



UNIVERSITY OF
GEORGIA

Indirect predictions based on SNP effects from ssGBLUP in large genotyped populations

Daniela Lourenco

A. Legarra, S. Tsuruta, Y. Masuda, T. Lawlor, I. Misztal

Why Indirect predictions (IP)?

- 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

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

SNP effects

GEBVs

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

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

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

$$\mathbf{GEBV}_{\text{young}} = w_1 \mathbf{PA} + w_2 \mathbf{GP} - w_3 \mathbf{PP}$$

$$\mathbf{GEBV}_{\text{young}} \approx \mathbf{GP}$$

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

$$\mathbf{GEBV}_{\text{young}} \approx \mathbf{GP} = \mathbf{Z}\hat{\mathbf{a}}$$

Problems with Indirect predictions

Genetic evaluation using single-step genomic best linear unbiased predictor in American Angus¹

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^{*}

© 2015 American Society of Animal Science. All rights reserved. J. Anim. Sci. 2015.93:2653–2662
doi:10.2527/jas2014-8836

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

$$\text{Avg}(\widehat{\text{GEBV}}) \approx 100$$



$$\text{Avg}(\mathbf{Z}\hat{\mathbf{a}}) \approx 0$$

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

- US Holstein 2014 type data
 - 8.3M animals in pedigree
 - No UPG
 - 9.2M Udder Depth (UD)
 - $h^2 = 0.33$
 - 9.2M Foot Angle (FA)
 - $h^2 = 0.11$
 - 105k genotyped
 - Training: 100k
 - Validation: 5k
- Complete
 - Phenotypes up to 2010
 - Genotypes up to 2010 (105k)
- Reduced
 - Phenotypes up to 2010
 - NO Genotypes for validation (100k)
- SNP effects and IP from Reduced
 - Compare with GEBV from Complete

Accuracy of SNP effects

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

\mathbf{G}^{-1}

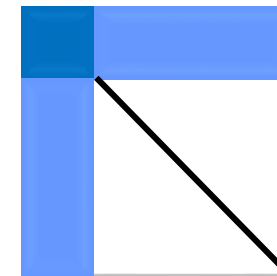


100k

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

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

APY \mathbf{G}^{-1}



15k core
85k noncore

$$\hat{\mathbf{a}}_{\mathbf{G}_{cc}^{-1}H} = \lambda \mathbf{D} \mathbf{Z}' \mathbf{G}_{cc_high_reliability}^{-1} \hat{\mathbf{u}}_{APY}$$

