Approximating genomic reliabilities for national genomic evaluation

The Working Group Genomic Reliability Calculation

Z. Liu, P. M. VanRaden, M. Lidauer, M. P. Calus, H. Benhajali, H. Jorjani, V. Ducrocq
Overview

- Current status of genomic reliabilities
- New solutions to GREL calculation
- A standardized method
- Implementation issues
- Future steps
- Usage of the software `snp_blup_rel`
Introduction

• Interbull introduced standardized procedures for calculating conventional EDC (2001)
  • Though the total reliabilities of EBV not fully harmonized
• Genomic reliabilities less comparable across countries
  • Lack of standard calculation procedure
  • Differences in GREL methods between countries
• GREL must be consistent with conventional REL
  • Between conventional and genomic evaluations
  • Animals in different life times: candidates, getting own phenotypes, entering reference population
Previous activities of GREL WG

- Interbull GREL Working Group established (2014)
- Two reports presented by Bevin Harris
  - Workshop in Verden, Feb 2015
  - Annual meeting in Orlando, July 2015
- Investigation on validation $R^2$ value and genomic reliability via simulation (M. Calus & B. Harris)
  - Conclusions: they are two different measures of accuracy of genomic prediction
  - As the validation $R^2$ increases, the difference between $R^2$ and genomic reliabilities reduces
New mission of GREL WG

• To develop standard procedures for approximating GREL for national genomic evaluation
  • Comparable GREL between countries
  • Consistent with conventional reliabilities
• Desired features of the standardized procedure
  • Account for residual polygenic effect
  • Feasible for any number of genotyped animals
  • Applicable to single-step genomic models
  • Efficient for frequent genomic evaluation
  • Consistent with the genomic validation $R^2$
Currently used GREL methods

• For multi-step genomic models
  • Harris and Johnson, 2010, JDS
  • Lidauer et al. 2016, GREL WG & EuroGenetics meetings
  • VanRaden et al. 2011, GSE
  • and other GREL methods

• For single-step genomic models
  • Misztal et al. 2013, JDS
  • Taskinen et al. 2013, Interbull Bulletin
Bottleneck of GREL calculation: inversion of large genomic relationship matrix \( G \)

- APY algorithm (Misztal et al. 2015)
- Calculating exact reliabilities of DGV for genotyped animals via `snp_blup_rel` (Mäntysaari & Strandén 2016)
  - Invert matrices using very efficient BLAS subroutines by parallel computing on multiple cores
  - No residual polygenic effect in the SNP BLUP model
  - Only \# SNP matters, NOT \# reference/genotyped animals
GREL WG activities

- GREL WG video conferences (in addition to emails)
  - 07 October 2016
    - 05 Oct. 2016 with Bevin for transition
  - 27 March 2017
  - 12 June 2017
- Adjusting theoretical genomic reliabilities using data from genomic validation (VanRaden, 2017)
  - GREL changes correspond to GEBV changes
  - Use GEBV Test data as candidates and AI bulls with daughters
• Information sources for conventional evaluation
  • Own data, progeny and parental contributions
• Information source method or EDC or daughter equivalent methods used for REL calculation
• **Genomic contribution** (single-step genomic BLUP model)

\[
H^{-1} = \begin{bmatrix}
A^{11} & A^{12} \\
A^{21} & G^{-1} + A^{22} - A^{-1}_{22}
\end{bmatrix} = A^{-1} + \begin{bmatrix}
0 & 0 \\
0 & G^{-1} - A^{-1}_{22}
\end{bmatrix}
\]
Calculating genomic contribution

- Reliability values of DGV for all genotyped animals
  - Using software `snp_blup_rel`
- For all genotyped animals, equivalent to
  \[
  \begin{bmatrix}
  \mathbf{Z}' \mathbf{R}^{-1} \mathbf{Z} & 0 \\
  0 & 0
  \end{bmatrix}
  + \sigma_g^{-2} \mathbf{G}^{-1}
  \]
  only reference animals provide phenotype data

- Conventional reliabilities for the genotyped animals
  \[
  \begin{bmatrix}
  \mathbf{Z}' \mathbf{R}^{-1} \mathbf{Z} & 0 \\
  0 & 0
  \end{bmatrix}
  + \sigma_g^{-2} \mathbf{A}^{-1}_{22}
  \]
  without inverting \( \mathbf{A}_{22} \)

- Pure genomic EDC gain:
  \[
  \varphi = \lambda \frac{\mathbf{R}_{DGV}}{1 - \mathbf{R}_{DGV}} - \lambda \frac{\mathbf{R}}{1 - \mathbf{R}_{A_{22}}}
  \]
Steps of the new GREL method (I)

