder should be used in routine evaluations, when different base populations are fitted into the model, both for BLUP and ssGBLUP.

Bermann *et al. Genetics Selection Evolution*   (2023) 55:6

Page 10 of 13

Appendix

Prediction error variances and reliabilities with a single metafounder

This Appendix shows that, under some assumptions, when there is a single metafounder that represents pedigree founders, the direct reliability of the breeding value has a lower numerical value than for the regular animal model with no metafounder. We also prove that the reliability of the contrast with a single metafounder is mathematically identical to the regular animal model with no metafounder.

A mixed model property

Let a mixed model be:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e},$$

$$E[\mathbf{y}] = \mathbf{X}\boldsymbol{\beta},$$

$$\mathrm{Var}\begin{pmatrix} \mathbf{u} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} \mathbf{K}\sigma_u^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{pmatrix}, \tag{7}$$

with the mixed model equations:

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \sigma_u^{-2}\mathbf{K}^{-1} \end{pmatrix}, \tag{8}$$

and inverse $\mathbf{C}^{-1} = \begin{pmatrix} \mathbf{C}^{11} & \mathbf{C}^{12} \\ \mathbf{C}^{21} & \mathbf{C}^{22} \end{pmatrix}$. Let $\mathbf{Z1} \in \mathcal{C}(\mathbf{X})$ (the column space of $\mathbf{X}$), then $\mathbf{C}^{22}\mathbf{K}^{-1}\mathbf{1} = \mathbf{1}\sigma_u^2$. That is, σ_u^2 is an eigenvalue of $\mathbf{C}^{22}\mathbf{K}^{-1}$, and its associated eigenvector is $\mathbf{1}$.

Proof By partitioned matrix inverse rules [42], $\mathbf{C}^{22} = (\sigma_u^{-2}\mathbf{K}^{-1} + \mathbf{Z}'\mathbf{R}^{-1}(\mathbf{I} - \mathbf{P_X})\mathbf{Z})^{-1}$, where $\mathbf{P_X} = \mathbf{X}(\mathbf{X}'\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{R}^{-1}$ is an oblique projector operator onto the column space of $\mathbf{X}$. Then,

$$\begin{aligned}
\mathbf{K}\mathbf{C}^{22^{-1}}\mathbf{1} &= \mathbf{K}\left(\sigma_u^{-2}\mathbf{K}^{-1}\mathbf{1} + \mathbf{Z}'\mathbf{R}^{-1}(\mathbf{I} - \mathbf{P_X})\mathbf{Z1}\right) \\
&= \mathbf{K}\left(\sigma_u^{-2}\mathbf{K}^{-1}\mathbf{1} + \mathbf{Z}'\mathbf{R}^{-1}(\mathbf{Z1} - \mathbf{P_X}\mathbf{Z1})\right) \\
&= \mathbf{K}\left(\sigma_u^{-2}\mathbf{K}^{-1}\mathbf{1} + \mathbf{Z}'\mathbf{R}^{-1}(\mathbf{Z1} - \mathbf{Z1})\right) \\
&= \sigma_u^{-2}\mathbf{K}\mathbf{K}^{-1}\mathbf{1} = \sigma_u^{-2}\mathbf{1}.
\end{aligned} \tag{9}$$

Since σ_u^{-2} is an eigenvalue of $\mathbf{K}\mathbf{C}^{22^{-1}}$ and $\mathbf{K}\mathbf{C}^{22^{-1}}$ is non-singular, then σ_u^2 is an eigenvalue of $\mathbf{C}^{22}\mathbf{K}^{-1}$, and its associated eigenvector is $\mathbf{1}$.