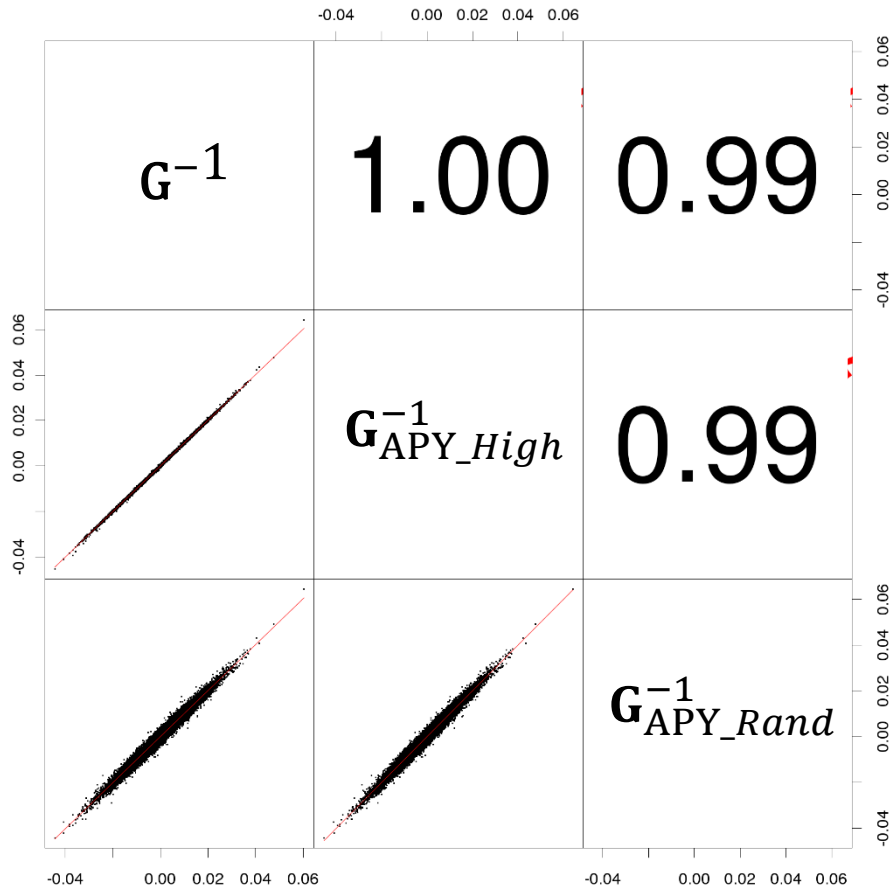
$$\hat{\mathbf{a}}_{\mathbf{G}_{cc}^{-1}R} = \lambda \mathbf{D} \mathbf{Z}' \mathbf{G}_{cc_random}^{-1} \hat{\mathbf{u}}_{APY}$$

