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

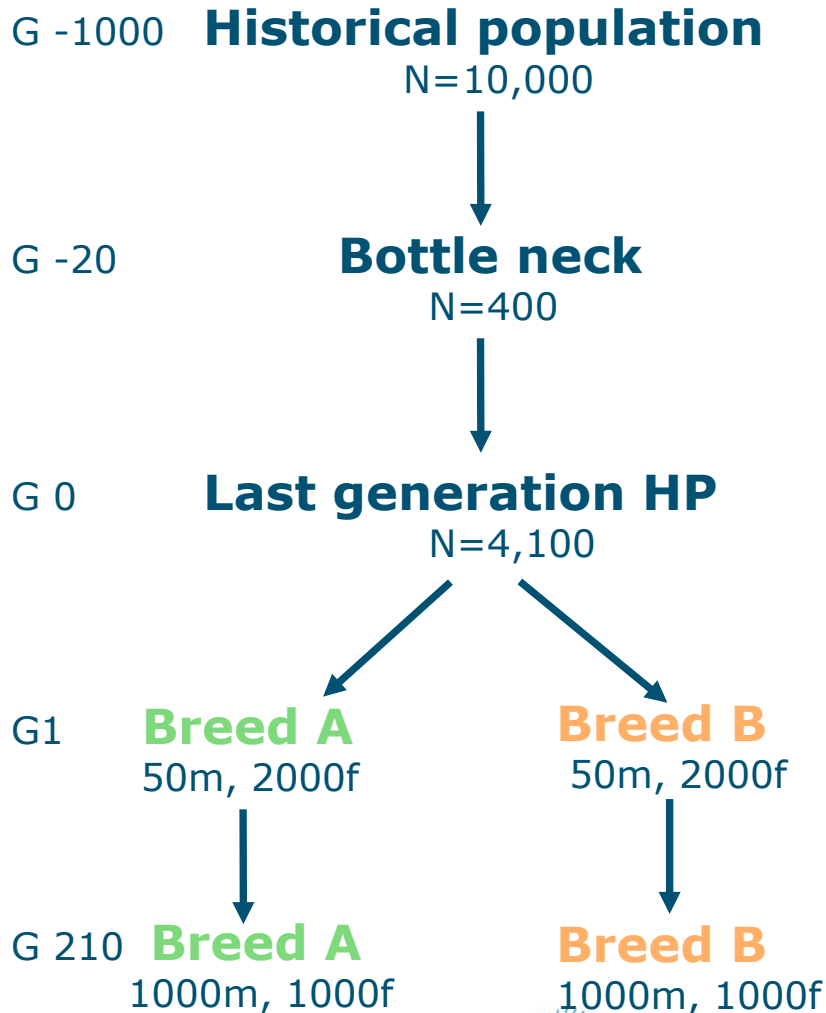- Issue: computing (1) & (2) from the same data may inflate large SNP-effects

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

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**   **Breed B**
50m, 2000f   50m, 2000f

↓   ↓

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

WAGENINGEN
UNIVERSITY & RESEARCH

100years
1918 — 2018

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		
	b_0	b_1
b_0	0.3	
b_1	0.05	0.025
Environmental variance 0.5		

