

One million animals genotyped

Genomic prediction models in genetic evaluation of large populations

J. ten Napel, G.C.B. Schopen, J. Vandenplas, A.R. Cromie, E.M. van Grevenhof,
R.F. Veerkamp



Introduction

- Genotyping bulls: order of 100k genotypes
- Genotyping cows: order of 1M genotypes
- How to utilise all available information for breeding value estimation?

- Aim of the study: Compare single-step methods to utilise all pedigree and genomic information in a large routine evaluation

Single-step methods

■ Breeding value model: ssGBLUP

- Using APY-inverse of the genomic relationship matrix
- Generally good approximation to avoid full inverse

■ Marker effect model: ssRRBLUP (after Fernando *et al.*, 2014, GSE 46:50)

- Using random regression on all available SNP simultaneously

■ Hybrid model (after Fernando *et al.*, 2016, GSE 48:96)

- Breeding value model for non-genotyped individuals
- Marker effect model for genotyped ones

Large routine evaluation

Beef cattle reproduction evaluation of ICBF, Ireland

	ICBF-2017	ICBF-2018
	N	N
Pedigree	12.0 million	15.2 million
Data: 1 trait	3.5 million	n/a
Data: any of 6 traits	11.1 million	13.6 million
Genotypes (50k SNP)	0.6 million	1.2 million

Analyses

■ Breeding value model

- APY inverse of G using 40,000 core animals sampled within breed groups
- ICBF-2017; MiXBBLUP & calc_grm

■ Marker effect model

- Imputation of non-genotyped individuals prior to analysis
- ICBF-2017; MiXBBLUP

■ Hybrid model

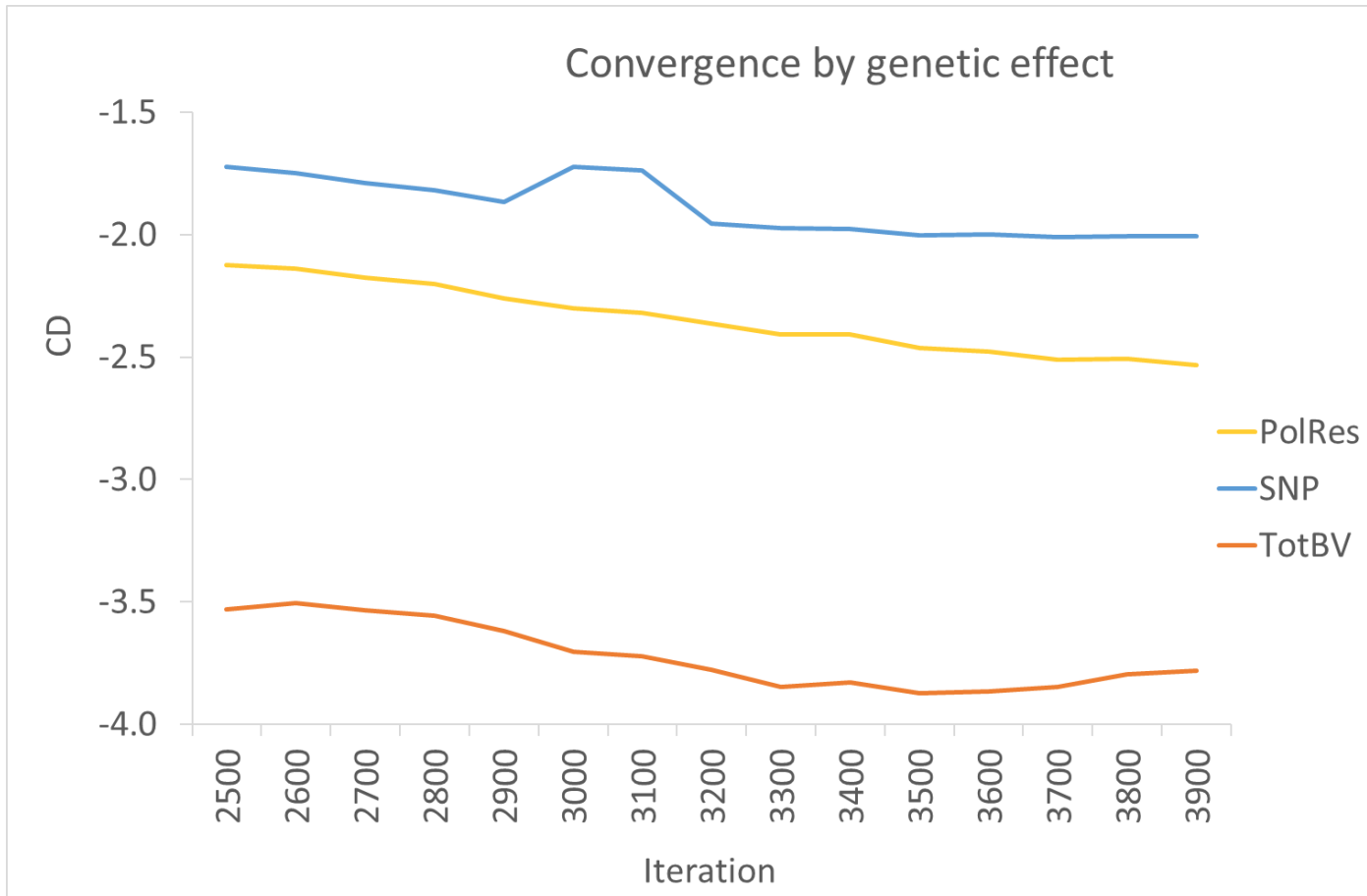
- Avoids imputation; keeps SNP covariates of genotyped individuals in memory in compressed format (4 SNP per byte); parallelized
- ICBF-2018; SNPBLUP software (J r mie Vandenplas, EAAP 2018, session 42)

