Approximation of reliabilities for single-step GEBVs

Hongding Gao, Matti Taskinen

MiX99 course on genomic prediction

COURSE DAY 3, March 11th, 2026

Contents

- Introduction
- Luke/MiX99 method for GEBV R^2
- Interbull method for GEBV R^2
- A case study: Luke method vs. Interbull method

Introduction

- Calculating exact reliabilities (R^2) by inverting LHS of MME is computationally infeasible for large datasets
- Efficient approximation methods are needed
- Single-step GEBV R^2 are usually approximated for genotyped and non-genotyped individuals separately
- It requires propagating the genomic information to the non-genotyped animals
- Residual polygenic (RPG) effect needs to be included

How MiX99 does

- Approximate GEBV R^2 for genotyped and non-genotyped individuals separately

- Use effective record contributions (ERC) as weights
  - ✓ For genotyped individuals via GBLUP/SNPBLUP
  - ✓ For non-genotyped individuals via PBLUP

Luke

Luke/MiX99 method: 7 steps

- Step1: compute EBV R^2 for all individuals in the pedigree
- Steps 2-3: compute GEBV R^2 for genotyped individuals
- Steps 4-7: compute GEBV R^2 for non-genotyped individuals

Step1: Compute EBV R^2

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{e}$$

- Input data: original phenotype, full pedigree, and (co)variance components
- This step produces EBV reliability $\mathbf{r}_p^2$
- We partition

$$\mathbf{r}_p^2 = \begin{bmatrix} \mathbf{r}_{p,n}^2 \\ \mathbf{r}_{p,g}^2 \end{bmatrix}$$

example.clm

```
DATAFILE  example.dat

INTEGER animal herd calvys1 calvys2 calvys3 cage1 cage2 cage3 hy1 hy2 hy3

REAL milk1 milk2 milk3

MISSING  -999.0

DATASORT BLOCK=herd PEDIGREECODE=animal

PEDFILE example.ped

PEDIGREE   G  am          # animal model

PARFILE    VC.var

PRECON n       # NO preconditioner need for apax

WITHINBLOCK G hy1

MODEL
  milk1 = calvys1   cage1    hy1   G(animal)
```

apax.dir


```
# Type of analysis: 4 = Tier and Meyer approach
4

# Maximum number of non-zeros in the sparse matrix
600000

# Original directive file given to mix99i (or -):
MiX99_DIR.DIR

# Absorption level effect
2
```

Run

```
mix99i example.clm > mix99i.log && apax99 < apax.dir > apax99.log
```

Step 2: reverse R2 for the genotyped individuals

- Use the reverse reliability method to compute the ERC_g for the genotyped individuals

- Input data: EBV R^2 for the genotyped individuals from step 1, i.e., $\mathbf{r}^2_{p,g}$ and full pedigree

Luke

apax.dir


```
# Reliability method (AccurType): Reversed reliability approximation,
# p = pedigree file given, no mix99i needed:
20 p

# Filename of the reliability information (id and r2s):
id_r2s.dat

# Pedigree file:
data/AM.ped

# Number of ERC values (and r2s):
1

# ERC parameters: tol  maxit smallest [h2]
                  0    0     0        0.3333333333333
```

Run

```
apax99 < apax.dir > apax99_step2_20p.log
```

New feature: we can combine step 1 and 2

- Currently these 2 steps done separately in ApaX:
  - Method 4: for Tier&Meyer reliability computation
  - Method 20: reversed reliability to get ERC

- New method: 24
  - Makes method 4 and then method 20 automatically

- Other new methods:
  - 22 & 23: like method 24 but uses methods 2 and 3 for the reliabilities and then method 20
  - 44: uses method 4 for reliabilities and method 40 for reversed reliabilities

Luke

Using method 24:

Like method 4 but also a file having ID codes of the genotyped need to be given

- ApaX99 with **methods 4 & 20**:
- Method 4

```
4           # Method number
100000      # Number of non-zeros
MiX99_DIR.DIR # Directive file
2           # Absorption level
```

