Tuning indirect predictions based on SNP effects from ssGBLUP

Daniela Lourenco

A. Legarra, S. Tsuruta, D. Moser, S. Miller, I. Misztal

Why Indirect predictions?

- ## Interim evaluations
  - Between official runs

- ## Not all genotyped animals are in the evaluations
  - Animals with incomplete pedigree increase bias and lower R^2

- ## Commercial products
  - e.g. GeneMax for non-registered animals

Indirect predictions in ssGBLUP

$$\begin{bmatrix} \mathbf{X}^{'}\mathbf{X} & \mathbf{X}^{'}\mathbf{W} \\ \mathbf{W}^{'}\mathbf{X} & \mathbf{W}^{'}\mathbf{W}+\mathbf{H}^{-1}\lambda \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^{'}\mathbf{y} \\ \mathbf{W}^{'}\mathbf{y} \end{bmatrix}$$

$$\mathbf{DGV} = \mathbf{Z}\hat{a}$$

SNP effects

GEBVs

$$\hat{a} = \lambda\mathbf{D}\,\mathbf{Z}^{'}\mathbf{G}^{-1}\hat{u}$$

$$\mathbf{GEBV_{young}} = w_1\mathbf{PA} + w_2\mathbf{DGV} - w_3\mathbf{PP}$$

$$\mathbf{GEBV_{young}} \approx \mathbf{DGV} = \mathbf{Z}\hat{a}$$

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}$$

$$\mathbf{G}_{\mathrm{APY}}^{-1}$$

Lourenco et al., 2015

3

Problems with Indirect predictions

Genetic evaluation using single-step genomic best linear unbiased predictor in American Angus[1]

D. A. L. Lourenco,*[2] S. Tsuruta,* B. O. Fragomeni,* Y. Masuda,* I. Aguilar,[†]
A. Legarra,[‡] J. K. Bertrand,* T. S. Amen,[§] L. Wang,[§] D. W. Moser,[§] and I. Misztal*

$$\text{COR}(\widehat{\text{GEBV}}, \mathbf{Z}\hat{a}) > 0.99$$

$$\text{Avg}(\widehat{\text{GEBV}}) \approx 100 \qquad \neq \qquad \text{Avg}(\mathbf{Z}\hat{a}) \approx 0$$

Objectives

1) Fine-tune indirect predictions to be compatible with GEBV

2) Investigate whether SNP effects are accurate when APY is used

- Possibly use subset of core animals

Dataset

- American Angus Association
  - 8.2M animals in pedigree
  - 6.2M birth weight (BW)
  - 6.8M weaning weight (WW)
  - 3.4M post-weaning gain (PWG)
  - 81k genotyped
    - born 1977-2012: 66k
    - born 2013-2014: 15k

- Complete
  - Phenotypes up to 2012
  - Genotypes up to 2014 (81k)
- Reduced
  - Phenotypes up to 2012
  - Genotypes up to 2012 (66k)

- 3-trait with mat and mpe
  - Results for PWG

Accuracy of SNP effects from $\mathbf{G}_{\mathrm{APY}}^{-1}$ or $\mathbf{G}_{\mathrm{CC}}^{-1}$

$$\hat{a}_{\mathbf{G}} = \lambda \mathbf{D} \, \mathbf{Z}' \mathbf{G}^{-1} \hat{u}$$

$$\hat{a}_{\mathbf{G}_{APY}^{-1}\mathbf{H}} = \lambda \mathbf{D} \, \mathbf{Z}' \mathbf{G}_{\mathrm{APY\_high\_reliability}}^{-1} \hat{u}_{\mathrm{APY}}$$

- Correlation between SNP effects

$$\hat{a}_{\mathbf{G}_{APY}^{-1}\mathbf{R}} = \lambda \mathbf{D} \, \mathbf{Z}' \mathbf{G}_{\mathrm{APY\_random}}^{-1} \hat{u}_{\mathrm{APY}}$$