\mathbf{G}^{-1} core

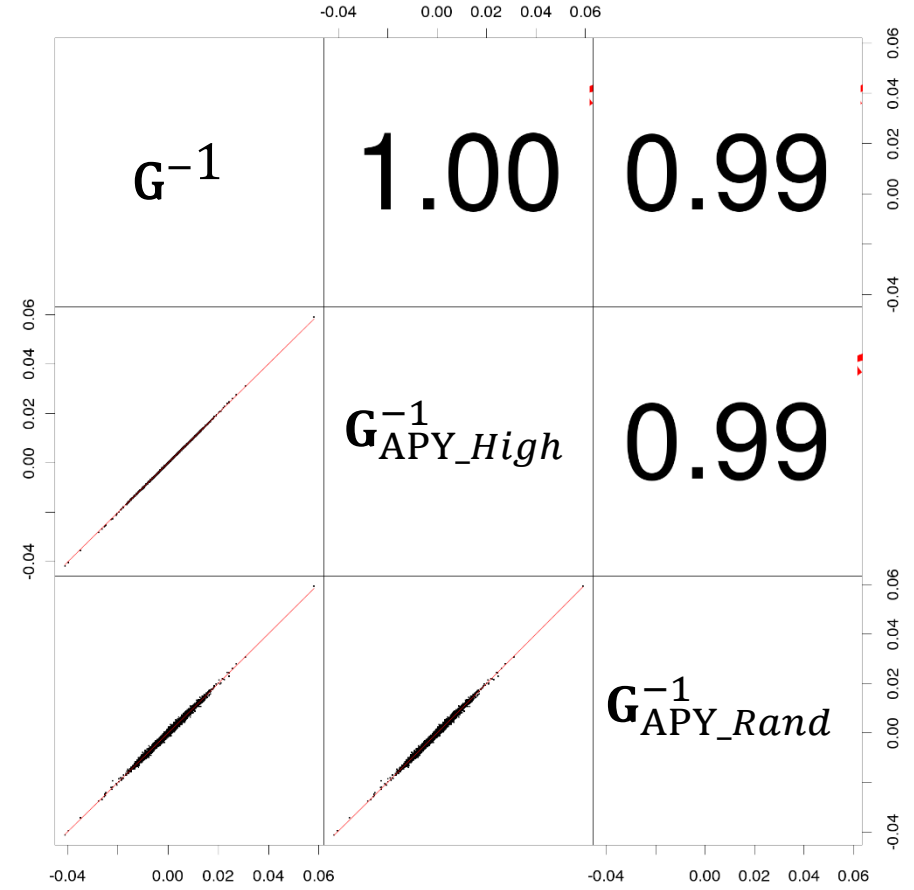


15k

Statistics for SNP effects - \mathbf{G}_{APY}^{-1}

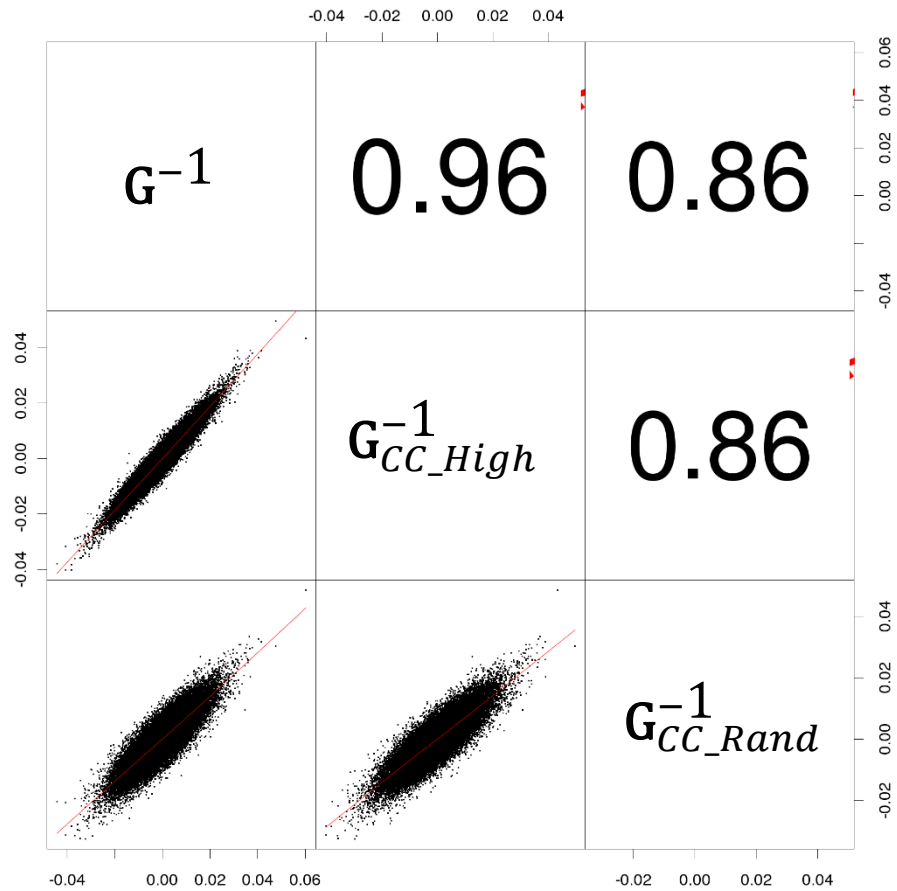


Udder Depth

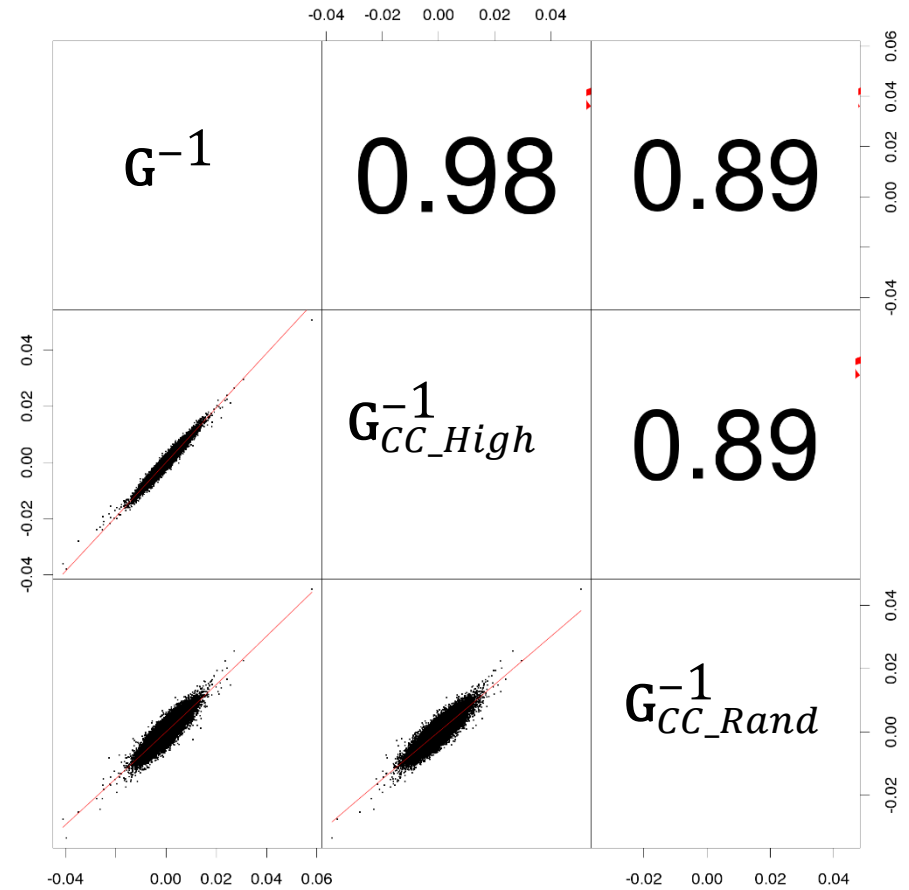


Foot Angle