Reliability for models with one metafounder

When one metafounder is included in an animal model, the assumed linear mixed model is:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \begin{bmatrix} \mathbf{0} & \mathbf{Z} \end{bmatrix}\mathbf{u}^* + \mathbf{e},$$

$$E[\mathbf{y}] = \mathbf{X}\boldsymbol{\beta},$$

$$\mathrm{Var}\begin{pmatrix} \mathbf{u}^* \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} \mathbf{A}_\gamma\sigma_{u\text{-related}}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{pmatrix} = \begin{pmatrix} \mathbf{A}_\gamma w\sigma_u^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{pmatrix}, \tag{10}$$

given that $w = \left(1 - \frac{\gamma}{2}\right)^{-1}$,

$$\text{and } \mathbf{A}_\gamma w = \begin{pmatrix} \gamma & \gamma\mathbf{1}' \\ \gamma\mathbf{1} & \mathbf{A}\left(1 - \frac{\gamma}{2}\right) + \gamma\mathbf{11}' \end{pmatrix}\left(1 - \frac{\gamma}{2}\right)^{-1}, \tag{11}$$

where γ is a scalar. For the genetic interpretation of γ, we refer the reader to Christensen [17] and Legarra et al. [13]. Using partitioned matrix inverse rules, it can be shown   that $\mathbf{A}_\gamma(w)^{-1} = \begin{pmatrix} \gamma^{-1}\left(1 - \frac{\gamma}{2}\right) + \mathbf{1}'\mathbf{A}^{-1}\mathbf{1} & -\mathbf{1}'\mathbf{A}^{-1} \\ -\mathbf{A}^{-1}\mathbf{1} & \mathbf{A}^{-1} \end{pmatrix}$.

The coefficient matrix of mixed model equations is:

$$\mathbf{M} = \begin{pmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} & \mathbf{M}_{13} \\ \mathbf{M}_{21} & \mathbf{M}_{22} & \mathbf{M}_{23} \\ \mathbf{M}_{31} & \mathbf{M}_{32} & \mathbf{M}_{33} \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{0} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{0} & \sigma_u^{-2}\left(\gamma^{-1}\left(1 - \frac{\gamma}{2}\right) + \mathbf{1}'\mathbf{A}^{-1}\mathbf{1}\right) & -\mathbf{1}'\mathbf{A}^{-1}\sigma_u^{-2} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & -\mathbf{A}^{-1}\mathbf{1}\sigma_u^{-2} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{A}^{-1}\sigma_u^{-2} \end{pmatrix}, \tag{12}$$

which thanks to the factorization of $\left(1 - \frac{\gamma}{2}\right)^{-1}$ is in terms of σ_u^{-2}, not of $\sigma_{u\text{-related}}^{-2}$, with inverse $\mathbf{M}^{-1} = \begin{pmatrix} \mathbf{M}^{11} & \mathbf{M}^{12} & \mathbf{M}^{13} \\ \mathbf{M}^{21} & \mathbf{M}^{22} & \mathbf{M}^{23} \\ \mathbf{M}^{31} & \mathbf{M}^{32} & \mathbf{M}^{33} \end{pmatrix}$.

Letting $\mathbf{C}^{22} = \left(\sigma_u^{-2}\mathbf{A}^{-1} + \mathbf{Z}'\mathbf{R}^{-1}(\mathbf{I} - \mathbf{P_X})\mathbf{Z}\right)^{-1}$, i.e., the prediction error variance of the breeding values under a regular animal model:

$$\mathbf{M}^{22} = \left(\mathbf{M}_{22} - \begin{pmatrix} \mathbf{M}_{21} & \mathbf{M}_{23} \end{pmatrix}\begin{pmatrix} \mathbf{M}_{11} & \mathbf{M}_{13} \\ \mathbf{M}_{31} & \mathbf{M}_{33} \end{pmatrix}^{-1}\begin{pmatrix} \mathbf{M}_{12} \\ \mathbf{M}_{32} \end{pmatrix}\right)^{-1}$$

$$= \left(\sigma_u^{-2}\left(\gamma^{-1}\left(1 - \frac{\gamma}{2}\right) + \mathbf{1}'\mathbf{A}^{-1}\mathbf{1}\right) - \mathbf{1}'\mathbf{A}^{-1}\mathbf{C}^{22}\mathbf{A}^{-1}\mathbf{1}\sigma_u^{-4}\right)^{-1}. \tag{13}$$

