



Optimizing SNP weights in weighted single-step GBLUP for genomic prediction and genome-wide association in catfish

Andre Garcia¹, A. Legarra², I. Aguilar³, B. Bosworth⁴, G. Waldbieser⁴,
S. Tsuruta, I. Misztal¹, D.A.L. Lourenco¹

¹ The University of Georgia, ² INRA, France, ³ Instituto Nacional de Investigación Agropecuaria, Uruguay,
⁴ USDA-ARS Warmwater Aquaculture Research Unit

Single-step genomic evaluation

- MME for ssGBLUP

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + H^{-1}\lambda \end{bmatrix} \begin{bmatrix} \beta \\ a \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

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

- A = pedigree relationship matrix
- A_{22} = pedigree relationship matrix for genotyped animals
- G = genomic relationship matrix
- $\lambda = \sigma_e^2 / \sigma_a^2$

Weighted ssGBLUP

- Some traits might benefit from different SNP weighting
- Weighted ssGBLUP (WssGBLUP) with BLUPf90 family of programs:

1. $D = I$

2. $G = \frac{ZDZ'}{2\sum p_i(1-p_i)}$ (Wang et al., 2012)

3. GEBV from ssGBLUP

4. $\hat{u} = \delta DZ'G^{-1}\hat{a}$

5. $d_i = \hat{u}_i^2 2p_i(1-p_i)$

6. Iterate from 2

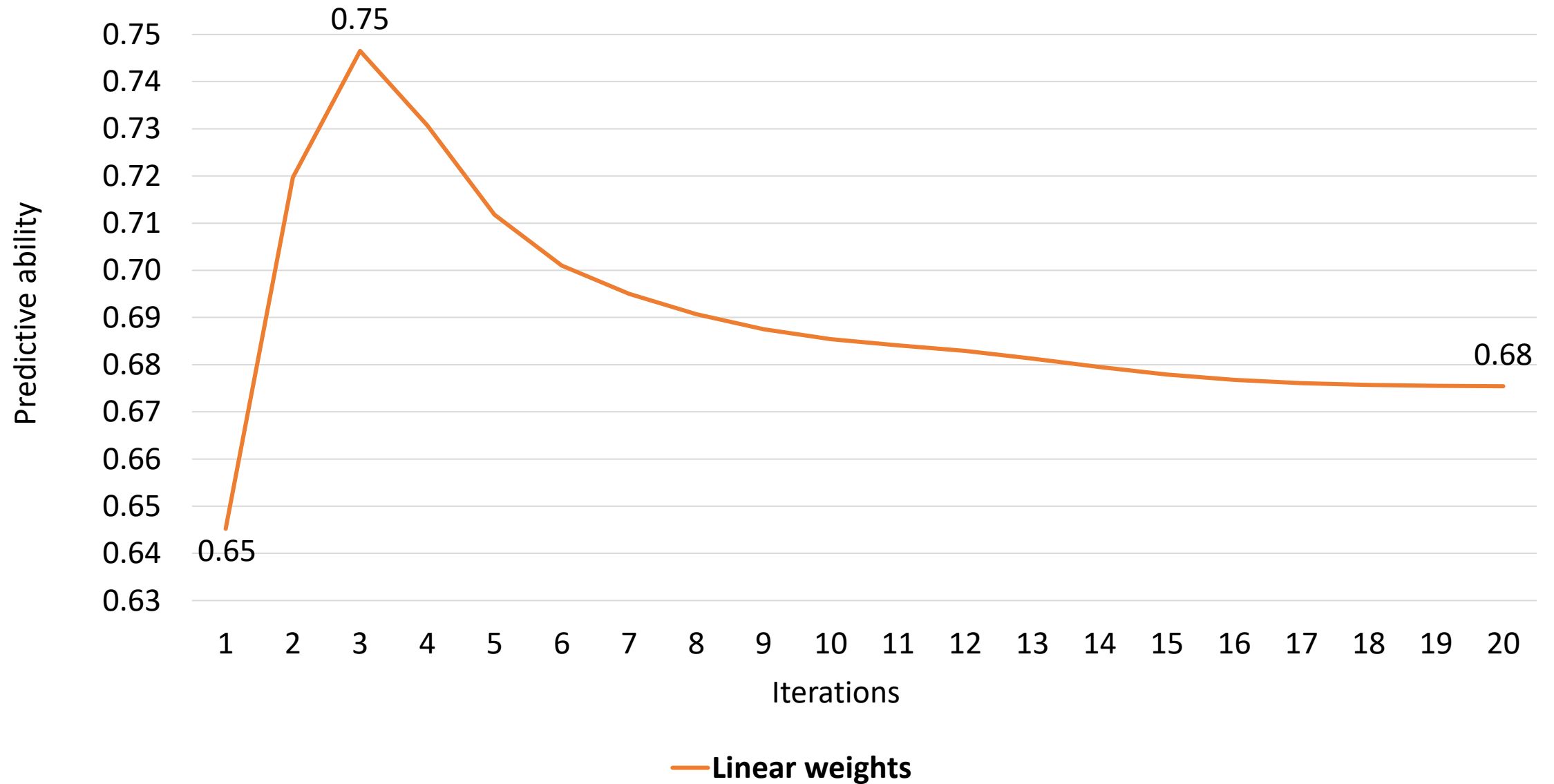
- Linear weights current implemented in postGSf90
- Some issues: decrease in accuracy and increase in inflation

