

Singular Value Decomposition of the Genotype Matrix

L. Ayres* M.P.L. Calus* J. Ødegård*[◇] T. Meuwissen[‡]

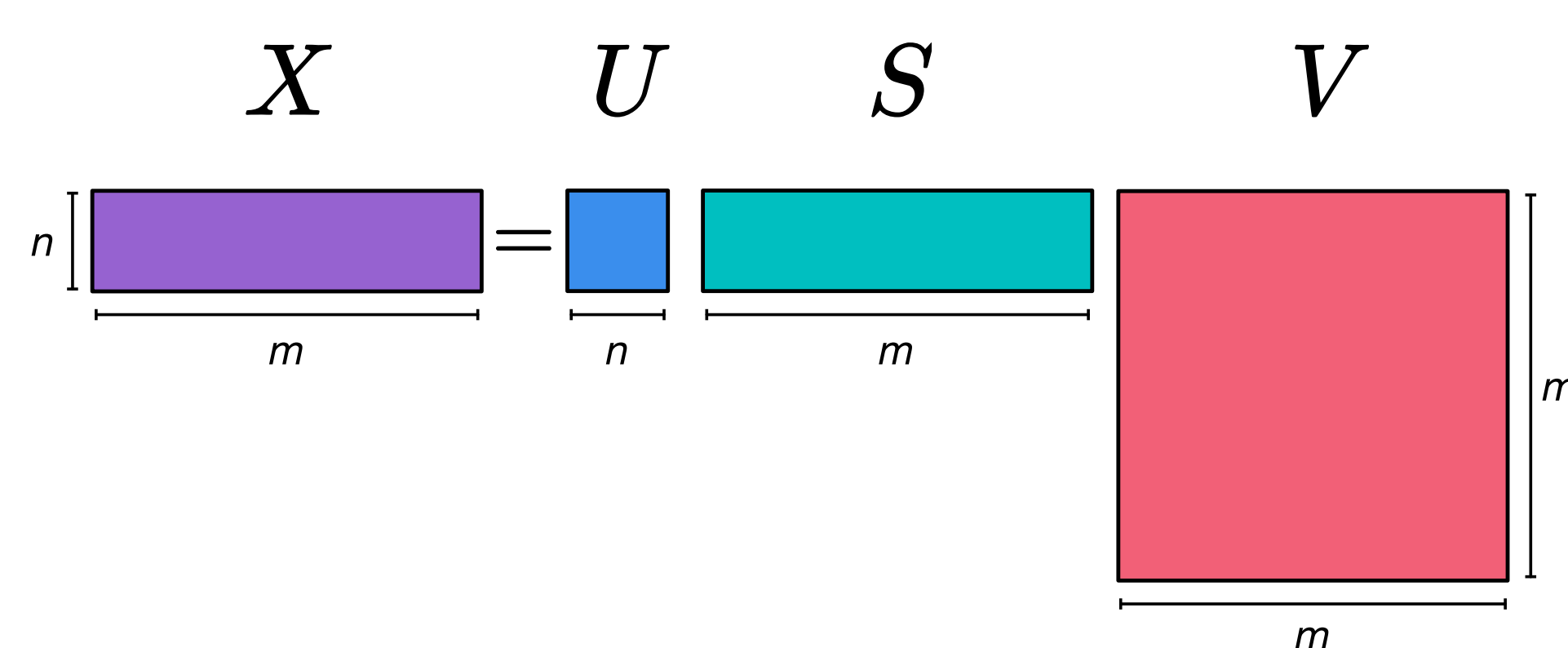
* Wageningen University & Research ‡ Norwegian University of Life Sciences ◇ AquaGen AS

