Comparison of algorithms for approximation of accuracies in ssGBLUP

Pedro Ramos, A. Garcia, K. Retallik, M. Bermann, I. Misztal, D. Lourenco

ASAS- July 2023

UNIVERSITY OF GEORGIA

AGI

Introduction

	Maternal						
	HP Acc % Daus	CEM Acc % Daus	Milk Acc %	MkH MkD	MW Acc % Prog	MH Acc % Prog	$EN %
EPD:	+19.4	+10	+19		+59	+.4	-11
ACC:	.29	.34	.35		.43	.45	
	2%	35%	90%		55%	45%	45%

	Production							
	CED Acc % Prog	BW Acc % Prog	WW Acc % Prog	YW Acc % Prog	RADG Acc % Prog	DMI Acc % Prog	YH Acc % Prog	SC Acc % Prog
EPD:	+5	+.8	+78	+140	+.27	+1.78	+.7	-.03
ACC:	.56	.86	.81	.65	.40	.40	.53	.60
	60%	40%	15%	15%	30%	95%	30%	95%
	41	361	260	64				15

Introduction

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \mathbf{A^{-1}\alpha} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{C_{11}} & \mathbf{C_{12}} \\ \mathbf{C_{21}} & \mathbf{C_{22}} \end{bmatrix} \longrightarrow \mathbf{C^{-}} = \begin{bmatrix} \mathbf{C^{11}} & \mathbf{C^{12}} \\ \mathbf{C^{21}} & \mathbf{C^{22}} \end{bmatrix}$$

$$\mathrm{PEV} = \mathrm{Var}\,(\mathrm{g} - \hat{\mathrm{g}}) = \mathbf{C^{22}}\,\boldsymbol{\sigma_e^2}$$

$$BIF_{accuracy} = 1 - \sqrt{\frac{PEV}{(1 + F_i)\sigma_u^2}}$$

Introduction

- Complex Models:
- Single-Step GBLUP:
  - Genomic and pedigree information combined ($\mathbf{H}^{-1}$) in a single method

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \boldsymbol{H^{-1}}\,\boldsymbol{\alpha} \end{bmatrix} \longrightarrow \mathbf{C}^{-} = \begin{bmatrix} \end{bmatrix}$$

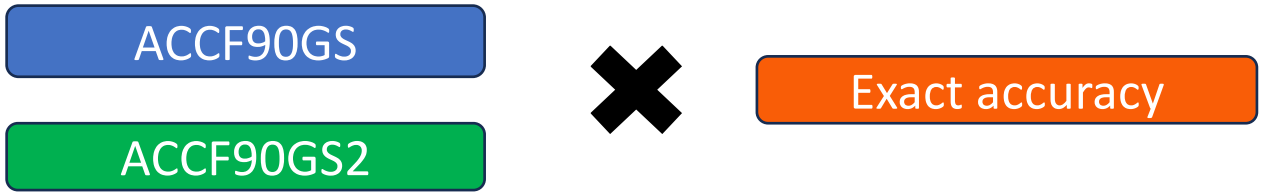
- Methods of approximating accuracy
  - Obtain the PEV without setting up the mixed model equations
  - Provide a feasible calculation of the genomic contribution

Objectives

- Compare two algorithms for approximation of accuracies:

1) Compare approximated accuracies between algorithms

2) Compare approximated and accuracies from the inverse of MME



3) Impact of adding new genotyped animals in the approximated accuracies

ACCF90GS

Genomic

$$\left\{\mathbf{Z}'\mathbf{Z}+\lambda\mathbf{A}^{-1}+\lambda\begin{bmatrix}0 & 0\\ 0 & \mathbf{G}^{-1}-\mathbf{A}_{22}^{-1}\end{bmatrix}\right\}$$

d_i^r d_i^p d_i^g

Adapted from Misztal and Wiggans (1988):

