

EAAP 2019 – 70th Annual Meeting, Ghent, Belgium, 26 August 2019

Generalised sparse-group lasso for whole-genome regression and genomic selection

Jan Klosa¹, Noah R. Simon², Volkmar Liebscher³, Dörte Wittenburg¹

¹ Leibniz Institute for Farm Animal Biology, Dummerstorf, Germany

² University of Washington, Seattle, WA, USA

³ University of Greifswald, Greifswald, Germany



LEIBNIZ INSTITUTE
FOR FARM ANIMAL BIOLOGY

DFG Deutsche
Forschungsgemeinschaft
German Research Foundation

Member of
Leibniz
Leibniz
Association

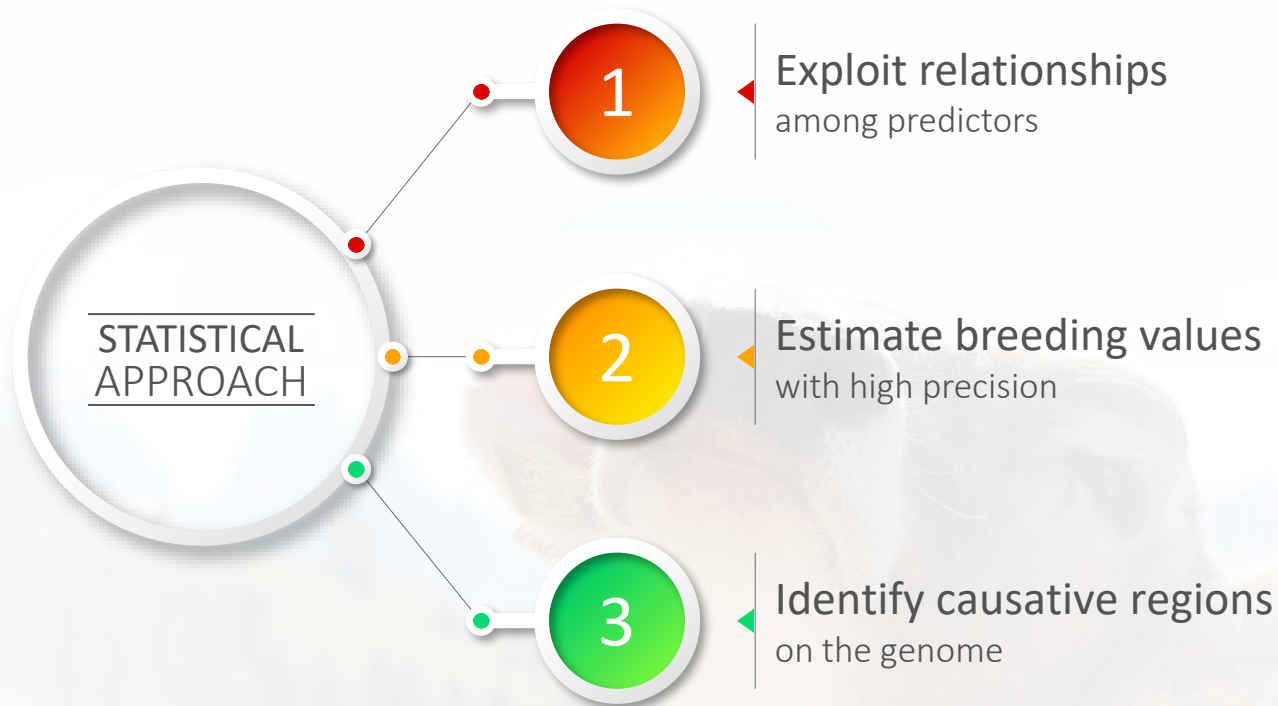
BACKGROUND

Phenotype:
Prediction



Genotype + environment:
Estimation

- “ $p > n$ ”
- Selection and shrinkage approach



WHOLE - GENOME REGRESSION

MODEL

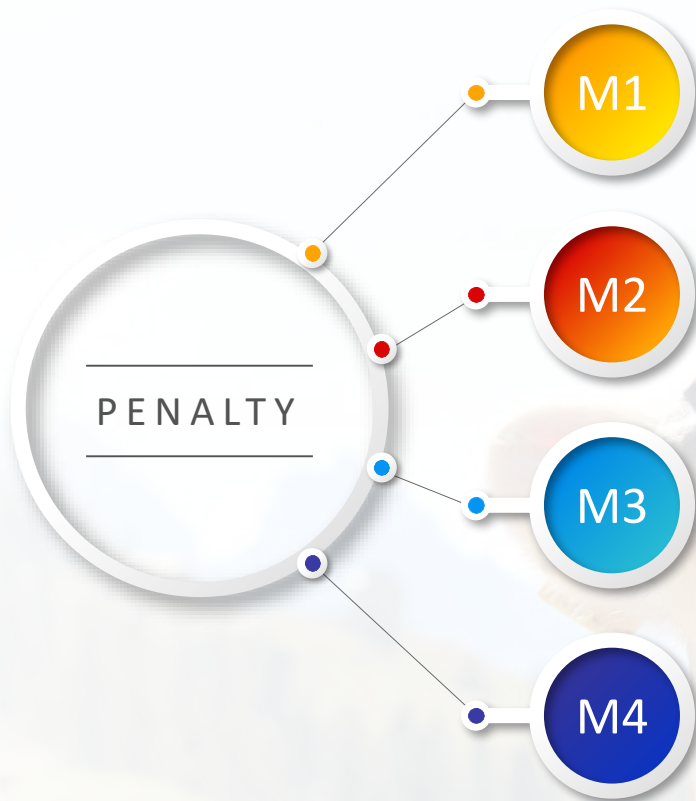
$$y = X\beta + e$$

y phenotype vector
 X marker-genotype matrix
 β vector of additive-genetic effects
 e vector of i.i.d. residuals

OPTIMI- SATION

$$\hat{\beta} = \arg \min_{\beta} \left\{ \frac{1}{2n} \|y - X\beta\|_2^2 + \text{penalty} \right\}$$