- Correlation between $\mathbf{Z}\hat{a}$

$$\hat{a}_{\mathbf{G}_{cc}^{-1}\mathbf{H}} = \lambda \mathbf{D} \, \mathbf{Z}' \mathbf{G}_{CC\_high\_reliability}^{-1} \hat{u}_{\mathrm{APY}}$$
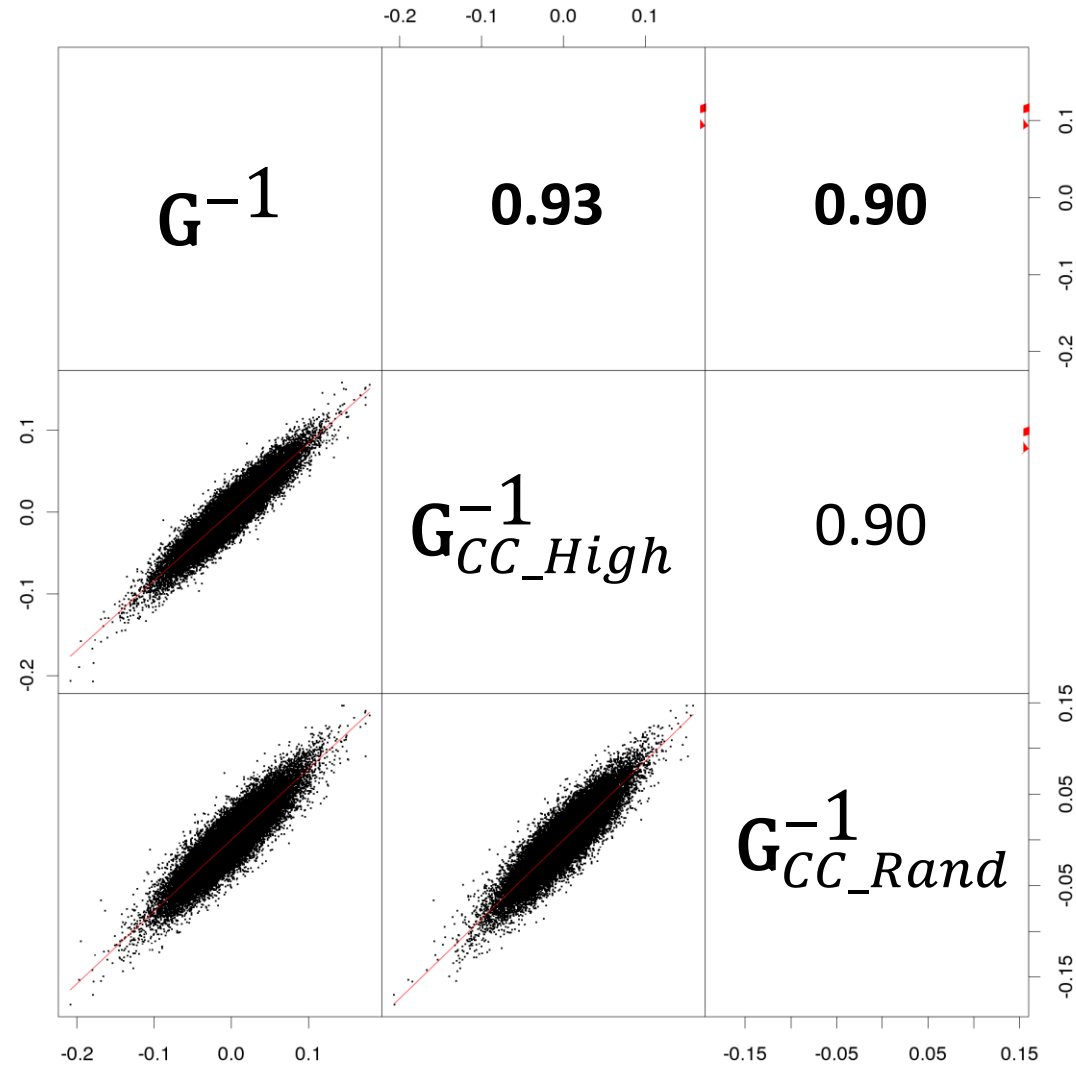
$$\hat{a}_{\mathbf{G}_{cc}^{-1}\mathbf{R}} = \lambda \mathbf{D} \, \mathbf{Z}' \mathbf{G}_{CC\_random}^{