Heritability across environments

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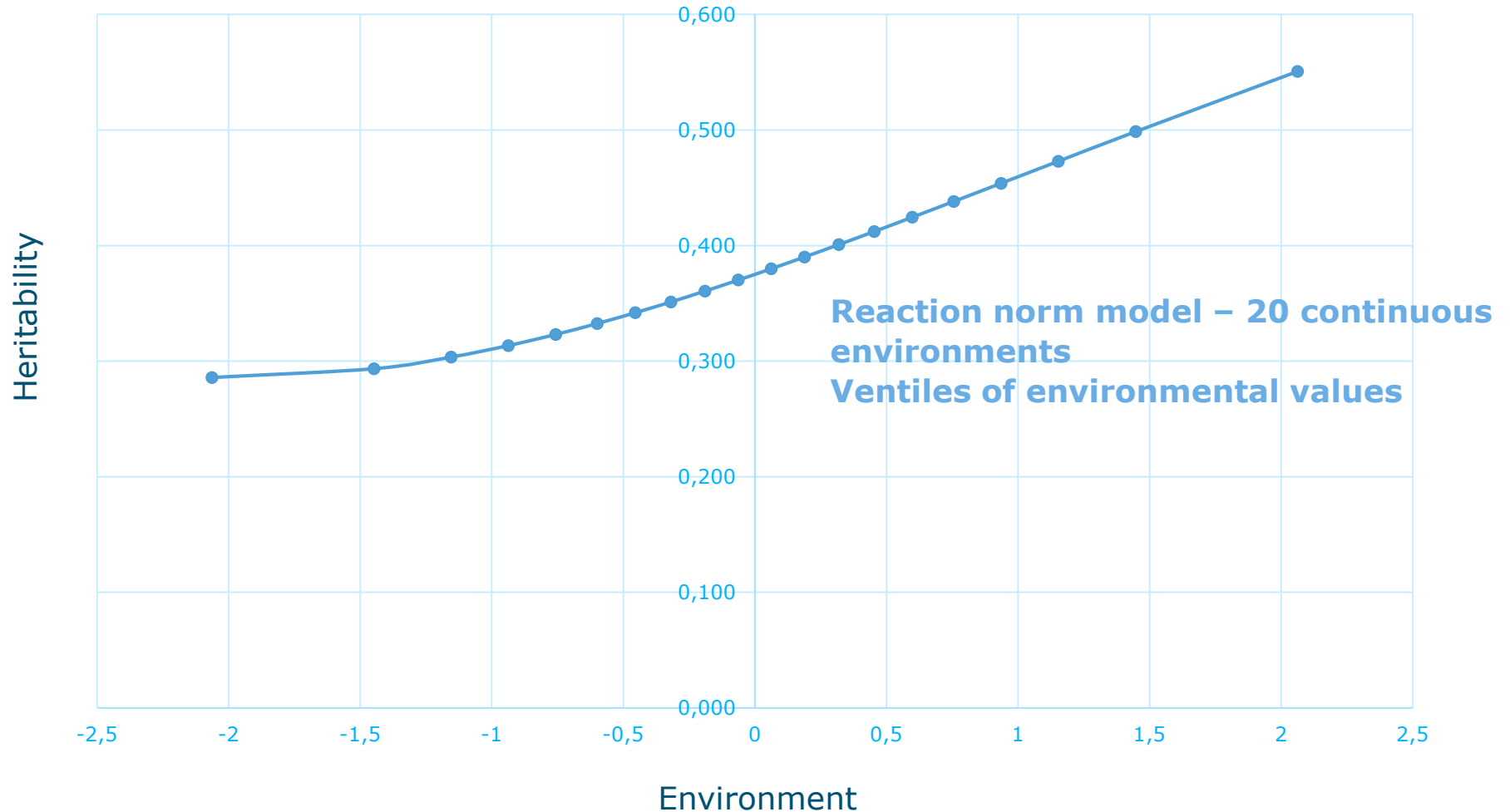
Validation study

Data set 1
Gen 205 + 206
4000 individuals
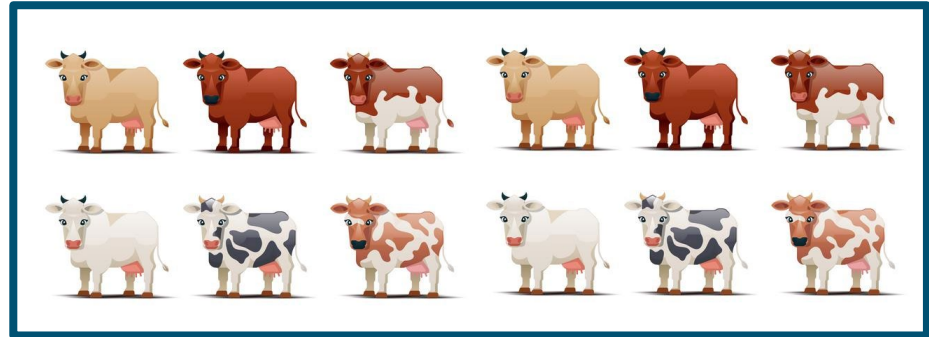
**Data set 2
Training**
Gen 207 + 208
4000 individuals

**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)

$$y = \mu + \beta_0 + \beta_1 * x + e$$

- 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

	b_0	b_1	b_0	b_1
b_0	0.3		0.35	
b_1	0.05	0.025	0.04	0.031

Results: Correlation between estimated GBV and TBV for b_0 and b_1

	Homogeneous SNP (co)variance	Heterogeneous SNP (co)variance
b_0	0.521	0.551
b_1	0.588	0.601

