Working Group on Genomic Reliability

February 2015

Mario Calus, Paul VanRaden, Martin Lidauer, Vincent Ducrocq, Zengting Liu and Bevin Harris
Background

• Genomic reliability is a measure of "precision" of the genomic breeding value

  • Measure of accuracy of prediction of young sires without daughter information

• It is reasonable to expect the validation $R^2$ and average genomic reliability of a cohort of young sires to be of a similar magnitude
Background

• Genomic reliability should be proportional to the expected change in the genomic breeding value as new information becomes available.

• Example: 99% interval, genetic standard deviation = 10
ITB Berlin 2014

- Need to undertake a simulation model
- Naive method had been developed
  - Undergone limited testing on national data sets
  - Showed some promise but difficult to assess without simulated data
Simulation

• Mimicking a dairy cattle population using QMSim (Sargolzaei and Schenkel, 2009)

• Recent generations:
  
  • 45,000 females and 5,000 males per generation
  
  • 1,000 males selected for mating, based on 64% reliable EBV or at random
  
  • Heritability = 0.3; Bull reliability is ~0.8 (with 45 daughters)
  
  • ~9,000 QTL with effects from normal distribution & ~42,000 SNPs
Simulation Data for analysis

- 2,000 training bulls from 2 generations
- 6,000 validation bulls from the next 3 generations
  - TBV, genotypes and pedigree are provided for all 8,000 bulls and DYDs on 2,000 training bulls
  - Heritability = 0.3; Bull reliability is ~0.8 (with 45 daughters)
Analysis: simulation data

- Very Preliminary Results: no replication

- Undertaken to ensure the simulated data behaves as expected

- Comparing traditional BLUP results with methods for genomic evaluation
Analysis: simulation data

- The data was checked for any big differences between G and A.

- For 8000 by 8000 bulls, only 1 pair of bulls (806434 and 880612) had a difference bigger 0.2 (related by 0.34 in G vs 0.13 in A)

- Genomic selection did not distort the genomic vs. pedigree relationships and shows that QMSim worked properly

- This is consistent with given the allele frequencies changed little, which is reasonable given many QTLs with small effects
Simulation

- No selection dataset
- 2000 training sires BLUP EBVs versus GEBVs using GBLUP

\[ \text{GEBV} = -0.02 + 0.98 \text{EBV} \quad R^2 = 0.95 \]
Simulation

- No selection dataset
- 3 Generations of test sires (1000 sires per generation)
- $R^2$ between TBV and both the BLUP EBVs and GEBVs

<table>
<thead>
<tr>
<th>Generation</th>
<th>EBV</th>
<th>GEBV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.19</td>
<td>0.51</td>
</tr>
<tr>
<td>2</td>
<td>0.07</td>
<td>0.43</td>
</tr>
<tr>
<td>3</td>
<td>0.04</td>
<td>0.40</td>
</tr>
</tbody>
</table>
Simulation

- No selection dataset
- Generation 1 of test sires
- Comparison of direct reliabilities from BLUP EBVs and GEBVs

Mean = 0.23

Mean = 0.58
Simulation

- No selection dataset
- Generation 3 of test sires
- Comparison of direct reliabilities from BLUP EBVs and GEBVs

Mean = 0.06

Mean = 0.50
Simulation

- No selection dataset
- 3 Generations of test sires (1000 sires per generation)
- $R^2$ between TBV and both the BLUP EBVs and GEBVs

<table>
<thead>
<tr>
<th>Generation</th>
<th>$R^2$ EBV</th>
<th>Mean Rel EBV</th>
<th>$R^2$ GEBV</th>
<th>Mean Rel GEBV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.19</td>
<td>0.23</td>
<td>0.51</td>
<td>0.58</td>
</tr>
<tr>
<td>2</td>
<td>0.07</td>
<td>0.11</td>
<td>0.43</td>
<td>0.52</td>
</tr>
<tr>
<td>3</td>
<td>0.04</td>
<td>0.06</td>
<td>0.40</td>
<td>0.50</td>
</tr>
</tbody>
</table>
Simulation