Results from simulated data

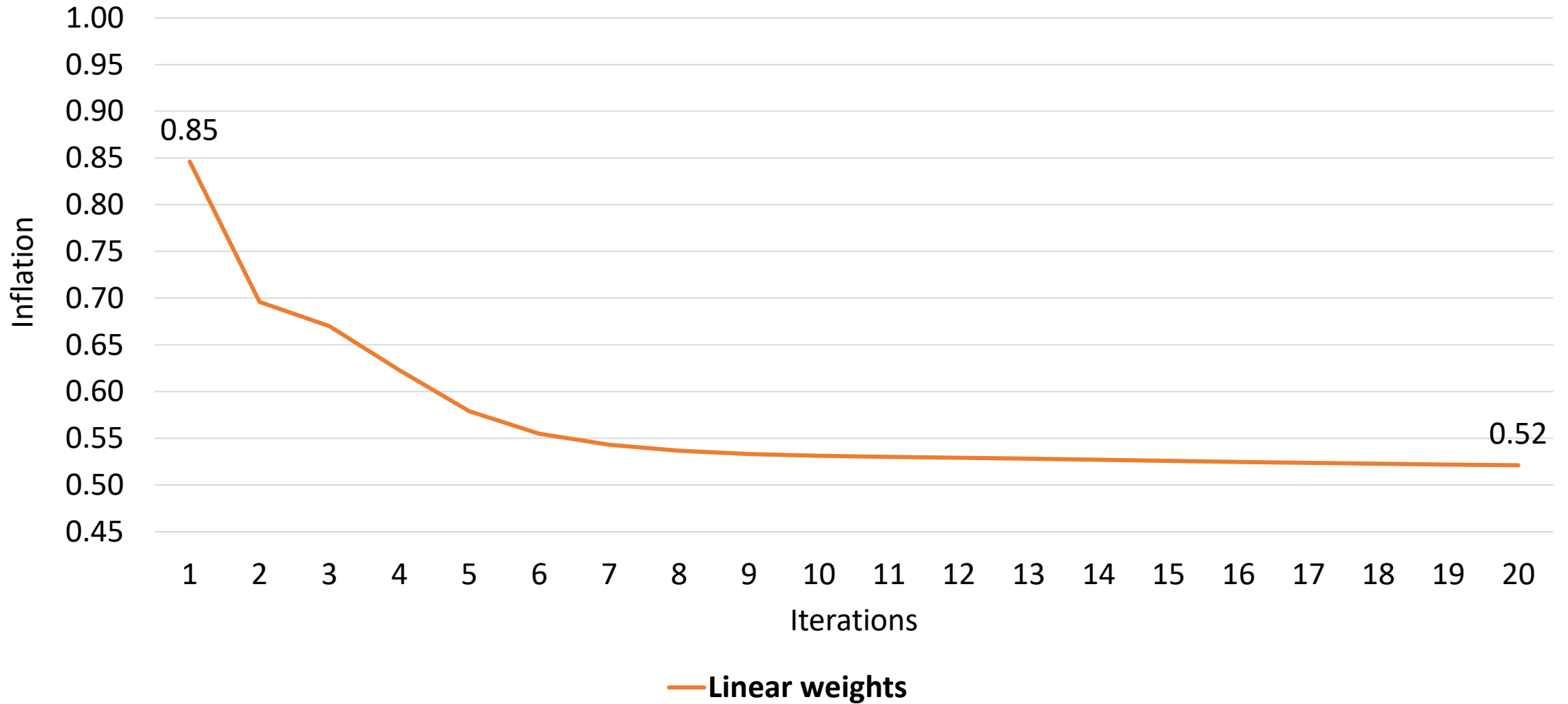
- Single trait
- $h^2 = 0.4$
- All genetic variance explained by 500 QTLs
- 45k markers
- 1000 genotyped animals as validation

- Predictive ability: $\text{cor}(\text{GEBV}, \text{TBV})$
- Inflation ($b_1 < 1.0$): $\text{TBV} = b_0 + b_1 \times \text{GEBV}$

Simulation Study



Simulation study



- There are other ways to calculate weights:
 - Non-linear weights:

$$d_i = CT \frac{|\hat{u}_i|}{sd(\hat{u})}^{-2} \quad (\text{VanRaden 2008})$$

- Implemented in postGSf90

Objective

- Test different weighting methods on WssGBLUP in channel catfish

Data

RESEARCH ARTICLE

Open Access



Development of genomic predictions for harvest and carcass weight in channel catfish

Andre L. S. Garcia^{1*} , Brian Bosworth², Geoffrey Waldbieser², Ignacy Misztal¹, Shogo Tsuruta¹ and Daniela A. L. Lourenco¹

- USDA-ARS Warmwater Aquaculture Research Unit
- Pedigree information: 36,365 animals
- Harvest weight (HW): 27,160 records ($h^2= 0.27$)
- Residual carcass weight (RCW)*: 6,020 records ($h^2= 0.34$)

