Bavarian State Research Center
for Agriculture

‘Gene dropping A’:

Avoiding classical computations of the numerator relationship matrix

C. Edel, E.C.G. Pimentel, R. Emmerling and K.-U. Götz

Institute for Animal Breeding

Motivation

- In genomic prediction the conventional **A** matrix is still in use, e.g.
  - genetic variance not accounted for by markers [$\lambda\mathbf{G} + (1-\lambda)\mathbf{A}$]
  - straightforward calculation of $\mathbf{H}^{-}$ in single step using [$\mathbf{A}_{22}^{-1}$]

- Calculating large **A** matrices
  - time consuming
  - additional level of complexity in marker based models
  - conceptual regression to the conventional animal model

Aim of the study

- Find a way to represent the information in **A** in a linear manner

LfL Animal Breeding

Basic idea

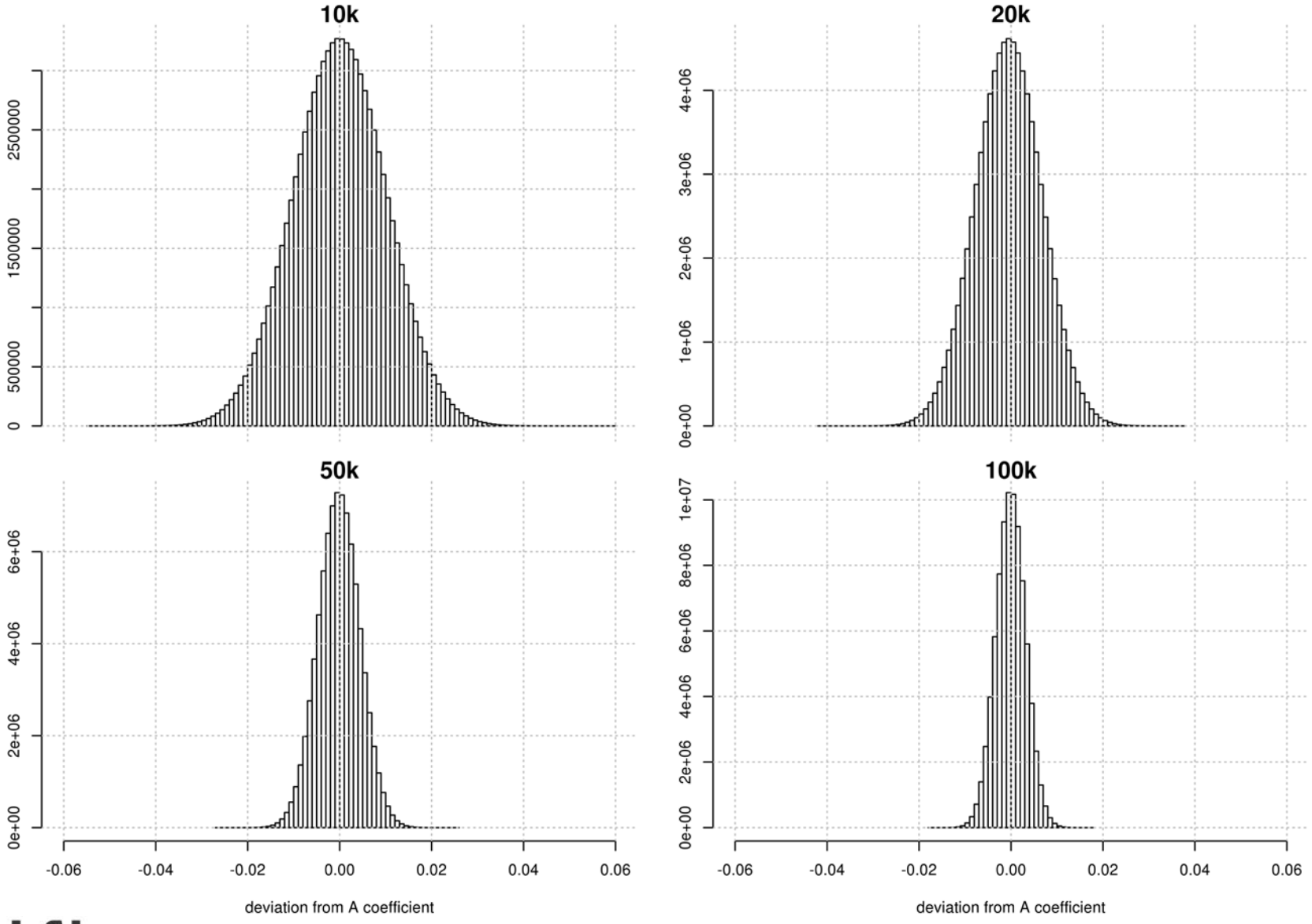
- ‘Gene-Dropping’ (MacCluer et al., 1986)
  - genes ‘dropped’ down the pedigree by a simulated gene flow
  - developed to simulate valid inheritance patterns
    - ✓ according to a known pedigree
    - ✓ compatible to observed genotypes/phenotypes

