

# Genome-Wide Breeding Value Estimation with Elastic Net

Xijiang Yu and Theo Meuwissen

Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, 1432, Ås, Norway



## OBJECTIVES

To test the Elastic Net on GWEBV estimation with simulated and real data.

## INTRODUCTION

- ▶ Genomic-wide breeding value (GWEBV) estimation is often facing with the problem of  $p \gg n$ 
  - ▷ I.e., the number of genotype values to be estimated ( $p$ ) is great greater than the number of phenotypic observations ( $n$ ).
- ▶ Several regularization methods were proposed to solve such problems, for example:
  - ▷ The Bayesian Lasso (Tibshirani, 1996).
  - ▷ The Elastic Net (Zou et al, 2005).

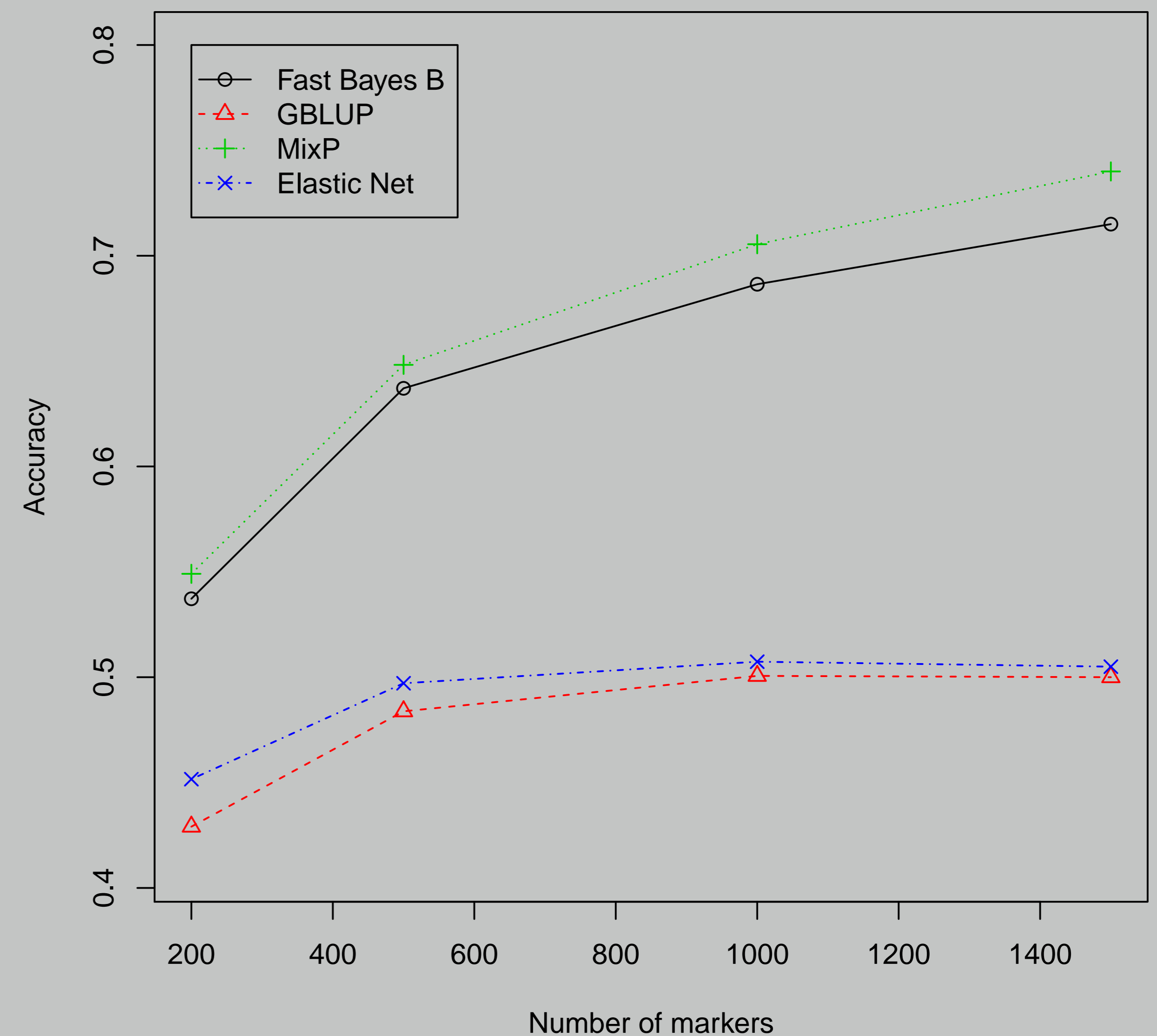
## MATERIALS

- ▶ Simulation
  - ▷ Based on a real pedigree of Norwegian Red Cattle
    - ▶ 8 generations
    - ▶ 19,523 individuals.
  - ▷ Ideal populations
    - ▶ Effective size  $N_e = 200$ .
    - ▶ One 1M chromosome,  $10^8$  base pairs.
    - ▶ Mutation rate  $10^{-8}/M/\text{meiosis}$ .
    - ▶ Random mating of  $10^4$  generations.
  - ▷ Genotypes of ideal populations were dropped through founders into the real pedigree.
  - ▷ Newest 250 individuals are as validation set.
  - ▷ 1,915 bulls are as training set.
  - ▷ 200 repeats.
- ▶ Wheat data set (Crossa et al, 2010).
  - ▷ 599 wheat lines, 1,447(1,279) markers
  - ▷ 10 repeats of random 10-fold cross-validation

## METHODS

- ▶ The elastic net methods
  - ▷ Data ( $\mathbf{X}, \mathbf{y}$ ).  $\mathbf{X}$  is the  $n \times p$  predictor matrix of standardized variables; and  $\mathbf{y}$  is the response vector.
 
