Genomic GxE approaches modelling heterogeneous SNP variances: applied to simulated data

Birgit Gredler-Grandl and Mario Calus (WUR)

Virtual Interbull Meeting 2021
Genomic GxE models

- Multi-trait approach and reaction norm model can be implemented relying on:
  - Genomic relationships (GREML)
  - Random regression on SNP genotypes (RR-REML)

- GREML and RR-REML are equivalent
  - Homogeneous (co)variance assumed for all SNPs

- Certain regions in genome may harbour QTL → assumption of equal (co)variances is violated
Genomic GxE models

Can we model heterogeneous SNP (co)variances and do those models improve accuracy of genomic prediction?
Model heterogeneous SNP variances

- Make SNP (co)variances heterogeneous by **weighing**
  - (1) Weights derived from estimated SNP effects
  - (2) Re-compute SNP-effects using those weights

- Issue: computing (1) & (2) from the same data may inflate large SNP-effects
Proposed solution: split data in two

- Estimate SNP-effects assuming equal (co)variances for all SNP
- Calculate SNP specific weights within environment

- Estimate GEBV using the 2nd subset, applying weights on SNP (co)variance matrix within environment

picture source: freepik.com
Simulation (1) (QMSim, Sargolzaei and Schenkel, 2009)

G -1000  **Historical population**  
N=10,000

G -20  **Bottle neck**  
N=400

G 0  **Last generation HP**  
N=4,100

G1  
**Breed A**  
50m, 2000f

**Breed B**  
50m, 2000f

G 210  
**Breed A**  
1000m, 1000f

**Breed B**  
1000m, 1000f

• Random mating and selection
• Genome 30 Chr
• 100 cM length
• 1700 markers per Chr
• 150 QTL per Chr
• ~ 51,000 markers
• ~ 4,500 QTL
• 5 replicates
Simulation of phenotypes

- Phenotypes follow a reaction norm model
- Input: environmental values, genetic & residual (co)variances
- QTL-effects are simulated for QTLs simulated in QMSim
- Phenotype: environmental value * TBV + residual error

**Gen cov matrix Reaction norm model**

<table>
<thead>
<tr>
<th></th>
<th>$b_0$</th>
<th>$b_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_0$</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>$b_1$</td>
<td>0.05</td>
<td>0.025</td>
</tr>
<tr>
<td><strong>Environmental variance</strong></td>
<td>0.5</td>
<td></td>
</tr>
</tbody>
</table>
Heritability across environments

Reaction norm model – 20 continuous environments
Ventiles of environmental values
Validation study

Data set 1
Gen 205 + 206
4000 individuals

Data set 2
Training
Gen 207 + 208
4000 individuals

Data set 2
validation
Gen 209 + 210
4000 individuals

Individuals in each generation are randomly assigned to environments
Model Data set 1

- Reaction norm model (mtg2)

\[ y = \mu + \beta_0 + \beta_1 \times x + e \]

- Backsolve SNP-effects (calc_grm)

- Calculate weights as:
Model Data set 2

- SNP-BLUP (MiXBLUP)
  \[ y = \mu + Z\beta_0 + ZQ\beta_1 + e \]

- Apply weights (D) on SNP (co)variance matrix:

- G
Results: Estimated genetic covariance matrix for $b_0$ and $b_1$ in data set 1

<table>
<thead>
<tr>
<th></th>
<th>$b_0$</th>
<th>$b_1$</th>
<th>$b_0$</th>
<th>$b_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_0$</td>
<td>0.3</td>
<td></td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td>$b_1$</td>
<td>0.05</td>
<td>0.025</td>
<td>0.04</td>
<td>0.031</td>
</tr>
</tbody>
</table>
Results: Correlation between estimated GBV and TBV for $b_0$ and $b_1$

<table>
<thead>
<tr>
<th></th>
<th>Homogeneous SNP (co)variance</th>
<th>Heterogeneous SNP (co)variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_0$</td>
<td>0.521</td>
<td>0.551</td>
</tr>
<tr>
<td>$b_1$</td>
<td>0.588</td>
<td>0.601</td>
</tr>
</tbody>
</table>
Application in Irish beef crossbred data set

- Trait: age at slaughter (Berry et al., 2017)
- 14,668 genotyped bulls, steers, heifers
- HD imputed genotypes (662,011 SNPs)
- Yield deviation as phenotypes
- CG-effects as continuous descriptor of environment
Age at slaughter in days

mean = 746.7
sd  = 123.5
min = 427.0
max = 1094.0
Breeds: PCA G-Matrix purebred and crossbred animals
How to define sets for analysis?

- K-means clustering approach (similar Saatchi et al., 2011)
- Distance matrix between individuals computed as follows:
  - Apply on herds
  - Set up GRM for herds
  - Define sets according to cluster results
Summary

- Analysis protocol to model heterogeneous SNP variances developed

- Slight increase in accuracy with heterogeneous SNP variances in reaction norm models in simulated data

- Currently investigating added value in real data
Thank you!

Alan Twomey and Donagh Berry

This project has received funding from the European Union’s Horizon 2020 research and innovation program under Grant Agreement No 727213