Statistics for SNP effects - \mathbf{G}_{CC}^{-1}

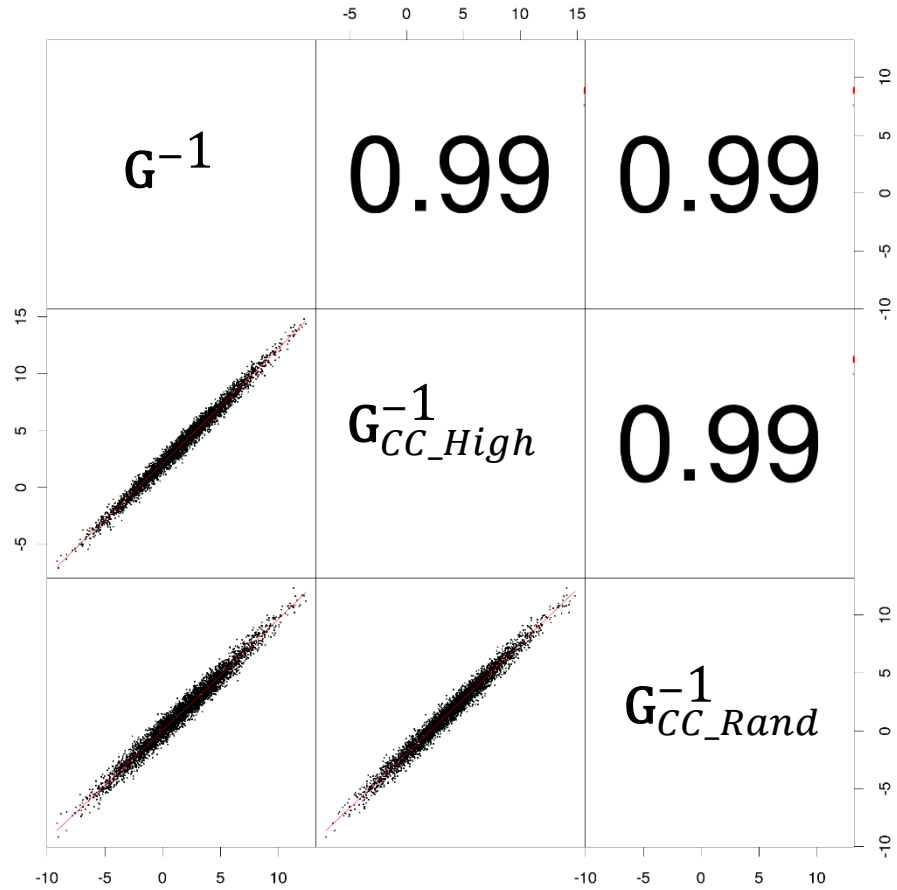


Udder Depth

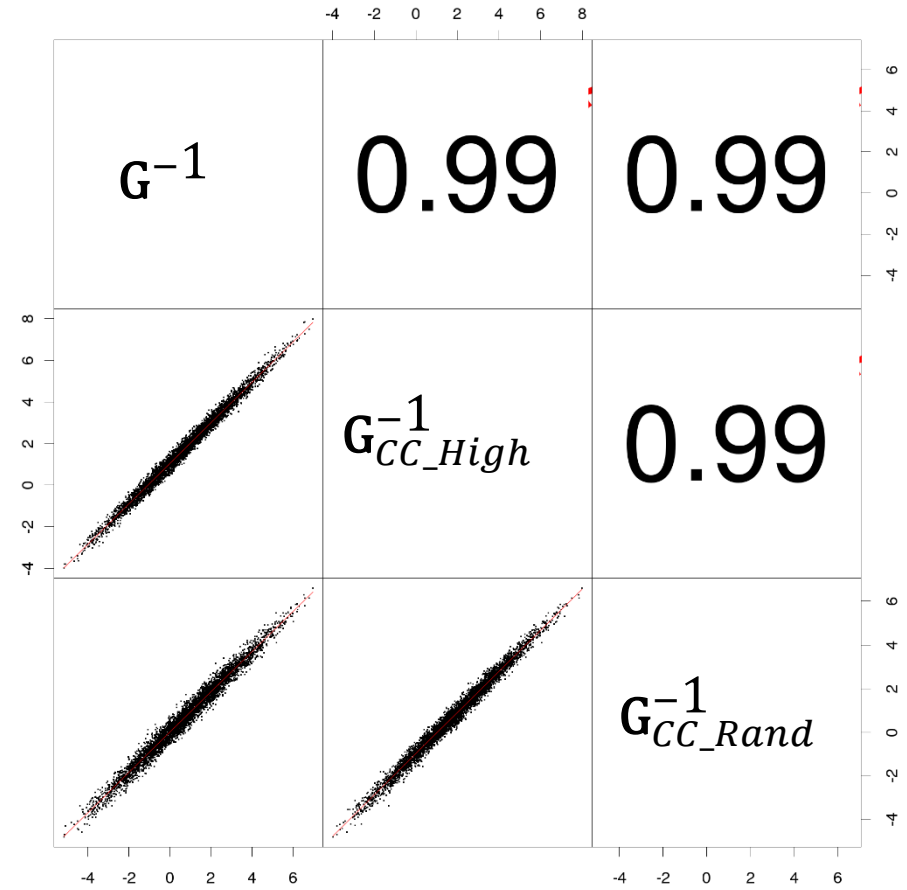


Foot Angle

Statistics for $\mathbf{Z}\hat{\mathbf{a}} - \mathbf{G}_{CC}^{-1}$



Udder Depth



Foot Angle

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) modeled as a mean for genotyped
 - $p(\mathbf{u}_g) = N(\mathbf{1}\mu, \mathbf{G})$
 - $\mu = (\text{Pedigree base}) - (\text{Genomic base})$

Fine-tuning indirect predictions from ssGBLUP

1) Formula in Legarra (2017)

