Approximation of reliabilities for random-regression single-step GBLUP models

M. Bermann, I. Aguilar, A. Alvarez Munera, J. Bauer, J. Pribyl, D. Lourenco, I. Misztal
June 2023
Introduction

• Random regression models (RRM) are used in many countries for longitudinal traits

• (G)EBV and its reliability are a function of random regression coefficients
  • Weighted average over three lactations of cumulative 305 days milk yield

• Genomics can be incorporated with single-step genomic best linear unbiased predictor

• Reliabilities must be approximated but no method exist when genomics is added

• Objective:
  • Develop an efficient and accurate method to approximate reliabilities in single-step RRM
Random regression models

$$y_{ijklm} = h t d_{ik} + \sum_{m=0}^{nr} \phi_{ijlm} \beta_m + \sum_{m=0}^{nr} \phi_{ijlm} u_{ijm} + \sum_{m=0}^{nr} \phi_{ijlm} p e_{ijm} + e_{ijklm}$$

- Record for $i^{th}$ lactation, $j^{th}$ cow, $k^{th}$ contemporary group, $l^{th}$ test day, $m^{th}$ polynomial coefficient
- Selection is made for $EBV_j = k' \hat{u}_j$
- Our method approximates the reliability of $EBV_j$
Approximation without genomics

• Accumulate reliability of different sources:
  • Own phenotype
  • Parents
  • Progeny

• Accumulation in blocks:
  • Traits
  • Correlated effects

• Read the pedigree twice
• Fast and robust
Approximation of accuracies in ssGBLUP

- Combine different sources in terms of ERC
- Convert pedigree reliabilities into ERC
- Use ERC as weights in GBLUP
- Obtain GBLUP reliabilities
- Combine reliabilities in terms of ERC and remove double-counting

Approximating Genomic Reliabilities for National Genomic Evaluation

Z. Liu¹, P. M. VanRaden², M.H. Lidauer², M. P. Calus³, H. Benhajali⁴, H. Jorjani⁵ and V. Ducrocq⁶

Efficient approximation of reliabilities for single-step genomic best linear unbiased predictor models with the Algorithm for Proven and Young

Matias Bemann,¹,© Daniela Lourenco, and Ignacy Misztal©
ACCF90GS2 flowchart

Sort pedigree

For each trait

- Calculate pedigree reliabilities
- Calculate genomic reliabilities
- Derive weights
- Remove double counting
- Propagate to ngt animals
- Combine sources of information

Multiple-trait adjustment

- Maximum number of generation
- Recursive quicksort
- Harris & Johnson (1998)
- Supplied by user
- Back-solve pedigree reliabilities with a root-finding technique
- Block sparse inversion
- Back-solve pedigree reliabilities and set to zero contributions from ngt animals
- Back-solve genomic reliabilities and use them as weights
- Effective daughter/records contribution
- Strabel & Misztal (2001)

mbermann@uga.edu
Approximation of accuracies with genomics

Accuracies of RR coefficients without genomics

Accuracies of EBV without genomics

ERC

Accuracy of single-trait ssGBLUP model

accf90rr

rrmebvf90

accf90GS2
Datasets

- Czech three lactation model for milk production
- Selection on average over the three lactations of 305 days cumulative milk production
- 4<sup>th</sup> order Legendre polynomials

Effects:
- General mean
- Herd-year-season
- Fixed regression for groups including:
  - Age in first lactation
  - Calving interval
  - Days open
  - Season
- Additive effect
- Permanent environmental effect
Datasets

• Small dataset:
  • 111,494 test-day records
  • 44,582 animals in the pedigree
  • 2,892 genotyped animals

• Large dataset:
  • 30,366,184 test-day records
  • 2,512,681 animals in the pedigree
  • 54,221 genotyped animals
Small dataset

Pedigree reliabilities

\[ \text{cor} = 0.99 \\
\text{b0} = 0.01 \\
\text{b1} = 0.96 \]

Genomic reliabilities

\[ \text{cor} = 0.99 \\
\text{b0} = 0.01 \\
\text{b1} = 0.95 \]
Small dataset

Genomic gain

- $\text{cor} = 0.98$
- $b_0 = 0.14$
- $b_1 = 0.83$

Genomic gain

- $\text{cor} = 0.97$
- $b_0 = 0.14$
- $b_1 = 0.85$

mbermann@uga.edu
Large dataset

Genomic gain

Estimated genomic rel

Estimated pedigree rel

cor = 0.84
b0 = 0.51
b1 = 0.41
Challenges

• Models with different traits: fat, protein, SCS

• Timepoint vs. cumulative EBV

• External information: Interbull
Conclusions

• Approximation of reliabilities worked well
• Scalable for large datasets
• Reliabilities of each random regressor with genomics are not needed
• Testing the method in different datasets is required
UGA Team
Questions?

mbermann@uga.edu