Guidelines for Approximating Genomic Reliabilities of the Single-Step Genomic Model

Z. Liu, I. Strandén, J. Vandenplas, H. Eding, M. Lidauer, K. Haugaard, and P. M. VanRaden

Interbull Working Group on Genomic Reliability Calculation
Reliability approximation for conventional and genomic EBV

Conventional reliability calculation for diverse genetic models
- A single trait repeatability model (VanRaden and Wiggans, 1991)
- Proven to be fairly accurate and highly efficient for large populations

Genomic reliability calculation methods
- For a multi-step genomic model (Liu et al. 2010; Wiggans and VanRaden 2010)
- For a single-step genomic model (Misztal et al. 2013)

Interbull working group on genomic reliability calculation
- Goal: Make national genomic reliabilities **comparable** across countries
- A standardized genomic reliability method (Liu et al. 2017)
- Applicable for the multi-step and single-step models
- Large-scale female animal genotyping just started in some countries
Main features of the Interbull genomic reliability method (Liu et al. 2017, Interbull Bulletin 51)

- Genotype data treated as an additional source of information contributing to total reliability

- Keep using **traditional reliability methods for conventional part** of single-step model
  - a random regression test-day model for milk production and somatic cell scores
  - a maternal-effect animal model for calving traits
  - a multi-parity, multi-trait animal model for fertility traits

- Include young animals and all genotyped animals
Features of the Interbull genomic reliability method (2)

Main features of the Interbull genomic reliability method

- Consider genomic relationship among **ALL** genotyped animals
- DGV reliabilities of young candidates depend on relationship to reference animals
- No longer approximating DGV reliabilities for young candidates (Liu et al. 2010, Wiggans & VanRaden, 2010)

- By applying the SNP BLUP model via the efficient Software **snp_blup_rel** (Strandén et al., LUKE, Finland)
- Allow to run multiple ‘single’ traits in parallel
- High efficiency of the software for extremely large data set with > 1.5 million genotyped animals
- Newest version v0.99 vs current v0.88 (Ismo Strandén and Zengting Liu, April 2024)
  - DEU 25 conformation traits: 1.32 mio genotyped animals, 386,062 reference animals
  - Peak RAM reduced by 56%
  - CPU time reduced by 45%
### April 2023 evaluation

<table>
<thead>
<tr>
<th>Frequency of</th>
<th>Test-day traits</th>
<th>Conformation traits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotyped Holstein animals</td>
<td>1,318,780 (1,138,039 females and 180,741 males)</td>
<td>25 traits in 3 sub-groups</td>
</tr>
<tr>
<td>Cows and bulls with phenotypes</td>
<td>13,528,444</td>
<td>3,144,366</td>
</tr>
<tr>
<td>--- Phenotypic records</td>
<td>263,673,267</td>
<td>3,144,366</td>
</tr>
<tr>
<td>Genotyped or phenotyped animals</td>
<td>14,402,662</td>
<td>4,131,336</td>
</tr>
<tr>
<td>Animals in pedigree</td>
<td>21,850,276</td>
<td>10,048,593</td>
</tr>
<tr>
<td>Reference animals (cows &amp; bulls)</td>
<td>524,187</td>
<td>386,062</td>
</tr>
</tbody>
</table>
Genomic reference populations for April 2023 evaluation

<table>
<thead>
<tr>
<th></th>
<th>Protein yield PKG</th>
<th>Stature STA</th>
<th>Locomotion LOC</th>
<th>Angularity ANG</th>
<th>Udder balance EUB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference cows</td>
<td>478,588</td>
<td>357,365</td>
<td>349,083</td>
<td>198,170</td>
<td>305,122</td>
</tr>
<tr>
<td>Reference bulls</td>
<td>45,591</td>
<td>28,635</td>
<td>27,696</td>
<td>27,748</td>
<td>27,205</td>
</tr>
<tr>
<td>Total</td>
<td>524,179</td>
<td>386,000</td>
<td>376,779</td>
<td>225,918</td>
<td>332,327</td>
</tr>
</tbody>
</table>
Theoretical DGV reliabilities for genotyped German Holstein AI bulls in April 2023 single-step evaluation

