

Indirect predictions based on SNP effects from GBLUP with increasing number of genotyped animals

Andre Garcia¹, Y. Masuda ¹, S. Miller², I. Misztal¹, D. Lourenco¹

¹ University of Georgia - Animal and dairy science

² Angus Genetics Inc. (AGI) - American Angus Association



Introduction

- Rapid increase of young genotyped animals into evaluations
- GBLUP or ssGBLUP

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



Becomes infeasible to invert directly

\mathbf{G}^{-1}



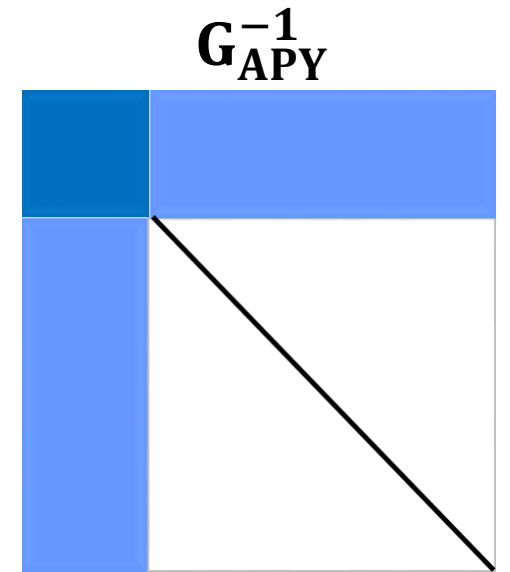
Introduction

- Algorithm for Proven and Young (APY)

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

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

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



Indirect predictions

Why do we may want to use IP?

- Interim evaluations
- Not all genotyped animals are included in evaluations
- Genomic predictions for non-registered animals

How to calculate IP?