Aim

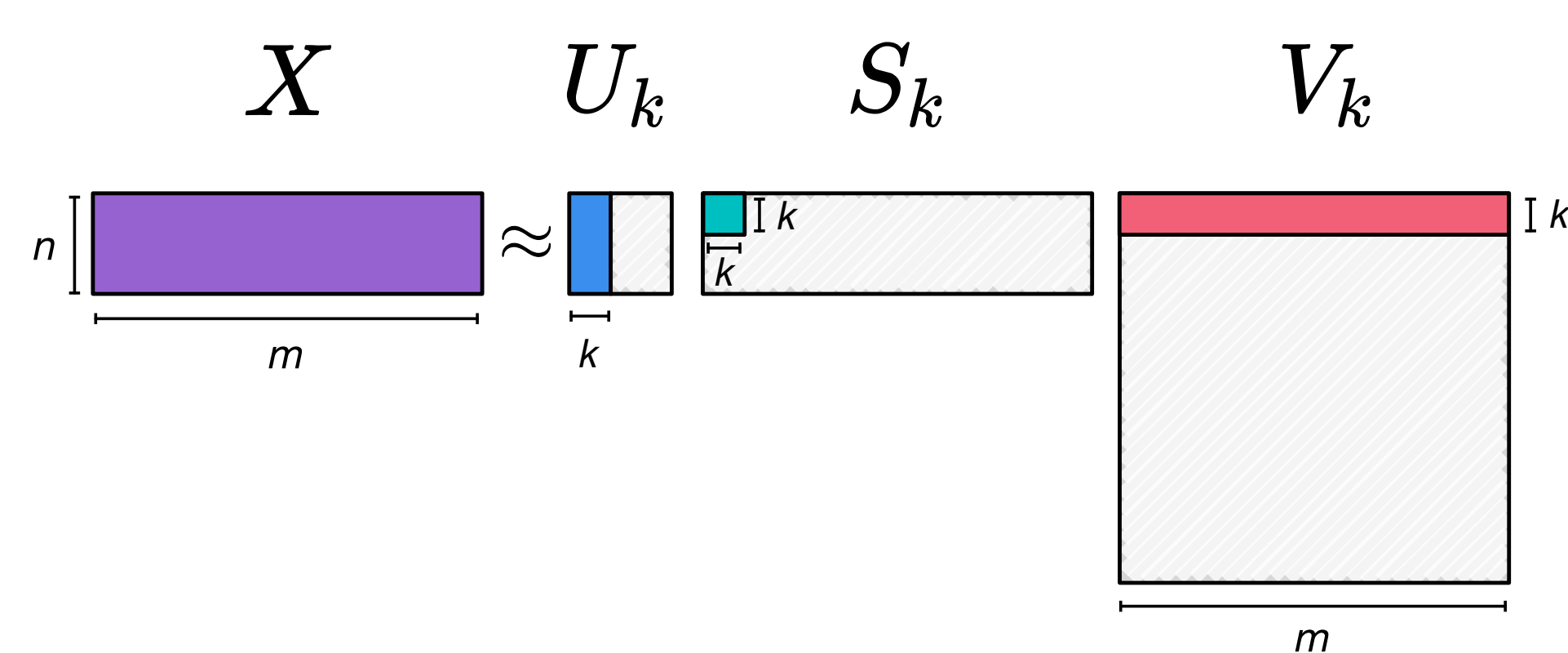
Evaluate the accuracy of prediction when varying the number of components in the truncated-SVD approximation of the genotype matrix.

SVD

Singular Value Decomposition



Truncated SVD



Simulation

- 1,000 QTL along 16,454 SNPs (1,927 Atlantic Salmon individuals)
- QTL effects sampled from normal distribution b_1, \dots, b_{1000} i.i.d. $b_i \sim N(0, V_m)$