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

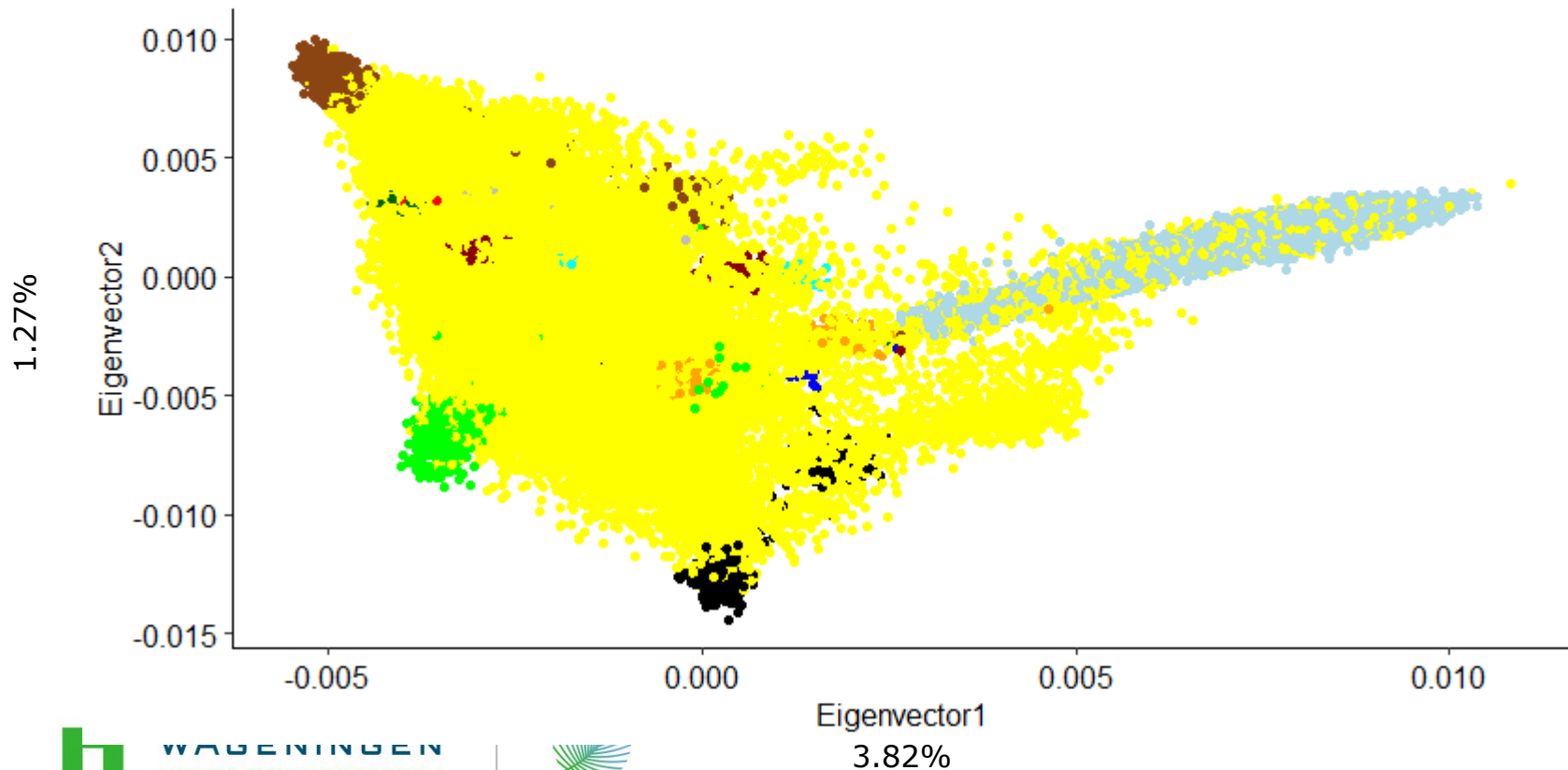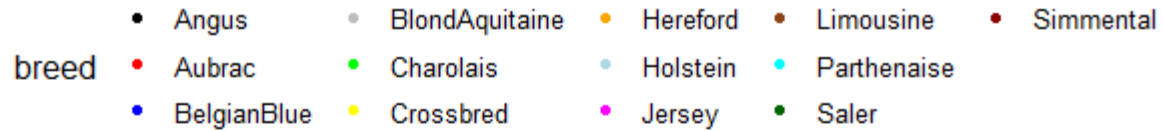- CG-effects as continuous descriptor of environment