- Method 20

```
20          # Method number
id_r2       # ID numbers and reliabilities
MiX99_DIR.DIR # Directive file
0           # Use defaults
```

- ApaX99 with **method 24**:

```
24          # Method number
id_genotyped # ID numbers of the genotyped
100000      # Number of non-zeros
MiX99_DIR.DIR # Directive file
0           # Use defaults for method 20
2           # Absorption level
```

Method 24 output

- PEVani: has reliabilities from method 4
- PEVani_genotyped: reliabilities of the genotyped from PEVani
- PEVani_ERCR2 ERC and reliabilities after method 20

Step 3: Compute GEBV R^2 for the genotyped individuals

- First compute the DGV R^2 for the genotyped individuals
- Avoid directly including the RPG effects in the SNPBLUP model

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}\mathbf{g} + \mathbf{e}$$

where

$$\mathbf{e} \sim N(\mathbf{0}, \mathbf{D}^{-1}\sigma_e^2) \quad (1)$$

$\mathbf{D}$ is a diagonal matrix with elements of ERC_g from Step 2

Luke

Run

```
snp_blup_rel_para \
     -nthr 10 -memhigh
     -m PvR1 -c 2pq -h2 0.443 \
     -a base_af.dat -s 0 0 3  -o Id_and_ERC.dat  -wt 2 markers.dat  SNP_rel.out \
     > snp_blup_rel.log
```

Here I only used every third markers instead of including all the SNPs

Blend with EBV R^2

$$r^2_{GEBV} = \frac{(1-\omega)\mathbf{G}_{ii} r^2_{DGV,i} + \omega \mathbf{A22}_{ii} r^2_{EBV,i}}{(1-\omega)\mathbf{G}_{ii} + \omega \mathbf{A22}_{ii}}$$

where

ω is the proportion of the RPG

$r^{2*}_{DGV,i}$ is the DGV R^2 from SNPBLUP without RPG for animal i

$r^2_{EBV,i}$ is the conventional EBV R^2 of animal i

$\mathbf{G}_{ii}$ is the diagonal element i of the $\mathbf{G}$ matrix

$\mathbf{A22}_{ii}$ is the diagonal element i of the $\mathbf{A}_{22}$ matrix

Step 4: compute weights for non-genotyped individuals

- Compute ERC_f (full) which are ERC corresponding to the EBV R^2 in step 1
- Using reversed reliability method
- We then partition

$$\mathbf{ERC}_f = \begin{bmatrix} \mathbf{ERC}_{f,n} \\ \mathbf{ERC}_{f,g} \end{bmatrix}$$

Luke

apax99.dir


```
# Reliability method (Accurtype): Reversed reliability approximation
20 p

# Filename of the reliability information (id and r2s):
/work/users/L2583/JDS2025/step4/reliability4_all.txt

# Pedigree file
/home/L2583/GEBV_R2/data305/new_pruned_pedigreeH.ped.s

# Number of ERC values (and r2s):
1

# ERC parameters: tol maxit smallest [h2]
     1e-8  10000 0.0001 0.4429984
```

Run

```
apax99 < apax99.dir > apax99_step4.log
```

Step 5: compute weights for genotyped individuals

- Compute $ERC_{g,g}$ which account for the genomic information and will be used in Step 7 as weights

$$ERC_{g,g} = ERC_{f,g} + \frac{1-h^2}{h^2} \times \left(\frac{r^2_{g,g}}{1-r^2_{g,g}} - \frac{r^2_{p,g}}{1-r^2_{p,g}}\right)$$

where

$ERC_{f,g}$ is extracted from ERC_f from step 4 for genotyped animals

$r^2_{p,g}$ is pedigree-based reliability from step 1 for genotyped animals

$r^2_{g,g}$ is the genomic reliability from step 3 for genotyped animals

Luke

Step 6: combine the weights for all individuals

- Aggregating the ERC values generated from Steps 4 and 5

$$\mathbf{ERC}_{ss} = \begin{bmatrix} \mathbf{ERC}_{f,n} \\ \mathbf{ERC}_{g,g} \end{bmatrix}$$

