Accounting for trait-specific genomic and residual polygenic covariances in multivariate single-step genomic evaluation

Karin Meyer

AGBU, a joint venture of NSW Department of Primary Industries and University of New England,
Armidale, 2351, Australia

WCGALP 2022

Introduction

- GBLUP: SNP markers available rarely explain all genetic variation
- Fit **R**esidual **P**oly**G**enic (RPG) effects
  - explicitly: augment model by RPG
    - double no. of equations for genetic effects (BV model)
    - slower convergence for iterative solution schemes
  - implicitly: weighted average of GRM & NRM for genotyped animals

$$\lambda \mathbf{G} + (1 - \lambda)\mathbf{A}_{22} \qquad \text{with} \qquad 0 \leq \lambda \leq 1$$

    - Univariate: $\lambda = \text{Var}_{SNP} / (\text{Var}_{SNP} + \text{Var}_{RPG})$
    - Multivariate: Single λ implies $\text{Cov}_{SNP} \propto \text{Cov}_{RPG}$
- Questions
  - How best to account for different λ for traits in MV analysis?
  - How much might we gain doing so?

Objectives

1. Show computing strategy for MV ssGBLUP with trait specific λ_i
    - avoid fitting separate RPG
    - exploit ‘canonical transformation’
2. Use simulation to evaluate effects on accuracy & bias of EBVs

Model

- Breeding value model for *t* traits:

$$\mathbf{y} = \underbrace{\mathbf{X}\mathbf{b}}_{\text{Fixed}} + \underbrace{\mathbf{Z}\mathbf{u}}_{\text{Genetic}} + \mathbf{e} \quad \text{with} \quad \begin{array}{ccccc} \mathbf{u} & = & \overbrace{\mathbf{g}}^{\text{SNP}} & + & \overbrace{\mathbf{a}}^{\text{RPG}} \\ \Sigma_U & = & \Sigma_G & + & \Sigma_A \end{array}$$

- Order $\mathbf{u}$: traits within animals; partition $\mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} \begin{array}{l} \text{Non-genot.} \\ \text{Genotyped; } n_2 \end{array}$
- Coefficient matrix in MME for ssGBLUP

$$\underbrace{\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} \end{bmatrix}}_{\text{'Data' part}} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \underbrace{\mathbf{A}^{-1} \otimes \Sigma_U^{-1}}_{\text{NRM part}} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \underbrace{\boxed{\Delta} - \mathbf{A}_{22}^{-1} \otimes \Sigma_U^{-1}}_{\text{'add-on' from } \mathbf{H}^{-1}} \end{bmatrix} \end{bmatrix}$$

Calculating Δ

- For $\boldsymbol{\Sigma}_G \propto \boldsymbol{\Sigma}_A$

$$\boldsymbol{\Delta} = \left(\lambda\mathbf{G} + (1-\lambda)\mathbf{A}_{22}\right)^{-1} \otimes \boldsymbol{\Sigma}_U^{-1} \quad \longrightarrow \text{need inverse of size } n_2 \times n_2$$

- Otherwise: $\boldsymbol{\Delta}$ not separable $\longrightarrow$ need inverse of size $tn_2 \times tn_2$

$$\boldsymbol{\Delta} = \left(\mathbf{G} \otimes \boldsymbol{\Sigma}_G + \mathbf{A}_{22} \otimes \boldsymbol{\Sigma}_A\right)^{-1}$$

Calculating Δ

- For $\Sigma_G \propto \Sigma_A$

$$\Delta = \left(\lambda \mathbf{G} + (1-\lambda)\mathbf{A}_{22}\right)^{-1} \otimes \Sigma_U^{-1} \quad \longrightarrow \text{ need inverse of size } n_2 \times n_2$$

- Otherwise: Δ not separable $\longrightarrow$ need inverse of size $tn_2 \times tn_2$

$$\Delta = \left(\mathbf{G} \otimes \Sigma_G + \mathbf{A}_{22} \otimes \Sigma_A\right)^{-1}$$

