Genomic reliabilities
Testing the protocol proposed by Interbull

H. Eding, J. Vandenplas, G. de Jong

Dubrovnik, August 26, 2018
OVERVIEW

• Introduction
• Six step procedure
• Test data and parameters used
• Some results
• Discussion
• Interbull started a working group Genomic Reliabilities

• Aim: Find a procedure to estimate animal genomics reliabilities
  – For multi- and single-step procedures
  – For genotyped and non-genotyped animals
  – Accounts for residual polygenic effects
  – Consistent results across countries (unified approach)
  – Feasible for large numbers if genotyped animals
  – Usable in routine genomic evaluations
INTRODUCTION

• AEU started development of a (D)PCG to solve single step models
  – Single step GBLUP
  – Single step SNP BLUP
  – Aimed for use in the national (genomic) evaluation
    • Replaces our current pseudo-trait methods

• In parallel new software to estimate GEBV reliabilities is needed
  – Incorporates traditional sources of information (parents, offspring, own)
  – Incorporates information from genotypes
  – Accounts for propagation (effect of genomic information on offspring, parents)
  – Proposal Interbull working group chosen as approach.
INTRODUCTION

• Project: Develop software/workflow
  – Implementing the Interbull protocol
  – Within frame work of new (D)PCG to solve single step SNP BLUP models
  – Produce correct GEBV reliabilities for genotyped and non-genotyped animals

• First test on existing flow: Milking speed and Temperament
  – To get a feel for the requirements
  – Compare results with current GEBV reliability estimates
SIX STEPS

1. Calculate SNP reliabilities
2. Derive DGV reliabilities
3. Adjust DGV reliabilities
4. Calculate genomic gain of reliabilities
5. Propagation (optional; non-gentyped animals)
6. Calculate final reliabilities
GETTING PARAMETERS

• Rate of imputation
  – ‘accuracy of genotype imputation’
  – All bulls on same chip, so $r_{\text{imp}} = 0.985$ (empirical mean)

• Theoretical to realized reliability factor $f$
  – Set to 1.0 for first testing
  – Unclear how to derive correct value

• Proportion of residual polygenic variance $k$
  – Expected values $\sim [0.05; 0.25]$
GETTING PARAMETERS: K

- Testing on cow reference

\[ \sigma_{SNP_{tot}}^2 = \left( \left( \frac{N - \sum_j pp_j}{100} \right) + \sum_j pp_j \right) \sigma_{SNP}^2 \]

\[ k = \frac{\sigma_{poly}^2}{(\sigma_{SNP_{tot}}^2 + \sigma_{poly}^2)} \]

- Variance components from current genomic selection flows
FIRST TEST: TEMPERAMENT AND MILKING SPEED

• Animal ID’s
  – Pedigree       : 6,300,989 animals
  – Observations  : 4,021,612

• Genotypes
  – Most recent run (2018 05 01)
  – Number of genotypes : 115,378
  – Number of SNP       : 37,995
  – Converted to [012]-format
TEST DATA USED: TEMPERAMENT AND MILKING SPEED

- During genomic selection validation
  - Added EDC ($\Delta_{EDC}$) are calculated from DGV and BLUP results
  - Based on differences in reliability EBV and GEBV for young bulls (no daughters)
  - Assumed constant (single genotype in single animal)

- Overview traits:

<table>
<thead>
<tr>
<th>Trait</th>
<th>$h^2$</th>
<th>$k$</th>
<th>$EDC_{add}$ (val 2018)</th>
<th>Mean rel. (GEBV YB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milking Speed</td>
<td>0.230</td>
<td>0.084</td>
<td>24.1</td>
<td>0.71</td>
</tr>
<tr>
<td>Temperament</td>
<td>0.114</td>
<td>0.208</td>
<td>5.1</td>
<td>0.38</td>
</tr>
</tbody>
</table>
**RESULTS: RUNTIME**

<table>
<thead>
<tr>
<th>Step</th>
<th>Action</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Conventional rels</td>
<td>3m10</td>
</tr>
<tr>
<td>2</td>
<td>Read genotypes</td>
<td>1m20</td>
</tr>
<tr>
<td>3</td>
<td>Convert to [012]</td>
<td>3m32</td>
</tr>
<tr>
<td>4</td>
<td>Run luke software</td>
<td>53m20</td>
</tr>
<tr>
<td>5</td>
<td>Apply Liu protocol</td>
<td>2m00</td>
</tr>
</tbody>
</table>

- Peak memory usage: 52.4 Gb
  - Number of genotypes: 115,378
  - Number of SNP: 37,995
RESULTS: BULLS GEBV VS EBV MILKING SPEED
RESULTS: BULLS GEBV VS EBV MILKING SPEED