By the mixed model property derived before, $\mathbf{C}^{22}\mathbf{A}^{-1}\mathbf{1}\sigma_u^{-4} = \mathbf{1}\sigma_u^{-2}$. Thus:

$$\mathbf{M}^{22} = \left(\sigma_u^{-2}\left(\gamma^{-1}\left(1 - \frac{\gamma}{2}\right) + \mathbf{1}'\mathbf{A}^{-1}\mathbf{1}\right) - \mathbf{1}'\mathbf{A}^{-1}\mathbf{1}\sigma_u^{-2}\right)^{-1}$$

$$= \left(\sigma_u^{-2}\gamma^{-1}\left(1 - \frac{\gamma}{2}\right)\right)^{-1} = \sigma_u^2 \frac{2\gamma}{2 - \gamma}. \quad (14)$$

Hence, the prediction error variance of the metafounder effect is $\sigma_u^2 \frac{2\gamma}{2-\gamma}$. Moreover, its prediction error variance is equal to its variance as seen from Eq. (11), since $\sigma_u^2 \gamma \left(1 - \frac{\gamma}{2}\right)^{-1} = \sigma_u^2 \frac{2\gamma}{2-\gamma}$.

Then, for the animal effect block:

$$\mathbf{M}^{33} = \left(\mathbf{C}^{22^{-1}} - \mathbf{A}^{-1}\mathbf{1}\mathbf{1}'\mathbf{A}^{-1}k^{-1}\right)^{-1}, \quad (15)$$

where $k = \sigma_u^2\left(\gamma^{-1}\left(1 - \frac{\gamma}{2}\right) + \mathbf{1}'\mathbf{A}^{-1}\mathbf{1}\right)$. Applying a variant of the Sherman–Morrison formula ([42]; formula 14) to Eq. (15):

$$\mathbf{M}^{33} = \mathbf{C}^{22} + \frac{\mathbf{C}^{22}\mathbf{A}^{-1}\mathbf{1}\mathbf{1}'\mathbf{A}^{-1}\mathbf{C}^{22}k^{-1}}{1 - \mathbf{1}'\mathbf{A}^{-1}\mathbf{C}^{22}\mathbf{A}^{-1}\mathbf{1}k^{-1}}. \quad (16)$$

Using the mixed model property:

$$\mathbf{M}^{33} = \mathbf{C}^{22} + \frac{\sigma_u^4 k^{-1}}{1 - \mathbf{1}'\mathbf{A}^{-1}\mathbf{1}\sigma_u^2 k^{-1}}\mathbf{1}\mathbf{1}'$$

$$= \mathbf{C}^{22} + \frac{\sigma_u^4}{k - \mathbf{1}'\mathbf{A}^{-1}\mathbf{1}\sigma_u^2}\mathbf{1}\mathbf{1}' = \mathbf{C}^{22} + \frac{\sigma_u^4}{\sigma_u^2\left(\gamma^{-1}\left(1 - \frac{\gamma}{2}\right) + \mathbf{1}'\mathbf{A}^{-1}\mathbf{1}\right) - \mathbf{1}'\mathbf{A}^{-1}\mathbf{1}\sigma_u^2}\mathbf{1}\mathbf{1}'$$

$$= \mathbf{C}^{22} + \frac{2\gamma\sigma_u^2}{2 - \gamma}\mathbf{1}\mathbf{1}' = \mathrm{Var}(\widehat{\mathbf{u}} - \mathbf{u}) + \mathrm{Var}(\widehat{u}_{mf} - u_{mf})\mathbf{1}\mathbf{1}'. \quad (17)$$

Thus, the prediction error variance of the breeding values in a model with one metafounder is equal to the prediction error variance of the breeding values in a model without metafounders plus the prediction error variance of the metafounder effect.

Lastly, using the mixed model property:

$$\mathbf{M}^{32} = \mathbf{C}^{22}\mathbf{A}^{-1}\mathbf{1}\sigma_u^4 \frac{2\gamma}{2 - \gamma} = \mathbf{1}\sigma_u^2 \frac{2\gamma}{2 - \gamma}. \quad (18)$$

Hence, the prediction error covariance between an animal effect and the metafounder effect is equal to the prediction error variance of the metafounder.

