289. Definition of reliabilities for models with metafounders

M. Bermann[1], I. Misztal[1], D. Lourenco[1], I. Aguilar[2] and A. Legarra[3*]

[1]*Animal and Dairy Science, University of Georgia, Athens, GA 30605, USA;* [2]*INIA, Montevideo, Uruguay;* [3]*INRAE, 31526 Castanet Tolosan, France;* andres.legarra@inrae.fr

Abstract

For models with several base populations (Unknown Parent Groups or Metafounders), the usual definition of reliability is ill-posed. Here we propose to define reliability based on contrasts with one or several metafounders, leading to a sounder genetic interpretation. In the case of a single metafounder, our definition equals the definition of reliability for the classical animal model. This definition also allows expressing the reliability of contrasts of metafounders, which may be of interest to decide if their setup is estimable with sufficient reliability. All desired quantities can be obtained from elements of the inverse of the MME.

Introduction

Reliabilities (*Rel*) are usually defined as the square of the correlation between estimated and true breeding values (EBV and TBV) on conceptual repeated sampling. When different base populations exist, they are modelled as Unknown Parent Groups (UPG; fixed or random) or Metafounders (MF; random). The total EBV (u) can be seen as a sum of genetic levels of the different base populations plus random deviations. An EBV has a reference point with a value of '0'. In the animal model with neither UPG nor MF, the base population is infinite with a mean breeding value of 0, so the EBVs already refer to this base population. In fixed UPG, EBVs are non-estimable functions and the '0' depends on the particular generalized inverse used for solving. There is thus no meaningful base population the EBVs refer to. In the case of MF, the EBVs refer to an ideal population of maximum heterozygosity, which has a value of 0.

Reliabilities need to refer to a base population, which is the '0' point when using some form of BLUP (Tier *et al.*, 2018). In the case of UPGs, the fact that they are fixed effects means that the reliability cannot be properly defined or computed. For MF, reliabilities obtained referring to the ideal population made little genetic sense as this population doesn't actually exist. We show that a sensible thing to do is to refer reliabilities to one (or several) MF using a contrast. In the case of a single base population, these reliabilities are identical to the classical ones.

On the other hand, UPGs or MFs are essential for correct genetic evaluations when animals lack pedigrees or when crosses of different breeds / populations / countries exist. Some genetic evaluations fit as many as 400 UPGs. In practice, UPG/MF may include few animals with phenotypes; therefore, they may not be accurately estimated, leading to suboptimal selection decisions (Kennedy, 1981). We therefore propose to routinely evaluate reliabilities of contrasts among MF, to ascertain if MF are correctly set up and estimated.

In (with neither UPG nor MF) BLUP and ssGBLUP, 'classical' reliabilities for individual i can be obtained using explicit inverses of the Mixed Model Equations, e.g. $Rel^c = 1 - PEV/(A_{ii}\sigma_u^2)$ using prediction error variance (PEV). However, including UPG or MF in the model leads to practical and conceptual problems. When UPG are fit as fixed effects, the usual equation for reliability does *not* hold, because $Var(\hat{u}) \neq Cov(\hat{u}, u)$ when u includes the UPG (Henderson, 1984). The expression is approximately correct *only* when UPG are estimated with very high precision. This is the case (roughly) when a large number of first-generation offspring from UPG have records, and UPG are not confounded. When UPG are fit as random effects with variance σ^2_{UPG}, the equation holds, but the UPG variance component should be included in the denominator e.g. $Rel_i^c = 1 - PEV/(\sigma^2_{UPG} + A_{ii}\sigma_u^2)$.

R.F. Veerkamp and Y. de Haas (eds)

DOI: 10.3920/978-90-8686-940-4_289,

On the other hand, MF are a comprehensive solution to model genetic relationships (Γ) between and within different base populations, for instance the increase of overall relationship within a breed over time (Legarra *et al.*, 2015). These are reflected in modified relationships $A^{(\Gamma)}$ and genetic variance $\sigma^2_{u(\Gamma)}$. By construction, MF are deemed to be more adequate than UPG, in particular for ssGBLUP applications. Because MF are random effects, in principle, one could use $1 - PEV^{(\Gamma)}/(A^{(\Gamma)}_{ii}\,\sigma^2_{u(\Gamma)}))$ to obtain reliabilities. However, empirical assessment showed that this equation underestimated reliabilities for proven bulls.