First we need SNP effects

$$\hat{a} = \mathbf{DZ}'\mathbf{G}^{-1}\hat{u}$$

SNP effects

GEV



Then calculate IP

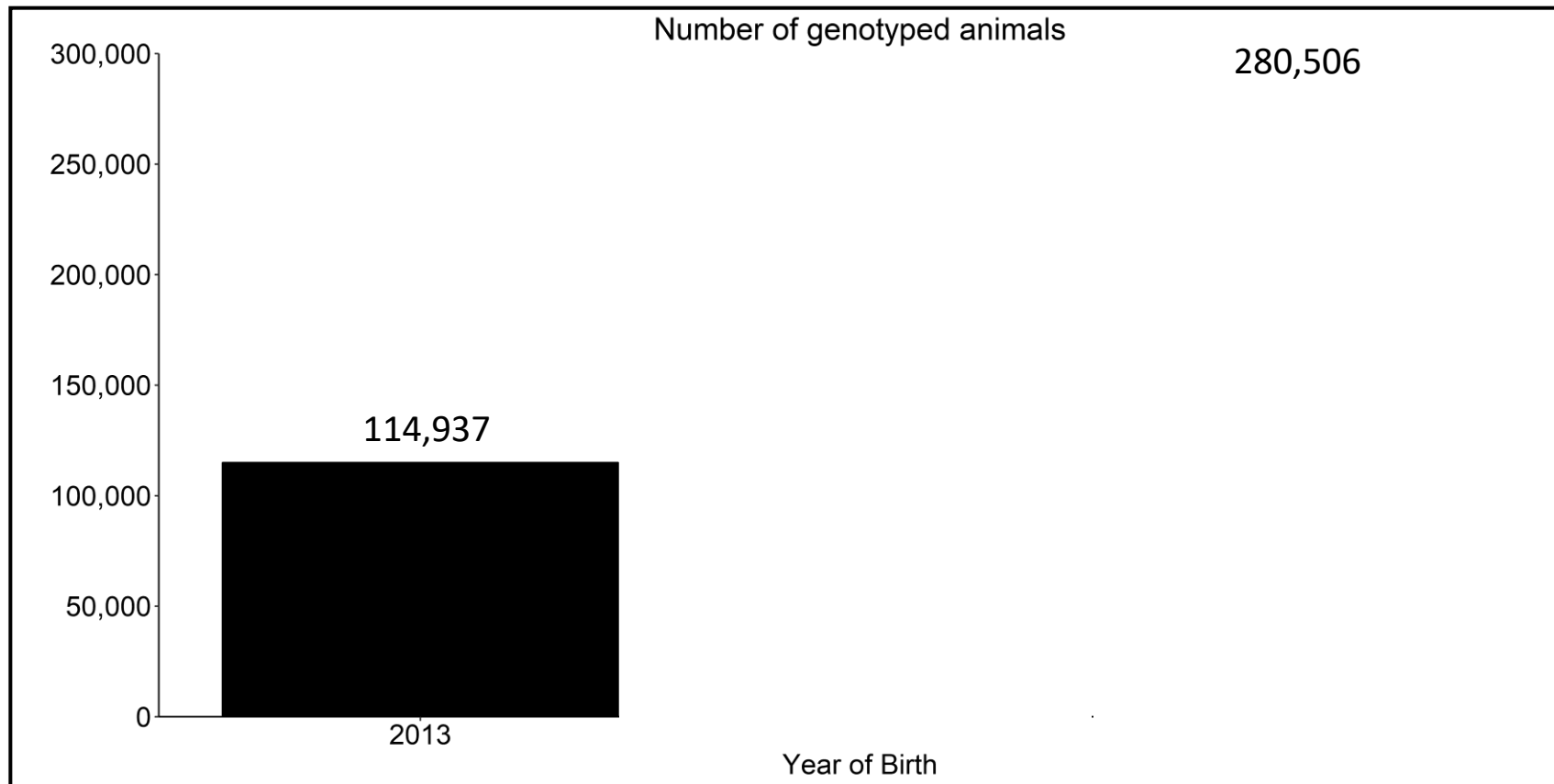
$$\text{IP} = \mathbf{Z}\hat{a}$$

VanRaden 2008
Stranden & Garrick 2009
Wang et al. 2012

What happens with IP when the number of genotyped animals keep increasing?

Data

- American Angus Association
- BW, WW and PWG



IP and core animals

- GBLUP with APY to get GEBV

- SNP effects :

$$\hat{\mathbf{a}} = DZ' G_{APY}^{-1} \hat{\mathbf{u}} \quad \longrightarrow \quad \text{All genotyped animals} \quad \bullet \quad \text{How do we choose core animals?}$$

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

- Correlations

- $\text{cor}(\text{GEBV}, \text{IP})$
- $\text{cor}(\text{GEBV}, \text{IP}_{core})$

IP and core animals

- Should reflect the dimensionality of G
 - Explain 98-99% of the variation in G
 - Often randomly selected
-
- 99% of variation = 19k core animals
-
1. Born up to 2013 (core13)
 2. Resampled as more genotypes were added (core14 and core15)

Results

$$\hat{a} = \mathbf{DZ}' \mathbf{G}_{APY}^{-1} \hat{u}$$

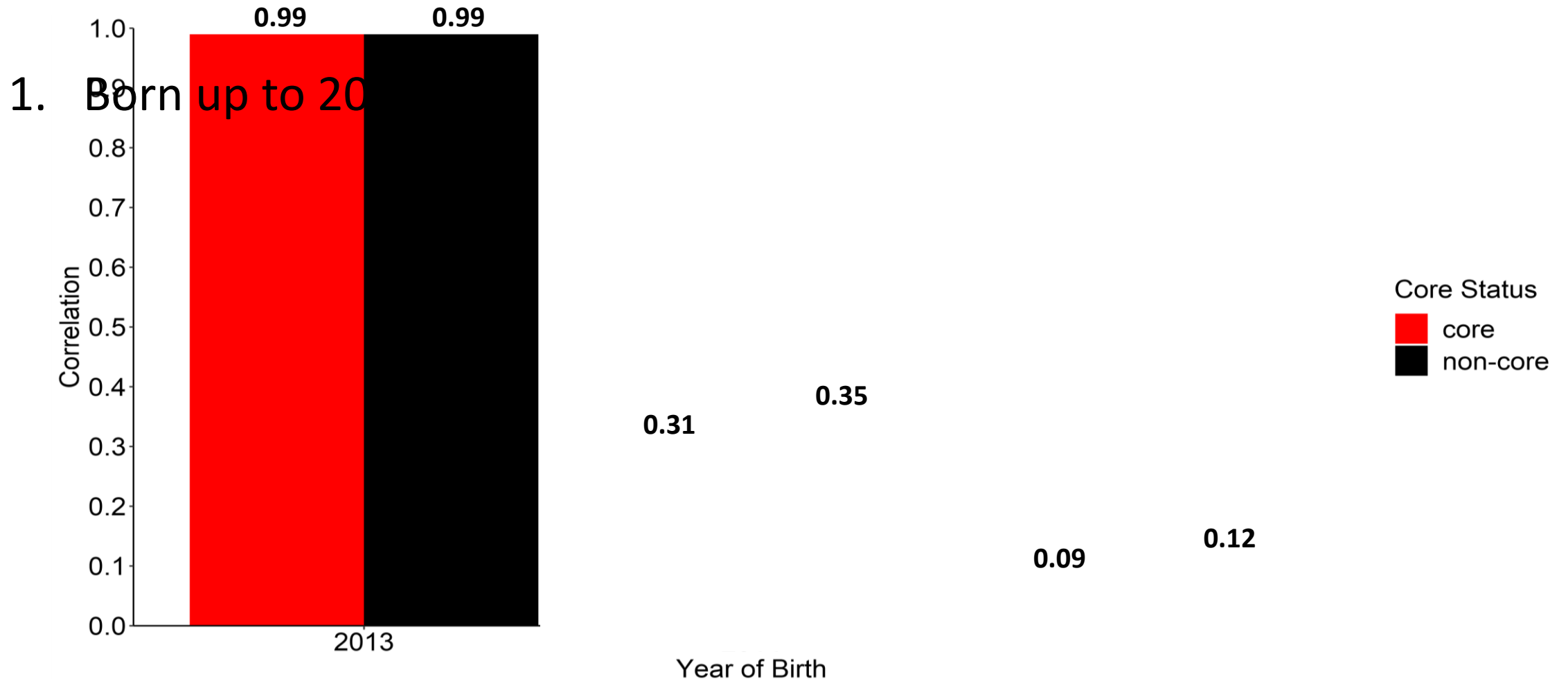
- All genotyped animals for each year class with APY
- Correlations between IP and GEBV ≥ 0.99
- All traits and all core definitions
- GEBV successfully retrieved from IP