Age at slaughter in days
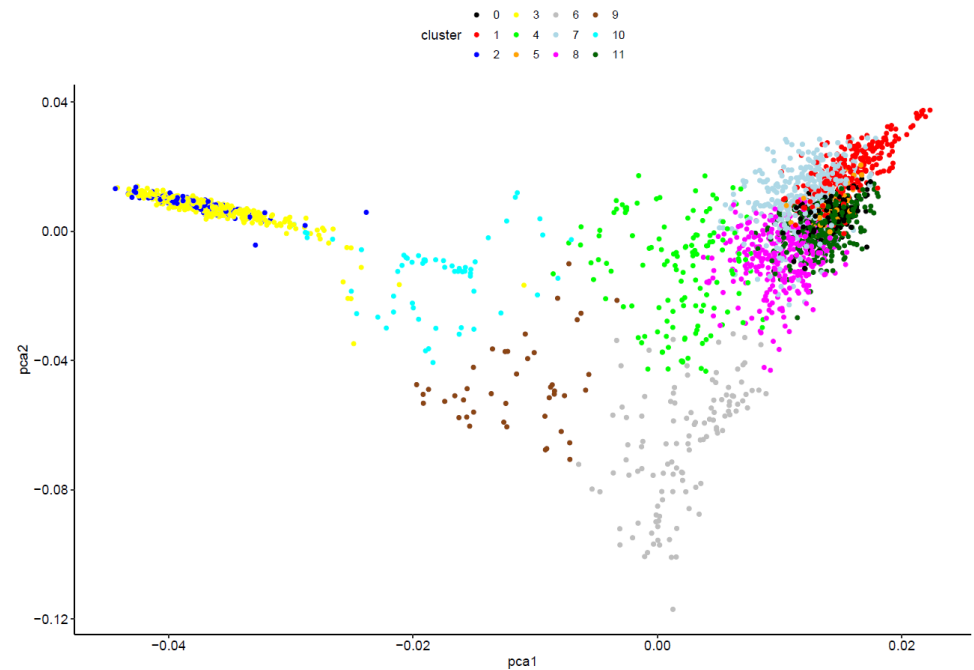
mean = 746.7

sd = 123.5

min = 427.0

max = 1094.0
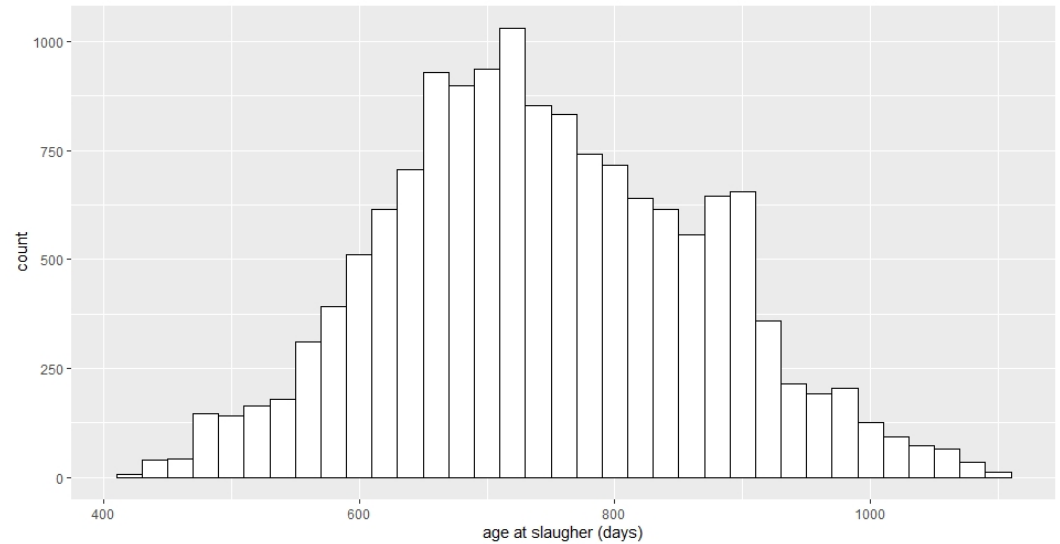
Breeds: PCA G-Matrix purebred and crossbred animals
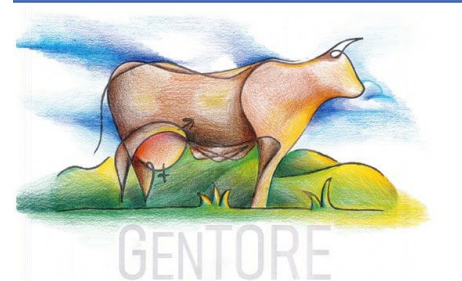
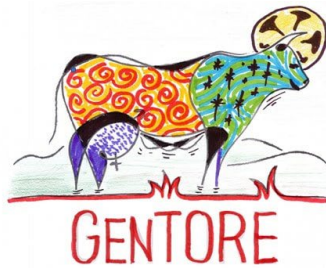# How to define sets for analysis?

- 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