- Selection dataset
  Validation Regression

<table>
<thead>
<tr>
<th>Generation</th>
<th>Bayes A 20%A</th>
<th>GBLUP 20%A</th>
<th>GBLUP 0%A</th>
<th>EBV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.95</td>
<td>0.96</td>
<td>0.87</td>
<td>0.68</td>
</tr>
<tr>
<td>2</td>
<td>0.93</td>
<td>0.94</td>
<td>0.82</td>
<td>0.67</td>
</tr>
<tr>
<td>3</td>
<td>0.90</td>
<td>0.92</td>
<td>0.79</td>
<td>-0.13</td>
</tr>
</tbody>
</table>

Validation $R^2$

<table>
<thead>
<tr>
<th>Generation</th>
<th>Bayes A 20%A</th>
<th>GBLUP 20%A</th>
<th>GBLUP 0%A</th>
<th>EBV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.30</td>
<td>0.30</td>
<td>0.32</td>
<td>0.07</td>
</tr>
<tr>
<td>2</td>
<td>0.25</td>
<td>0.26</td>
<td>0.27</td>
<td>0.03</td>
</tr>
<tr>
<td>3</td>
<td>0.21</td>
<td>0.22</td>
<td>0.23</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Simulation

- Selection dataset
  - The highest R-square was with no polygenic variance, but this analysis fails the regression test
  - The lack of daughter proofs for low GEBV bulls can cause low squared correlations
  - An adjustment may be required from the reduced R-square to get the correct PEV
  - The Interbull genomic validation group may have to derive formulas to adjust reliabilities for genomic pre-selection
Simulation

- With Selection
  Validation $I = R^2$ between TBV and (G)EBV
  Model = ASREML PEV using assumed variance

<table>
<thead>
<tr>
<th>Animals</th>
<th>Validation</th>
<th>Reliabilities</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BLUP</td>
<td>GBLUP</td>
<td>BLUP</td>
</tr>
<tr>
<td>Tr. gen. 1</td>
<td>0.857</td>
<td>0.861</td>
<td>0.949</td>
</tr>
<tr>
<td>Tr. gen. 2</td>
<td>0.841</td>
<td>0.849</td>
<td>0.949</td>
</tr>
<tr>
<td>Val. gen. 1</td>
<td>0.126</td>
<td>0.472</td>
<td>0.280</td>
</tr>
<tr>
<td>Val. gen. 2</td>
<td>0.048</td>
<td>0.398</td>
<td>0.138</td>
</tr>
<tr>
<td>Val. gen. 3</td>
<td>0.000</td>
<td>0.349</td>
<td>0.066</td>
</tr>
</tbody>
</table>
Simulation

- With Selection
  Validation I = R2 between TBV and (G)EBV
  Model = ASREML PEV using estimated variances from the data

<table>
<thead>
<tr>
<th>Animals</th>
<th>Validation I</th>
<th></th>
<th>Reliabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>REML</td>
<td>GREML</td>
<td>REML</td>
</tr>
<tr>
<td>Tr. gen. 1</td>
<td>0.818</td>
<td>0.814</td>
<td>0.538</td>
</tr>
<tr>
<td>Tr. gen. 2</td>
<td>0.788</td>
<td>0.806</td>
<td>0.535</td>
</tr>
<tr>
<td>Val. gen. 1</td>
<td>0.110</td>
<td>0.431</td>
<td>0.158</td>
</tr>
<tr>
<td>Val. gen. 2</td>
<td>0.045</td>
<td>0.362</td>
<td>0.078</td>
</tr>
<tr>
<td>Val. gen. 3</td>
<td>0.000</td>
<td>0.307</td>
<td>0.038</td>
</tr>
</tbody>
</table>
Future Work

• Finish validating simulation model
  • Produce multiple replicates from the simulation model
  • Extend simulation model to produce data sets useful single step models
• Provide data-sets to members to test their own genomic reliability software
• Provide recommendations on calculation methods
• Provide recommendations on detecting genomic reliability over-estimation – relative validation $R^2$