-1} \hat{u}_{\mathrm{APY}}$$
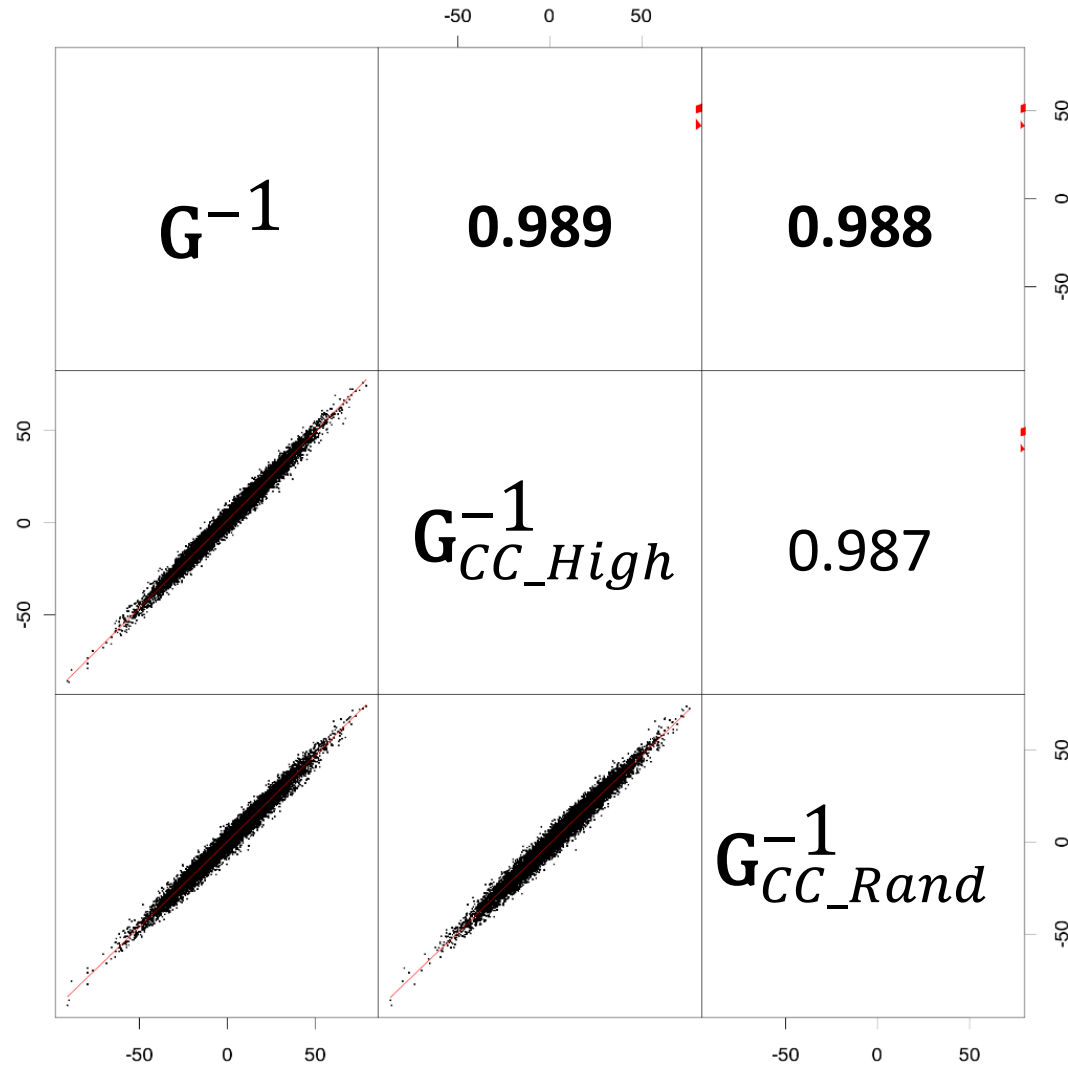
Statistics for SNP effects
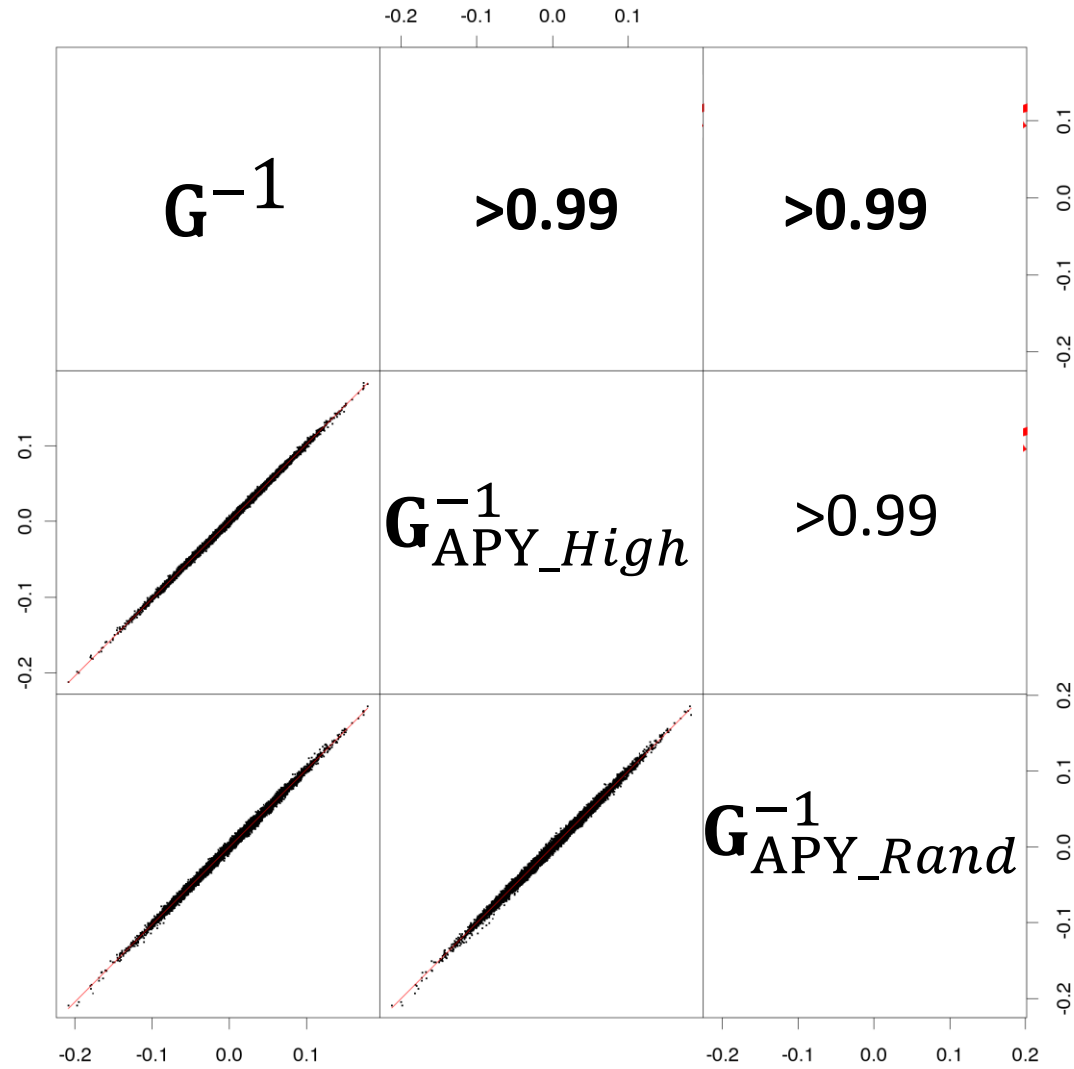
Statistics for SNP effects

