Efficiency of BLUP genomic prediction models that use pre-computed SNP variances

Mario Calus

Chris Schrooten

Roel Veerkamp

LIVESTOCK RESEARCH
WAGENINGEN UR

CRV

Animal Breeding &
Genomics Centre

Introduction

- Different genomic prediction models, i.e.:

Type of model	Comp. time		SNP variances	
BLUP	Fast	☺	Assumed	☹
Bayes	Slow	☹	Estimated	☺

- *Objective:* Can we develop BLUP models that use pre-estimated SNP-variances?

LIVESTOCK RESEARCH
WAGENINGEN UR

CRV

Animal Breeding &
Genomics Centre

Models & Scenarios

Model	Description
BayesC	1 distribution of SNP effects
Bayes SSVS	2 distributions of SNP effects
RR-BLUP	Equal variance per SNP
BLUP-C	Uses variances from BayesC
BLUP-SSVS	Uses variances from Bayes SSVS

Scenario	Animals used in Bayes (to estimate SNP variances used in BLUP)
FULL	All
RAN50	Random 50%
TOP50	Best 50%
BOT50	Worst 50%

BLUP-C & BLUP-SSVS always uses all animals

LIVESTOCK RESEARCH WAGENINGEN UR

CRV
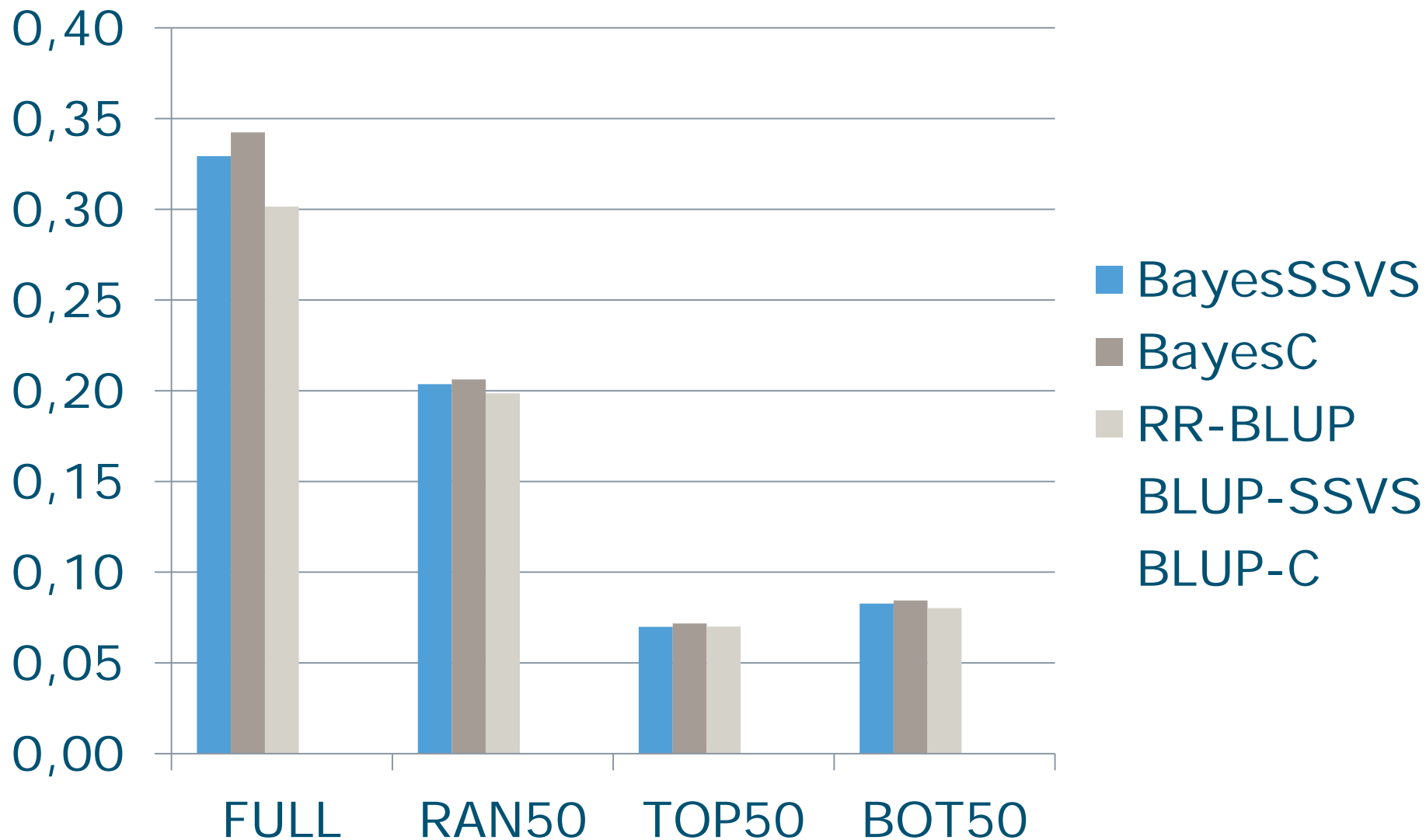
Animal Breeding & Genomics Centre

Simulated data *(Hickey & Gorjanc, G3, 2012)*

- 2,000 animals in training

- 3 generations of 500 selection candidates

- ~57,500 SNPs

- 900 QTL with effects from a Gamma distribution

- Settings Bayesian models:
  - 100,000 iterations, 10,000 burn-in
  - Polygenic effect in Bayes models (not in BLUP)