$$V_m = \frac{h^2}{2 \sum_{i=1}^{1000} p_i(1 - p_i)}$$

- QTLs masked from genotype matrix \mathbf{X}

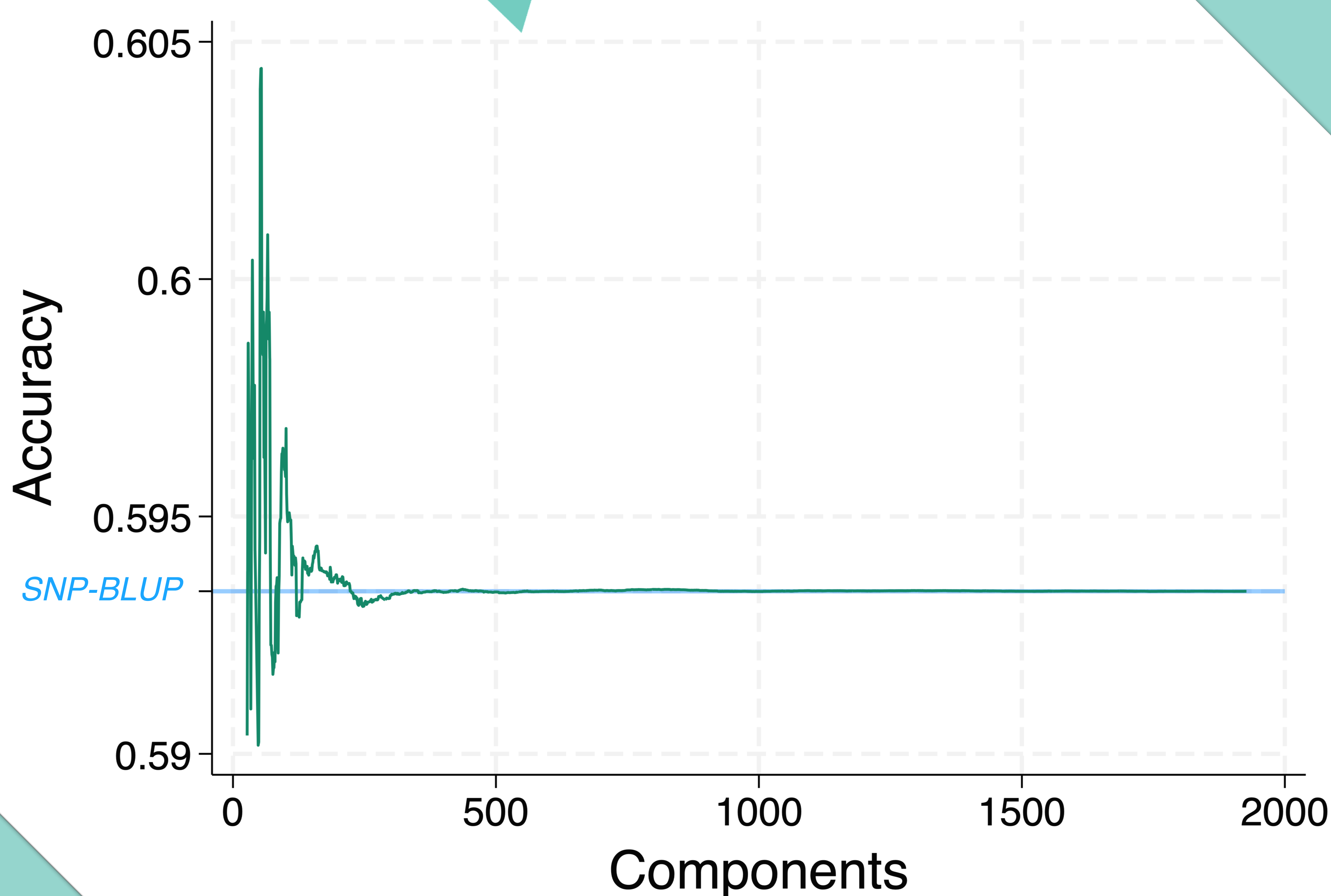
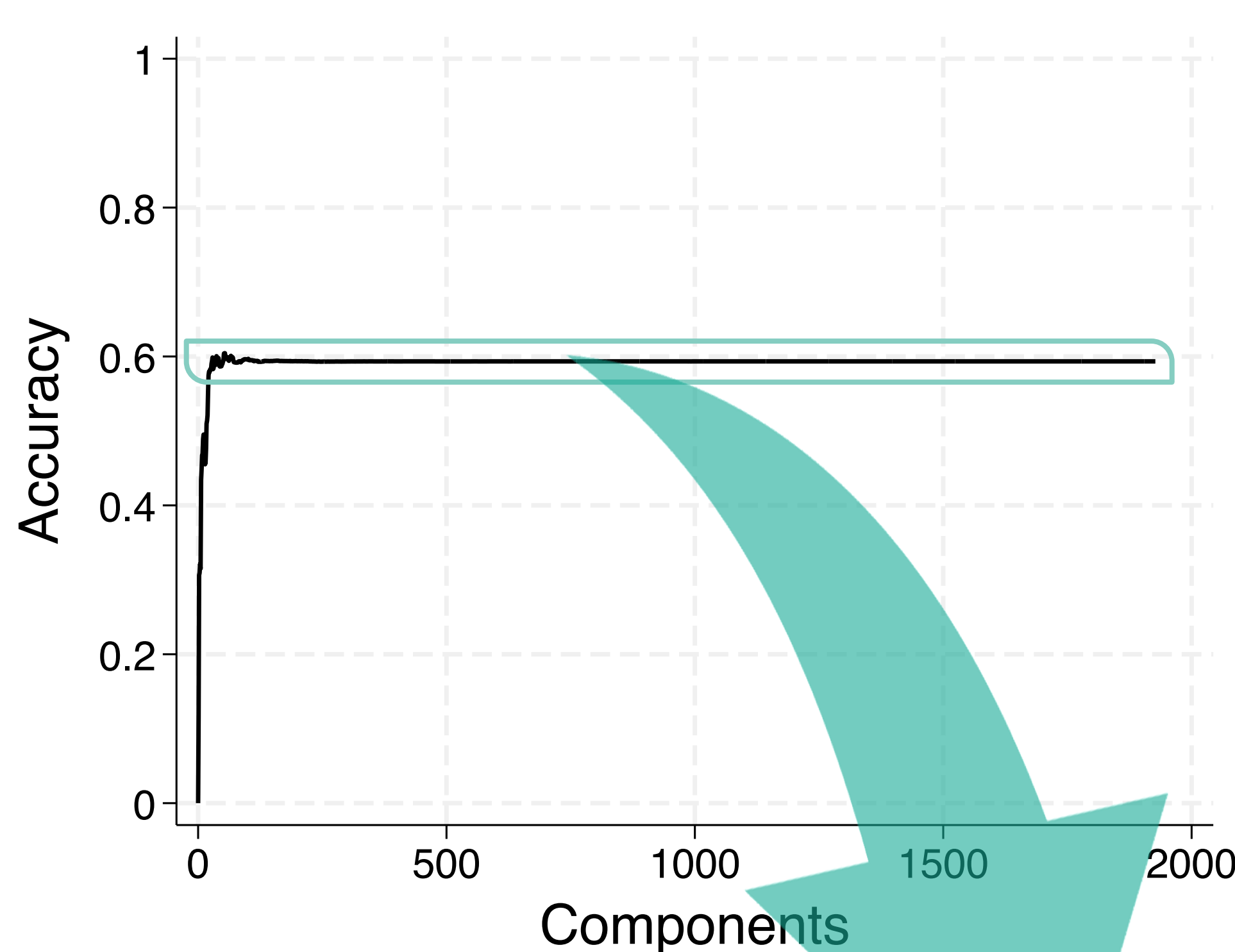
- True breeding values \mathbf{g} calculated $\mathbf{g} = \mathbf{X}\mathbf{b}$

- Breeding values predicted by PCRR $\hat{\mathbf{g}} = \mathbf{X}\hat{\mathbf{b}} = \mathbf{X}\mathbf{V}_k\hat{\mathbf{S}}_k$ (10-fold cross-validation)

- Correlation coefficients estimated

$$\hat{r} = \frac{\text{Cov}(g, \hat{g})}{\hat{\sigma}_g \hat{\sigma}_{\hat{g}}}$$

Results



Conclusions

- SVD is useful for data reduction of genotype matrices
- Good accuracies with few components
- SVD can yield higher accuracies than SNP-BLUP

QR code for supplement



lucas.ayres@wur.nl