Use of resources

	BVM-1	MEM	BVM-6	HM
Data	ICBF-2017	ICBF-2017	ICBF-2017	ICBF-2018
# traits	1	1	6	6
Genomic info @ solving	On disk, 4 b/element	On disk, 4 b/SNP	On disk, 4 b/element	In memory, 1 b / 4 SNP
# genotyped	0.6 million	0.6 million	0.6 million	1.2 million
Size genomic info	93 Gb	686 Gb*	93 Gb	85 Gb
Total time	3d 20:07	61d 7:18	> 5 d 20:42	1d 17:00
# iterations	1,270	3,937	> 677	2,500
Time/iteration	0:02:47	0:21:20	0:08:51	0:00:54

* Observed or imputed genotype for each data record

Convergence of marker effect models

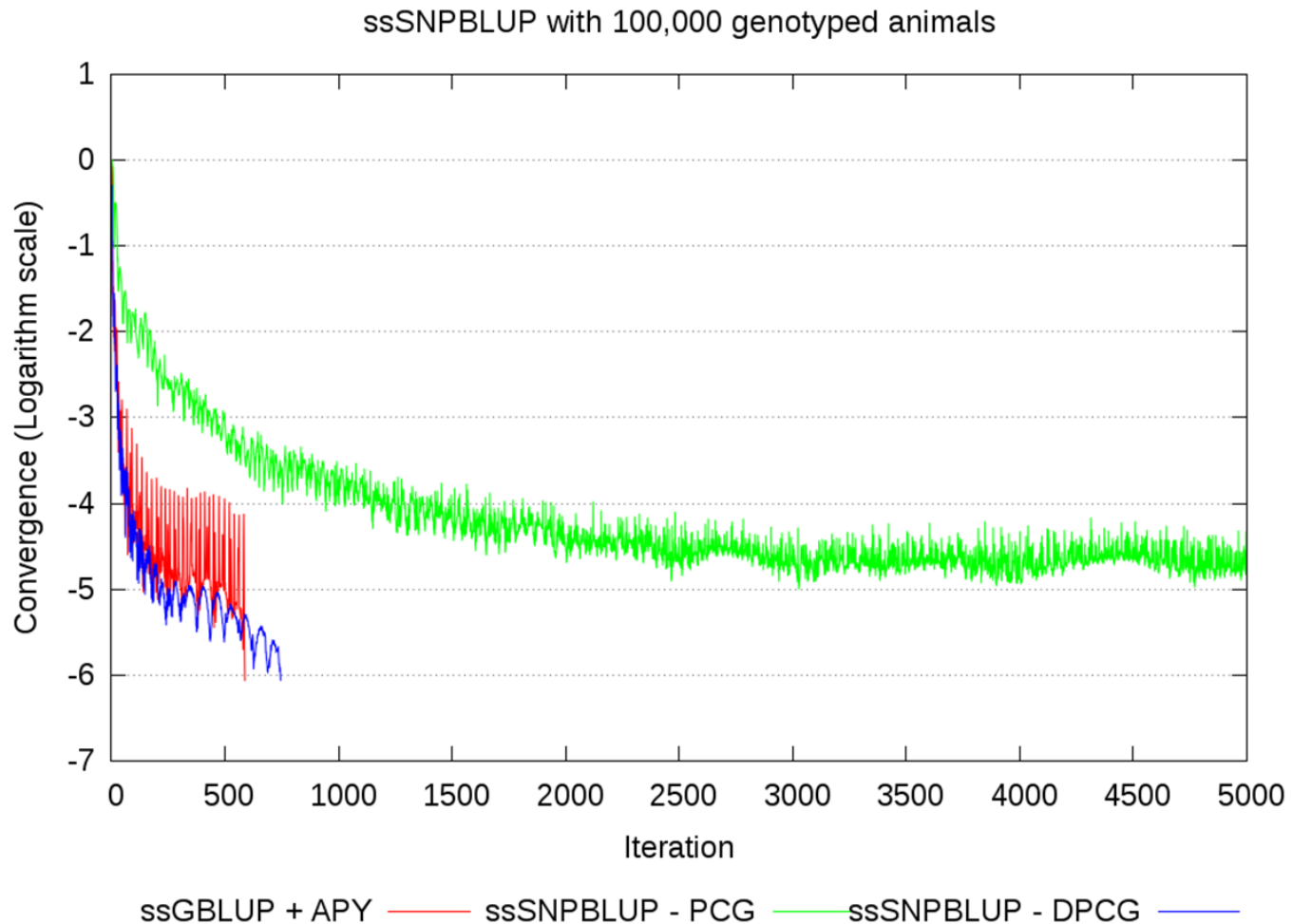


$$CD = \sqrt{\frac{\sum(sol_t - sol_{t-100})^2}{\sum(sol_t)^2}}$$

Performance

- Marker effect model is **slower** than breeding value model
 - More time per iteration
 - More iterations until full convergence
- Solving of marker effect model
 - Iteration-on-data with SNP covariates read from disk is **unsuitable** for marker effect models
- Slow convergence of marker effect model
 - **Sum** of genomic breeding value and polygenic residual **converges faster** than SNP effects and polygenic residuals separately
 - **Potential solutions** being explored: better convergence criterion, alternative solver

Promise: convergence using Deflated PCG solver



Conclusions

- Utilising a million genotyped individuals in routine evaluations is feasible
- For solving with iteration-on-data, a breeding value model using an APY-inverse of G is much more efficient than a marker effect model
- With the many more non-genotyped than genotyped individuals in this evaluation, the hybrid model is more efficient than the marker effect model
- There is scope for dramatic improvement of convergence of marker effect models and hybrid models

Acknowledgments



Breed4Food is dedicated to be the leading research consortium in animal breeding, genetics and genomics enabling the Breed4Food partners to breed better products to benefit society's needs.



WWW.BREED4FOOD.COM