Results

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

Weaning Weight

- Subset of core animals (19k)

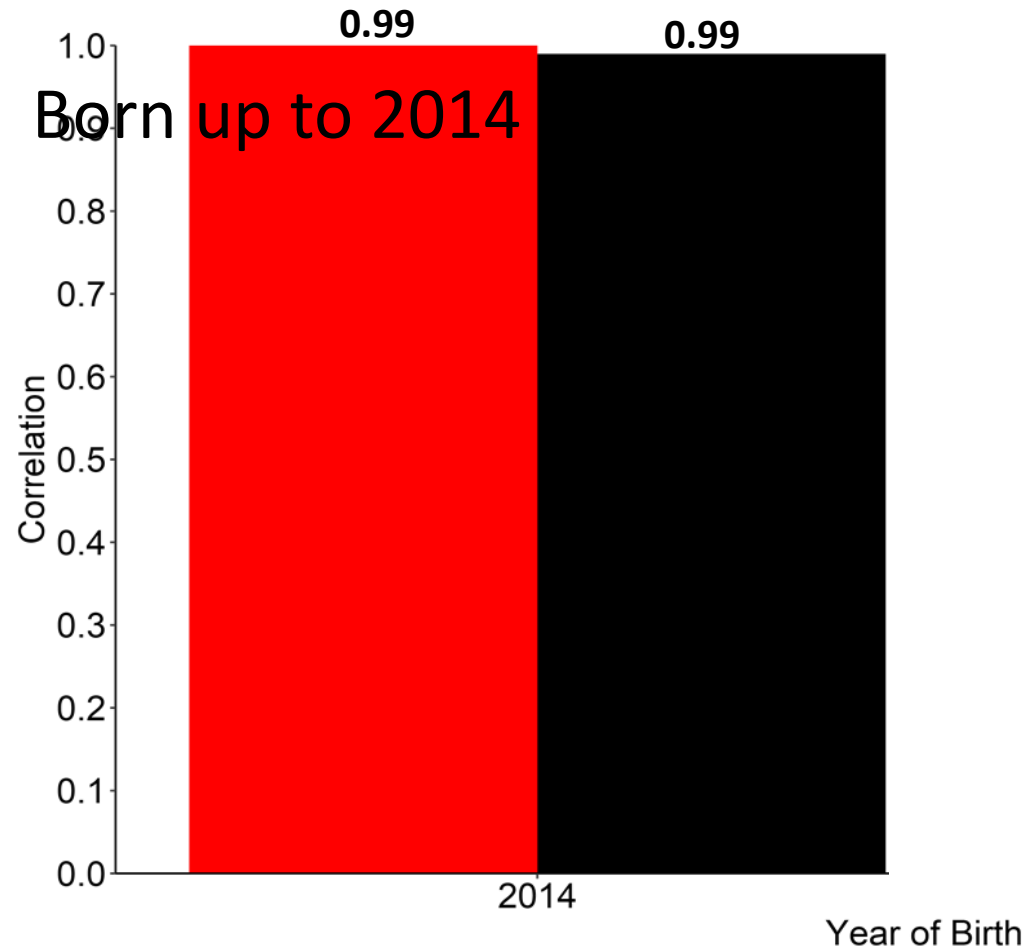


Results

Weaning Weight

- Subset of core animals (19k)