- Use ‘canonical transformation’: $\underbrace{\mathbf{Q}\Sigma_G\mathbf{Q}' = \mathbf{I} \quad \text{and} \quad \mathbf{Q}\Sigma_A\mathbf{Q}' = \mathbf{D}}_{\text{Simultaneously diagonalise both } \Sigma}$

$$\Delta = \left[\mathbf{I} \otimes \mathbf{Q}'\right]\left(\mathbf{G} \otimes \mathbf{I} + \mathbf{A}_{22} \otimes \mathbf{D}\right)^{-1}\left[\mathbf{I} \otimes \mathbf{Q}\right]$$

Reorder $\mathbf{u}_2$: Animals within traits $\longrightarrow$ need t inverses of size $n_2 \times n_2$

$$\Delta = \left[\mathbf{Q}' \otimes \mathbf{I}\right]\underbrace{\left(\mathbf{I} \otimes \mathbf{G} + \mathbf{D} \otimes \mathbf{A}_{22}\right)^{-1}}_{\text{Diagonal blocks for traits}}\left[\mathbf{Q} \otimes \mathbf{I}\right]$$

Efficient calculation of inverse matrices

- Use Woodbury matrix identity to compute t inverses

Mäntysaari et al. (2017)

For $\mathbf{G} = \mathbf{M}\mathbf{M}'/s$ with $\mathbf{M}$: matrix of centred marker counts

$$\left(\mathbf{G} + d_k\mathbf{A}_{22}\right)^{-1} = \frac{1}{d_k}\left[\mathbf{A}_{22}^{-1} - \mathbf{A}_{22}^{-1}\mathbf{M}\left(s\, d_k\mathbf{I} + \mathbf{M}'\mathbf{A}_{22}^{-1}\mathbf{M}\right)^{-1}\mathbf{M}'\mathbf{A}_{22}^{-1}\right]$$

For $\mathbf{M}'\mathbf{A}_{22}^{-1}\mathbf{M} = \mathbf{E}\mathbf{T}\mathbf{E}'$ ⟵ Eigen-decomposition

$$\left(\mathbf{G} + d_k\mathbf{A}_{22}\right)^{-1} = \frac{1}{d_k}\left[\mathbf{A}_{22}^{-1} - \mathbf{A}_{22}^{-1}\mathbf{M}\mathbf{E}\underbrace{\left(s\, d_k\mathbf{I} + \mathbf{T}\right)}_{\text{Diagonal}}{}^{-1}\mathbf{E}'\mathbf{M}'\mathbf{A}_{22}^{-1}\right]$$

⇒ Need single eigen-decomp. of matrix of size $m \times m$ (m: no. of markers)

Mäntysaari, E. A., Evans, R. D., and Strandén, I. (2017). Efficient single-step genomic evaluation for a multibreed beef cattle population having many genotyped animals. *J. Anim. Sci.*, 95(11):4728–4737.

Simulation

Data

- 21,000 animals; 8 gen.s
- 3 traits: $\lambda_i \approx 0.90, 0.85, 0.20$
- All $r_G = 0.7$, $r_E = 0.3$
- Randomly select 10, 30, 40, 50% of genotypes for animals in gen.s 5 to 8
- Omit records in gen. 8
- 30 replicates

REML analyses

FULL	Estimate $\boldsymbol{\Sigma}_G$ and $\boldsymbol{\Sigma}_A$
ONE	Estimate $\boldsymbol{\Sigma}_U$ and single λ
FX0.5	Estimate $\boldsymbol{\Sigma}_U$ for $\lambda = 0.5$