Very high theoretical DGV reliabilities for youngest AI bulls

Important to adjust genomic reliabilities based on genomic validation results
Variation in theoretical DGV reliabilities for genotyped DEU Holstein AI bulls in April 2023 single-step evaluation

Small variation of theoretical DGV reliabilities within youngest birth years

1. Less critical, if variability in individual DGV reliabilities is ignored
2. A constant genomic EDC gain may give a reasonable approximation
3. Avoid the time-consuming part of theoretical DGV reliability calculation in routine evaluation
GREL method optimization, modification and changes since 2017 (I)

- Variation in genomic reliabilities among young animals becomes smaller
  - Due to a high number of genotyped animals and more complete ancestry in reference population
  - Level of genomic reliabilities for young animals more important to ascertain
  - A constant of genomic EDC gain by genotype data to be determined

- Separation of the GREL steps between routine single-step evaluation and genomic validation
  - Too long computation for DGV reliabilities due to millions of genotyped animals
    - Large-scale female genotyping just started back in 2016/2017
    - Even for the highly efficient software `snp_blup_rel` (Luke, Finland)

- Two separate Guidelines for routine genomic evaluation and deriving genomic EDC gain
A SNP BLUP model without a residual polygenic effect for calculating exact reliabilities of DGV
- A posterior consideration of residual polygenic variance: $\text{REL}_{\text{DGV}} = (1-k) \text{REL}_{\text{SNP}}$
  - GEBV reliability as a weighted function of $\text{REL}_{\text{DGV}}$ and conventional reliability $A22$
  - Too many reference cows with negative genomic EDC gain particularly for traits with low heritability
- A SNP BLUP model with RPG requires more computing time than available (Ben Zaabza, et al. 2020)

Reducing SNP markers by selecting equidistant markers for faster calculation of DGV reliabilities
- At least 15,000 SNP markers were shown to be needed
- Similar effect may be achieved by the adjustment of genomic reliability

Propagated genomic reliabilities for non-genotyped relatives of genotyped animals following the concept of genotype confidence (Eding, 2022)
- Instead of using a fixed value of upper limit

The new GEBV test software 2024A (Sullivan, 08.05.2024) provides for validation bulls:
- $\text{var} (\hat{\mu}_E - \hat{\mu}_L)$
Two Guidelines for single-step genomic reliability calculation and adjustment

Guideline for GREL Calculation

Guideline for GREL Adjustment

Approved by Interbull Steering Committee
When a genomic EDC constant needs to be updated / GREL to be adjusted

- A constant value for genomic EDC gain is used for all genotyped animals
  - Properly determine the level of genomic reliability especially for young candidates
  - Ignores differences in DGV reliabilities among genotyped animals
  - Small difference for large reference population
  - For smaller populations like Jersey or a new trait
    - Low number of validation bulls
    - Special rules may be needed

- Updating the genomic EDC gain parameter whenever a GEBV test is required
  - Implementation of a new model
  - Introduction of major model changes
  - Routine genomic validation every 2 years

- Apply the same rules of data truncation of the GEBV test for the updating
Too high genomic reliabilities in case of inflation $b_1 < 1$?