Materials & methods

The reliability from PEV in the MF case computed as $1 - PEV^{(\Gamma)}/(A^{(\Gamma)}_{ii}\,\sigma^2_{u(\Gamma)}$ refers to this ideal population, but we are not interested in this reliability as it has no genetic interpretation or use. It is more meaningful to refer the reliability of the EBV to a single, genetically well-defined population (Tier *et al.*, 2018). This population can simply be one of the MF. In a purebred population with MF modelling missing pedigree, it can simply be the oldest MF. Thus, instead of considering the EBV $\hat{u}_i^{mf}$, we consider the contrast $\hat{u}_i^{mf} - \hat{u}_{mf}^{mf}$. To clarify, we use the super-index *mf* for an EBV modelled with metafounders, whereas subindices refer generally either to individual *i* or to the reference metafounder *mf*.

Reliability of EBV with metafounders. Given the contrast of EBVs, we can define the reliability of the contrast as $Rel_i^{mf} = Rel\left(\hat{u}_i^{mf} - \hat{u}_{mf}^{mf}\right)$. This is equal to:

$$Rel_i^{mf} = Rel\left(\hat{u}_i^{mf} - \hat{u}_{mf}^{mf}\right) = 1 - \frac{Var\left(\left[u_i^{mf} - u_{mf}^{mf}\right] - \left[\hat{u}_i^{mf} - \hat{u}_{mf}^{mf}\right]\right)}{Var\left(\left[u_i^{mf} - u_{mf}^{mf}\right]\right)}$$
$$= 1 - \frac{PEV\left(\hat{u}_i^{mf}\right) + PEV\left(\hat{u}_{mf}^{mf}\right) - 2PEC\left(\hat{u}_i^{mf}, \hat{u}_{mf}^{mf}\right)}{\left(A^{(\Gamma)}_{ii} + A^{(\Gamma)}_{mf,mf} - 2A^{(\Gamma)}_{i,mf}\right)\sigma^2_{u(\Gamma)}}$$

This is the general equation to obtain the reliability of the contrast 'individual – reference MF' $(\hat{u}_i^{mf} - \hat{u}_{mf}^{mf})$ from the inverse of the MME. In the equation, $PEC\left(\hat{u}_i^{mf}, \hat{u}_{mf}^{mf}\right)$ is the prediction error covariance, and $A^{(\Gamma)}_{ij}$ are corresponding relationship coefficients (which in ssGBLUP will be from matrix $H^{(\Gamma)}$. Interestingly, we can define different reliabilities with different metafounders; for instance, in a complex, unbalanced cross, for the same animal, one can potentially define different reliabilities for each origin.

In the particular case of a single metafounder, it can be shown that $Rel_i^{mf} = Rel\left(\hat{u}_i^{mf} - \hat{u}_{mf}^{mf}\right)$ is equal to the classical 'animal model, no UPG' reliability $Rel_i^c = 1 - PEV/(A_{ii}\sigma^2_u)$ where PEV is from the 'classical' animal model, i.e. $Rel_i^{mf} = Rel_i^c$. This is intuitively expected.

Reliability of metafounders' contrasts estimation. Following the same logic, we define a reliability of contrasts of MF estimation, namely

$$Rel^{mf}_{mf(i),mf(j)} = Rel\left(\hat{u}^{mf}_{mf(i)} - \hat{u}^{mf}_{mf(j)}\right)$$
$$= 1 - \frac{PEV\left(\hat{u}^{mf}_{mf(i)}\right) + PEV\left(\hat{u}^{mf}_{mf(j)}\right) - 2PEC\left(\hat{u}^{mf}_{mf(i)}, \hat{u}^{mf}_{mf(j)}\right)}{\left(\Gamma_{i,i} + \Gamma_{j,j} - 2\Gamma_{i,j}\right)\sigma^2_{u-related}}$$

this allows to appraise the quality of MF estimates in a scale from 0 to 1, something very difficult to do with UPG. Indeed, Kennedy (1981) suggested to check the standard error of contrasts of UPG to avoid wrong selection decisions. It is worth noting that the definition of Rel^{mf} above is strictly the same as for animals.

Computing aspects. To obtain reliabilities in 'classical' animal models (neither UPG nor MF), only the PEV from the diagonal of the inverse of the MME are needed, usually from the sparse inverse. In our

proposal, to obtain Rel_i^{mf}, the vector of $PEC\left(\hat{u}_i^{mf}, \hat{u}_{mf}^{mf}\right)$ is needed. This vector can be obtained by solving the equation $(MME)x = y$, where y contains 1 in the position of the reference MF and 0 elsewhere; on the output, x contains PEC of the reference MF versus all the other unknowns in the MME. Because MF are random effects, this vector of PECs (and all PEVs in the procedure) is invariant to the usual lack of full rank of the MME.