2. Born up to 2014



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

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

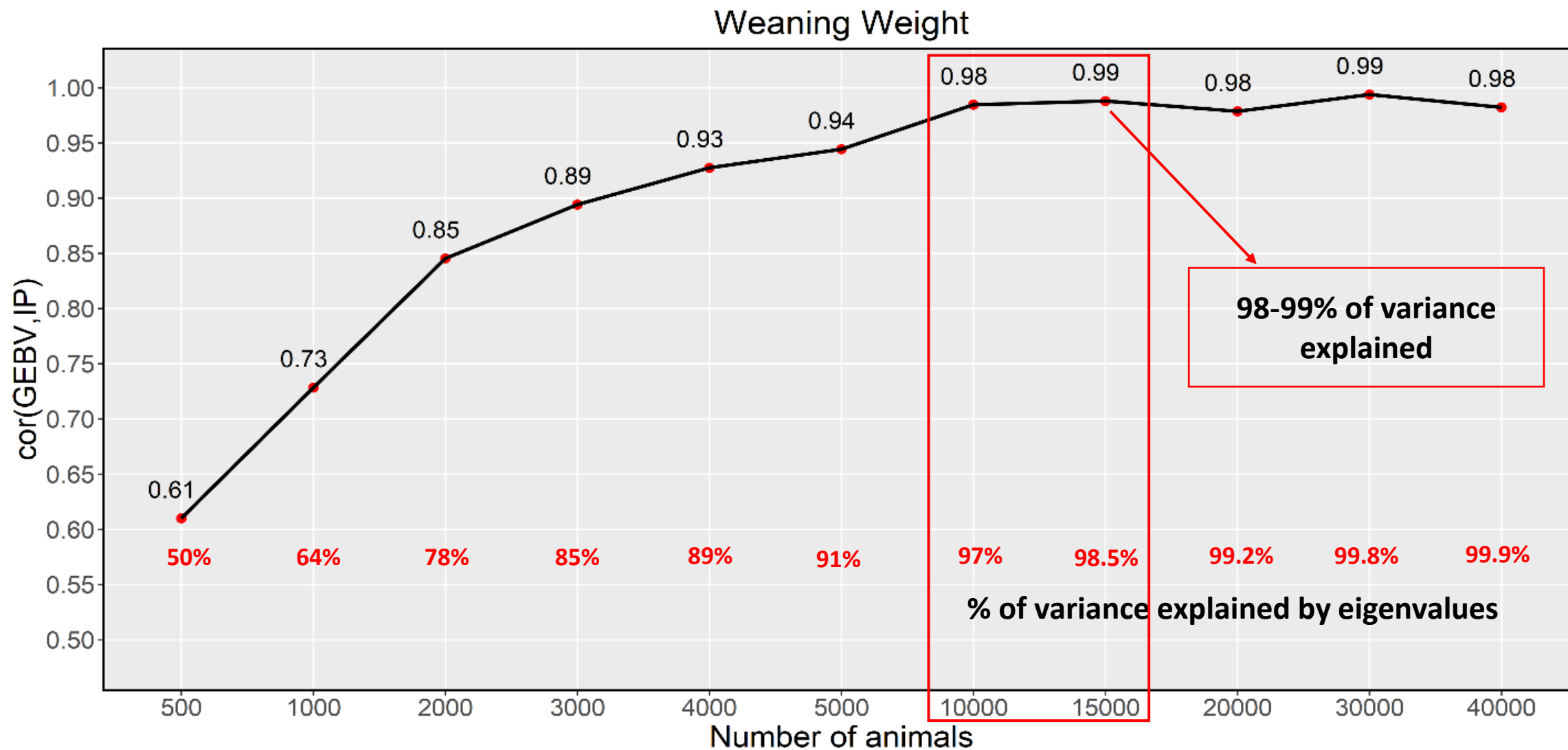
0.99

0.99

How many animals are needed to obtain reliable indirect predictions?

Data up to 2015

Correlation between GEBV and IP



Remarks

- Using all genotyped animals to calculate SNP effects:
 - Indirect predictions are robust
 - Independent of the choice of core animals for APY
- Using the subset of core animals:
 - Robust indirect predictions with updated set of core animals
- Using only a small portion of the data:
 - Indirect predictions are less accurate

Acknowledgements