(n -invariant formulation)

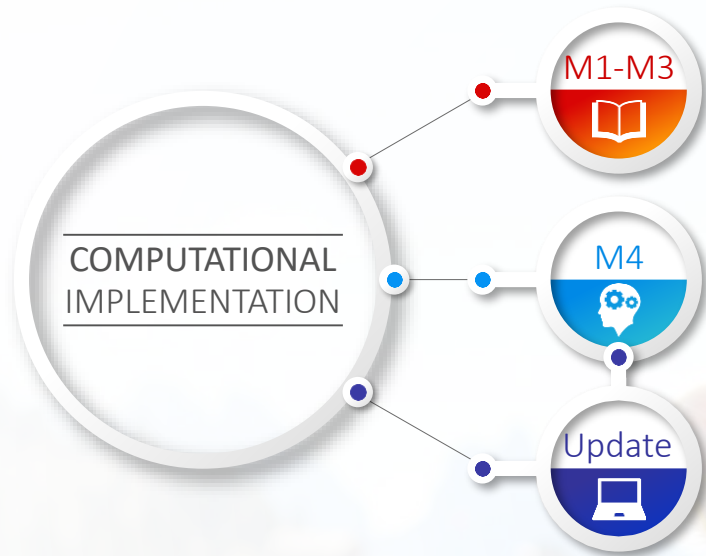


lasso: (Tibshirani, 1996)
 $\lambda \|\beta\|_1$

group lasso: (Yuan and Lin, 2006)
 $\lambda \sum_m \sqrt{p_m} \|\beta_m\|_2$

sparse-group lasso: (Simon et al., 2013)
 $\alpha \lambda \|\beta\|_1 + (1 - \alpha) \lambda \sum_m \sqrt{p_m} \|\beta_m\|_2$

generalised sparse-group lasso:
 $\alpha \lambda \|\mathbf{D}\beta\|_1 + (1 - \alpha) \lambda \sum_m \sqrt{p_m} \|\mathbf{E}_m \beta_m\|_2$



Proximal gradient descent

(backtracking line search, warm starts)

ADMM

(new variables: $\gamma = D\beta$, $\delta = E\beta$, w , new parameter ρ)

- ▶ Straight forward for $\beta^{(k+1)}$, $\gamma^{(k+1)}$, and $w^{(k+1)}$
- ▶ Assumption for $\delta^{(k+1)}$: group structure in $(D + E)\beta$
- ▶ $D = I$
- ▶ E block diagonal, $E_m^T E_m = (I_{m \times m} K)^{-1}$, K ... covariance matrix

1000 individuals

10 families
100 half sibs per family

466 SNPs

2 chromosomes
233 SNPs per chromosome

100 experiments



3 different numbers of QTL:

1, 3, or 9 groups per chromosome

3 traits with different heritabilities

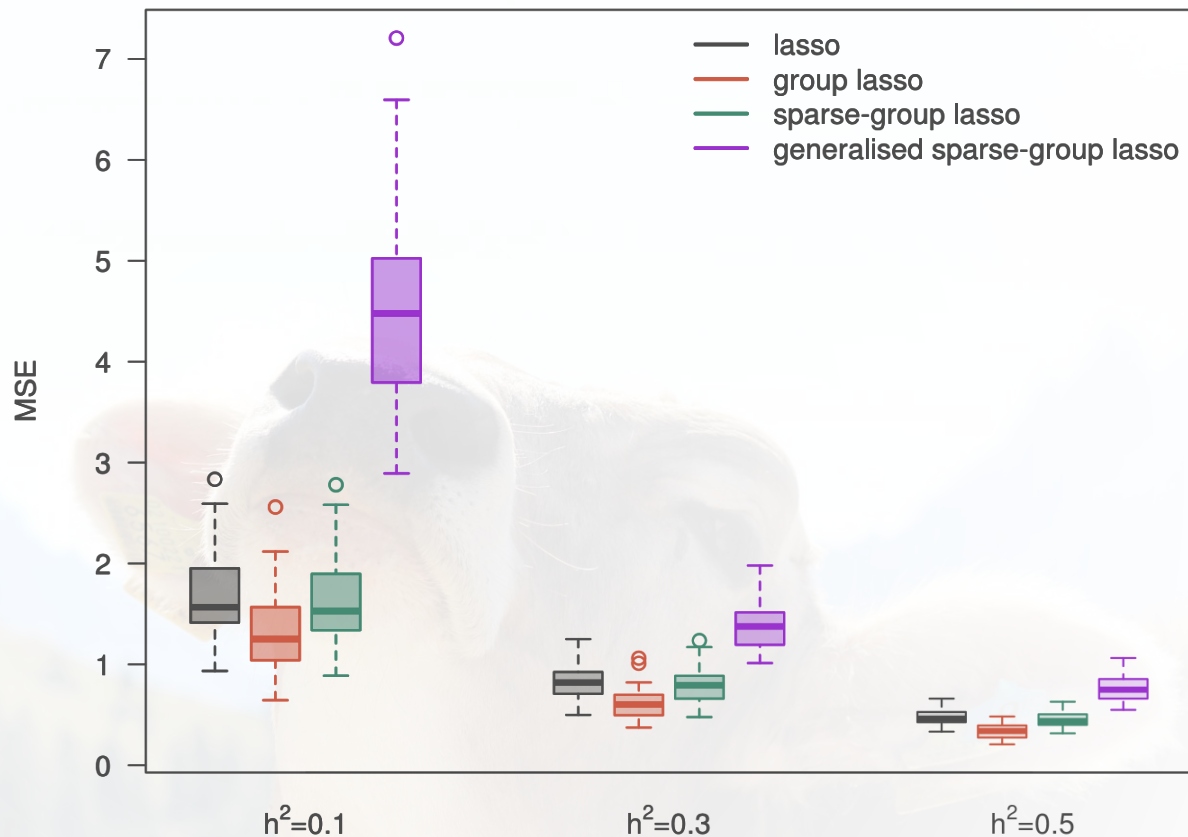
$h^2 \in \{0.1, 0.3, 0.5\}$

RESULTS MEAN SQUARED ERROR

1 group of QTL
per chromosome

Equal effect sizes
within groups
("simplified scenario")