Statistics for $\mathbf{Z}\hat{a}$

Fine-tuning indirect predictions from ssGBLUP

Understanding genetic and genomic bases

- Base of BLUP: *founders of the pedigree*

- Base of GBLUP: *genotyped* animals

- Base of SSGBLUP: Vitezica et al. (2011) modelled as a mean in genotyped animals

  - $p(\boldsymbol{u}_g) = N(\mathbf{1}\mu, \mathbf{G})$

  - μ = (Pedigree base) – (Genomic base)

Fine-tuning indirect predictions from ssGBLUP

1) Formula in Legarra (2017)

$$\hat{\boldsymbol{u}}_{ip} = \hat{\mu} + 0.95\mathbf{Z}\hat{a} + 0.05\,\hat{\boldsymbol{u}}_{parents}$$

SNP and pedigree fractions

2) Double fitting

   a) fit a regression using genotyped animals in the evaluation

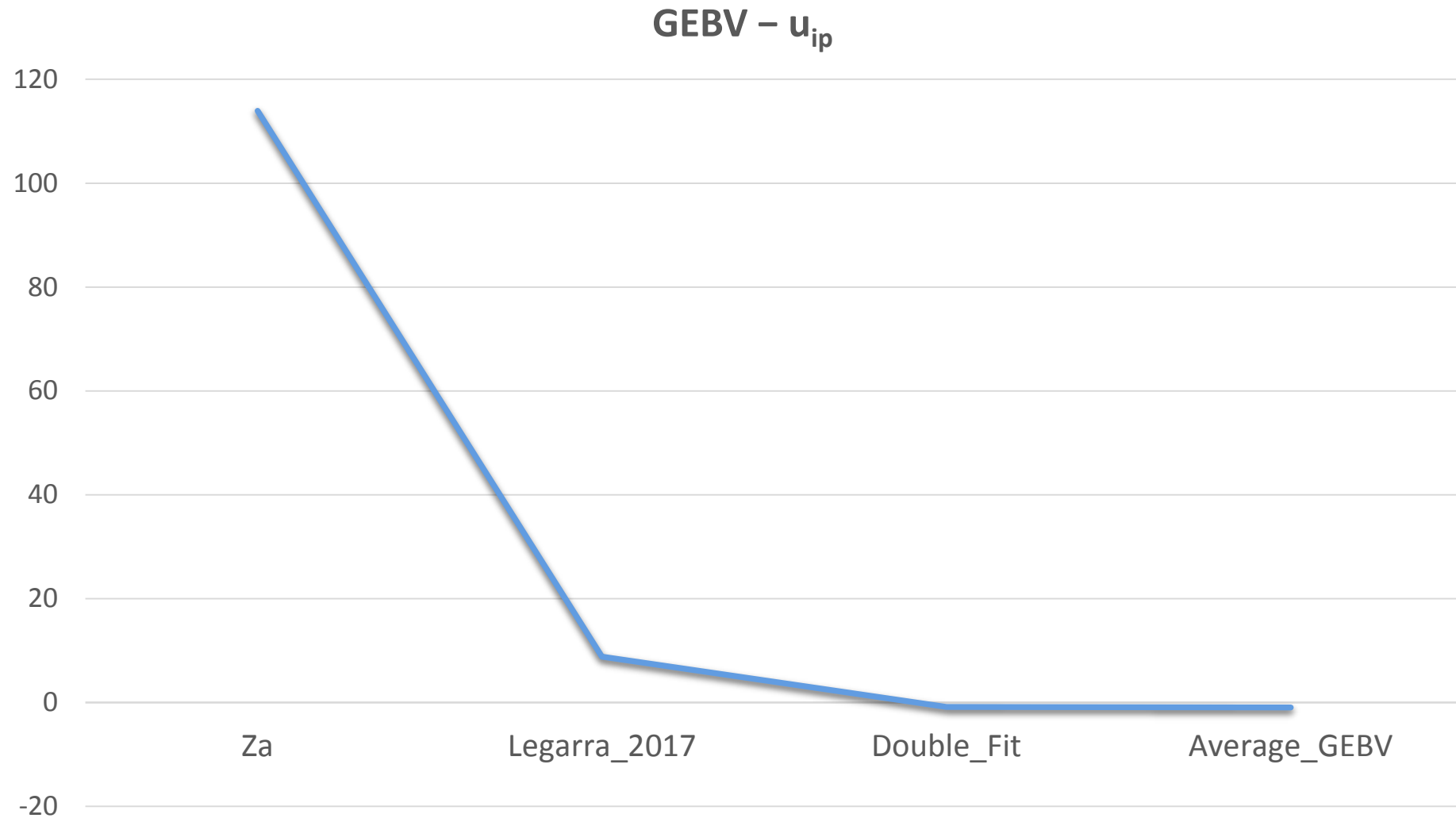 $$\mathrm{DGV}_{eval} = b_0 + b_1\mathbf{Z}\hat{a}$$