Using Eqs. (14) and (17) and $\mathrm{Var}(\widehat{u}_i - u_i) = \mathbf{A}_{ii}(1 - rel_i)$, the reliability of an animal under a model with one metafounder is:

$$rel_i^* = 1 - \frac{\mathrm{Var}(\widehat{u}_i - u_i) + \mathrm{Var}(\widehat{u}_{mf} - u_{mf})}{\mathbf{A}_{ii}\sigma_u^2 + \gamma\left(1 - \frac{\gamma}{2}\right)^{-1}\sigma_u^2}$$

$$= 1 - \frac{(2 - \gamma)\mathbf{A}_{ii}(1 - rel_i) + 2\gamma}{(2 - \gamma)\mathbf{A}_{ii} + 2\gamma} = \frac{(2 - \gamma)\mathbf{A}_{ii}}{(2 - \gamma)\mathbf{A}_{ii} + 2\gamma}rel_i. \quad (19)$$

Given the range of γ, $rel_i^* < rel_i$.

The reliability of the contrast between the ith animal and the metafounder (rel_{c_i}) is:

$$rel_{c_i} = 1 - \frac{\mathrm{Var}\left((\widehat{u}_i^* - u_i^*) - (\widehat{u}_{mf} - u_{mf})\right)}{\mathrm{Var}(u_i^* - u_{mf})}$$

$$= 1 - \frac{\mathrm{Var}(\widehat{u}_i^* - u_i^*) + \mathrm{Var}(\widehat{u}_{mf} - u_{mf}) - 2\,\mathrm{cov}\left((\widehat{u}_i^* - u_i^*), (\widehat{u}_{mf} - u_{mf})\right)}{\mathrm{Var}(u_i^*) + \mathrm{Var}(u_{mf}) - 2\mathrm{cov}(u_i^*, u_{mf})}$$

$$= 1 - \frac{\mathbf{M}_{ii}^{22} + \mathbf{M}^{33} - 2\mathbf{M}_i^{23}}{\left(\mathbf{A}_{\gamma ii} + \mathbf{A}_{\gamma_{mf,mf}} - 2\mathbf{A}_{\gamma_{i,mf}}\right)\sigma_u^2}$$

$$= 1 - \frac{\mathbf{C}_{ii}^{22} + \frac{2\gamma\sigma_u^2}{2-\gamma} + \frac{2\gamma\sigma_u^2}{2-\gamma} - 2\frac{2\gamma\sigma_u^2}{2-\gamma}}{\left(\mathbf{A}_{ii}\sigma_u^2 + \gamma\left(1 - \frac{\gamma}{2}\right)^{-1}\sigma_u^2\right) + \left(\gamma\left(1 - \frac{\gamma}{2}\right)^{-1}\right)\sigma_u^2 - 2\left(\gamma\left(1 - \frac{\gamma}{2}\right)^{-1}\right)\sigma_u^2}$$

$$= 1 - \frac{\mathbf{C}_{ii}^{22}}{\mathbf{A}_{ii}\sigma_u^2} = 1 - \frac{\mathrm{Var}(\widehat{u}_i - u_i)}{\mathrm{Var}(u_i)} = rel_i. \quad (20)$$

Thus, the reliability of the contrast (rel_{c_i}) in a model with one metafounder and the reliability of the regular animal model (rel_i) are identical.

For multiple-trait models, the necessary mixed-model property is $\left((\mathbf{K} \otimes \mathbf{G}_0)\mathbf{C}^{22^{-1}}\right)(\mathbf{1} \otimes \mathbf{I}) = (\mathbf{1} \otimes \mathbf{I})$, where $\mathbf{K}$ is the covariance matrix for random effects and $\mathbf{G}_0$ is the covariance matrix between traits. Then, Eq. (20) holds because it can be proved that the results for Eqs. (14), (17), and (18) are $\mathbf{G}_0 \frac{2\gamma}{2-\gamma}$, $\mathbf{C}^{22} + \mathbf{1}\mathbf{1}' \otimes \mathbf{G}_0 \frac{2\gamma}{2-\gamma}$, and $\mathbf{1} \otimes \mathbf{G}_0 \frac{2\gamma}{2-\gamma}$, respectively.