1. Reliabilities of SNP markers, \( \text{REL}_{\text{SNP}} \), via \text{snp\_blup\_rel}
   - Assumption: SNP markers explain all genetic variation

2. Reliabilities of direct genomic values (DGV)
   - Proportion of residual polygenic variance \((k)\)
   - Accuracy of imputation \((r_{\text{IMP}})\): preferably allele dosage
     \[
     \text{REL}_{\text{DGV}} = (1 - k) \times r_{\text{IMP}}^2 \times \text{REL}_{\text{SNP}}
     \]
   - For reference animals
     \[
     \text{REL}_{\text{DGV}} = r_{\text{IMP}}^2 \times \text{REL}_{\text{SNP}}
     \]
   - EDC of DGV for a genotyped animal:
     \[
     \text{EDC}_{\text{DGV}} = \lambda_4 \times \text{REL}_{\text{DGV}} / (1 - \text{REL}_{\text{DGV}})
     \]
     \[
     \text{where } \lambda_4 = (4 - h^2) / h^2
     \]
3. Adjusting to realized reliabilities of DGV

\[ \text{EDC}_{\text{real}}^{\text{DGV}} = s \times \text{EDC}_{\text{DGV}} \]

- A constant EDC adjustment factor determined by realized GEBV variations via GEBV Test

4. Genomic EDC gain \((G-A_{22})\) for each genotyped animal

- Calculate reliabilities \(\text{REL}_{A_{22}}\) as in conventional evaluation
- Only reference animals provide phenotypes

\[ \text{EDC}_{A_{22}} = \hat{\lambda}_4 \text{REL}_{A_{22}} / (1 - \text{REL}_{A_{22}}) \]

- Genomic EDC gain for a genotyped animal

\[ \text{EDC}_{\text{gain}} = \text{EDC}_{\text{real}}^{\text{DGV}} - \text{EDC}_{A_{22}} \]

\[ \text{EDC}_{\text{gain}} = 0, \text{ if } \text{EDC}_{\text{gain}} < 0 \]
5. (Optional) Propagation to non-genotyped relatives

- Involving potentially tens of millions of animals
- $EDC_{\text{gain}}$ of only reference animals as data for propagation
- In 2 directions of pedigree for progeny & parental contributions

\[ EDC_{T\text{gain}} = \lambda_4 \frac{REL_{\text{propg}}}{(1 - REL_{\text{propg}})} \]

- As propagation does not account for LD break-down

\[ EDC_{T\text{gain}} \leq \max(EDC_{\text{gain}} \text{ of candidates}) \]

- For all genotyped animals set:

\[ EDC_{T\text{gain}} = EDC_{\text{gain}} \text{ from Step 4} \]
6. Final reliabilities enhanced with genomic information
   - Total conventional reliability by phenotype data and pedigree
   - Calculated from a single-step model or a conventional model
     \[ EDC_{\text{CONV}} = \lambda_4 \cdot \text{REL}_{\text{CONV}} / (1 - \text{REL}_{\text{CONV}}) \]
   - Final EDC of the animal
     \[ EDC_{\text{final}} = EDC_{\text{CONV}} + EDC_{\text{Tgain}} \]
   - Final reliability enhanced with genomic information
     \[ \text{GREL}_{\text{final}} = EDC_{\text{final}} / (EDC_{\text{final}} + \lambda_4) \]
• GEBV differences btw 2 evaluations (VanRaden, 2017)
• Use validated data from Interbull’s GEBV Test
• Calculate using the standardized method
  • GRELearly for an early, truncated evaluation
  • GRELlater for a later, complete evaluation
• Expected change in genomic reliabilities (a constant)

\[ \exp(\text{GREL}_{\text{chng}}) = \frac{\text{Var}(\text{GEBV}_{\text{later}} - \text{GEBV}_{\text{early}})}{\text{Var}(\text{BV})} \]

• Expected average reliability in the early evaluation

\[ \exp(\text{GREL}_{\text{early}}) = \text{avg}(\text{GREL}_{\text{later}}) - \exp(\text{GREL}_{\text{chng}}) \]
Adjusting genomic reliabilities

• Convert genomic reliabilities of early evaluation to EDC
  
  \[
  \text{avg}(\text{EDC}_{\text{early}}) = \lambda_4 \frac{\text{avg}(\text{GREL}_{\text{early}})}{1 - \text{GREL}_{\text{early}}}
  \]
  
  \[
  \exp(\text{EDC}_{\text{early}}) = \lambda_4 \frac{\exp(\text{GREL}_{\text{early}})}{1 - \exp(\text{GREL}_{\text{early}})}
  \]