*Carcass weight adjusted to a common body weight

- 2911 genotyped animals
- 54k markers after quality control

Weighted ssGBLUP

- BLUPf90 family of programs
- Linear weights

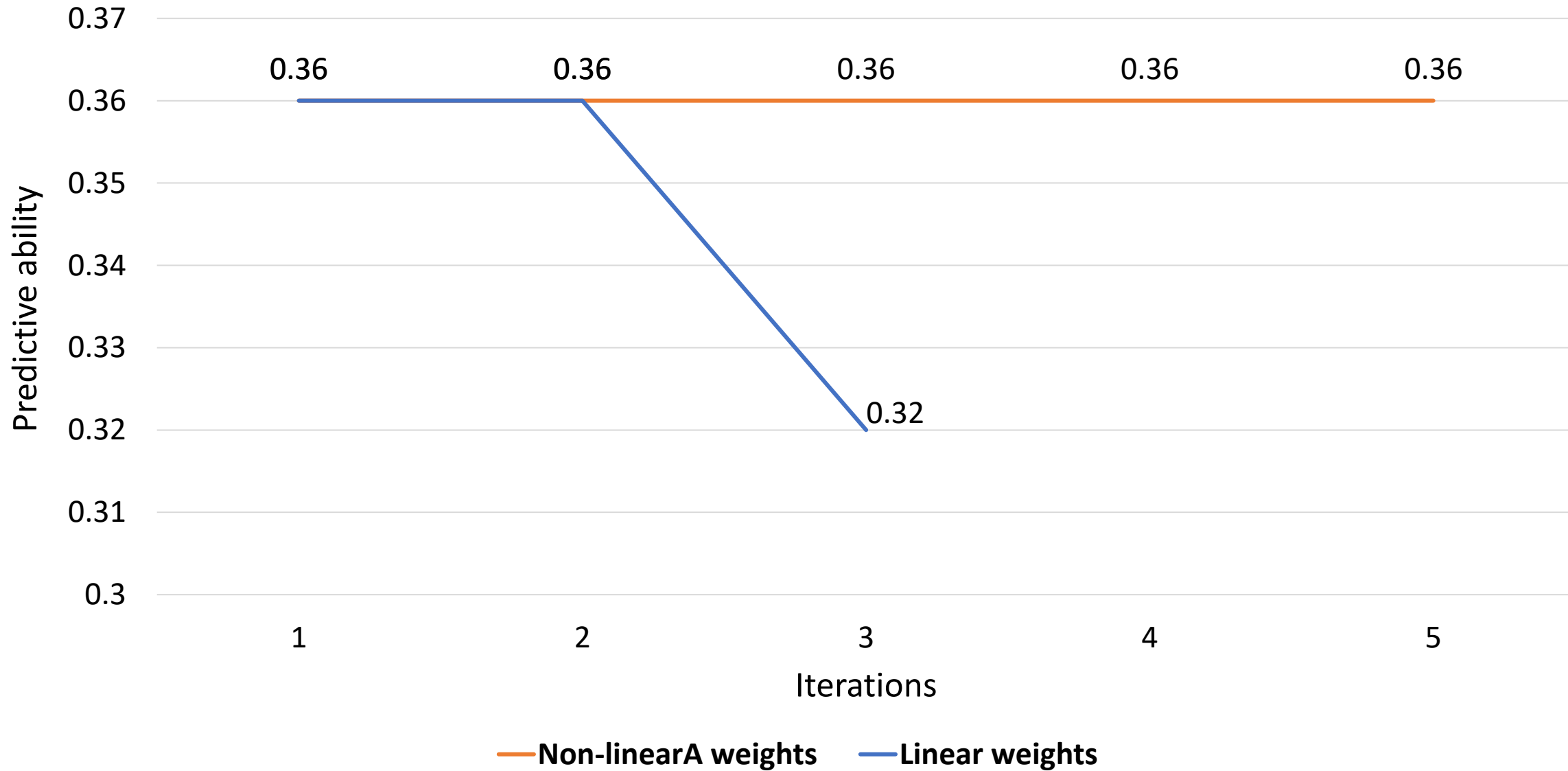
$$d_i = \hat{u}_i^2 2p_i(1 - p_i)$$

- Non-linear weights

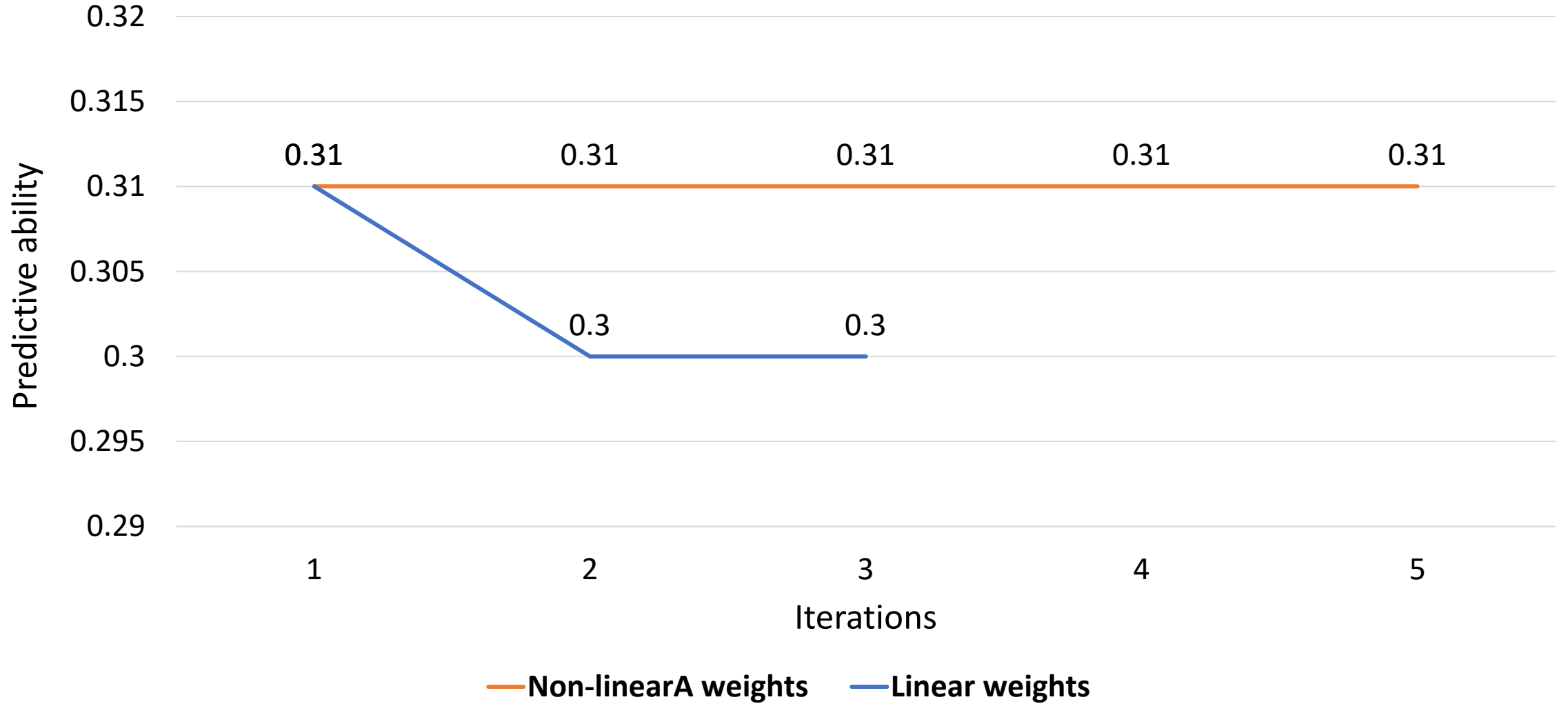
$$d_i = CT \frac{|\hat{u}_i|}{sd(\hat{u})}^{-2}$$

