Study of Single-Step SNP BLUP in a Japanese Holstein Population

T.Osawa¹ & J.Saburi¹

¹National Livestock Breeding Center, Fukushima, 961-8511, Japan
t0ohsawa@nlbc.go.jp(Corresponding Author)

Summary

Single-step SNP BLUP (ssSNP) with a residual polygenic effect is theoretically equivalent to single-step genomic BLUP (ssGBLUP). Our objectives were to implement ssSNP in a Japanese Holstein population and to compare the ssSNP and ssGBLUP. Data from 2,221,477 cows (4,126,586 305-days milk yield records for the first three lactations) were used. There were 9,786 sires with daughters; 4,080 of the sires were genotyped. The dataset also included 1,588 young bulls. 42,275 SNP markers were considered. A repeatability model was used. ssSNP used values of 0.5, 0.4, 0.3, and 0.2 for the residual polygenic variance parameter (k). Pearson correlations between ssSNP and ssGBLUP and regression slopes of ssGBLUP on ssSNP for SNP effects were lower than expected. These results are thought because many small estimates of SNP effects are noisy. Genotyped sires with daughters showed high correlations when k was 0.3 or more. k=0.2 resulted in slightly lower correlations. In the case of young bulls, correlations and regression slopes were lower than those for reference animals. k=0.2 in ssSNP resulted in unexpected estimates of SNP effects and genomic EBV. This result may be due to insufficient number of reference animals.

Keywords: single-step SNP BLUP, residual polygenic effect, genomic evaluation, dairy cattle

Introduction

Currently, genetic evaluations of dairy cattle in many countries include large scale SNP marker information, which is known as genomic evaluation. Most countries have applied a multi-step model for the genomic evaluation of dairy cattle (VanRaden., 2008), which utilizes deregressed EBV obtained via prior conventional genetic evaluation. The multi-step model is relatively easy to implement in conventional genetic evaluation systems. The major advantage is that only selected animals, which are reliable progeny-tested bulls, can be included in the reference population. However, the separate steps of estimating SNP effects and EBV during conventional evaluation cannot fully account for genomic preselection; therefore, genomic EBV (GEBV) will be biased (Patry and Ducrocq., 2011). In contrast to the multi-step model, the single-step genomic BLUP (ssGBLUP) can evaluate genotyped and non-genotyped animals jointly, thus providing unbiased predictive value (Aguilar et al., 2010). However, the ssGBLUP do not directly estimate SNP effects. Liu et al. (2014 & 2016) developed a single-step SNP BLUP (ssSNP) with a residual polygenic effect that directly estimates the SNP effects. The single-step GBLUP and ssSNP are equivalent in theory.

The objectives of our study were to implement ssSNP in a Japanese Holstein population and to compare the ssSNP and ssGBLUP.

Materials and Methods
The dataset comprised 4,126,586 305-day milk yield records in the first three lactations of 2,221,477 cows in a Japanese Holstein population. There were 9,786 sires with daughters; of those, 4,080 sires were genotyped. The dataset also included 1,588 young bulls. In this analysis, 42,275 SNP markers were considered. Variance components used were permanent environmental variance \( \sigma^2_D = 277,730 \), additive genetic variance \( \sigma^2_A = 533,170 \), and residual variance \( \sigma^2_E = 758,850 \).

**Single-step SNP BLUP**

The ssSNP used in this analysis is based on the method proposed by Liu et al. (2014 & 2016). A repeatability model was used as follows:

\[
y = Xb + Z_p p + W u + e,
\]

where \( y \) is a vector of 305-day milk yields, \( b \) is a vector of the fixed effects of herd-calving year-parity, calving age, and region-calving month-calving year, \( p \) is a vector of permanent environmental effects of cows, \( u \) is a vector of additive genetic effects, \( e \) is a vector of residual effects. \( X, Z_p, \) and \( W \) are design matrices for effects \( b, p, \) and \( u \), respectively.

Two groups of animals are assumed; group 1, without genotyped animal (\( u_1 \)) and group 2, with genotyped animal (\( u_2 \)). Thus, additive genetic effects of the genotyped animals are obtained by \( \hat{u}_2 = Z_g + \hat{u}_1 \), where, \( g \) is a vector of additive genetic effects of the SNP markers, \( \hat{a}_2 \) is residual polygenic effects, and \( Z \) is a design matrix of regression coefficients on genotyped animals at all SNP markers: 2-2\( p_j \), 1-2\( p_j \), or -2\( p_j \), for genotype AA, AB, or BB of the \( j \)th SNP marker (VanRaden, 2008), where \( p_j \) represents allele frequency of the \( j \)th SNP marker. A (co)variance structure of the SNP marker effects is \( \text{var}(g) = B \sigma^2_g \), where \( B \) is \( \sum_{j=1}^{\infty} 2p_j(1-p_j)I^{-1} \).

