

# Improving genomic prediction by weighted G-BLUP

Guosheng Su, Ole F. Christensen, Luc Janss and Mogens S. Lund



# Background

- **Bayesian VSM often predict more accurate GEBV than G-BLUP**
- **G-BLUP has a low computational demand**
- **Weighted G-BLUP may achieve both**
- **Various weighting factors for building G-matrix (estimated variance, effect, P-value)**

# Objective

- **Find optimal weights and strategy for weighted G-BLUP**
  - Assessing alternative weighting factors from Bayesian VSM output
  - Weighing single markers or intervals
  - The time intervals when weights need to be updated.

# Data

- **5221 progeny-tested Nordic Holstein bulls genotyped with the 54K chip**
- **Bulls born after 2004 as validation data (~20%)**
- **De-regressed proofs (DRP) derived from genetic evaluations in January 2013.**
- **Traits: milk, fat, protein, fertility and mastitis.**

# Model comparisons

## • Unweighted G-BLUP

- $G = MDM'$ ,
- $m_{ij} = 0 - 2p_j, 1 - 2p_j$  and  $2 - 2p_j$
- $$d_{jj} = \frac{1}{n_m 2p_j(1 - p_j)}$$

## • Bayesian VSM

## • Weighted G-BLUP

- $G = MD^*M'$ ,  $D^* = DT$
- $T$  is a diagonal matrix of weights  $t_{jj}$  derived from Bayesian VSM
- Weight was standardized to be mean weight=1

# Bayesian VSM

$$y = 1\mu + Xq + e$$

- $\mathbf{x}_j = \mathbf{m}_j / \sqrt{2p_j(1-p_j)}$
- $m_j$  with element of  $0-2p_j$ ,  $1-2p_j$  or  $2-2p_j$ .
- $q_j \sim \pi_1 N(0, \sigma_1^2) + \pi_2 N(0, \sigma_2^2) + \pi_3 N(0, \sigma_3^2) + \pi_4 N(0, \sigma_4^2)$

# Weighted G-BLUP

- **Weights for G-BLUP from Bayesian VSM**

- Posterior variance of SNP effects ( $WV_q = \sum_{i=1}^4 \hat{\pi}_i \hat{\sigma}_i^2$ )
- Square of posterior mean of SNP effect ( $Wq^2$ )
- P values from a t-test for SNP effect  $WP_q = -\log_{10}(P_{qj})$

- **Data used to derive weights**

- Bulls born before 2005, 2004, 2002, 2000

- **Number of markers in weighted interval**

- 1, 5, 10, 30, 50, 70, 100, 150

# Unweighted G-BLUP and BVSM

Trait	Reliability	
	G-BLUP	BVSM
Milk	0.483	0.516
Fat	0.468	0.508
Protein	0.462	0.478
Fertility	0.446	0.451
Mastitis	0.395	0.404
Mean	0.451	0.471



# Unweighted G-BLUP and Bayesian VSM

Trait	Reliability		Regression coefficient	
	G-BLUP	BVSM	G-BLUP	BVSM
Milk	0.483	0.516	0.872	0.878
Fat	0.468	0.508	0.842	0.830
Protein	0.462	0.478	0.814	0.817
Fertility	0.446	0.451	0.980	0.970
Mastitis	0.395	0.404	0.900	0.902
Mean	0.451	0.471	0.882	0.879

# Reliability

Weight	M_1	M_10	M_30	M_50	M_100
$WV_q$	0.464	0.467	0.468	0.467	0.466
$Wq^2$	0.446	0.453	0.456	0.456	0.454
$WP_q$	0.457	0.459	0.460	0.460	0.459
Mean	0.456	0.459	0.460	0.459	0.458

# Regression coefficient of DRP on GEBV

Weight	M_1	M_10	M_30	M_50	M_100
$WV_q$	0.832	0.871	0.878	0.880	0.882
$Wq^2$	0.761	0.826	0.852	0.867	0.870
$WP_q$	0.822	0.869	0.879	0.882	0.884
Mean	0.818	0.856	0.867	0.872	0.874

# Reliabilities with different intervals of weight calculations

Trait	Years from prediction to weight calculation			
	0	1	3	5
<b>Milk</b>	0.511	0.511	0.508	0.506
<b>Fat</b>	0.505	0.505	0.508	0.499
<b>Protein</b>	0.472	0.470	0.469	0.467
<b>Fertility</b>	0.449	0.448	0.446	0.445
<b>Mastitis</b>	0.401	0.402	0.403	0.403
<b>Mean</b>	0.468	0.467	0.467	0.464

# Conclusions

- **Weighted G-BLUP improves the accuracy of genomic prediction**
- **Posterior variances of marker effects from a Bayesian VSM are appropriate weights**
- **A common weight on intervals reduces bias**
- **Weights can be updated once per three years**