- CT= 1.025, 1.125 and 1.25
- Limit for weights: 5, 10 and 20
- Predictive ability: $\text{cor}(\text{GEBV}, \hat{Y})$
- Inflation ($b_1 < 1.0$): $\hat{Y} = b_0 + b_1 \times \text{GEBV}$

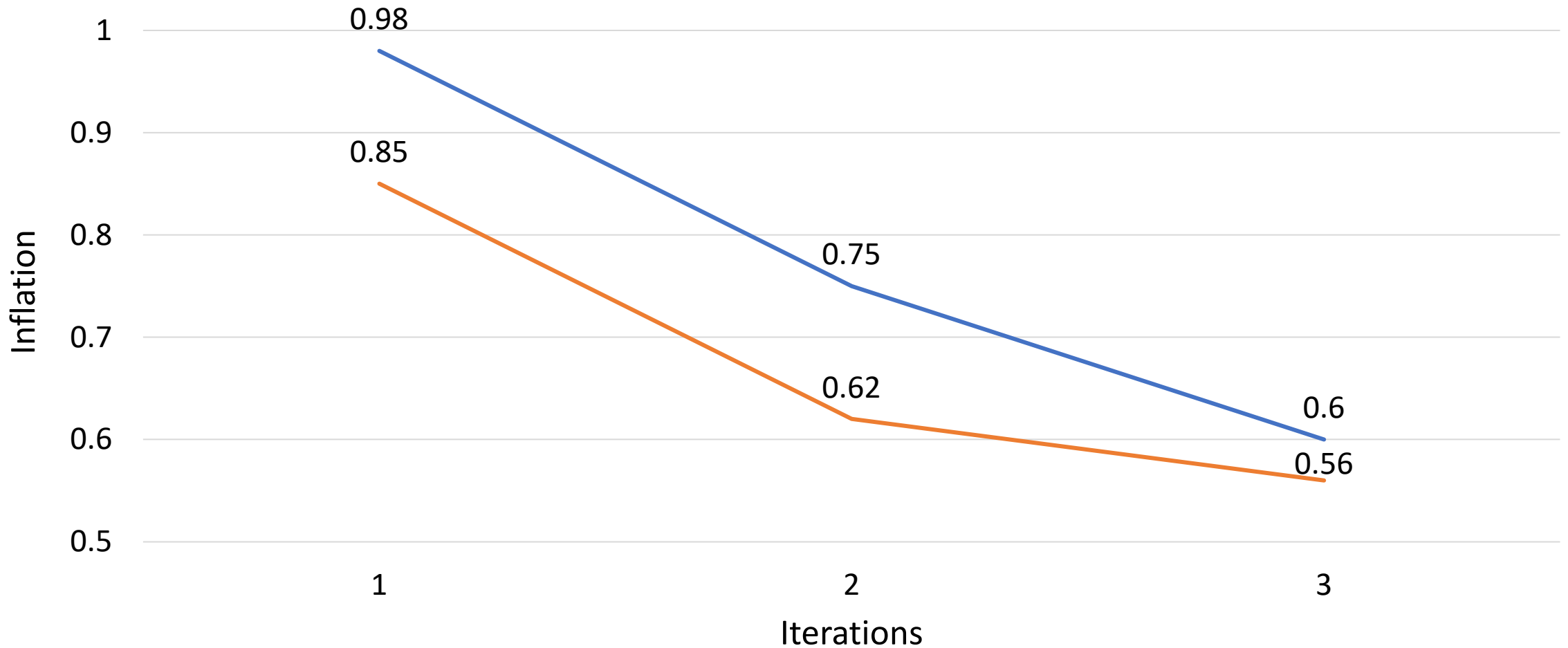
Predictive ability - Harvest weight



Predictive ability - Residual carcass weight

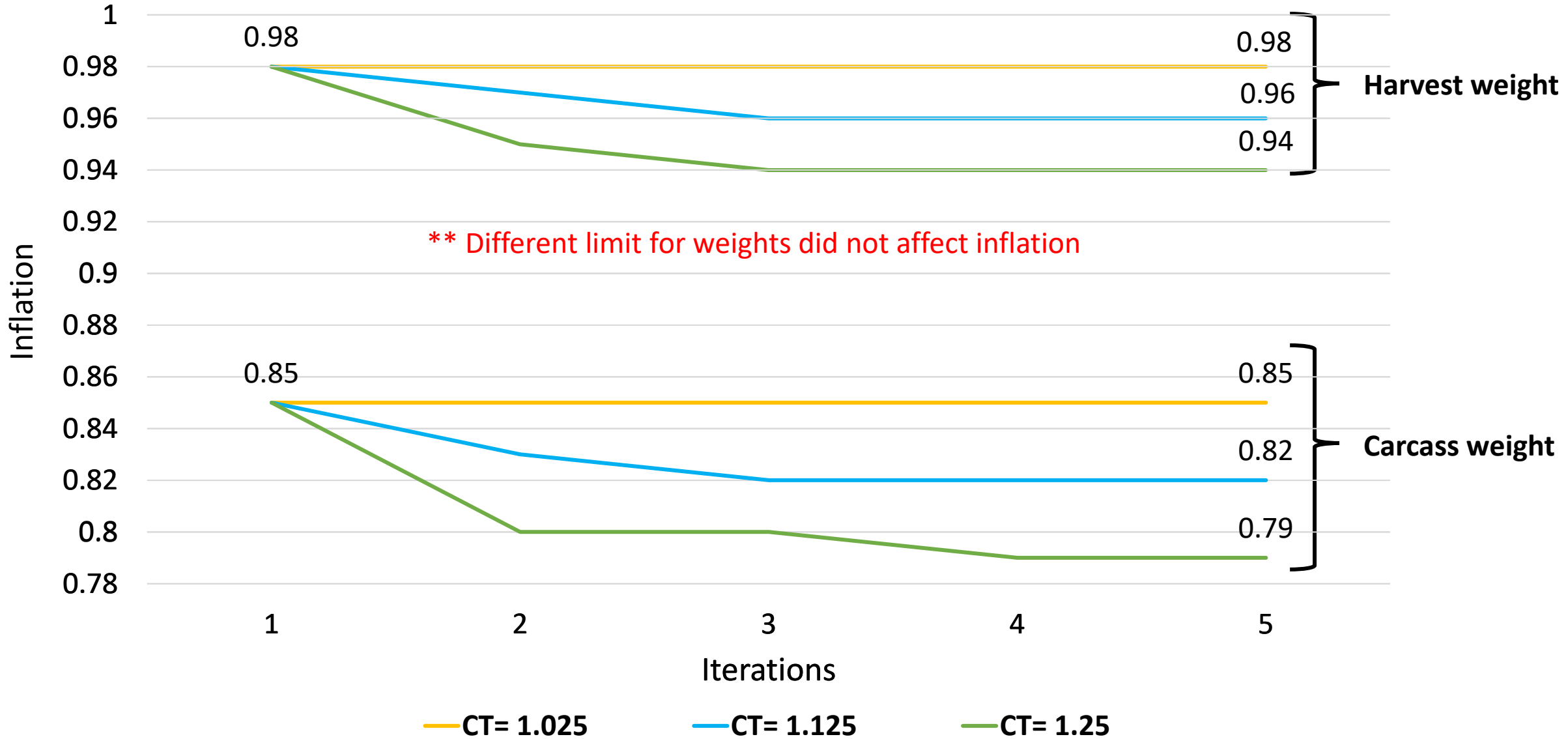


Inflation - Linear weights