- **GEBV test based on validation bulls**
  \[
  \hat{u}_L = b_0 + b_1 \hat{u}_E + \varepsilon
  \]

- **Regression and correlation coefficients**
  
  \[
  b_1 = r \frac{\sqrt{\text{var}(\hat{u}_L)}}{\sqrt{\text{var}(\hat{u}_E)}} \implies \text{var}(\hat{u}_L) = \frac{b_1^2}{r^2} \text{var}(\hat{u}_E)
  \]

- **GREL adjustment according to GEBV from the full/later and truncated/early evaluations**
  
  \[
  \text{var}(\hat{u}_E - \hat{u}_L) = \text{var}(\hat{u}_L) - (2b_1 - 1)\text{var}(\hat{u}_E) \text{ versus } \text{var}(\hat{u}_L) - \text{var}(\hat{u}_E)
  \]

- $b_1 = 1$ suggesting no under- or overestimation, \[\text{var}(\hat{u}_E - \hat{u}_L) = \text{var}(\hat{u}_L) - \text{var}(\hat{u}_E)\]

- $b_1 < 1$ suggesting $\text{var}(\hat{u}_E)$ too high, \[\text{var}(\hat{u}_E - \hat{u}_L) > \text{var}(\hat{u}_L) - \text{var}(\hat{u}_E), \text{ expected average reliability of early evaluation } E(\Re_E) \text{ would be lower than the case of } b_1 = 1\]

- $b_1 > 1$ suggesting $\text{var}(\hat{u}_E)$ too low, \[\text{var}(\hat{u}_E - \hat{u}_L) < \text{var}(\hat{u}_L) - \text{var}(\hat{u}_E), \text{ expected average reliability of early evaluation } E(\Re_E) \text{ would be higher than the case of } b_1 = 1\]

- No, the adjusted genomic reliabilities will NOT be too high, when $b_1 < 1$
Implementation issues to consider

- Applicability for small genotyped / reference populations
  - For a new trait, like dry matter intake
  - For a small breed, like Jersey
  - Next project for the WG

- Same GREL adjustment for all sub-trait of a trait group?
  - Sub-traits have similar data and heritability values

- A multi-breed genomic evaluation system
  - e.g. Holstein + Jersey
  - Breed-specific genomic reliability adjustment

- Interbull Mendelian Sampling variance test for the single-step model
  - Use genomic reliabilities instead of conventional reliabilities
Future research topics

- Extension to a SNP BLUP model with a residual polygenic effect (Ben Zaabza, et al. 2020; 2021)
  - A Monte Carlo sampling based approach
  - To further improve the computational efficiency

- Multi-trait models for modelling genomic information at ALL the steps
  - Currently, a single trait model is assumed for the genomic information

- Fewer steps for single-step genomic reliability approximation
  - Conventional reliabilities using a multi-trait or single-trait model
  - Divide the whole population into genotyped and non-genotyped sub-populations
    - Quantify ‘added value’ of genotyping based on reference population
  - Propagate the genomic information gain to the non-genotyped relatives
  - Adjust the theoretical genomic reliability levels using a truncated evaluation

- Merge the steps of conventional reliability calculation with those of genomic reliability
  - Explore the structure of LHS of the MME of the single-step model
Summary and Conclusions (I)

- The two GREL Guidelines for routine evaluation and validation
  - Developed by WG and approved by the Steering Committee
  - Separate Guidelines for routine evaluation and for deriving GREL adjustment via GEBV test
  - Feasible for large populations with millions of genotyped animals in routine evaluation
  - Ensure a realistic level of genomic reliabilities of young candidates

- DEU implemented the Guidelines in all 10 trait groups
  - Using data from a full evaluation April 2023 and a truncated data set

- Update genomic EDC gain parameter whenever a GEBV test is required
  - Adjusting genomic reliabilities is linked to the GEBV test

- Implementation issues like for small populations to be considered

- Future R&D projects for further optimization of the Interbull GREL method
Summary and Conclusions (II)

- All member countries are encouraged to apply the Interbull GREL method

- The Interbull GREL WG can provide technical support

- We are one step closer to our goal:
  - To make genomic reliabilities comparable across countries
  - To make genomic reliabilities comparable between traits within country
  - To make genomic reliabilities comparable between candidates and bulls with many daughters
Thank you for your attention!