- Why not generate several thousand virtual SNP genotypes by ‘gene-dropping’ as a proxy for **A** in genomic models?

Methods I

- As an illustration: calculation of **A** matrix from virtual SNPs
  - pedigree of Fleckvieh reference population
  - 10/20/50/100k virtual SNPs
  - genotypes randomly assigned to pedigree base animals (MAF=.5)
  - dropped through the pedigree
    - ✓ function ‘gen.simuSample’, R-library GENLIB
    - ✓ easily to parallelize
    - ✓ finally: matrix calculation (VanRaden type 1)
  - deviations from true **A** calculated and plotted

Results I



LfL Animal Breeding

Conclusion I

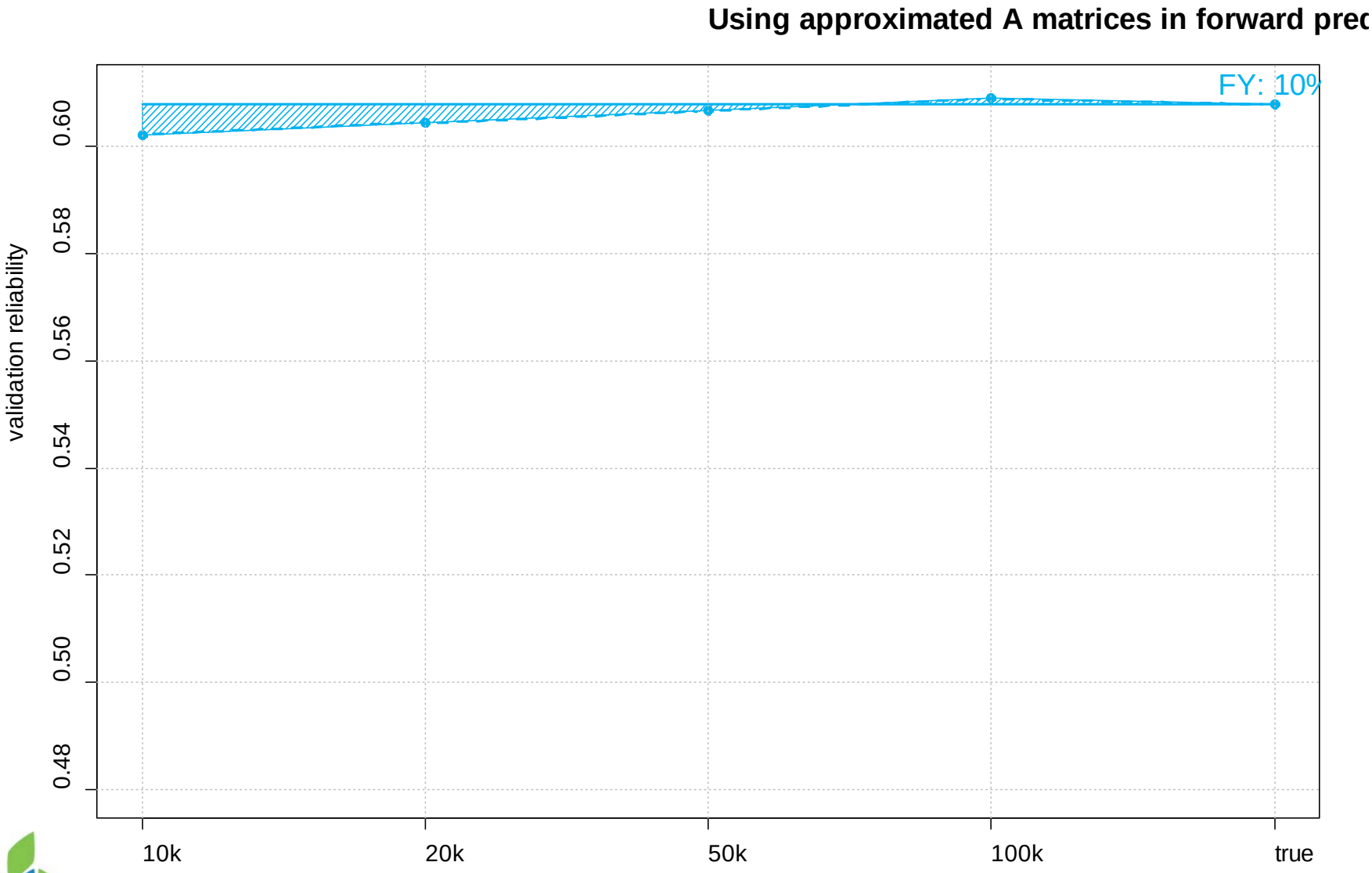
- Calculation of **A** matrix by gene-dropping
  - feasible and easy to parallelize
  - faster than most standard algorithms
- 50k and 100k dummy SNP
  - random deviations from true **A** matrix are small
- Is reliability and unbiasedness of genomic predictions affected?