Liu et al. (2014) showed that two equations (A1 and A2; Appendix) solved the ssSNP. To estimate SNP effects, equation (A2) should include reference animals only. Liu et al. (2014) proposed a filter matrix \( F = diag\{1,0,0,\ldots,1,1,0\} \) to define genotyped animals as reference animals (=1) or not (=0). The \( F \) matrix insert can be inserted into equation (A2):

\[
\left( \frac{1}{1-z}B^{-1} + \frac{1}{z}Z^tFA_{zz}^{-1}Z \right)_{zz}^{-1} = \frac{1}{z}Z^tFA_{zz}^{-1}\hat{a}_2,
\]

where \( A_{zz}^{-1} \) is the inverse of the pedigree relationship matrix including only genotyped animals, and \( k \) is the residual polygenic variance parameter. In this study, \( k=0.5, 0.4, 0.3, \) and 0.2 were used.

**Computation procedure**

The ssSNP needs to iterate equations (A1) and (2) to obtain the final solutions. In our study, Step-1: equation (A1) is solved by preconditioned conjugate gradients. Step-2: substituting \( \hat{u}_2 \) calculated in Step-1 for right-hand-side (RHS) in equation (2), equation (2) is solved by...
multiplying the inverse matrix of the left-hand-side (LHS) by RHS. Step-3: the genomic term of RHS in equation (A1) corrected with \( \hat{\mathbf{G}} \) calculated in Step-2, and returns to Step-1. For effect estimates of equation (A1) and equation (2), convergence criteria were defined as the sum of squared differences in the effect estimates between two consecutive rounds of iteration divided by the sum of squares of the estimates from the current round of iteration.

Single-step GBLUP

GEBV and SNP effects in ssGBLUP were estimated by BLUPF90 and POSTGSF90 (Aguilar et al., 2010 & 2014). The following scaling parameters were applied to the \( \mathbf{H}^{-1} \) matrix:

\[
\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & (\alpha \mathbf{G} + \beta \mathbf{A}_u)^{-1} - \mathbf{A}_u^{-1} \end{bmatrix},
\]

where, \( \mathbf{G} \) is the genomic relationship matrix, \( \alpha \) and \( \beta \) are scaling parameters, with \( \alpha=0.5; \beta=0.5, \alpha=0.6; \beta=0.4, \alpha=0.7; \beta=0.3, \) and \( \alpha=0.8; \beta=0.2. \)

Results and Discussion

Comparison of SNP effects

Pearson correlations between ssSNP and ssGBLUP, and regression slope coefficients of ssGBLUP on ssSNP for SNP effects are shown in Table 1. Pearson correlations and regression slopes were lower than expected. In addition, \( k=0.2 \) exhibited much lower correlations than the other values, and the SNP effects resulted in unexpected estimates. This may be because many small estimates of SNP effects are noisy.

\[\text{Table 1. Pearson correlations between ssSNP}^1 \text{ and ssGBLUP}^2 \text{ and regression slope coefficients of ssGBLUP on ssSNP for SNP effects.}\]

<table>
<thead>
<tr>
<th>( \alpha ) ( \beta )</th>
<th>( k=0.5 )</th>
<th>( k=0.4 )</th>
<th>( k=0.3 )</th>
<th>( k=0.2 )</th>
<th>( k=0.5 )</th>
<th>( k=0.4 )</th>
<th>( k=0.3 )</th>
<th>( k=0.2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5, 0.5</td>
<td>0.635</td>
<td>0.604</td>
<td>0.561</td>
<td>0.345</td>
<td>0.464</td>
<td>0.519</td>
<td>0.570</td>
<td>0.692</td>
</tr>
<tr>
<td>0.6, 0.4</td>
<td>0.651</td>
<td>0.621</td>
<td>0.578</td>
<td>0.352</td>
<td>0.471</td>
<td>0.530</td>
<td>0.584</td>
<td>0.702</td>
</tr>
<tr>
<td>0.7, 0.3</td>
<td>0.660</td>
<td>0.633</td>
<td>0.591</td>
<td>0.356</td>
<td>0.467</td>
<td>0.527</td>
<td>0.582</td>
<td>0.693</td>
</tr>
<tr>
<td>0.8, 0.2</td>
<td>0.661</td>
<td>0.637</td>
<td>0.598</td>
<td>0.357</td>
<td>0.446</td>
<td>0.505</td>
<td>0.561</td>
<td>0.662</td>
</tr>
</tbody>
</table>

\(^1\) single-step SNP BLUP  
\(^2\) single-step genomic BLUP

Comparison of GEBV

Genotyped sires with daughters, which are the reference animals, showed high correlations when \( k \) was 0.3 or more. \( k=0.2 \) resulted in slightly lower correlations (Table 2). When the size of the \( k \) parameter decreased, except for \( k=0.2 \), the regression slopes tended to decrease. This seemed to be due to the unexpected estimates of SNP effects at \( k=0.2 \). In the case of young bulls, correlations and regression slopes were lower than those for reference animals (Table
Our results were not entirely consistent. This was because the number of reference animals may not have been sufficient. In Japan, there are about 5,000 foreign genotyped sires with multiple-across country evaluation (MACE) EBV without domestic daughters. In the future, inclusion of those sires should be considered.