— Harvest weight — Residual carcass weight

Inflation - Non-linearA weights



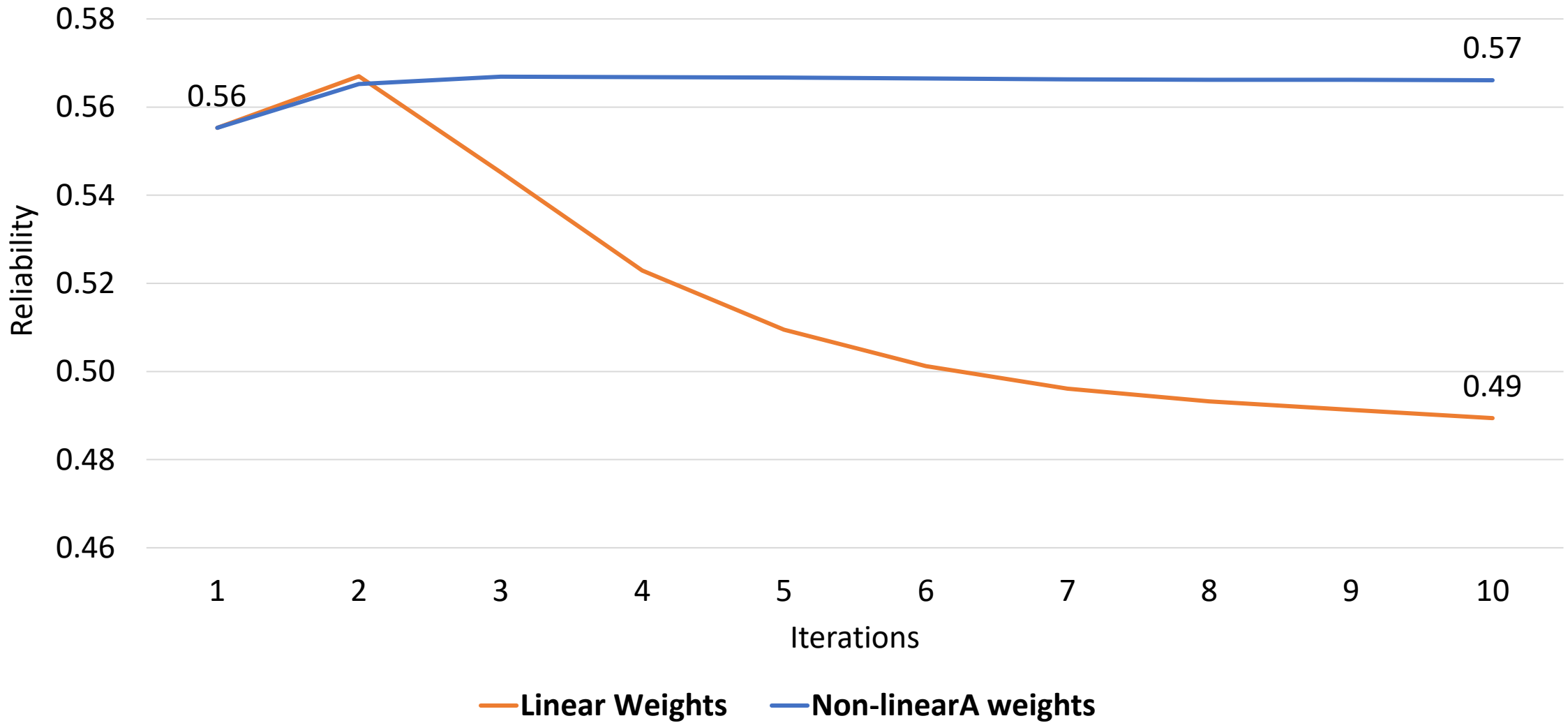
Remarks

- Non-linearA weighting method is more stable and converges faster
- Smaller CT values or unweighted ssGBLUP seems to give best results
- Different limits for the SNP weights 5, 10 or 20 did not affect predictive ability or inflation