Methods II

- Using true and approximated **A** matrices in a forward prediction
  - 5 traits: MY, FY, PY, STA and UD
  - ~6,700 reference bulls, ~2,200 validation bulls
- Investigated
  - validation reliability

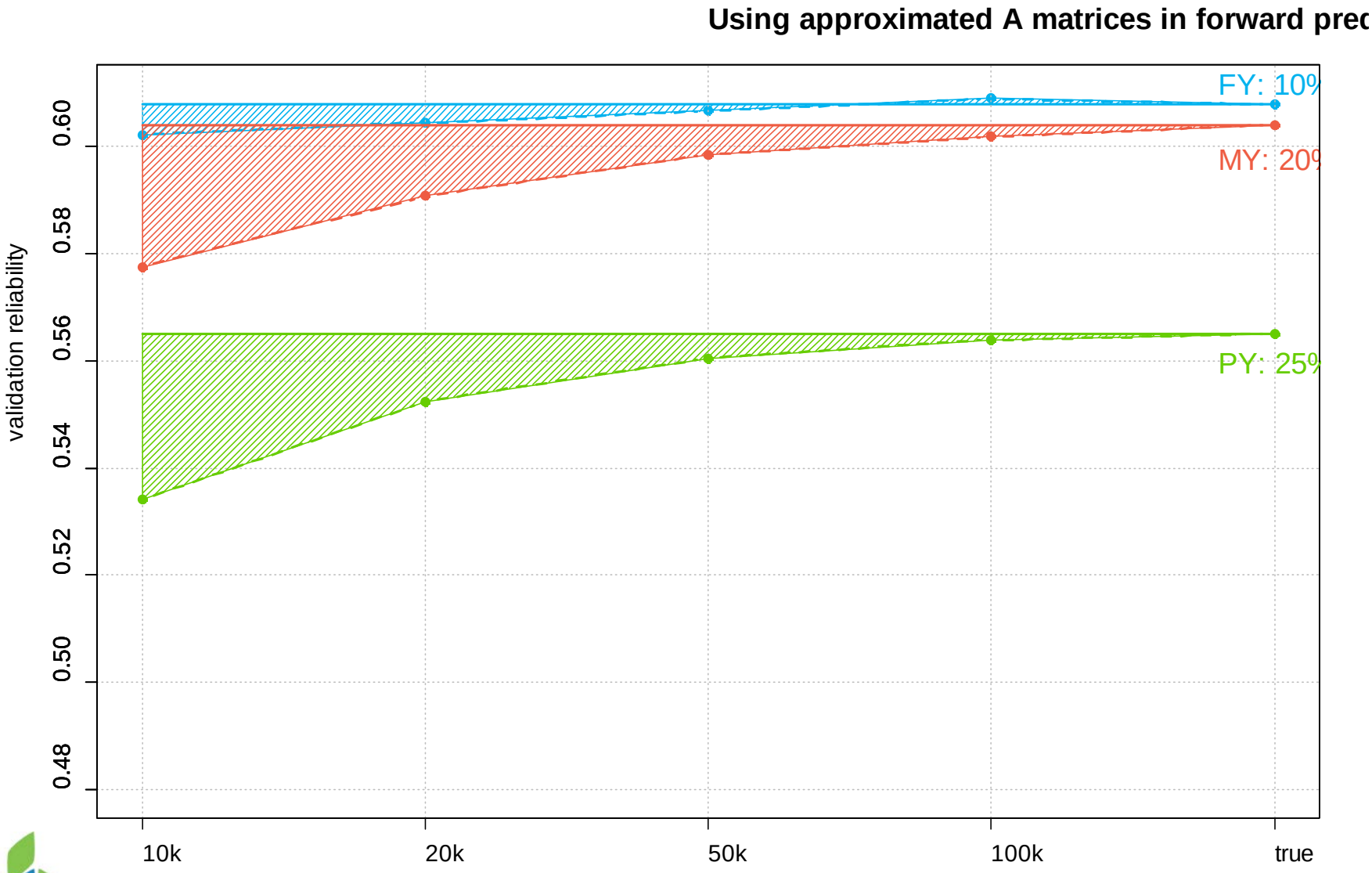
LfL Animal Breeding

Results II: Validation Reliabilities