Young genotyped bulls (no daughters)

- Mean GEBV reliability: 0.81
- Mean added EDC: 71.0
RESULTS: BULLS GEBV VS EBV TEMPERAMENT
RESULTS: BULLS GEBV VS EBV TEMPERAMENT

- Young genotyped bulls (no daughters)
  - Mean GEBV reliability: 0.72
  - Mean added EDC: 70.9
## Test Data Used: Temperament and Milking Speed

- **Overview traits:**

<table>
<thead>
<tr>
<th>Trait</th>
<th>$\text{EDC}_{\text{add}}$ (val 2018)</th>
<th>Mean rel. (GEBV YB aug '18)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milking speed</td>
<td>24.1</td>
<td>0.71</td>
</tr>
<tr>
<td>Temperament</td>
<td>5.1</td>
<td>0.38</td>
</tr>
</tbody>
</table>
TEST DATA USED: TEMPERAMENT AND MILKING SPEED

• Overview traits:

<table>
<thead>
<tr>
<th>Trait</th>
<th>$EDC_{\text{add}}$ (val 2018)</th>
<th>$EDC_{\text{add}}$ (ITB YB)</th>
<th>Mean rel. (GEBV YB aug ‘18)</th>
<th>Mean rel. (ITB YB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milking speed</td>
<td>24.1</td>
<td>71.0</td>
<td>0.71</td>
<td>0.81</td>
</tr>
<tr>
<td>Temperament</td>
<td>5.1</td>
<td>70.9</td>
<td>0.38</td>
<td>0.72</td>
</tr>
</tbody>
</table>

• Conclusion: Correction/scaling seems unavoidable
DISCUSSION: RUNTIME

• Getting SNP reliabilities was most costly in terms of run time
  – ~ 55 minutes per trait
• Most evaluations are multiple trait
  – Example: Fertility evaluation has 40 traits ~ 35h (at 120,000 genotypes of 37,995 SNP)
  – Conventional rels fertility ~ 20h
  – Solution: run traits in parallel

• Applying protocol to obtain GEBV rels requires little time
  – Includes propagation in 6.2 mln non-genotype animals

• Results suggest the protocol is fit for use in routine evaluation.
DISCUSSION: RESULTS

• Propagation results in increased Grel vs conventional reliability
  – VanRaden&Wiggans algorithm gives satisfactory results

• Grel seems to overestimate reliability of GEBV
  – Possible cause 1: No deregression (yet) of conventional EDC
  – Possible cause 2: No scaling with ‘realized’ reliability

• Deregression not expected to reduce overestimation much

• Conclusion: Scaling of REL_{SNP} or EDC_{SNP} is necessary
  – Most important factor in estimation procedure
  – Possibly use ‘added EDC’ statistic from genomic validation
DISCUSSION: SCALING

• Genomic validation produces two main statistics:
  – $R^2_{DGV}$: Mean genomic reliability of validation bulls (no daughters)
  – $R^2_{BLUP}$: Mean conventional reliability of validation bulls

• Additional statistic produced: mean added EDC
  – $\Delta_{EDC} = EDC(R^2_{DGV}) - EDC(R^2_{BLUP})$
  – Estimate of EDC’s added to information by single genotype of bull without genotyped relatives

• Alternative scaling (will be tested):
  – $REL_{SNP} \Rightarrow EDC_{SNP}$
  – Get $EDC_{SNP}$ for bulls in validation
  – $f = \Delta_{EDC} / \text{mean}(EDC_{SNP, \text{val}})$
  – $EDC_{dgvi} = EDC_{snpi} \times f \times (1 - r_{imp})$
    • Residual additive genetic variance already accounted for.
DISCUSSION: CONCERNS

• Interbull protocol relies on a number of ‘outside’ parameters
  – Proportion of residual additive genetic variance $k$
  – Ratio of realized versus estimated reliability $f$
• Possibly these can be derived from validation results
• But: many more traits in routine evaluation than in genomic evaluation
  – E.g. lactation specific traits, predictor traits
• Values for $k$ and $f$ not readily available for all
  – Especially $f$ seems pivotal to correct estimation of $grel$. 
• Ideal: An approach that estimates correct $\text{REL}_{\text{SNP}}$ or $\text{EDC}_{\text{SNP}}$
  – Would make possible a self-contained reliability estimation
FINAL REMARKS

- Reliability calculations using Interbull protocol seems feasible for routine use.
- Given parameters used we get reasonable estimates
- Correct estimates are contingent on correct $f$ value