• Calculate adjustment factor in genomic EDC
  
  \[
  f = \frac{\exp(\text{EDC}_{\text{early}})}{\text{avg}(\text{EDC}_{\text{early}})}
  \]

• \( f < 1 \) (\( \geq 1 \)) indicates over- (under) estimated GREL

• Applicable to any two evaluations, as long as GEBV are validated via GEBV Test
• Allele frequencies of SNP markers
  • Estimates of base population (Gengler 2007)
  • 0.5 for all SNP markers
    • Too low $\text{REL}_{\text{SNP}}$ for some reference bulls with extreme diagonals of $\mathbf{G}$ matrix (not blended with $\mathbf{A}_{22}$)
  • Frequencies of current population
    • Reference animals or all genotyped animals
• **Recommendation:** use allele frequencies of the current population of ALL genotyped animals
• Conventional reliability $\text{REL}_{\text{A22}}$ for genotyped animals
  • Data from reference pop., progeny and parental contributions
Implementation issues (II)

- Frequencies of calculation of $REL_{SNP}$
  - $REL_{SNP}$ most time-consuming
  - MACE/national evaluation $\rightarrow$ invert LHS ($snp\_blup\_rel$) & $REL_{SNP}$ calculation for all genotyped animals
- Monthly / weekly genomic evaluation $\rightarrow$ only for new candidates
- Simplification for just-in-time continuous genomic evaluation (Alkhoder et al. 2014)
- Frequencies of updating GREL adjustment factor
  - Same as GEBV Test
Test application to German Holsteins

- Genotype & phenotype data from May 2017 evaluation
  - 35,533 EuroGenomics Holstein reference bulls
  - 314,608 genotyped animals & 45,613 SNP markers

- Computing resources used for running snp_blup_rel

  - Step 1: inverting MME using reference animals
    - Total clock time c.a. 60 minutes on 10 cores
    - Peak RAM c.a. 38 Gb

  - Step 2: calculating REL_{SNP} for all genotyped animals
    - Total clock time c.a. 82 minutes on 10 cores
    - Peak RAM c.a. 121 Gb (RAM intensive option)
• Phenotypes from April 2017 MACE evaluation
• Genotypes from Apr 17 DEU HOL genomic evaluation
• 35,533 EuroGenomics reference bulls
  • 31,428 Holstein bulls born before 2010
  • 894 DEU bulls born in 2010 to 2012 as validation bulls
• Interbull GEBV Test for all traits
  • GEBVearly from the truncated and GEBVlater from the full evaluations are validated
• GREL calculation for the two evaluations: GRELearly for the truncated and GRELlater for the full evaluation
Further development

- Second- or third-generation candidates
  - Use validated GREL of 1st generation from a later eval.
- Shrinkage factor (regression coefficient $b_1$) of DGV
- Multi-trait genomic models
  - Same as for conventional evaluation
Verification and Validation

• Accuracy of the new GREL method
  • Reliability calculation by matrix inversion
• Comparison to the other GREL methods
• True reliabilities from the previous simulation studies
• YOU ARE INVITED TO DO THE VALIDATION AND COMPARISON!
Next steps

- Countries to test the `snp_blup_rel` software
- Countries to test the new GREL method
- Country feedback for fine-tuning & further development
  - `snp_blup_rel` and the new GREL method
- Official implementation by all NGECs
Summary

- `snp_blup_rel` an efficient tool for $\text{REL}_\text{SNP}$ calculation in an unified way across countries

- Limiting factor for $\text{REL}_\text{SNP}$ calculation is # of SNPs, no longer # reference/genotyped animals

- The GREL method makes GREL comparable across countries & consistent with conventional REL

- The adjustment to realized reliability ensures GREL changes corresponding to GEBV changes

- The new GREL method is efficient and feasible for any number of genotyped animals

- Verification and validation are needed
Use of the software `snp_blup_rel`

- Developed and kindly provided by LUKE, Finland
- NGECs with national genomic evaluation received a copy of software `snp_blup_rel` on 29.03.2017
- Ensures that all countries calculate DGV reliabilities in the same way
- Verified to give equal results with own programs
- `snp_blup_rel` is very efficient with many options
- NGECs must not use it for other purposes than just the DGV reliability calculation!
- NGECs must not distribute it to any other institutions!
Acknowledgements

- Bevin Harris, former convener of the GREL WG
- Interbull Centre
- Esa Mäntysaari and Ismo Strandén, LUKE, Finland