LIVESTOCK RESEARCH
WAGENINGEN UR

CRV

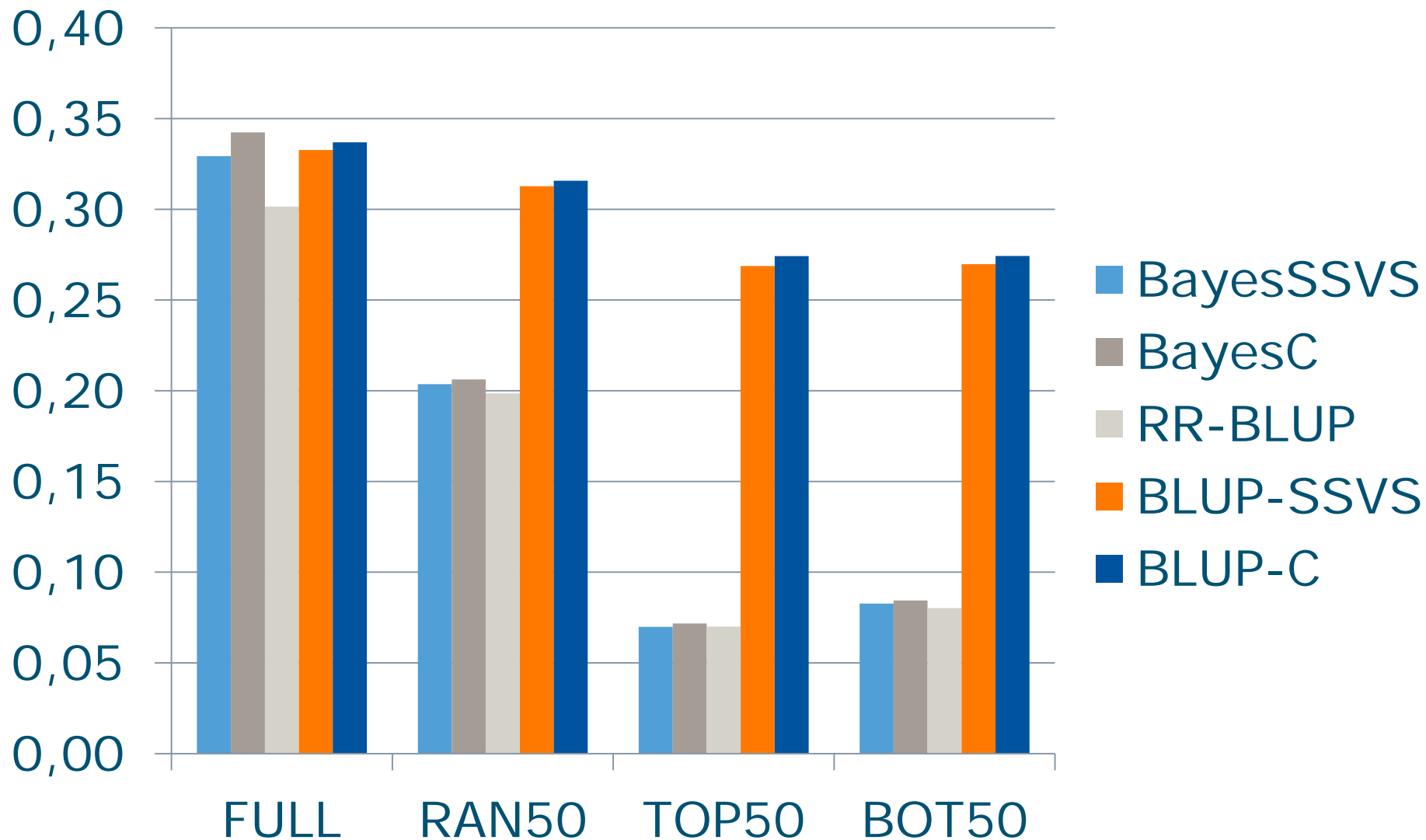Animal Breeding &
Genomics Centre

Dairy cattle data (CRV)

- 2 traits (kg protein & udder depth)