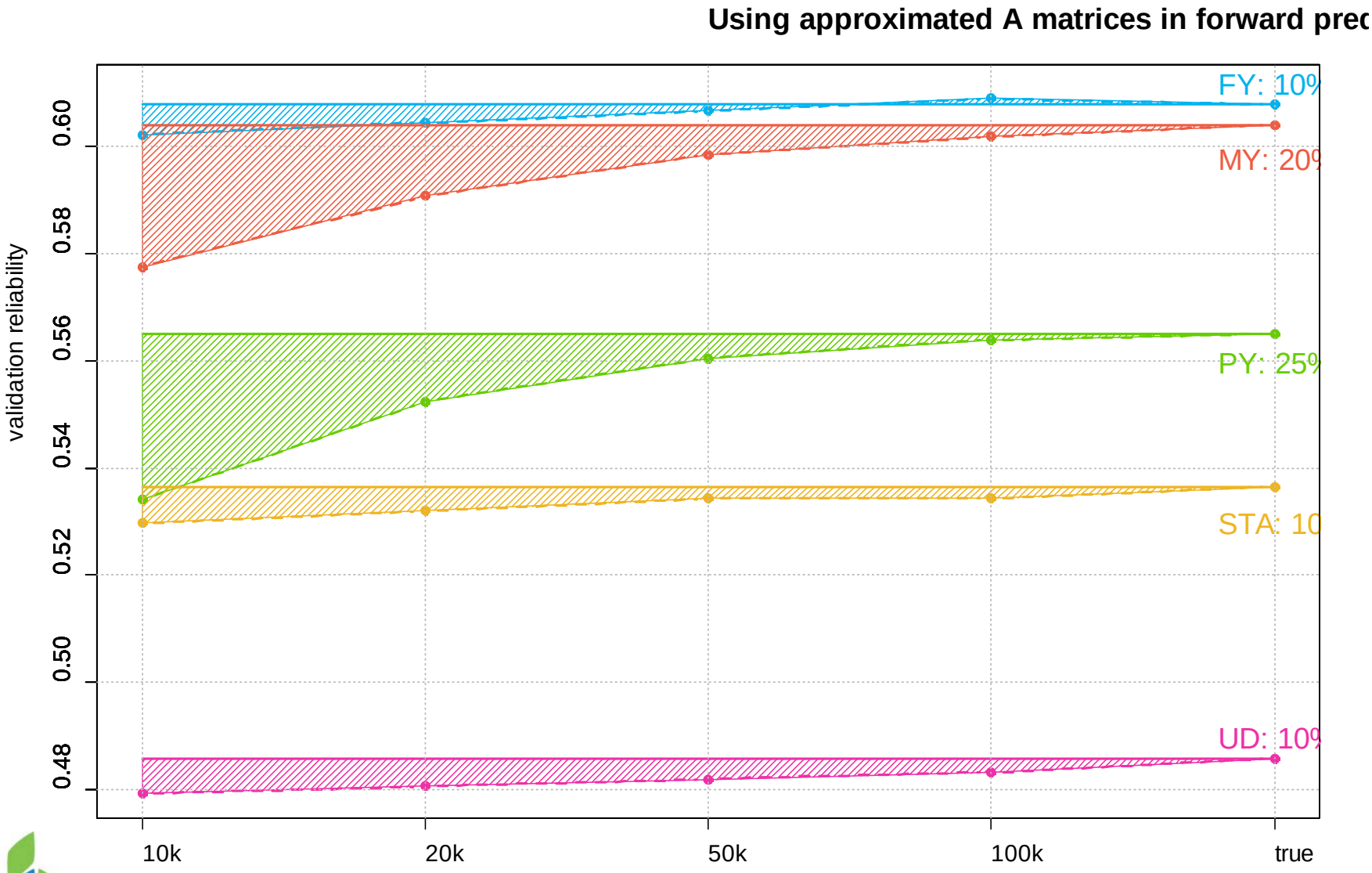
LfL Animal Breeding

Results II : Validation Reliabilities



LfL Animal Breeding

Results II : Validation Reliabilities



Conclusion II

- If **A** matrix is required and marker dimension is not crucial
  - more markers are better but…
  - 100k seems to be sufficient in most cases

Extending the concept of virtual SNPs