   b) apply regression for indirectly predicted animals

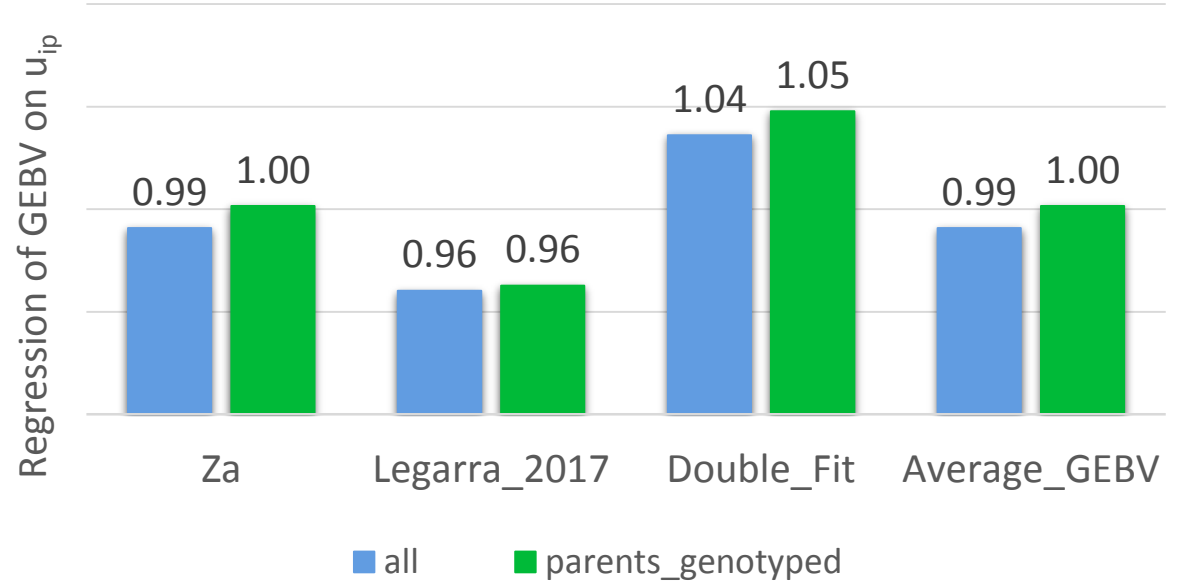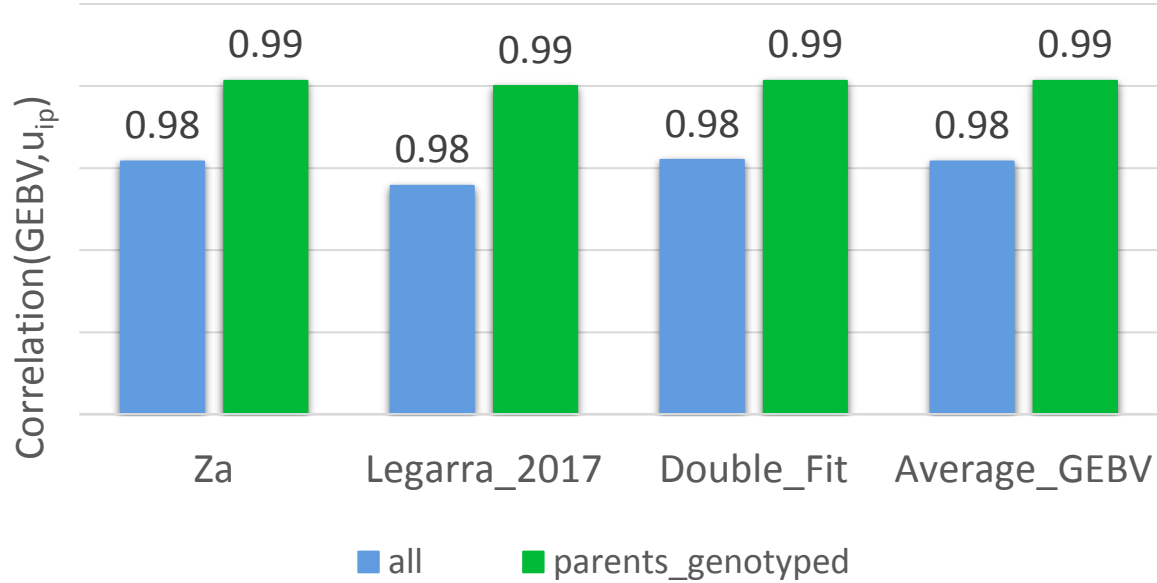 $$\hat{\boldsymbol{u}}_{ip} = b_0 + b_1\mathbf{Z}\hat{a}$$