Author contributions
AL conceived the idea of the study and conducted the analysis. MB formalized the theory. MB, AL, and IA took part in the discussion and co-wrote the manuscript. DL and IM edited the manuscript. All authors read and approved the final manuscript.

Data availability
The data used in this work is owned by the organizations responsible for the Lacaune breeding program (Upra Lacaune, Ovitest, Confederation Générale de Roquefort), to whom it may be requested.

Declarations

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

References
1. Sorensen DA, Kennedy BW. Estimation of response to selection using least-squares and mixed model methodology. J Anim Sci. 1984;58:1097–106.
2. Kennedy BW, Schaeffer LR, Sorensen DA. Genetic properties of animal models. J Dairy Sci. 1988;71:17–26.
3. Quaas RL. Additive genetic model with groups and relationships. J Dairy Sci. 1988;71:91–8.
4. Robinson GK. Group effects and computing strategies for models for estimating breeding values. J Dairy Sci. 1986;69:3106–11.
5. Westell RA, Quaas RL, Van Vleck LD. Genetic groups in an animal model. J Dairy Sci. 1988;71:1310–8.
6. Kennedy BWCR. Henderson: the unfinished legacy. J Dairy Sci. 1991;74:4067–81.
7. Van Vleck LD. Variance of prediction error with mixed model equations when relationships are ignored. Theor Appl Genet. 1993;85:545–9.
8. Henderson CR. Applications of linear models in animal breeding. Guelph: University of Guelph; 1984.
9. Kennedy BW. Bias and mean square error from ignoring genetic groups in mixed model sire evaluation. J Dairy Sci. 1981;64:689–97.
10. VanRaden PM. Accounting for inbreeding and crossbreeding in genetic evaluation of large populations. J Dairy Sci. 1992;75:3136–44.
11. Lo LL, Fernando RL, Grossman M. Covariance between relatives in multibreed populations: additive model. Theor Appl Genet. 1993;87:423–30.
12. García-Cortés LA, Toro MA. Multibreed analysis by splitting the breeding values. Genet Sel Evol. 2006;38:601–15.
13. Legarra A, Christensen OF, Vitezica ZG, Aguilar I, Misztal I. Ancestral relationships using metafounders: finite ancestral populations and across population relationships. Genetics. 2015;200:455–68.
14. Tier B, Meyer K, Swan A. On implied genetic effects, relationships and alternate coding. In: Proceedings of the 11th world congress on genetics applied to livestock production: 11–16 February 2018; Auckland. 2018.
15. Legarra A, Aguilar I, Misztal I. A relationship matrix including full pedigree and genomic information. J Dairy Sci. 2009;92:4656–63.
16. Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ. Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J Dairy Sci. 2010;93:743–52.
17. Christensen OF. Compatibility of pedigree-based and marker-based relationship matrices for single-step genetic evaluation. Genet Sel Evol. 2012;44:37.
18. Garcia-Baccino CA, Legarra A, Christensen OF, Misztal I, Pocrnic I, Vitezica ZG, Cantet RJ. Metafounders are related to F_{st} fixation indices and reduce bias in single-step genomic evaluations. Genet Sel Evol. 2017;49:34.
19. Xiang T, Christensen OF, Legarra A. Technical note: genomic evaluation for crossbred performance in a single-step approach with metafounders. J Anim Sci. 2017;95:1472–80.
20. Macedo FL, Astruc JM, Meuwissen THE, Legarra A. Removing data and using metafounders alleviates biases for all traits in Lacaune dairy sheep predictions. J Dairy Sci. 2022;105:2439–52.
21. Angarita BK, Han J, Cantet RJC, Chewning SK, Wurtz KE, Siegford JM, Ernst CW, Steibel JP. Estimation of direct and social effects of feeding duration in growing pigs using records from automatic feeding stations. J Anim Sci. 2021;99:skab042.
22. Poulsen BG, Ask B, Nielsen HM, Ostersen T, Christensen OF. Prediction of genetic merit for growth rate in pigs using animal models with indirect genetic effects and genomic information. Genet Sel Evol. 2020;52:58.
23. Misztal I, Wiggans GR. Approximation of prediction error variance in large-scale animal models. J Dairy Sci. 1988;71:27–32.
24. Tier B, Meyer K. Approximating prediction error covariances among additive genetic effects within animals in multiple-trait and random regression models. J Anim Breed Genet. 2004;121:77–89.
25. Meuwissen T, Luo Z. Computing inbreeding coefficients in large populations. Genet Sel Evol. 1992;24:305–13.
26. Liu Z, Goddard ME, Reinhardt F, Reents R. A single-step genomic model with direct estimation of marker effects. J Dairy Sci. 2014;97:5833–50.
27. Ben Zaabza H, Mäntysaari EA, Strandén I. Using Monte Carlo method to include polygenic effects in calculation of SNP-BLUP model reliability. J Dairy Sci. 2020;103:5170–82.
28. Bermann M, Lourenco D, Misztal I. Efficient approximation of reliabilities for single-step genomic best linear unbiased predictor models with the algorithm for Proven and Young. J Anim Sci. 2022;100:skab353.
29. Legarra A, Aguilar I, Colleau JJ. Short communication: methods to compute genomic inbreeding for ungenotyped individuals. J Dairy Sci. 2020;103:3363–7.
30. Fernando RL, Dekkers JC, Garrick DJ. A class of Bayesian methods to combine large numbers of genotyped and non-genotyped animals for whole-genome analyses. Genet Sel Evol. 2014;46:50.
31. Colleau JJ. An indirect approach to the extensive calculation of relationship coefficients. Genet Sel Evol. 2002;34:409–21.
32. Colleau JJ, Palhière I, Rodríguez-Ramilo ST, Legarra A. A fast indirect method to compute functions of genomic relationships concerning genotyped and ungenotyped individuals, for diversity management. Genet Sel Evol. 2017;49:87.
33. Lourenco D, Tsuruta S, Masuda Y, Bermann M, Legarra A, Misztal I. Recent updates in the BLUPF90 software suite. In: Proceedings of the 12th world congress on genetics applied to livestock production: 3–8 July 2022; Rotterdam. 2022.
34. Strandén I, Christensen OF. Allele coding in genomic evaluation. Genet Sel Evol. 2011;43:25.
35. Gengler N, Mayeres P, Szydlowski M. A simple method to approximate gene content in large pedigree populations: application to the myostatin gene in dual-purpose Belgian Blue cattle. Animal. 2007;1:21–8.