Step 7: Compute GEBV R^2 for non-genotyped individuals

- Use a simplified weighted-PBLUP model

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{a} + \mathbf{e}$$

where

$$\mathbf{e} \sim N(\mathbf{0}, \mathbf{D}_p^{-1}\sigma_e^2)$$

$\mathbf{D}_p$ is a diagonal matrix with elements of ERC from Step 6

example.clm

```
DATAFILE  /work/users/L2583/JDS2025/step56/schemeE/pheno.txt

INTEGER animal ones
REAL ERC1 milk1

MISSING  -999.0
DATASORT PEDIGREECODE=animal

PEDFILE /home/L2583/GEBV_R2/data305/new_pruned_pedigreeH.ped.s
PEDIGREE   G  am  # animal model
PARFILE  /work/users/L2583/JDS2025/step1/parin.dat

MODEL
  milk1 =  ones    G(animal) !weight=ERC1
```

apax.clm


```
1  # Type of analysis: 1= Interbull accuracies
2  4
3  # Maximum number of non-zeros in the sparse matrix
4  600000
5  # Original directive file given to mix99i (or -):
6  MiX99_DIR.DIR
7  # Absorption level effect
8  2
```

Run

```
mix99i example.clm > mix99i.log && apax99 < apax99.dir > apax99_step7.log
```

More details

The cookbook for approximating reliabilities in large-scale single-step genomic evaluation using MiX99

Hongding Gao Ismo Strandén

January 26, 2023

RESEARCH ARTICLE Open Access

A computationally efficient method for approximating reliabilities in large-scale single-step genomic prediction

Hongding Gao*, Andrei A. Kudinov, Matti Taskinen, Timo J. Pitkänen, Martin H. Lidauer, Esa A. Mäntysaari and Ismo Strandén

Luke

Interbull method

- Step 1: calculate genomic ERC/EDC gain φ_c
- Step 2: propagate genomic information
- Step 3: combine the genomic R^2 gain with the conventional R^2

Step 1: calculate genomic ERC/EDC gain φ_c

- This step requires the Interbull setup:
  - ✓A list of validation bulls for Interbull GEBV test (in order to compute *f*)
  - ✓Adjustment factor (*f*): ratio of the expected and theoretical EDC
  - ✓GEBV from full and reduced dataset
  - ✓Theoretical R2 from both full and reduced dataset
  - ✓Expected EDC

Step 1: calculate genomic ERC/EDC gain φ_c

$$\varphi_i^{adj} = \frac{1 - h^2}{h^2} \times \left(\frac{r_{DGV}^2}{1 - r_{DGV}^2} \times f - \frac{r_{EBV}^2}{1 - r_{EBV}^2} \right)$$

$$\varphi_c = \frac{1}{n} \sum_{1}^{n} \varphi_i^{adj}$$

Step 2: propagate genomic information (φ_i^{prog})

- Passing φ_c via pedigree first from youngest to oldest animals
- Then from oldest to youngest via pedigree