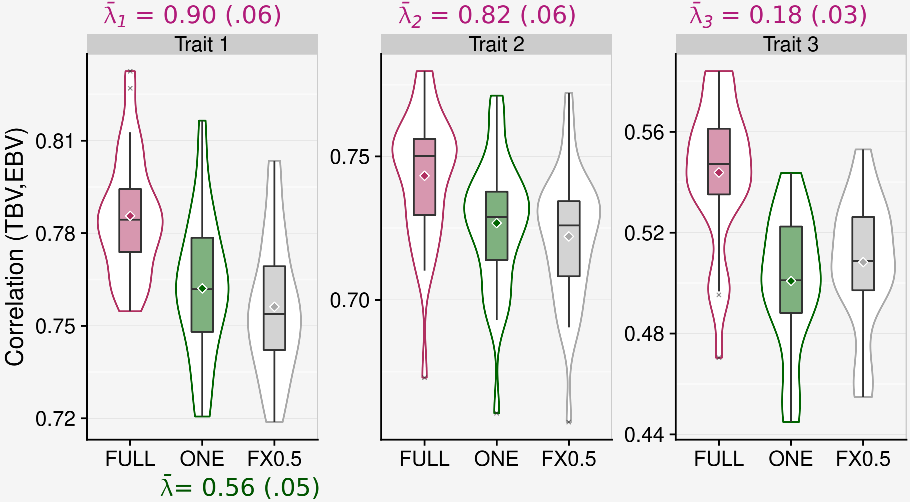
Summary statistics

Distribution over replicates

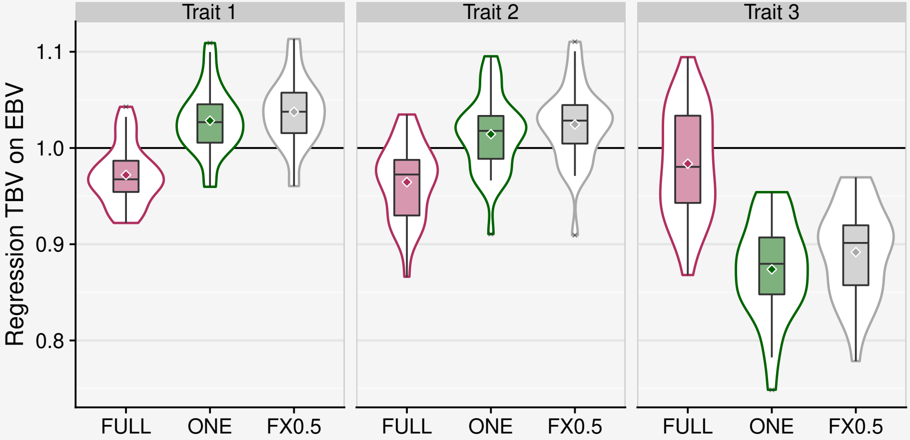
- Correlation TBV and EBV
- Regression of TBV on EBV

for genotyped animals in gen. 8

Accuracy of predicted BV

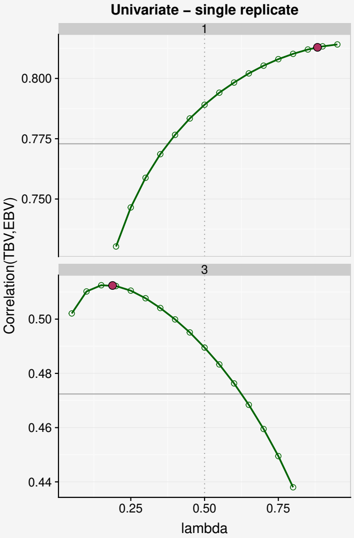


Dispersion of predicted BV



Comments

- Modest gains in accuracy for FULL
  - in spite of wide range in true λ_i
- Reasons?
  - Small simulation
  - Sampling errors in $\hat{\Sigma}_G$ and $\hat{\Sigma}_A$
  - EBVs insensitive against choices of λ_i deviating from ‘optimal’ values
    - reported in literature for univariate analyses
    - seems to hold for MV fitting single, ‘average’ λ



Conclusions

- Modelling trait-specific genetic covariances in ssGBLUP feasible
- Canonical transformation
    - Efficient calculation of matrix inverse
    - Avoid splitting EBVs into genomic & RPG parts
    - Substantially better convergence rates in iterative solution schemes (not shown)
- In practice
    - Trade-off potential gains *vs.* computational demands
    - Limited gains $\longrightarrow$ evaluate carefully

agbu

mla
MEAT & LIVESTOCK AUSTRALIA