Bermann *et al. Genetics Selection Evolution*   (2023) 55:6

Page 13 of 13

36. Harris B, Johnson D. Approximate reliability of genetic evaluations under an animal model. J Dairy Sci. 1998;81:2723–8.
37. Da Y, Grossman M, Misztal I. Prediction error variance and restricted maximum likelihood estimation for animal model with relationship grouping. J Dairy Sci. 1989;72:2125–35.
38. Hickey JM, Keane MG, Kenny DA, Cromie AR, Mulder HA, Veerkamp RF. Estimation of accuracy and bias in genetic evaluations with genetic groups using sampling. J Anim Sci. 2008;86:1047–56.
39. Strandén I, Aamand GP, Mäntysaari EA. Single-step genomic BLUP with genetic groups and automatic adjustment for allele coding. Genet Sel Evol. 2022;54:38.
40. Poulsen BG, Ostersen T, Nielsen B, Christensen OF. Predictive performances of animal models using different multibreed relationship matrices in systems with rotational crossbreeding. Genet Sel Evol. 2022;54:25.
41. Aldridge MN, Vandenplas J, Bergsma R, Calus MPL. Variance estimates are similar using pedigree or genomic relationships with or without the use of metafounders or the algorithm for proven and young animals. J Anim Sci. 2020;98:skaa019.
42. Henderson HV, Searle SR. On deriving the inverse of a sum of matrices. SIAM Rev. 1981;23:53–60.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.