Results from dairy data

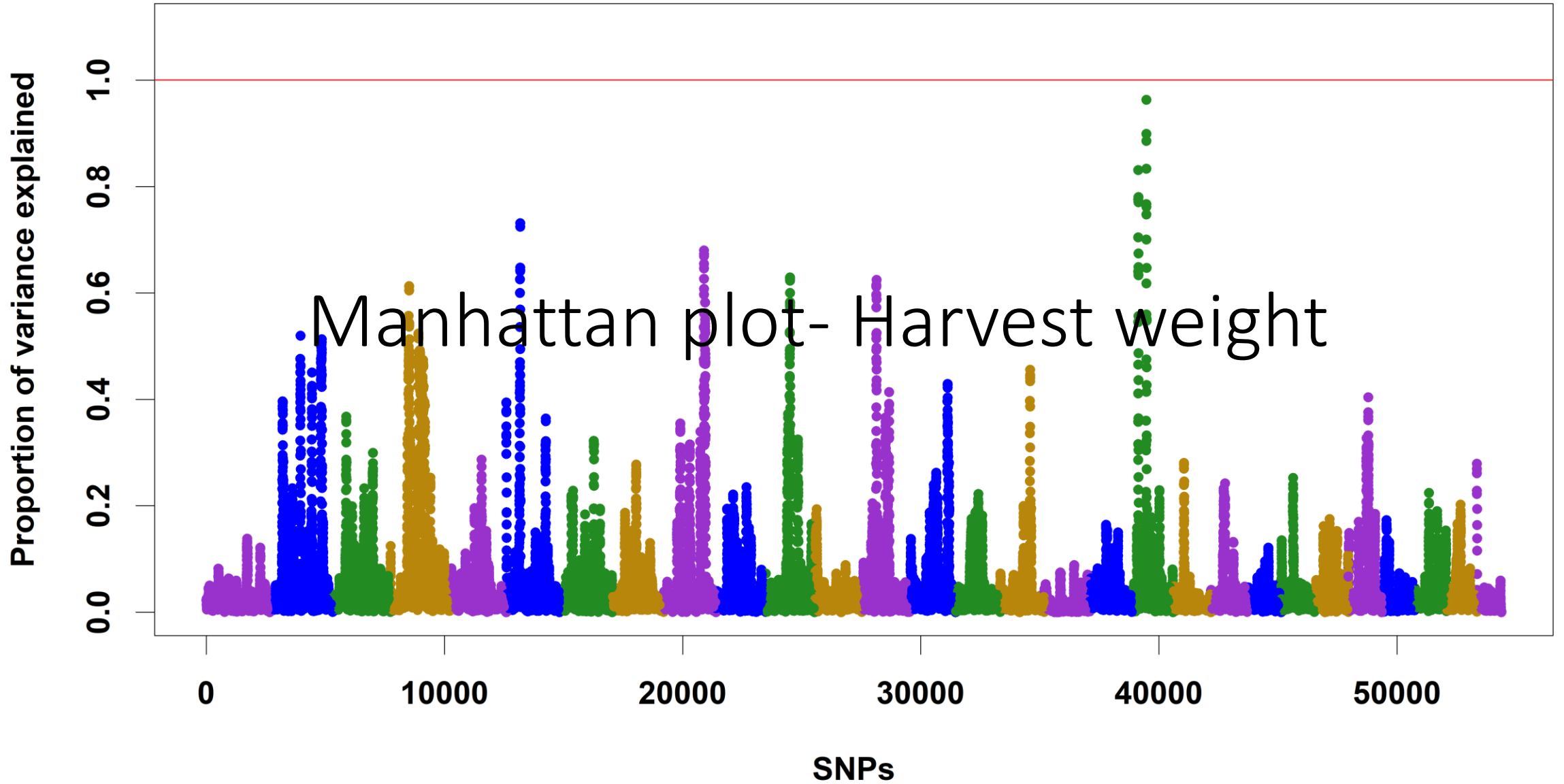
- Study done by Breno Fragomeni (Fragomeni et al., 2019)
- Holstein Association
- \approx 4 million records for stature
- 26,877 genotyped bulls
- 54k markers
- 2521 bulls for validation
- Weighted GBLUP with linear and non-linear weights