$$LHS_{uu}^{ii} = {}^{1}\Big/{}_{(\lambda + d_i^r + d_i^p + d_i^g)}$$

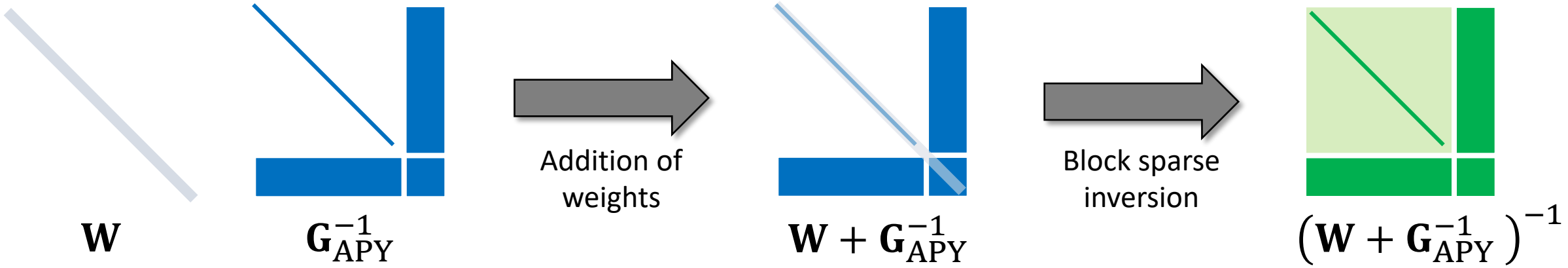
ACCF90GS

$$d_i^g = \alpha * \left[\overline{REL} + (1 - g_{ii}) + Z * \left(\overline{REL - REL_{pa}}\right)\right]$$

- α = ratio of residual variance to animal genetic variance
- REL = average of reliability based on pedigree
- (1-gii) = genomic information for the animal *i*
- Z = adjustment factor defined as: $\left(\frac{\text{Genotyped individuals}}{100} * 0.1\right)$
- REL_{pa} = average reliability based on the pedigree of genotyped animals with phenotype
- $REL - REL_{pa}$ = contribution from phenotypes of genotyped animals

ACCF90GS2

- Block sparse inversion of $\mathbf{G}^{-1}$computed by the APY method:



$$diag(\mathbf{Z'Z} - \mathbf{Z'X}(\mathbf{X'X})^{-1}\mathbf{X'Z} + \mathbf{G}_{APY}^{-1})^{-1} \approx diag\left(\mathbf{W} + \mathbf{G}_{\mathrm{APY}}^{-1}\right)^{-1}$$