Additionally, the relationships of the metafounders with the other individuals can be obtained by solving $\boldsymbol{A}^{(\Gamma)^{-1}}\boldsymbol{x} = \boldsymbol{y}$ where $y = 1$ in the location of the metafounder and 0 otherwise. The solution vector x contains the values of the row of $A^{(\Gamma)}$ corresponding to the reference MF, and hence all $A_{i,mf}^{(\Gamma)}$. A similar procedure can be used for SSGBLUP $H^{(\Gamma)}$, combined with Legarra *et al.* (2020) to obtain $diag(H^{(\Gamma)})$.

Example. The 12-individuals pedigree with 2 MF described in Legarra *et al.* (2015) (Figure 1) was considered, with:

$$h^2 = 0.5\ ,\ \boldsymbol{\Gamma} = \begin{pmatrix} 0.55 & 0.48 \\ 0.48 & 0.77 \end{pmatrix},$$

and one record per individual. The model includes an overall (fixed) mean. We computed Rel^{mf1}, Rel^{mf2} and Rel^{c} (in the last case, ignoring MF in the model).

Results

Numbers from the example are shown in Table 1. It can be seen that Rel_i^{mf} is roughly similar to Rel_i^{c} in this (rather artificial) example. It is interesting to note that the reliability is higher with respect to the MF who is *not* the ancestor of the individual, and this can be explained because the phenotype of an individual contributes both to its EBV and to the EBV of its ancestral MF. It is also of interest to see that the reliability of the contrast of metafounders is not very high.

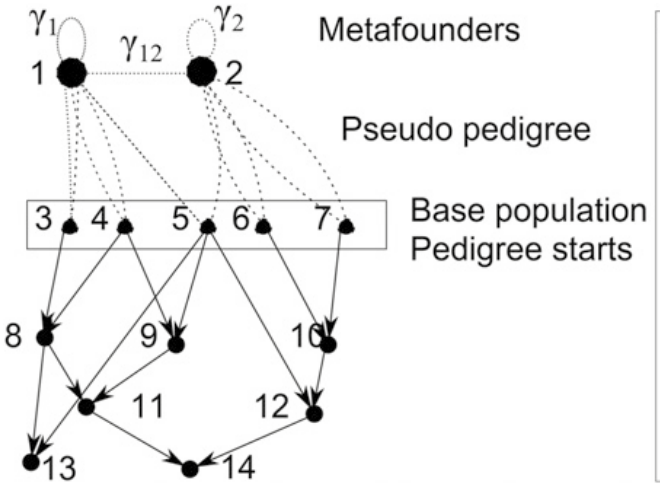

Figure 1. Example pedigree with metafounders.

Table 1. Reliability of EBV based on the contrast with each metafounder (Rel^{mf}) and from the 'classical' animal model with no MF or UPG (Rel^{c}).[1]

Individual	Rel^{mf1}	Rel^{mf2}	Rel^{c}
1		0.23*	
2	0.23*		
3	0.39	0.49	0.43
4	0.40	0.50	0.44
5	0.40	0.41	0.45
6	0.46	0.35	0.43
7	0.46	0.35	0.43
8	0.36	0.51	0.42
9	0.34	0.45	0.41
10	0.48	0.32	0.41
11	0.37	0.50	0.43
12	0.41	0.31	0.39
13	0.30	0.42	0.37
14	0.33	0.37	0.37

[1] An asterisk indicates reliability of the contrasts among the two metafounders.

Discussion

This work provides a rationale for the definition and computation of reliabilities in a model with several founder populations – something that had not been done before. Our proposal should give more meaningful reliabilities in populations with missing pedigrees or crossbreds. In addition, we emphasize that the computations involved are doable for medium-size data sets, and approximations can be used for larger ones. Moreover, the reliability of contrasts among MF may be used to verify whether differences across MF are estimated with enough precision, and therefore to help defining number and definition of MF.

Acknowledgements

This work received financing from European Unions' Horizon 2020 Research & Innovation Programme, Grant/Award Number: N°772787, Smarter.

References

Henderson C.R. (1984) Applications of Linear Models in Animal Breeding. University of Guelph, Guelph.

Kennedy B.W. 1981. J Dairy Sci 64:689–697. https://doi.org/10.3168/jds.S0022-0302(81)82631-8.

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

Tier B., Meyer K., and Swan A.. (2018) Proc. of the 11th WCGALP, Auckland, New Zealand. http://www.wcgalp.org/proceedings/2018/implied-genetic-effects-relationships-and-alternate-allele-coding