Reliability - US Holsteins

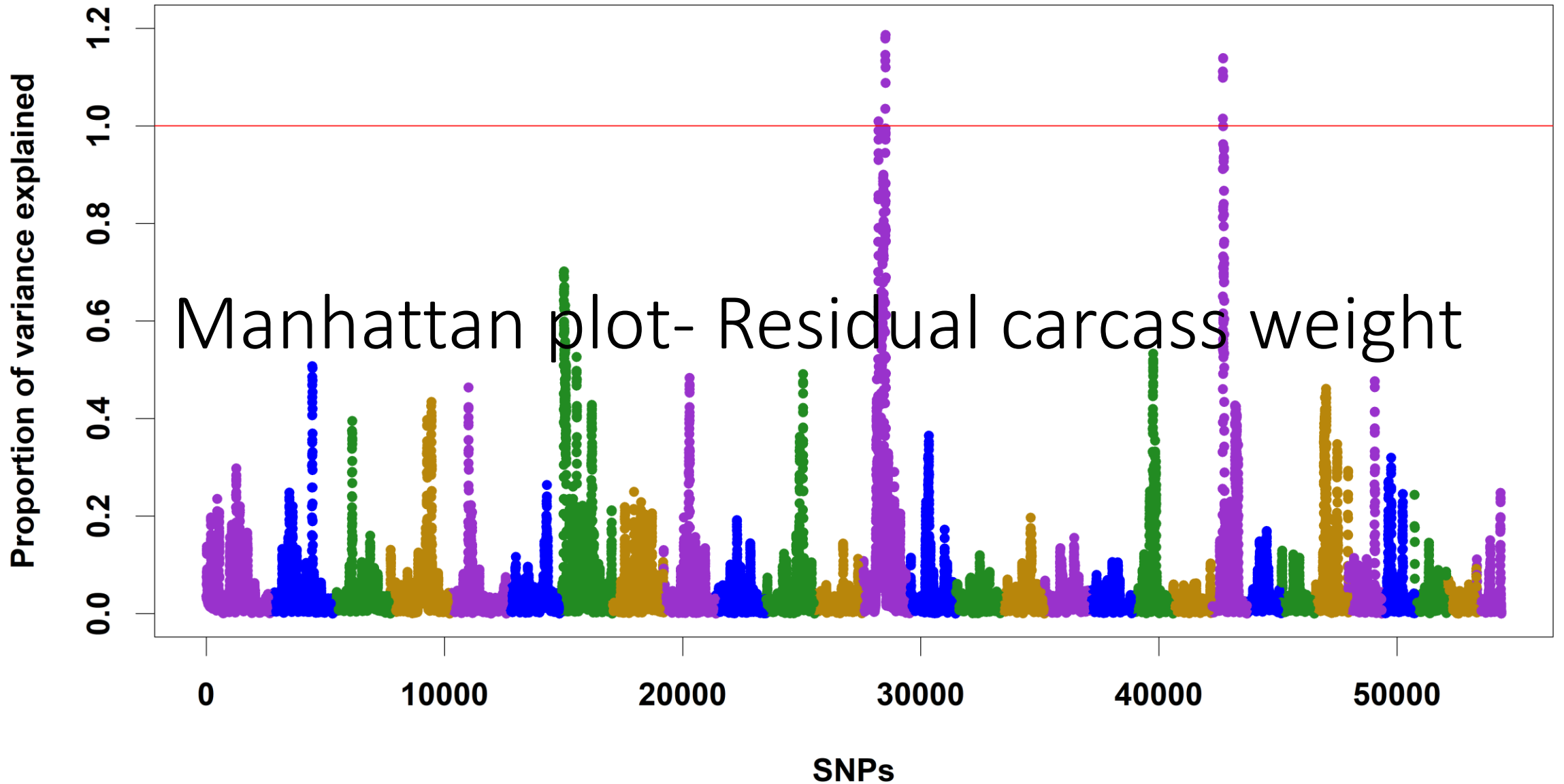


GWAS results

Variance explained by 20 adjacent SNP



Variance explained by 20 adjacent SNP



- We can get SNP effects from ssGBLUP
- We can also perform GWAS
 - Proportion of variance explained
- How about P-values?

Already implemented on postGSf90 wait for the update release!

- P-values are calculated as:

- $pvalue_i = 2 \left(1 - \Phi \left(\left| \frac{\hat{u}_i}{SD(\hat{u}_i)} \right| \right) \right),$

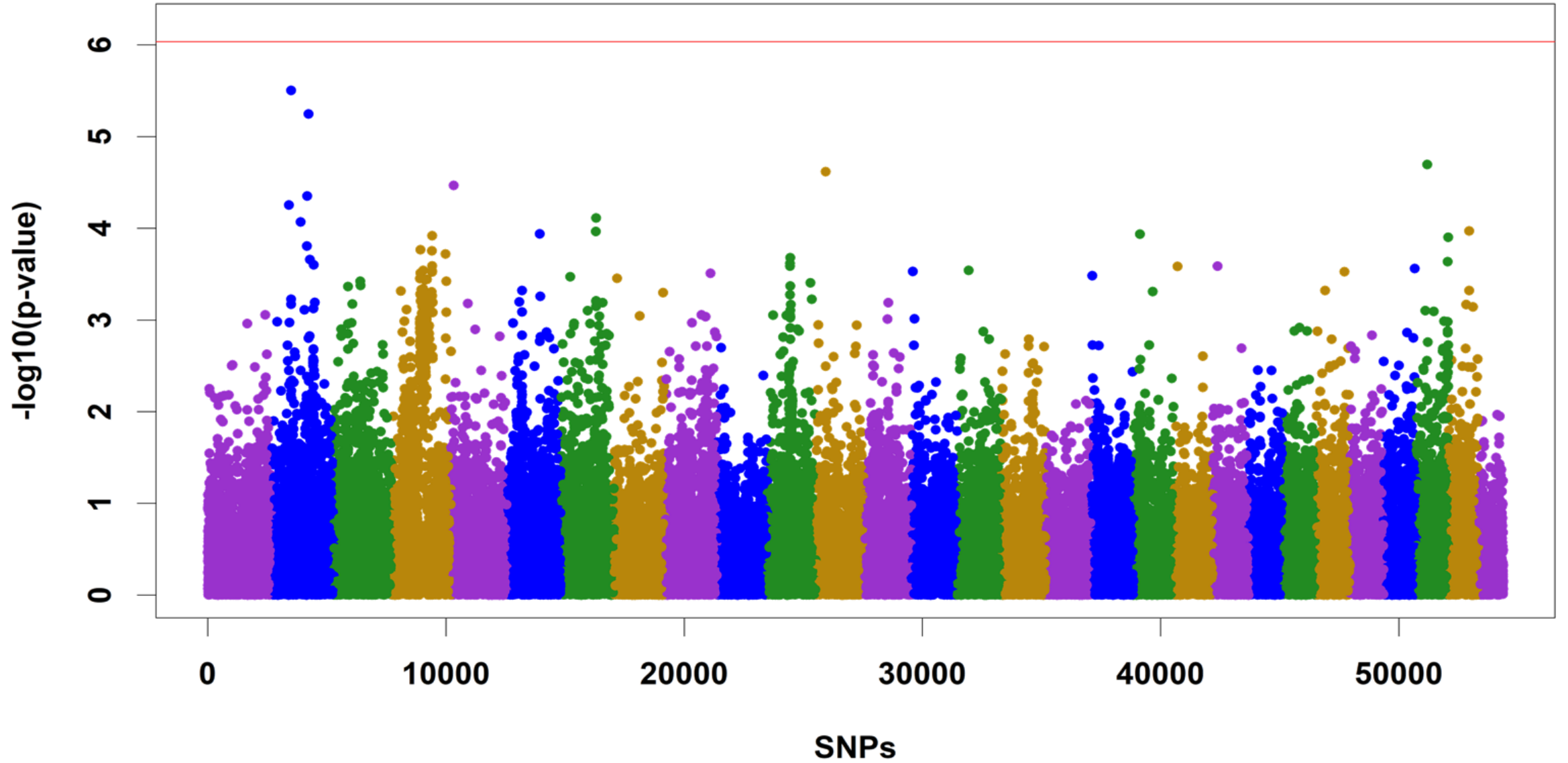
(Aguilar et al., 2019)