3) Add average GEBV

$$\hat{\boldsymbol{u}}_{ip} = \overline{GEBV}_{eval} + \mathbf{Z}\hat{a}$$

Bias of indirect predictions

GEBV − u_{ip}

Correlation & Regression Coefficient

Fine-tuning indirect predictions in ssGBLUP

$$\mathrm{E}(\boldsymbol{u}|\boldsymbol{a}) = \widehat{\boldsymbol{u}}|\widehat{\boldsymbol{a}} = \mu + \mathbf{Z}\frac{\mathbf{1}}{\mathbf{2}\sum\boldsymbol{p}(\mathbf{1}-\boldsymbol{p})}\left(\mathbf{I}\frac{\mathbf{1}}{\mathbf{2}\sum\boldsymbol{p}(\mathbf{1}-\boldsymbol{p})}\right)^{-\mathbf{1}}(\widehat{\boldsymbol{a}}-0)$$

$$\widehat{\boldsymbol{u}}|\widehat{\boldsymbol{a}} = \mu + \mathbf{Z}\widehat{\boldsymbol{a}}$$

$$\approx$$

$$\widehat{\boldsymbol{u}}|\widehat{\boldsymbol{a}} = \overline{GEBV} + \mathbf{Z}\widehat{\boldsymbol{a}}$$

Final Remarks

- Indirect predictions are unbiased after corrections

  - Can be used as interim evaluation

- Indirect predictions based on core animals are slightly less accurate

  - Reduction in computing time (no $\mathbf{G}_{nc}^{-1}$ and $\mathbf{G}_{nn}^{-1}$)

- SNP effects from ssGBLUP may be useful for SNP MACE

Acknowledgements