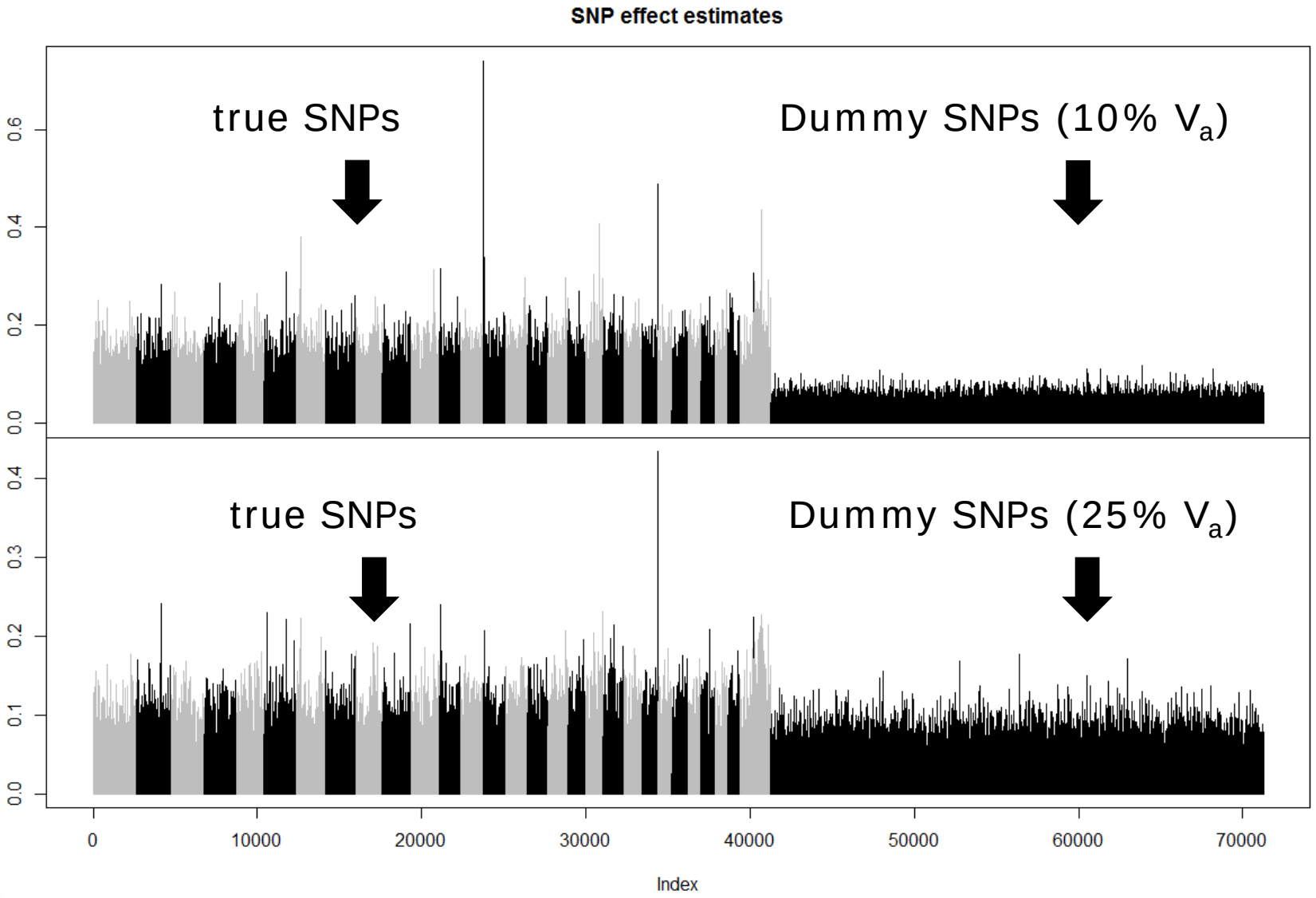
- Aim: use virtual SNPs directly to represent polygenic component
  - leaner model
  - simplified and fast prediction via SNP effects
- Conceptual problem
  - generation of polygenic component by gene-dropping gives slightly different polygenic relationship in each run