$$\hat{\beta} = \arg \min_{\beta} \|\mathbf{y} - \mathbf{X}\beta\|^2 + \lambda_2 \|\beta\|^2 + \lambda_1 \|\beta\|_1$$
- ▶ Other methods for comparison
  - ▷ Fast Bayes B (Meuwissen et al, 2009)
  - ▷ GBLUP (Meuwissen et al, 2001)
  - ▷ MixP (Yu et al, 2011)

## RESULTS: Simulation Data



**comparison of accuracy of Fast-Bayes-B, GBLUP, MixP and Elastic Net**  
 Number of QTL simulated per chromosome is 5, which explain 2% of total phenotypic variance.  
 Accuracy is measured with the correlation coefficient between GWEBV and true values.

## RESULTS: Wheat Data

	Yield	1	2	3	4
Fast Bayes B	0.52	0.49	0.38	0.44	
GBLUP	0.53	0.50	0.39	0.46	
MixP	0.53	0.50	0.40	0.46	
Enet	0.53	0.50	0.38	0.45	

Accuracy is measured with correlation coefficient between GWEBV and phenotypic values.  
 No significant accuracy difference of different methods was observed.

## CONCLUSIONS

- ▶ With few QTL, the elastic net is better than GBLUP, but less accurate than MixP and Fast Bayes B.
- ▶ Its computation speed is at the same order of GBLUP and MixP.
- ▶ No accuracy difference between estimation methods was found in real wheat data.

## REFERENCES

- ▶ Crossa J et al., Genetics 2010, 186(2):713-724
- ▶ Meuwissen TH et al., Genetics 2001, 157(4):1819-29
- ▶ Meuwissen TH et al., Genet Sel Evol 2009, 41:2
- ▶ Tibshirani R, J R Statist Soc B 1996, 58(1):267-288
- ▶ Zou H et al., J R Statist Soc B 2005, 67(2):301-320

## ACKNOWLEDGMENT

- ▶ Norwegian research council
- ▶ GENO AS
- ▶ CIMMYT's Global Wheat Program