Step 3: GEBV R^2

- For genotyped animals

$$\varphi_i^{total} = \varphi_i^{conv} + \varphi_c$$

- For non-genotyped animals

$$\varphi_i^{total} = \varphi_i^{conv} + \varphi_i^{prog}$$

- Final GEBV R^2

$$R_i^2 = \frac{\varphi_i^{total}}{\varphi_i^{total} + \lambda}$$

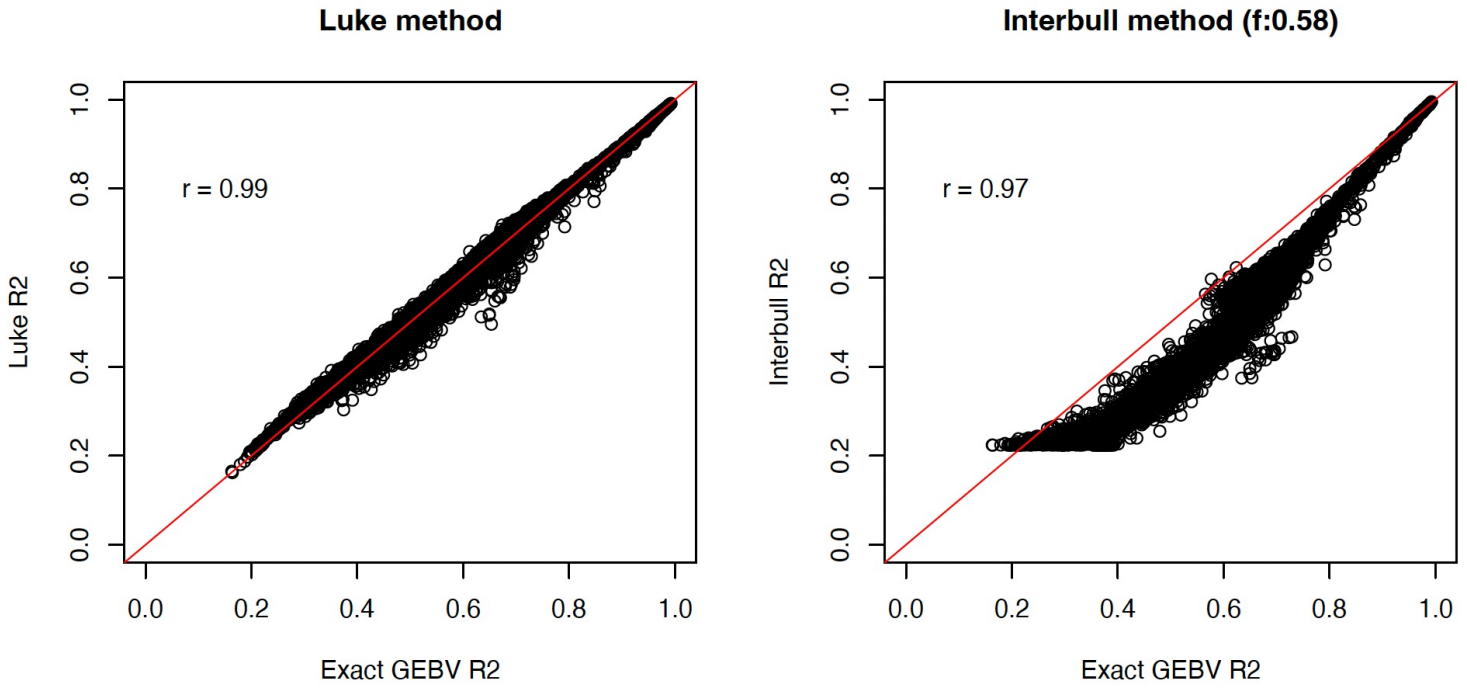
A case study

- Data
  - 47K Finnish Red cows with 305-day milk yield records from lactation one
  - 19k genotyped animals
  - 50K SNPs
  - h^2: 0.44
  - RPG: 0.3