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

SNP and
pedigree
fractions

2) Double fitting

a) fit a regression using genotyped animals in the evaluation

$$GEBV_{eval} = b_0 + b_1\mathbf{Z}\hat{\mathbf{a}}$$

b) apply regression for indirectly predicted animals

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

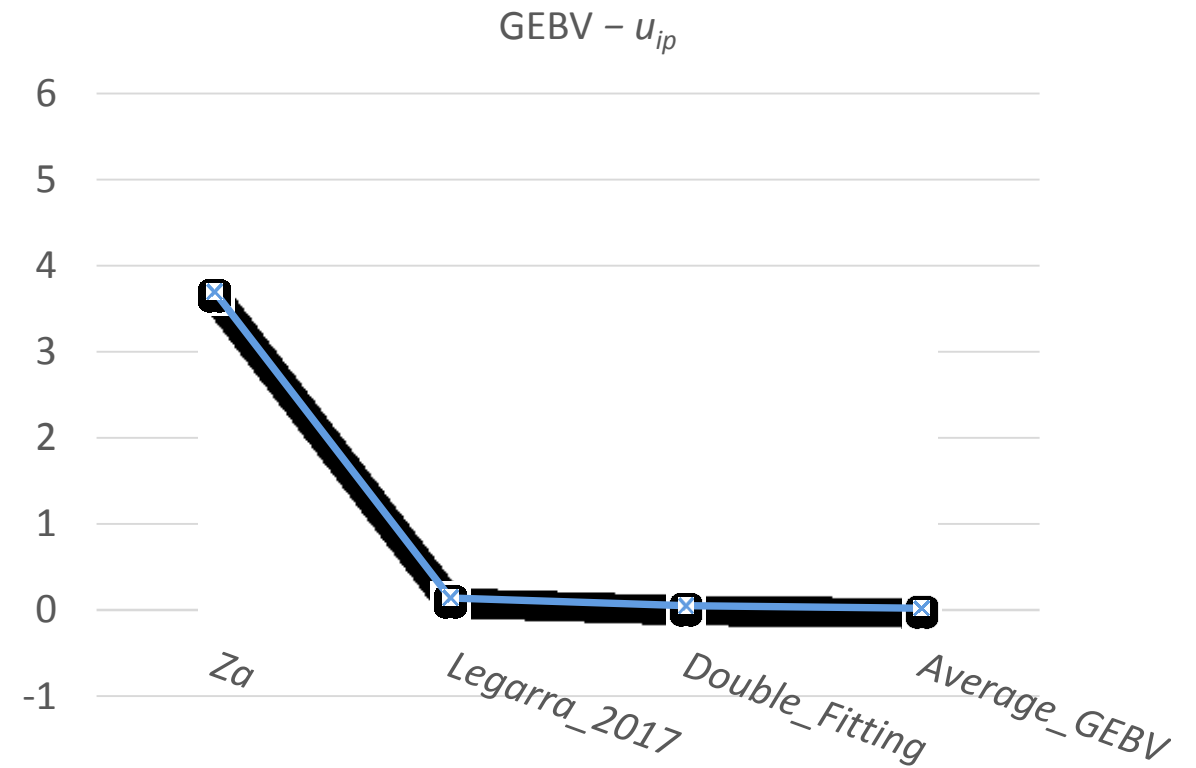
3) Add average GEBV

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

Bias of indirect predictions

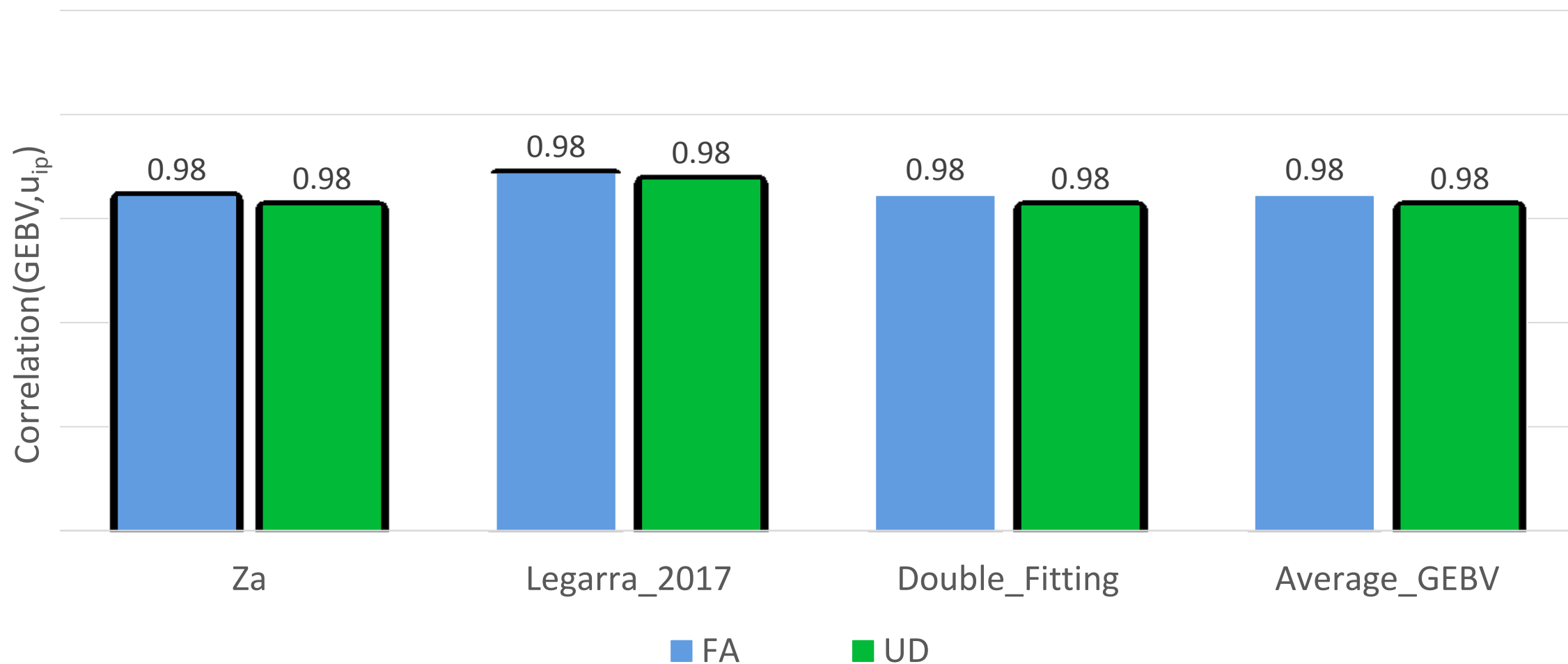


Udder Depth



Foot Angle

Correlation between GEBV and indirect predictions



Fine-tuning indirect predictions in ssGBLUP

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

$$E(\hat{\mathbf{u}}|\hat{\mathbf{a}}) = \mu + \mathbf{Z}\hat{\mathbf{a}}$$

\approx

$$E(\hat{\mathbf{u}}|\hat{\mathbf{a}}) = \overline{GEBV} + \mathbf{Z}\hat{\mathbf{a}}$$

Final Remarks

- SNP effects can be calculated based on APY \mathbf{G}^{-1} or core \mathbf{G}^{-1}
 - Slightly less accurate with core
- Indirect predictions based on core animals are accurate
 - Reduction in computing time compared to \mathbf{G}^{-1}
 - Similar computing time as APY \mathbf{G}^{-1}
- Indirect predictions are unbiased after corrections
 - Average GEBV, Legarra (2017) or double fitting
 - Can be used as interim evaluation

Acknowledgements