Methods III

- Investigation context
  - single-Step SNP-BLUP model (Fernando et al., 2014)
  - 50k gene-dropped SNP representing polygenic component
  - routine data for FY and PY

- Subject of investigation: repeatability of solutions obtained with the gene-dropping method
  - rank correlations of genomic EBV

Results III



LfL Animal Breeding

Results III

- Rank correlations of GEBV between 3 repeated runs, 50k **A**-SNPs

FY: 10 % V_a

reference animals

	Run1	Run2	Run3
Run1	1.00	**1.00**	**1.00**
Run2		1.00	**1.00**
Run3			1.00

prediction animals

	Run1	Run2	Run3
Run1	1.00	**0.98**	**0.97**
Run2		1.00	**0.98**
Run3			1.00

But: 50k to true >.99

PY: 25 % Va

reference animals

	Run1	Run2	Run3
Run1	1.00	**1.00**	**1.00**
Run2		1.00	**1.00**
Run3			1.00

prediction animals

	Run1	Run2	Run3
Run1	1.00	**0.92**	**0.92**
Run2		1.00	**0.93**
Run3			1.00

But: 50k to true >.98

LfL Animal Breeding

Conclusion III

- Working with 50k dummy SNP gives satisfying results
- In consecutive runs
  - deviations from true **A** are not correlated
  - additional variation arises
- Can be alleviated by
  - keeping an arbitrary proportion of SNPs generated
  - dropping them further, if pedigree is extended
  - using more SNPs

LfL Animal Breeding

General Conclusion

- Linearization of **A** matrix by virtual SNPs generated by gene-dropping might be helpful in many contexts
  - simpler models in SNP-BLUP applications
  - easy and fast prediction via SNP estimates
  - in standard single-step GBLUP: [$\mathbf{A}_{22}^{-1}$] via APY?
  - nice illustration:
    - ✓ unlinked markers do catch pedigree relationship

LfL Animal Breeding

Thank you for your attention

We gratefully acknowledge:

- Arbeitsgemeinschaft Süddeutscher Rinderzucht- und Besamungsorganisationen

LfL Animal Breeding