A supplementary document to the Interbull genomic reliability method

A standardised statistical method was developed by the Interbull Working Group (Liu et al. 2017) to approximate genomic reliability values for national genomic evaluations. The aim of the current document is to give further recommendations on how to clearly define a genomic reference population, required in some calculation steps of the method, and how to avoid double counting of performance data contribution of cows in case of a mixed reference population.

I. Definition of the genomic reference population

Genomic reference animals are defined as genotyped animals that provide primary information for the estimation of SNP marker effects. The reference animal may be a genotyped bull with some phenotyped non-genotyped daughters or a genotyped cow with own performance records. The phenotype information of the reference animal must be directly measured on the animal itself, like cow's own performance records in national animal model evaluations, or non-genotyped daughters' performance data of bulls like for Interbull bull MACE evaluation.

II. Estimation of reliabilities of the performance data contribution of genomic reference animals

In early years of genomic selection, most reference animals have been genotyped bulls with daughters. Even in a large-scale cow genotyping scheme, not all daughters of genotyped bulls are genotyped and qualified for being included in reference population. Usually, countries do not have access to phenotype and genotype data of foreign daughters of domestic or international reference bulls. Therefore, a mixed set of reference bulls and cows may represent the most common form of national genomic reference population for a long time. In case of such a mixture, a double counting of the performance data contribution of these reference cows, contributions to their own effective records as well as to their sires' effective daughter contributions, must be avoided. Following is a set of recommendations on how to estimate the accuracy of the performance information of the reference animal expressed as effective daughter contribution (EDC) for reference bulls or effective record contribution (ERC) for reference cows. We denote EDC or ERC as φ throughout this document,

following the convention by Liu et al. (2017). The EDC or ERC may be expressed on a sire model basis, φ_s, or on an animal model basis, φ . Both types of EDC or ERC, φ and φ_s, result in equal reliability with their respective variance ratio: $\lambda = \frac{1-h^2}{h^2}$ for the animal model and $\lambda_s = \frac{4-h^2}{h^2}$ for the sire model. In this document we express the EDC or ERC on the animal model basis φ as Liu et al. (2017).

II.1. Estimation of effective record contribution (ERC) for reference cows:

For each animal with own performance records in the national genetic evaluation, $R_i(o)$ is estimated. Estimation of $R_i(o)$ depends on the genetic evaluation model.

a) Single trait (repeatability) model for the national genetic evaluation (where individual lactations are considered as the same trait):

$$R_i(o) = \frac{m \cdot h^2}{1+(m-1)r} \qquad (1)$$

where:

$$m = \sum_{j=1}^{l} w_{ij}$$

$$w_{ij} = x_{ij} \cdot (x_{\bullet j} - \sum_{k}^{n_{sj}} x_{kj}) / x_{\bullet j}$$

l = number of lactations for animal i

x_{ij} = weight that j^{th} lactation received in the national evaluation for the i^{th} animal

n_{s_j} = number of daughters of the sire of the i^{th} animal in the animal's contemporary group for the j^{th} lactation

$x_{\bullet j} = \sum_{k=1}^{ncg_j} x_{kj}, \quad i \in k$, i.e. the sum of x_{kj}'s over all ncg_j animals in the contemporary group for the animal's j^{th} lactation

ncg_j = number of animals in animal's j^{th} lactation contemporary group

r = repeatability of lactation estimate considered in the national evaluation

h^2 = heritability estimate considered in the national genetic evaluation

b) Multiple trait model for the genetic evaluation (where each lactation, part of lactation or test day observation are treated as different traits):

Let k'EBV be the estimated breeding value or transmitting ability of the bull for the trait of interest (milk, fat or protein yield), where EBV is a vector with multiple trait (lactation, part-lactation,

test day) estimates of breeding value or transmitting ability, and k a vector with weights given to each estimate.

$$R_i(o) = \mathbf{k'G'P^{-1}Gk} / \mathbf{k'Gk} \qquad (2)$$

where:

$$\mathbf{G} = \begin{bmatrix} \sigma_a^2 & \cdots & r_{a_{1t}}\sigma_{a_1}\sigma_{a_t} \\ \vdots & \ddots & \vdots \\ r_{a_{t1}}\sigma_{a_t}\sigma_{a_1} & \cdots & \sigma_{a_t}^2 \end{bmatrix}$$

$$\mathbf{P} = \begin{bmatrix} \sigma_1^2 & \cdots & \sigma_{1t} \\ \vdots & \ddots & \vdots \\ \sigma_{t1} & \cdots & \sigma_t^2 \end{bmatrix}$$

t = number of genetically distinct traits (lactation, part-lactation or test day observations)

$r_{a_{jj'}}$ = genetic correlation estimates between genetically distinct traits j and j' considered in the national genetic evaluation

σ_{a_j} = genetic standard deviation estimate for genetically distinct trait j considered in the national genetic evaluation

Diagonal elements of **P** are:

$$\sigma_j^2 = \frac{1+(m_j-1)r_j}{m_j}\sigma_{p_j}^2$$

where:

$\sigma_{p_j}^2$ = phenotypic variance estimate for genetically distinct trait j considered in the national genetic evaluation

m_j = defined as m above for the single trait national genetic evaluation model, computed for genetically distinct trait j

r_j = repeatability of observations on genetically distinct trait j considered in the national genetic evaluation

Off-diagonal elements of **P** are:

$$\sigma_{jj'} = r_{a_{jj'}} \sigma_{a_j} \sigma_{a_{j'}} + r_{e_{jj'}} \sigma_{e_j} \sigma_{e_{j'}}$$

where:

$r_{e_{jj'}}$ = non-genetic correlation between genetically distinct traits *j* and *j'* considered in the national genetic evaluation

σ_{e_j} = **non-genetic standard deviation for genetically distinct trait *j* considered in the national genetic evaluation**

- for missing traits, computation of the nominator in (2) can be done in two ways: 1) set the rows and columns corresponding to the missing trait to zero in the P matrix, or 2) remove the corresponding rows and columns in the P matrix and the corresponding rows in the G matrix. Either way should give the same results. However, the first method is recommended since it is less ambiguous (i.e. the G matrix is not affected and the P matrix is of same dimension under all circumstances). The denominator in (2) is the same for all animals, whether observations are missing or not.

Please note that Formula [1] and [2] are identical to Formula [1] and [2] described in the weighting factor procedure (Interbull, 2000).

EDC for the bull s , as sire of the cow i , is calculated as:

For a cow *i*, denote the reliability based on her own performance records as $R_i(o)$, which was calculated as described above. Her reliability $R_i(o)$ is converted to ERC using the animal-model variance ratio λ :

$$\varphi_i = \lambda R_i(o) / \left(1 - R_i(o)\right) . \qquad [3]$$

II.2. Estimation of effective daughter contribution (EDC) for reference bulls:

Once $R_i(o)$ is computed for all animals with own records in the genetic evaluation, information from the cow and her dam is combined and the cow contrbution to her sire, bull s, EDC is as follows:

$$\varphi_i^d = EDC_i(o+d) = \lambda R_i(o) / \left(4 - R_i(o).\left[1 + R_{dam}(o)\right]\right) \qquad [4]$$

Where :

$R_{dam}(o)$ is the reliablitity contributed by own records of dam of the reference cow.

EDC for the bull *s* , as sire of the cow *i* , is computed as:

$$\varphi_s = \sum_{i=1}^{n_s} \varphi_i^d \qquad [5]$$

where n_s is the number of daughters with performance records for the bull *s*.

We assume that this bull *s* and some of his daughters are all included in a mixed genomic reference population. Among all the n_s daughters of the bull *s*, n_g daughters are genotyped and qualified for being included in the genomic reference population, $n_n = n_s - n_g$ daughters have either no genotypes or are excluded from the genome reference population. To avoid double counting the

contribution of the n_g reference daughters' performance records, let the bull *s* represent only the n_n non-reference daughters, his EDC is modified as:

$$\varphi_s^{mod} = \sum_{i=1}^{n_n} \varphi_i^d = \varphi_s - \sum_{i=1}^{n_g} \varphi_i^d \qquad [6]$$

where φ_s^{mod} is the modified EDC for the reference bull *s* with n_g reference daughters.

For the reference cow *i*, φ_i from Formula [3] is used as ERC for the genomic reliability calculation. For the reference bull *s*, his EDC is calculated using Formula [6], representing the contrition by his n_n non-reference daughters. ERC of reference cows or the modified EDC of reference bulls are described as n_e in Equation [2] of the genomic reliability method (Liu et al. 2017).

If all his n_s daughters of the example reference bull *s* are all included in the mixed genomic reference population, this bull *s* would no longer contribute any information to the estimation of SNP effects.

If the reference bull *s* has daughters also in foreign countries and his MACE evaluation is used in a single-step or multi-step genomic evaluation, Formula [6] can be extended to calculate a modified EDC for the example reference bull under the assumptions that his foreign daughters are not included in the mixed reference population:

$$\varphi_s^{mod} = \varphi_s^{MACE} - \sum_{i=1}^{n_g} \varphi_i^d \qquad [7]$$

where φ_s^{MACE} represents his EDC contributed by his domestic daughters and foreign daughters worldwide. The EDC φ_s^{MACE} can be calculated using his EDC values from all countries, country heritability values and genetic correlations between countries. Note that $\varphi_s^{MACE} \geq \varphi_s$.

References

Interbull, 2000. New weighting factors for the international evaluation, revised July 2000. https://wiki.interbull.org/public/CoP_AppendixIV?action=SlideShow

Liu, Z., P. M. VanRaden, M.H. Lidauer, M. P. Calus, H. Benhajali, H. Jorjani, and V. Ducrocq. 2017. Approximating genomic reliabilities for national genomic evaluation. Interbull Bulletin: 51: 75-85.