

¹Andre Garcia, ¹Y. Masuda, ²S. Miller, ¹I. Misztal, ¹D. Lourenco

¹ University of Georgia – Animal and Dairy Science department, Athens GA 30605, USA

² Angus Genetics Inc., St. Joseph MO 64506, USA

Introduction

- Rapid increase of the number of genotyped animals
- APY algorithm can be used to compute G^{-1} in GBLUP and ssGBLUP for large-scale genomic data
- Do we need to use all genotyped animals?
- Indirect Predictions (IP) can be used for:
 - Interim evaluations
 - Predictions for animals not in the main evaluation

Objectives

1. Evaluate the quality of indirect predictions when the number of genotyped animals increase
2. Investigate how many genotyped animals are needed to obtain reliable IP

Materials and Methods

- American Angus Association data
- Birth weight, weaning weight, and post weaning gain

Objective 1

- Genotyped animals were divided in 3 groups by year of birth
 - Up to 2013: 114,937
 - Up to 2014: 183,847
 - Up to 2015: 280,506

- GBLUP model with APY
- BLUPf90 family of programs

- G inverse with APY

$$G = \begin{bmatrix} G_{cc} & G_{cn} \\ G_{nc} & G_{nn} \end{bmatrix} \quad \begin{array}{l} c = \text{core} \\ n = \text{non-core} \end{array}$$

$$G_{APY}^{-1} = \begin{bmatrix} G_{cc}^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -G_{cc}^{-1}G_{cn} \\ I \end{bmatrix} M_{nn}^{-1} \begin{bmatrix} -G_{nc}G_{cc}^{-1} & I \end{bmatrix}$$

$$m_{nn,i} = g_{ii} - G_{ic}G_{cc}^{-1}G_{ci}$$

- All genotypes used to calculate GEBV
- SNP effects were obtained using all genotyped animals:

$$\hat{a} = DZ'G_{APY}^{-1}\hat{u} \quad \Rightarrow \quad IP = Z\hat{a} \quad \Rightarrow \quad \text{All genotyped animals}$$

- SNP effects were obtained using only core animals:

$$\hat{a}_{core} = DZ'G_{core}^{-1}\hat{u}_{core} \quad \Rightarrow \quad IP = Z\hat{a}_{core} \quad \Rightarrow \quad \text{Subset of core animals}$$

- Number of core animals was computed as the number of eigenvalues of G explaining 99% of the variance
- Different sets of core animals (19k) randomly sampled:
 - Animals born up to 2013 (core13)
 - Re-sampled as more genotyped animals added (core14 and core15)

- Correlations between IP and GEBV were used to evaluate the quality of the predictions for each scenario

Objective 2

- Increasing number of genotyped animals to calculate SNP effects and IP (no APY)
- 500 to 40k animals

Results

- Results were similar for all traits, only weaning weight is presented

All genotyped animals:

When using all genotyped animals with APY, correlations between IP and GEBV were >0.99 for all traits and scenarios, regardless of the core choice

Subset of core animals (core13):

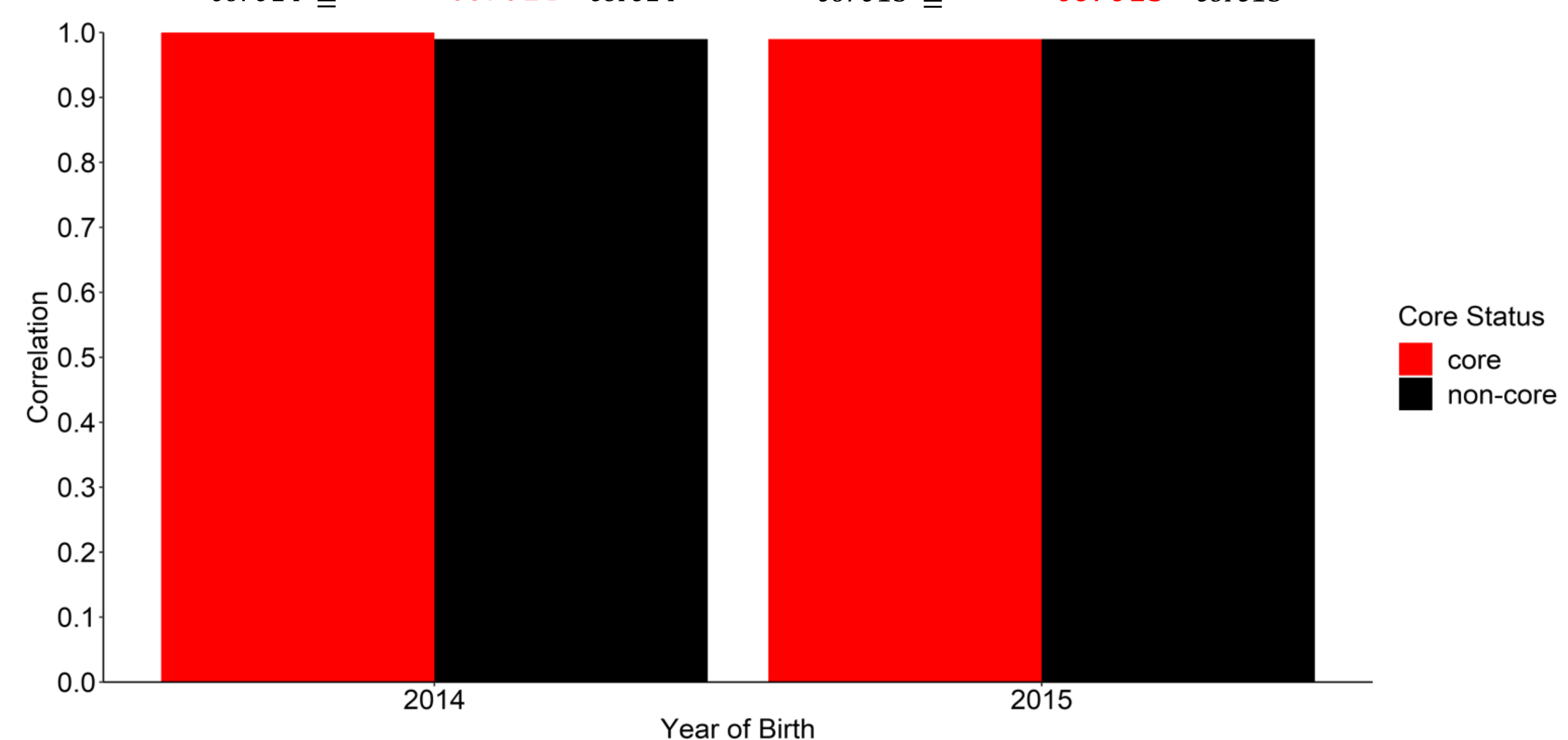
$$\hat{a}_{core13} = DZ'G_{core13}^{-1}\hat{u}_{core13}$$