Data sets

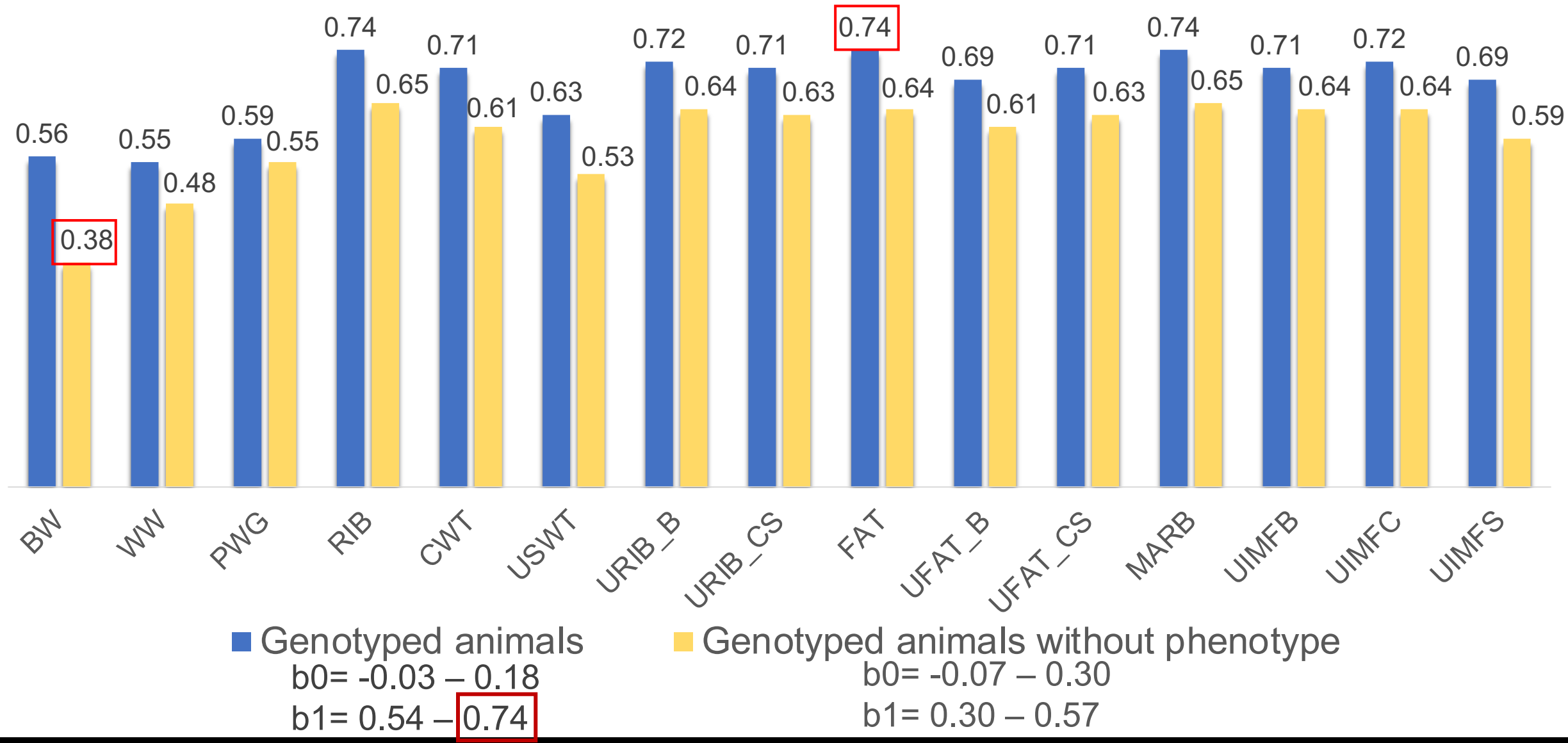
- Dataset:
  - 9.7M Birth weight (BW)
  - 10.1M Weaning weight (WW)
  - 4.9M Postweaning gain (PWG)
  - 132k Carcass weight (CWT)
  - 132k Fat (FAT)
  - 132k Marbling (MARB)
  - 132k Ribeye area (RIB)
  - 2.2M Ultrasound fat (UFAT)
  - 2.2M Ultrasound Ribeye (URIB)
  - 2.6M Weight at ultrasound time (USWT)
  - 2.2M Ultrasound intramuscular fat (UIMF)

  - 1.2M Genotyped animals
  - 4.3M – 12.5M animals in pedigree

- Subset of data:
- 31k – 132k records per trait
- 177k -534k animals in pedigree
- 30k – 50k Genotyped animals