- 4,245 - 4,271 bulls in training

- 730 bulls in validation

- 37,995 SNPs

- Settings Bayesian models:
  - BayesSSVS; 30,000 iterations & 5,000 burn-in
  - BayesC;   60,000 iterations & 10,000 burn-in
  - 3 chains per analysis
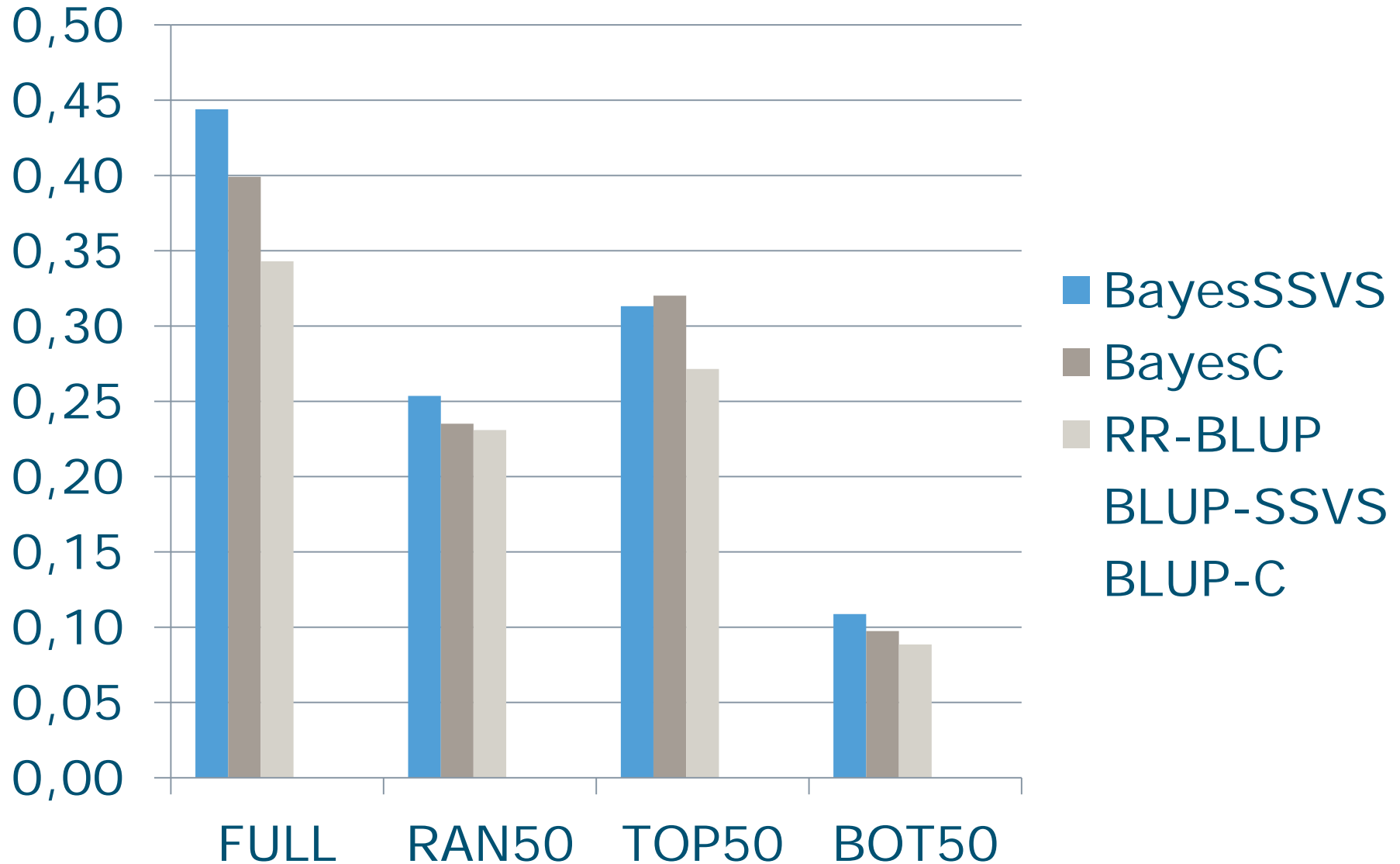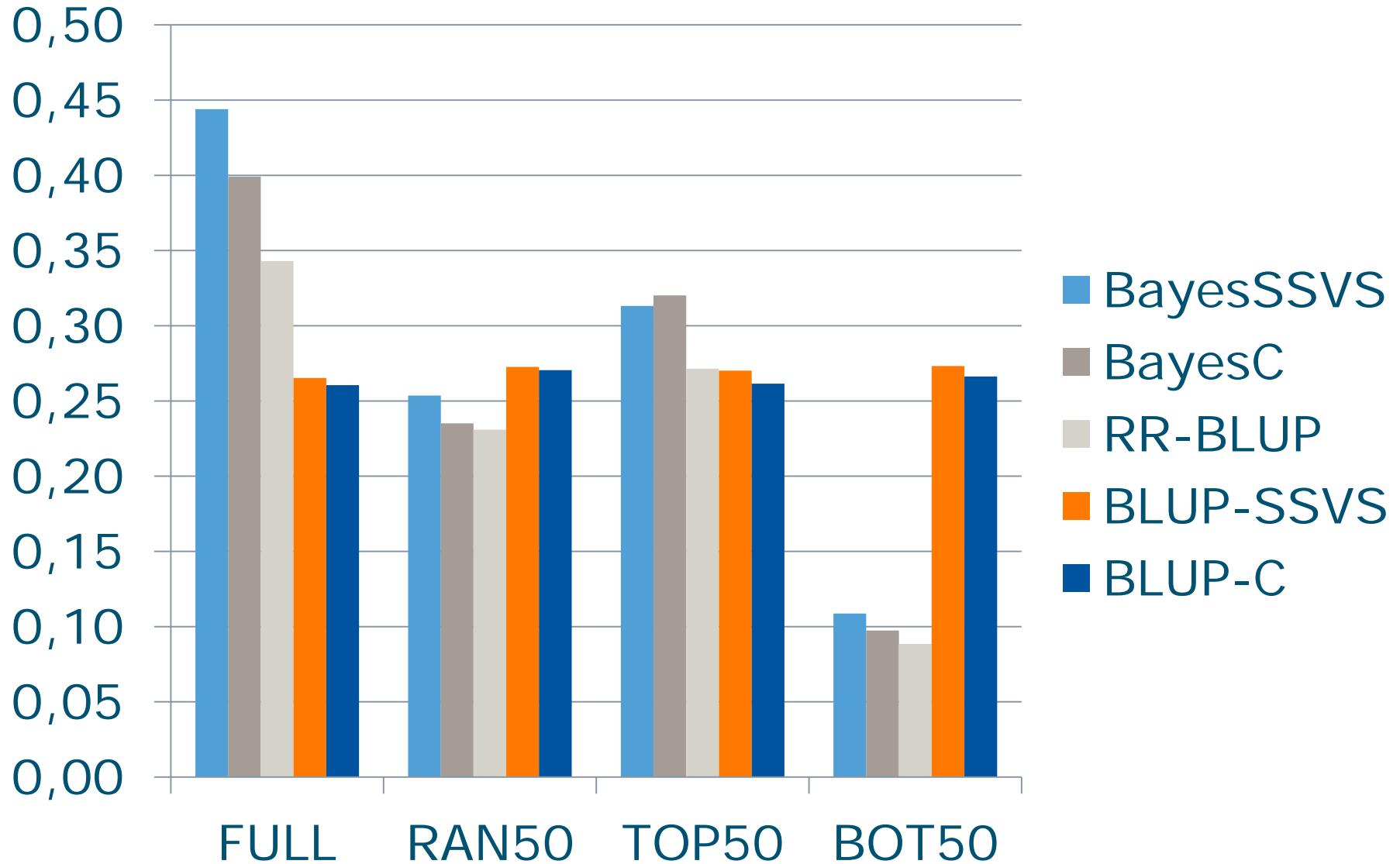 - Polygenic effect in Bayes models (not in BLUP)