$\alpha = 0.95$

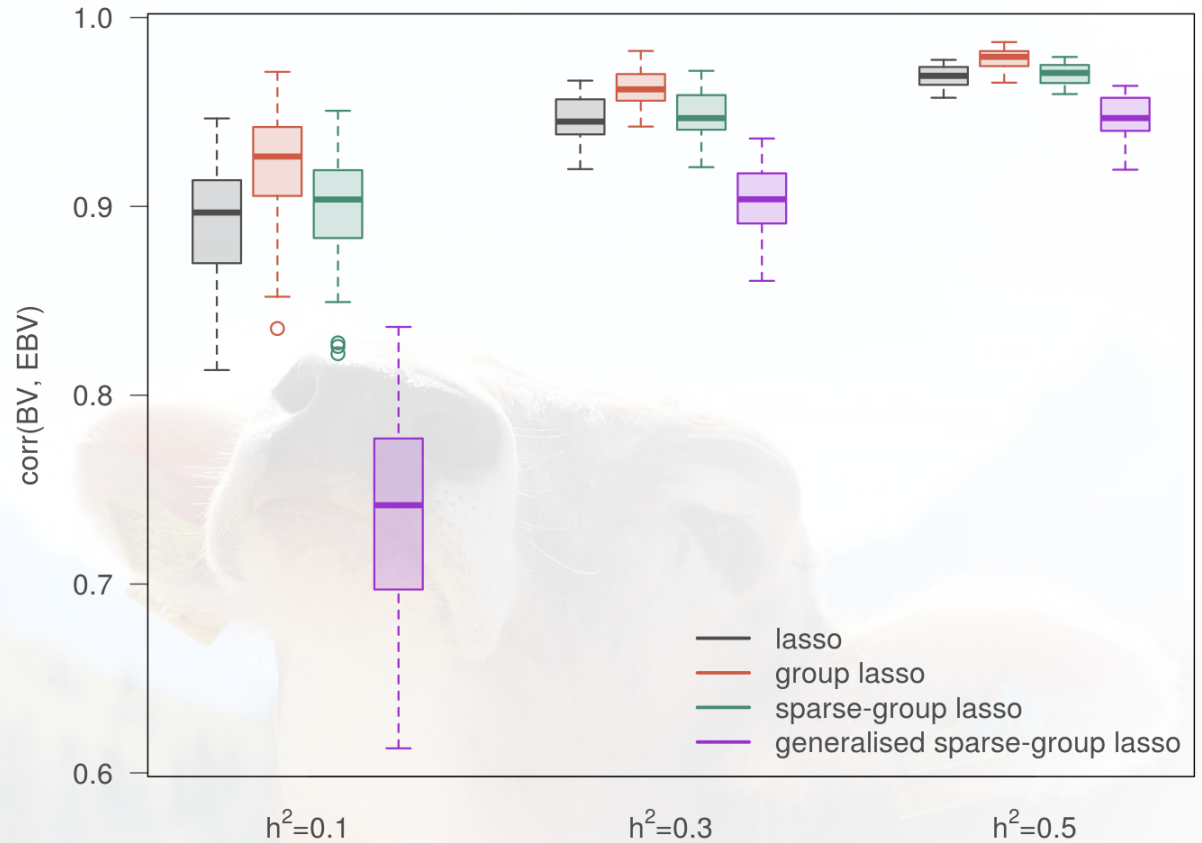


CORRELATION BREEDING VALUES

1 group of QTL per chromosome

Equal effect sizes within groups ("simplified scenario")

$\alpha = 0.95$



DISCUSSION
AND OUTLOOK

Thank you very much.

Gene
Chan

the grouping:
age

$$\alpha(D\beta - \gamma) + (1 - \alpha)(E\beta - \delta) = 0$$

$$\rightarrow \begin{pmatrix} D \\ E \end{pmatrix} \beta - \begin{pmatrix} \gamma \\ \delta \end{pmatrix} = 0$$

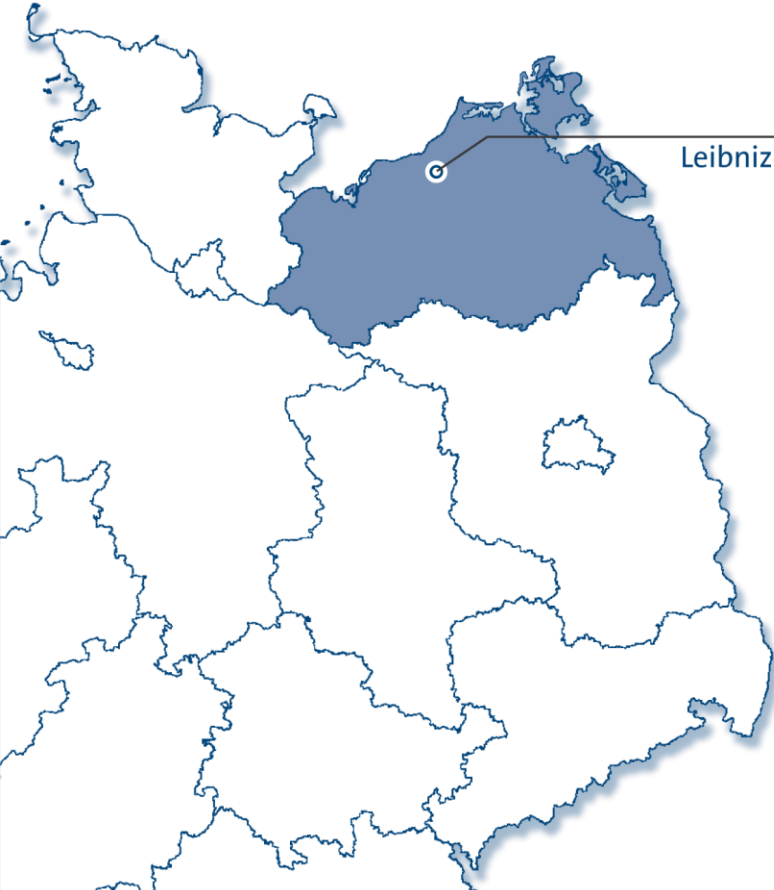
Decoupled updates for γ and δ

Theoretical correlation
(Wittenburg et al., 2016)





LEIBNIZ INSTITUTE
FOR FARM ANIMAL BIOLOGY



Dummerstorf

Leibniz Institute for Farm Animal Biology FBN

Leibniz Institute for Farm Animal Biology FBN

Wilhelm-Stahl-Allee 2
18196 Dummerstorf
Germany

Contact

Jan Klosa

Phone: +49 38208 68 918

Fax: +49 38208 68 902

E-Mail: klosa@fbn-dummerstorf.de

Internet: www.fbn-dummerstorf.de