Results

Correlation Betwen Algorithms

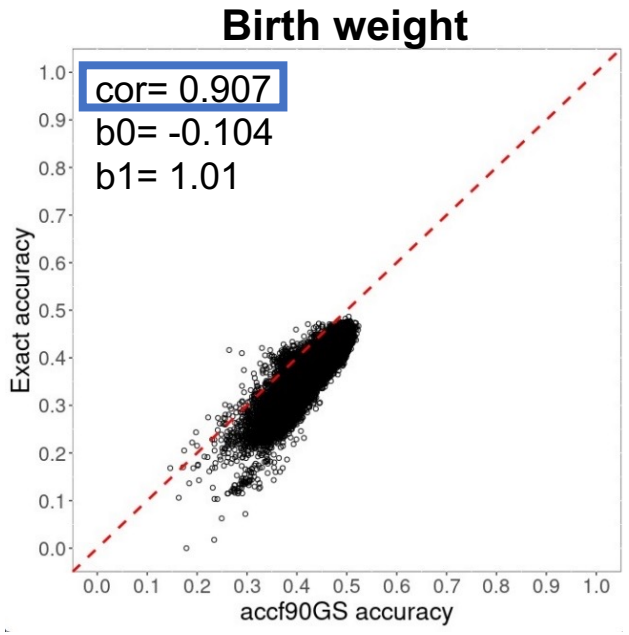


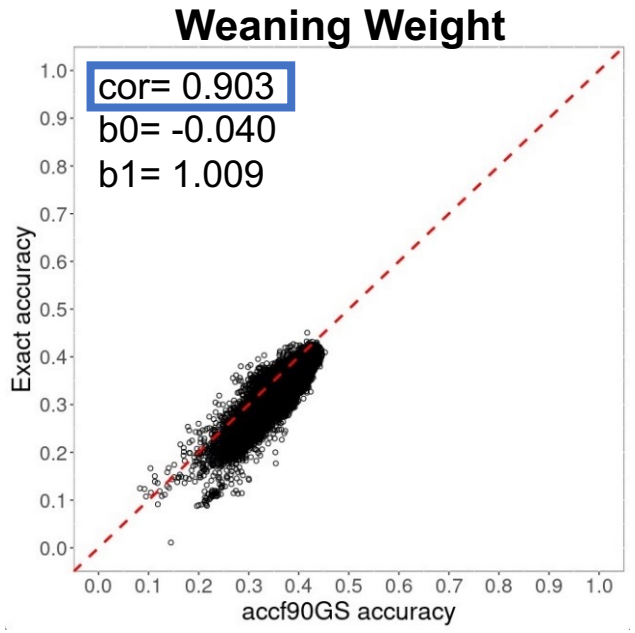
b0= -0.03 – 0.18
b1= 0.54 – 0.74

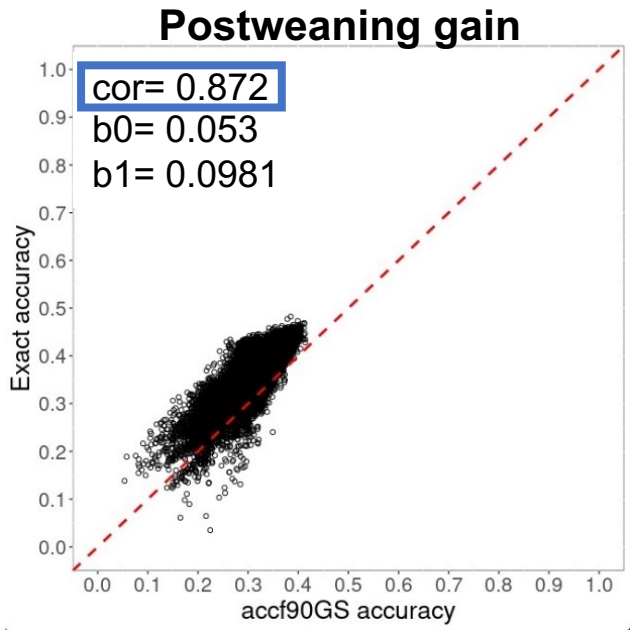
b0= -0.07 – 0.30
b1= 0.30 – 0.57

Comparison between approximated and exact accuracies

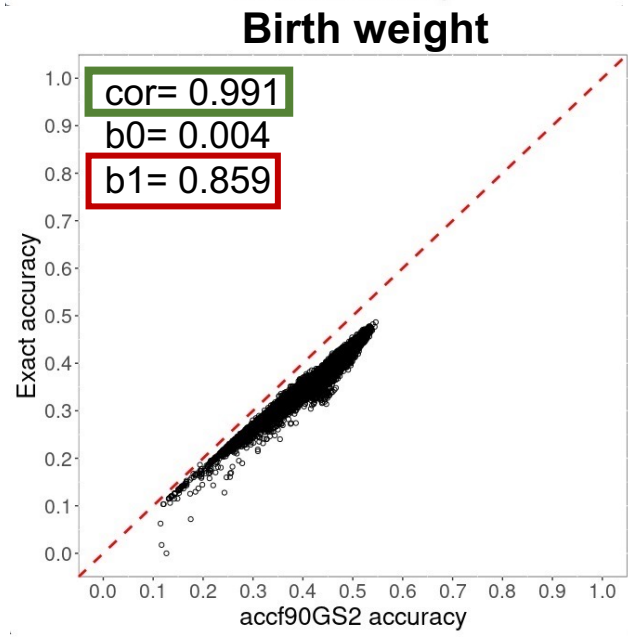
ACCF90GS:

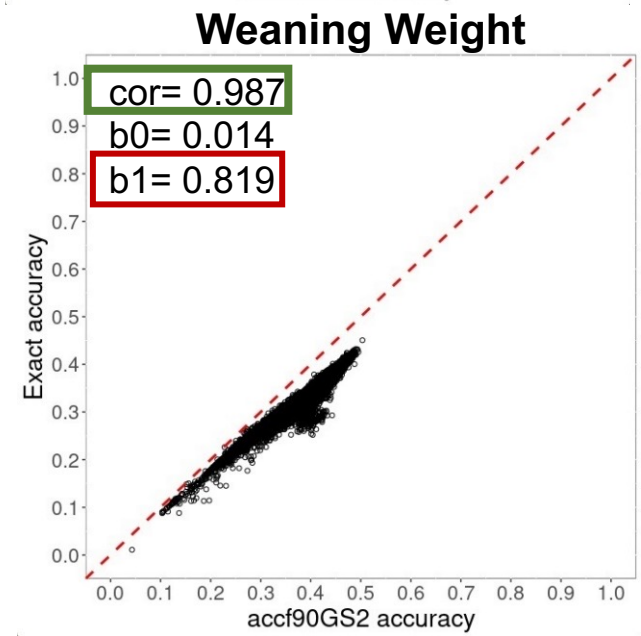


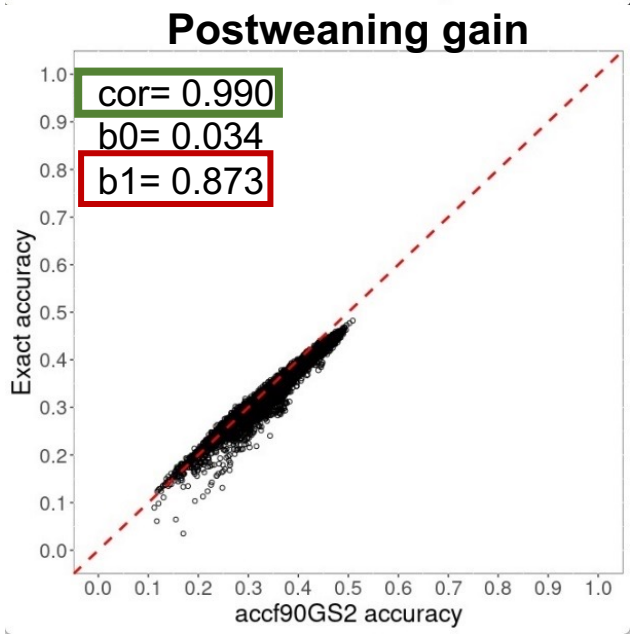




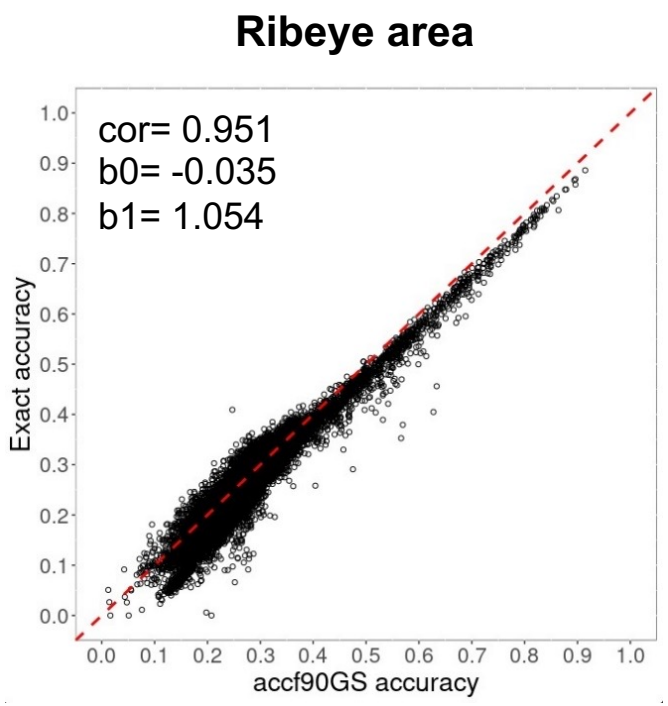
ACCF90GS2:

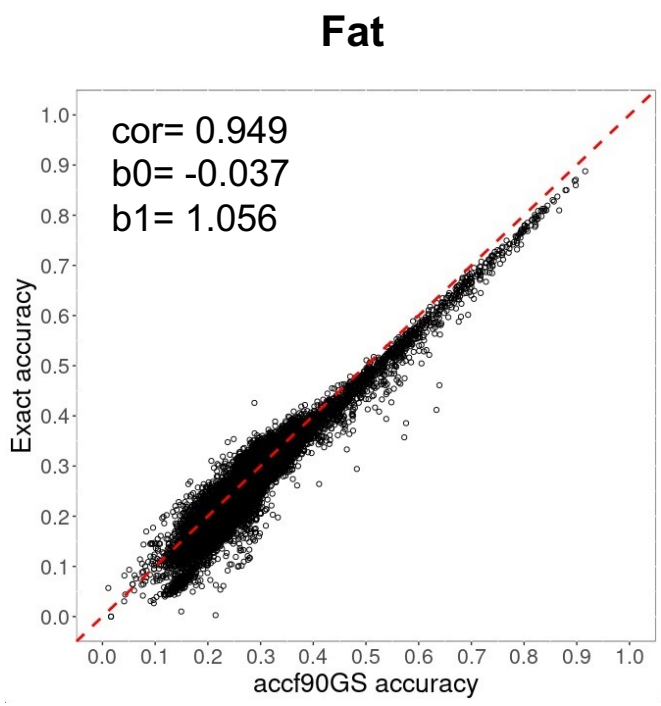


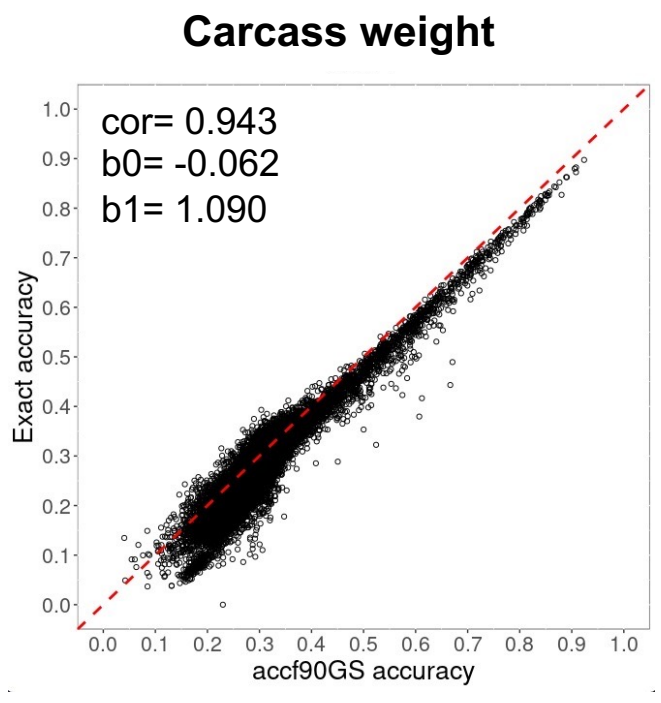




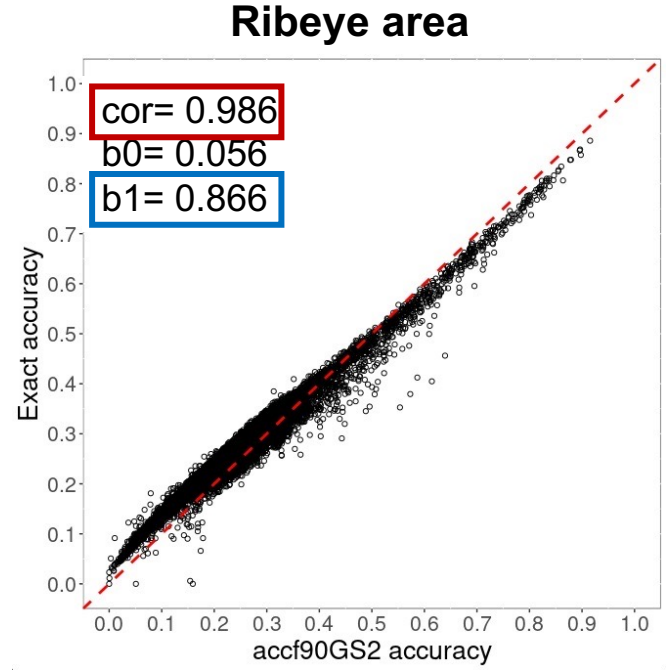
ACCF90GS:

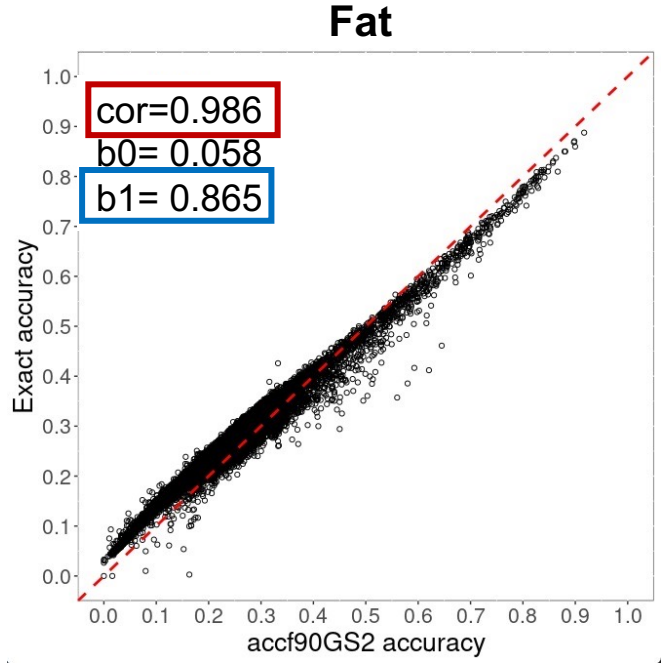


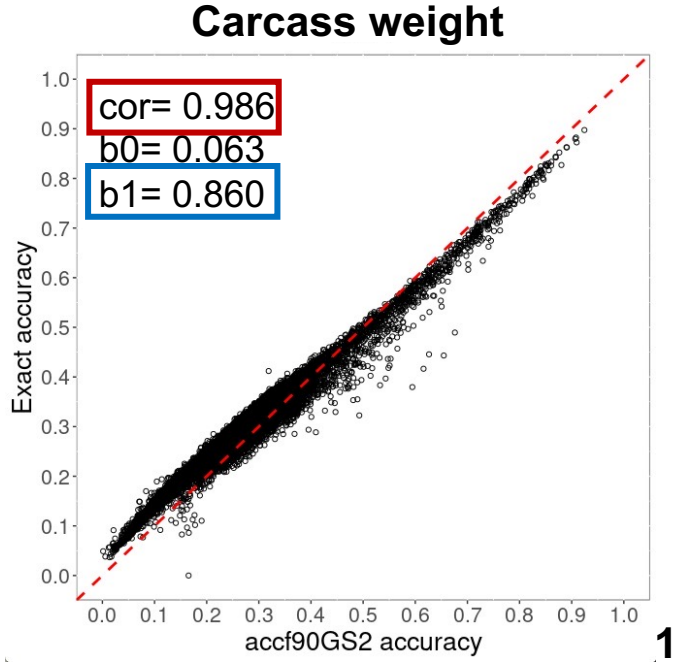




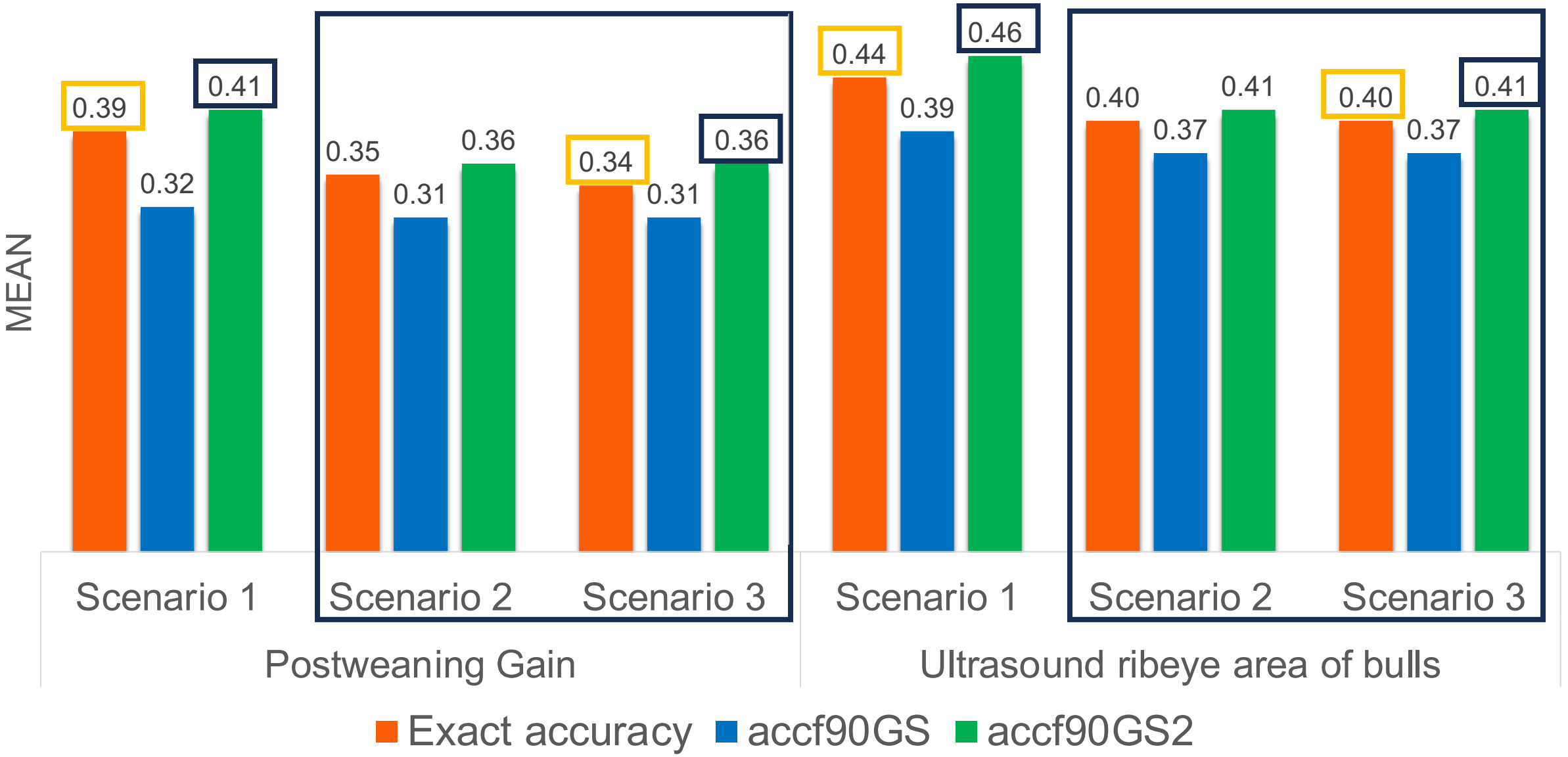
ACCF90GS2:







Evaluation of the impact of adding genotyped animals in the approximated accuracies



Scenario 1= inclusion of 10,000 genotyped animals with phenotype; Scenario 2= inclusion of 10,000 genotyped animals without phenotype; Scenario 3= no inclusion of animals.

Take home messages

- Approximated accuracies from accf90GS2 are closer to the exact accuracies
- More suitable for approximating accuracies in large-scale routine evaluations
- Computing time may still be challenging
- Adding genotyped animals without phenotypes slightly increases overall accuracy
- Largest impact on the mean is when adding genotyped animals with phenotypes

Acknowledgments

UNIVERSITY OF GEORGIA

ANGUS
THE BUSINESS BREED
AMERICAN ANGUS ASSOCIATION
3201 Frederick Avenue • St. Joseph, MO 64506 • (816) 383-5100 • Fax (816) 233-9703 • E-mail: angus@angus.org

THANK YOU!

Pedro Vital Brasil Ramos

pv11222@uga.edu