LIVESTOCK RESEARCH
WAGENINGEN UR

CRV

Animal Breeding &
Genomics Centre

Reliabilities (R²) *simulated* data

Reliabilities (R²) *simulated* data

Reliabilities (R^2) dairy cattle – kg protein

Legend: BayesSSVS, BayesC, RR-BLUP, BLUP-SSVS, BLUP-C

Categories: FULL, RAN50, TOP50, BOT50

Reliabilities (R²) dairy cattle – kg protein

Reliabilities (R^2) dairy cattle – udder depth

BLUP models – Convergence (#iterations)

Using different SNP variances:

Scenario	Model	Simulated data			Dairy cattle data		
		Av.	Min.	Max.	Av.	Min.	Max.
FULL	BLUP-SSVS	7	7	7	217	43	397
	BLUP-C	7	7	7	286	43	544
TOP50	BLUP-SSVS	7	7	7	335	42	628
	BLUP-C	7	7	7	283	49	519

Using non-random subsets of data:

Scenario	Model	Simulated data			Dairy cattle data		
		Av.	Min.	Max.	Av.	Min.	Max.
FULL	RR-BLUP	7	7	7	435	393	477
TOP50	RR-BLUP	21332	921	67917	510	440	583

Convergence criterion = 1.E-06

LIVESTOCK RESEARCH
WAGENINGEN UR

CRV

Animal Breeding &
Genomics Centre

Results summarized

- Bayesian versus BLUP counterpart models:
    - Equivalent in simulated data
    - But *not* in real cattle data

$\Rightarrow$ Poor performance BLUP models in cattle data needs further investigation

- Impact using "suboptimal" SNP variances in BLUP:

Data	Accuracy	Convergence
Simulated	--	0
Dairy cattle	0	0

Conclusions

BLUP models are highly efficient

Simulated data:

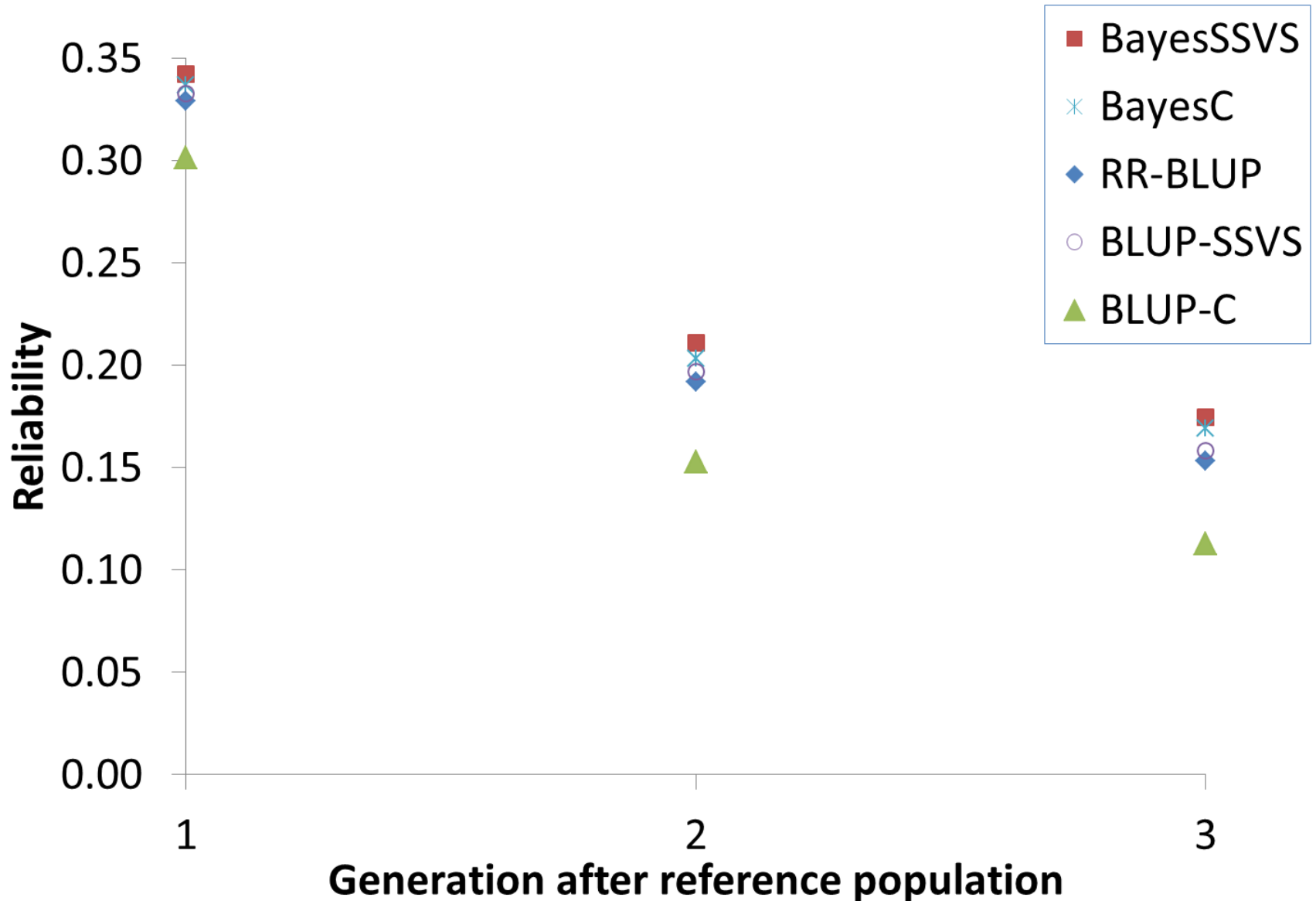- The BLUP models capture same features as variable selection methods

Dairy cattle data:

- Bayes SSVS significantly outperforms other methods
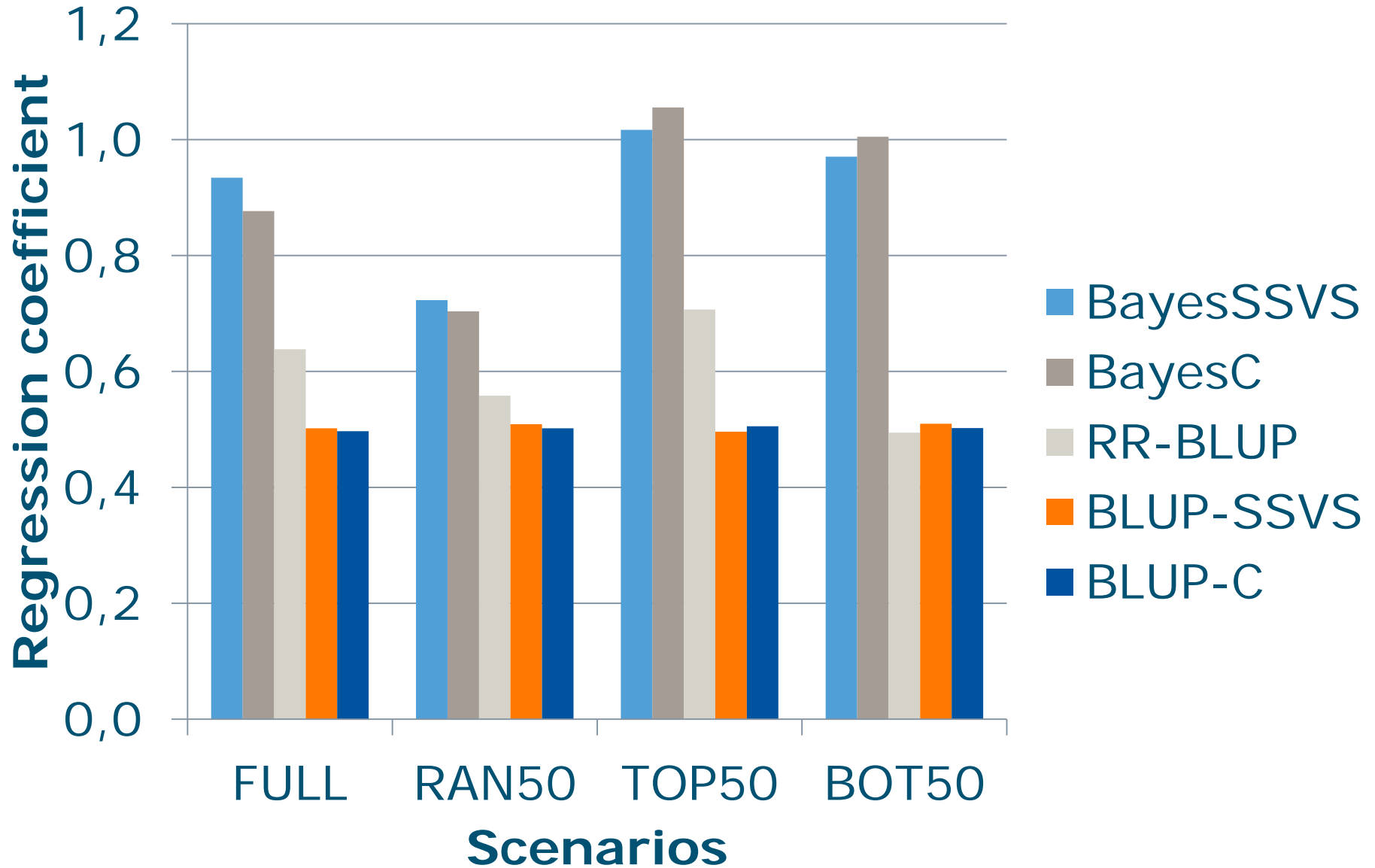- BLUP models are inferior *=> Needs further investigation*

LIVESTOCK RESEARCH
WAGENINGEN**UR**

CRV

Animal Breeding &
Genomics Centre

Acknowledgements (Funding)

LIVESTOCK RESEARCH
WAGENINGEN UR

Animal Breeding &
Genomics Centre

Results – using all (simulated) data

Kg protein

Legend:
- BayesSSVS
- BayesC
- RR-BLUP
- BLUP-SSVS
- BLUP-C

X-axis (Scenarios): FULL, RAN50, TOP50, BOT50

Y-axis (Regression coefficient): 0,0 — 1,2

Udder depth