Luke

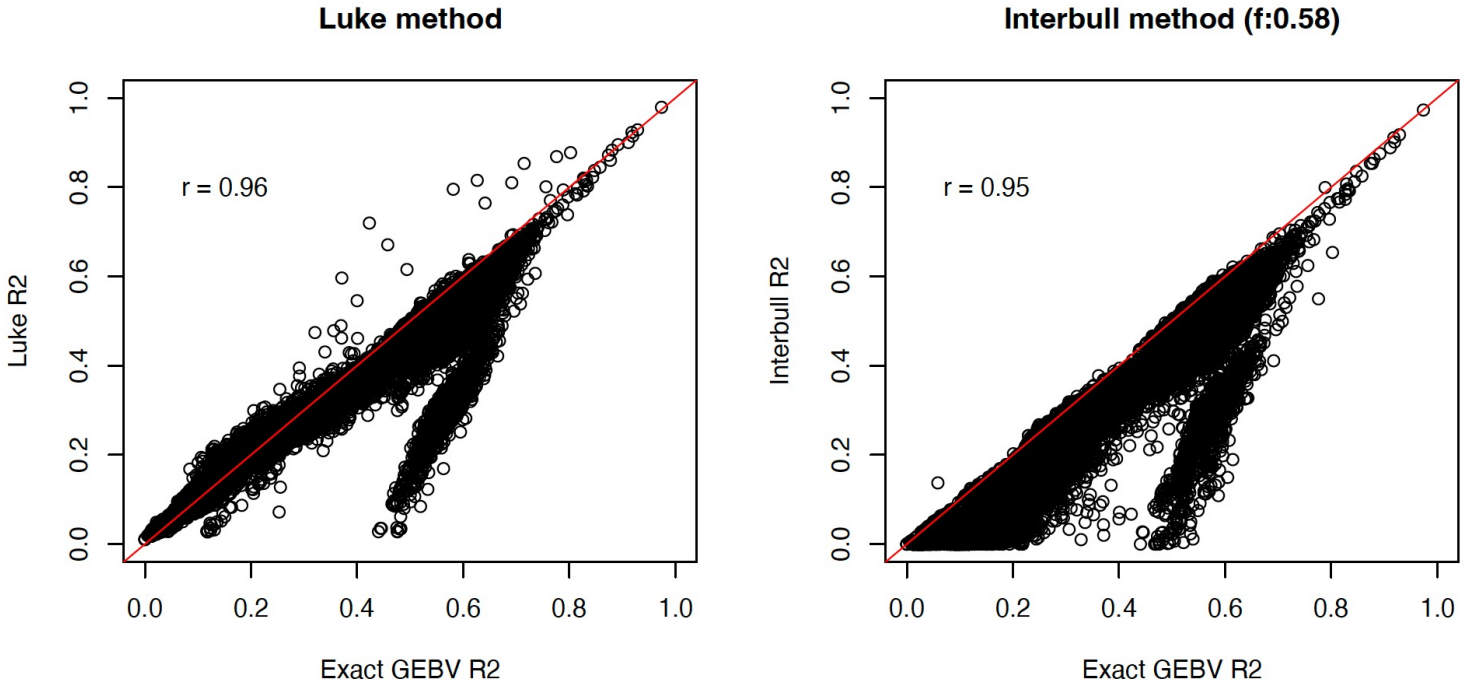
Genotyped animals

- Mean R^2: 0.66, 0.63, 0.57 for exact, Luke, and Interbull method



Non-genotyped animals

- Mean R2: 0.48, 0.44, 0.43 for exact, Luke, and Interbull method



Conclusions

- Both methods provided effective strategy for approximating GEBV R^2 from single-step model in practice

- The approximated GEBV R^2 were in good agreement with the exact GEBV R^2

More details

INTERBULL BULLETIN NO. 61. 21-22 June 2025, Louisville, Kentucky, USA

Comparing methods for approximating reliabilities in large-scale single-step genomic evaluations

H. Gao[1], I. Strandén[1] and Z. Liu[2]

INTERBULL BULLETIN NO. 60. 20-21 May 2024, Bled, Slovenia

Guidelines for Approximating Genomic Reliabilities of the Single-Step Genomic Model

Z. Liu[1], I. Strandén[2], J. Vandenplas[3], H. Eding[4], M. Lidauer[2], K. Haugaard[5], and P. M. VanRaden[6]

Luke

Thank you!