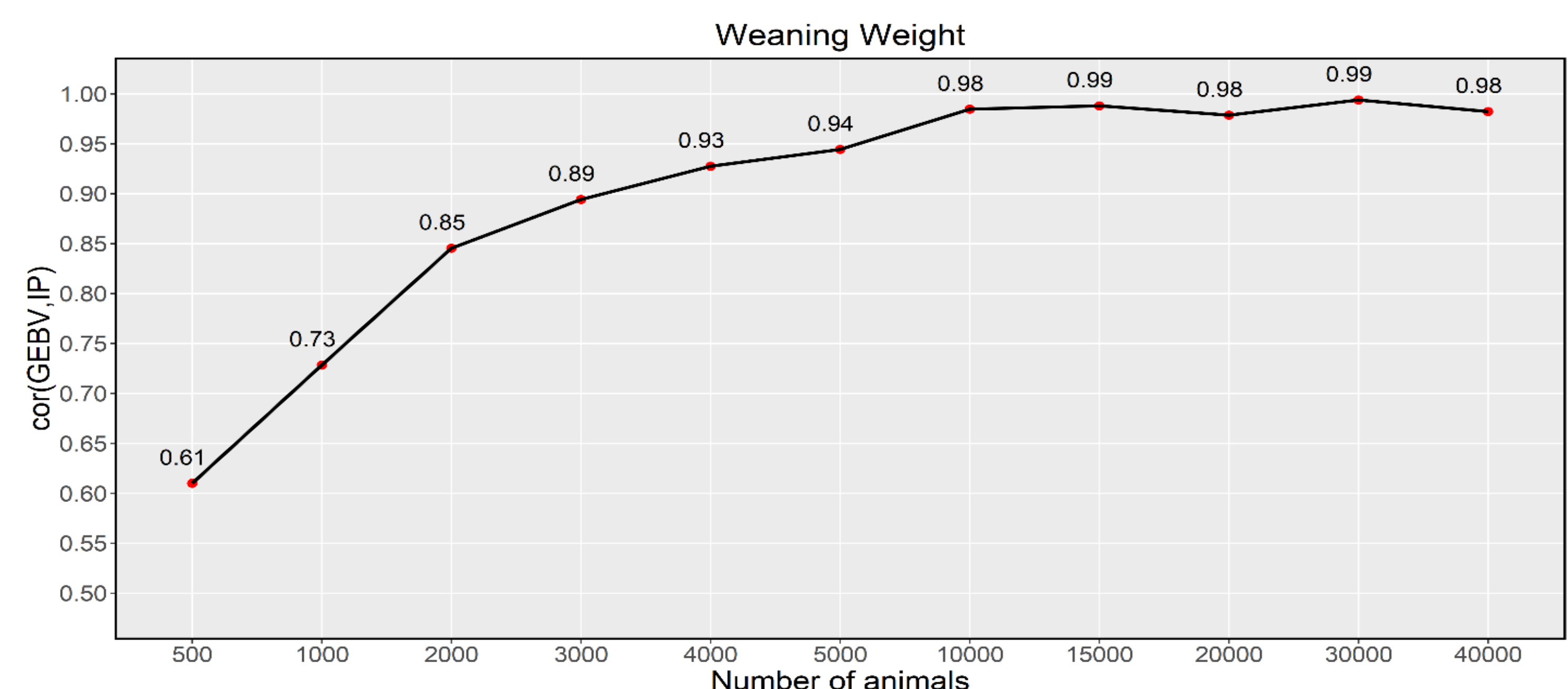
Subset of core animals (core14 and core 15):

$$\hat{a}_{core14} = DZ'G_{core14}^{-1}\hat{u}_{core14}$$

$$\hat{a}_{core15} = DZ'G_{core15}^{-1}\hat{u}_{core15}$$



Correlation between IP and GEBV as the number of genotyped animals increases



Conclusions

- IP are robust when all genotyped animals are used to compute SNP effects
- An updated core set is needed to obtain good IP when a subset of genotyped animals is used to compute SNP effects
- IP are less accurate if the number of genotyped animals is less than the dimensionality of the genomic information