where Φ is the cumulative standard normal function

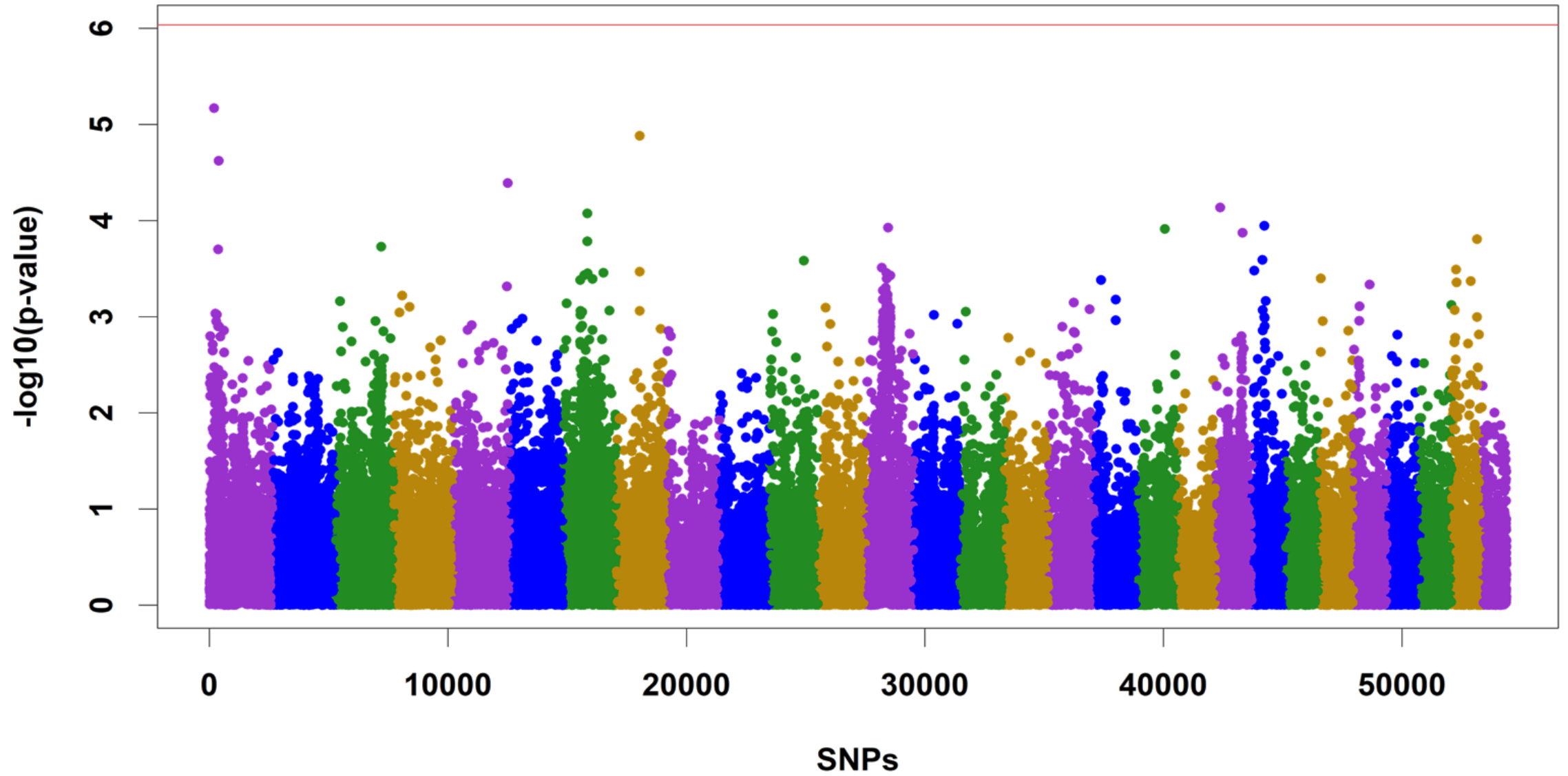
$$\hat{u} = \delta \mathbf{DZ}' \mathbf{G}^{-1} \hat{a}$$

$$Var(\hat{u}_i) = \delta \mathbf{DZ}' \mathbf{G}^{-1} (\mathbf{G} \sigma_a^2 - \mathbf{C}^{a_2 a_2}) \mathbf{G}^{-1} \mathbf{ZD} \delta$$

Harvest Weight - SNP P_VALUES
Iteration 1 | CT= 1.025 | Limit= 5



Residual Carcass Weight - SNP P_VALUES
Iteration 1 | CT= 1.025 | Limit= 5



Conclusion

- One iteration gives the best predictive ability results for harvest weight and residual carcass weight
- When weights are needed, the use on non-linearA method is recommended for WssGBLUP and GWA under ssGBLUP framework
- P-values for SNP can be calculated based on all available phenotypes, pedigree and genotypes in ssGBLUP

Acknowledgements

- ARS-USDA research agreement number 58-6066-6-025
- The University of Georgia animal breeding and genetics group

