Genomic prediction using a G-matrix weighted with SNP variances from Bayesian mixture model

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

- Bayesian variable selection models often predict more accurate genomic breeding value (GEBV) than GBLUP
- GBLUP has low computational demand
- One approach to combine the advantages of both models is to use weighted G-matrix (Zhang et al. 2010)
- Various weighting factors for building G-matrix
Objective

- Achieve the benefits of both GBLUP and Bayesian variable selection models by using weighted G-matrix
- Assess weighting factors and methods to build G-matrix.
- Investigate the time-interval to update weights
Statistical models and methods

1. **GBLUP**: GBLUP model using original G-matrix

   \[ y = \mathbf{1}\mu + \mathbf{Z}g + e \]

   \( g \) is the vector of additive genetic effects, \( g \sim N(0, \mathbf{G}\sigma_g^2) \)

   \( \mathbf{G} = \mathbf{M}\mathbf{M}'/n_{\text{marker}} \) (VanRaden 2008)

   \[ m_j = \frac{0 - 2p_j}{\sqrt{2p_jq_j}}, \frac{1 - 2p_j}{\sqrt{2p_jq_j}}, \frac{2 - 2p_j}{\sqrt{2p_jq_j}} \]

   for genotype \( A_1A_1, A_1A_2 \) and \( A_2A_2 \), \( p=p_{A_2} \)
2. BMIX: Bayesian mixture model assuming a mixture of 4 distributions for SNP effects (Gao et al. 2013)

\[ y = 1\mu + Mq + e \]

**q** is the vector of SNP effects.

\[ q_i \sim \pi_1 N(0, \sigma_1^2) + \pi_2 N(0, \sigma_2^2) + \pi_3 N(0, \sigma_3^2) + \pi_4 N(0, \sigma_4^2) \]

\[ \pi_1 = 0.889, \pi_2 = 0.1, \pi_3 = 0.01, \pi_4 = 0.001 \]

\[ \sigma_1^2 < \sigma_2^2 < \sigma_3^2 < \sigma_4^2 \]
3. \textbf{GBLUP}_w: GBLUP model using weighted G-matrix to account for heterogeneous variances of SNP effects

\[ y = 1\mu + Zg + e \]

\( g \) is the vector of additive genetic effects, \( g \sim N(0, G_w\sigma_g^2) \)

\( G_w = MTM'/n_{\text{marker}} \)

\( T \) is a diagonal matrix with \( t_{jj} = \) weight on SNP \( j \)
Weighting factors

1. Square of individual SNP effect from the BMIX ($Wq^2$)

2. Posterior variance of individual SNP from the BMIX model ($WV_q$), calculated as $\sum \pi_i \sigma_i^2$

Final weight is standardized so that mean weight = 1
Weighting methods

1. Single-marker weighting: one weight for one marker

2. Group-marker weighting: one common weight to a group of markers

   A common weight = mean weight of the makers in the same group

   Group sizes: 5, 10, 30, 50, 70, 100, and 150
Data analysis

Marker and Phenotypic Data – Nordic Holsteins

Marker data: 54K chip

Phenotypic data: De-regressed proofs (DRP) of milk, fat, protein, fertility, mastitis

Reference data: About 3950 bulls born before 2005

Test data: About 1150 bulls born in 2005 and later
Data to derive weight

1. Lag0y: The same as the reference data (before 2005) used for current prediction

2. Lag1y: Bulls born before 2004, i.e., 1 year before current prediction

3. Lag3y: Bulls born before 2002, i.e., 3 years before current prediction

4. Lag5y: Bulls born before 2000, i.e., 5 years before current prediction
Reliability of GEBV and regression coefficient of DRP on GEBV, obtained from GBLUP and BMIX4

<table>
<thead>
<tr>
<th>Trait</th>
<th>Reliability (%)</th>
<th>Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GBLUP</td>
<td>BMIX4</td>
</tr>
<tr>
<td>Milk</td>
<td>48.3</td>
<td>51.6</td>
</tr>
<tr>
<td>Fat</td>
<td>46.8</td>
<td>50.8</td>
</tr>
<tr>
<td>Protein</td>
<td>46.2</td>
<td>47.8</td>
</tr>
<tr>
<td>Fertility</td>
<td>44.6</td>
<td>45.1</td>
</tr>
<tr>
<td>Mastitis</td>
<td>39.5</td>
<td>40.4</td>
</tr>
<tr>
<td>Mean</td>
<td>45.1</td>
<td>47.1</td>
</tr>
</tbody>
</table>

BMIX gave 2.0% higher reliability
Reliability of GEBV from GBLUP with original (UNW) or weighted ($WV_q$ and $Wq^2$) G- matrices, averaged over 5 traits

$WV_q$ lead to reliability 1.7% higher than original GBLUP
Regression of DRP on GEBV from GBLUP with original (UNW) or weighted (WV\(_q\) and Wq\(^2\)) G-matrices, averaged over 5 traits

\[ M1 = a \text{ weight on a single marker}, \quad M5 \sim M150 = a \text{ common weight on 5 } \sim 150 \text{ markers} \]

WV\(_q\) leads to less bias than Wq\(^2\)
Group weighting has less bias than single-marker weighting
Reliability of GEBV from GBLUP with G-matrix weighted by \( WV_q \) and a common weight on a group of 30 markers, derived from 0, 1, 3, or 5 year old data

<table>
<thead>
<tr>
<th>Traits</th>
<th>Lag 0</th>
<th>Lag 1y</th>
<th>Lag 3y</th>
<th>Lag 5y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milk</td>
<td>51.1</td>
<td>51.1</td>
<td>50.8</td>
<td>50.6</td>
</tr>
<tr>
<td>Fat</td>
<td>50.5</td>
<td>50.5</td>
<td>50.8</td>
<td>49.9</td>
</tr>
<tr>
<td>Protein</td>
<td>47.2</td>
<td>47.0</td>
<td>46.9</td>
<td>46.7</td>
</tr>
<tr>
<td>Fert</td>
<td>44.9</td>
<td>44.8</td>
<td>44.6</td>
<td>44.5</td>
</tr>
<tr>
<td>Masti</td>
<td>40.1</td>
<td>40.2</td>
<td>40.3</td>
<td>40.3</td>
</tr>
<tr>
<td>Mean</td>
<td>46.8</td>
<td>46.7</td>
<td>46.7</td>
<td>46.4</td>
</tr>
</tbody>
</table>

Weight derived from 3y old data does not reduce reliability
Regression of DRP on GEBV using a G-matrix weighted by WVq derived from 0, 1, 3, or 5 year old data

<table>
<thead>
<tr>
<th>Trait</th>
<th>Group-marker(30) weight</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lag 0</td>
</tr>
<tr>
<td>Milk</td>
<td>0.87</td>
</tr>
<tr>
<td>Fat</td>
<td>0.82</td>
</tr>
<tr>
<td>Protein</td>
<td>0.82</td>
</tr>
<tr>
<td>Fert</td>
<td>0.98</td>
</tr>
<tr>
<td>Masti</td>
<td>0.90</td>
</tr>
<tr>
<td>Mean</td>
<td><strong>0.88</strong></td>
</tr>
</tbody>
</table>

Weights derived from old data did not increase bias
Conclusions

- Posterior variance of individual SNP from BMIX is a good alternative weighting factor.
- A common weight on a group of about 30 markers could be a good weighting method.
- Weights can be updated once per two or three years.

Thank you for your attention.