Table 2. Pearson correlations between ssSNP$^1$ and ssGBLUP$^2$ and regression slope coefficients of ssGBLUP on ssSNP for genotyped sires with daughter and young bulls.

<table>
<thead>
<tr>
<th>ssGBLUP</th>
<th>Pearson correlation</th>
<th>Regression slope</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$k=0.5$</td>
<td>$k=0.4$</td>
</tr>
<tr>
<td>$\alpha=0.5$, $\beta=0.5$</td>
<td>0.997</td>
<td>0.994</td>
</tr>
<tr>
<td>$\alpha=0.6$, $\beta=0.4$</td>
<td>0.997</td>
<td>0.994</td>
</tr>
<tr>
<td>$\alpha=0.7$, $\beta=0.3$</td>
<td>0.996</td>
<td>0.994</td>
</tr>
<tr>
<td>$\alpha=0.8$, $\beta=0.2$</td>
<td>0.995</td>
<td>0.993</td>
</tr>
</tbody>
</table>

**Acknowledgments**

I am grateful to Zengting Liu (VIT, Germany) for the advice provided on the implementation of single-step SNP BLUP.

**List of References**


APPENDIX

Liu et al., (2014) used equations (A1) and (A2) to solve single-step SNP BLUP with a residual polygenic effect:

\[
\begin{bmatrix}
X'R^{-2}X & X'R^{-2}Z_p & X'R^{-2}W_1 & X'R^{-2}W_x \\
Z_p'R^{-1}X & Z_p'R^{-1}Z_p + \sigma_{e}^2I & Z_p'R^{-1}W_1 & Z_p'R^{-1}W_x \\
W_1'R^{-2}X & W_1'R^{-2}Z_p & W_1'R^{-2}W_1 + \sigma_{e}^2A^{11} & \sigma_{e}^2A^{12} \\
W_x'R^{-2}X & W_x'R^{-2}Z_p & \sigma_{e}^2A^{21} & W_x'R^{-2}W_x + \sigma_{e}^2\left(A^{22} + \left(\frac{1}{2} - 1\right)A^{-1}_n\right)
\end{bmatrix}
\begin{bmatrix}
b \\
p \\
u_1 \\
u_2
\end{bmatrix} = \begin{bmatrix}
X'y \\
Z_p'y \\
W_1'y \\
W_x'y + \sigma_{e}^2A^{11}Z_p'
\end{bmatrix}, \quad \text{(A1)}
\]

\[
\left(\frac{1}{1-k}B^{-1} + \frac{1}{k}Z'\sigma_{A_{n}}^{-1}Z\right)\hat{u}_2 = \frac{1}{k}Z'\sigma_{A_{n}}^{-1}\hat{u}_2, \quad \text{(A2)}
\]

where \(\hat{b}\) is a vector of fixed effects, \(\hat{p}\) is a vector of permanent environmental effects of the cow, \(\hat{u}_1\) is a vector of additive genetic effects of the non-genotyped animals, \(\hat{u}_2\) is a vector of additive genetic effects of the genotyped animals, \(\sigma_{e}^2\) is a permanent environmental variance, \(\sigma_{g}^2\) is an additive genetic variance, \(\sigma_{r}^2\) is a residual variance, \(X\) is a design matrix for effect \(\hat{b}\), \(Z_p\) is a design matrix for \(\hat{p}\), \(W_1\) is a design matrix for effect \(\hat{u}_1\), \(W_2\) is a design matrix for effect \(\hat{u}_2\), \(R^{1}\) is \(\sigma_{g}^{-2}I\), \(A^{11}, A^{12}, A^{21}\), and \(A^{22}\) are submatrices of the inverse of the pedigree relationship matrix \(A^{-1}\) that include all animals, and subscripts 1 and 2 refer to groups of non-genotyped and genotyped animals, respectively. \(A^{-1}_{n}\) is the inverse of the pedigree relationship matrix \(A_{22}\) including only genotyped animals, \(Z\) is a design matrix of regression coefficients on genotyped animals at all SNP markers \((2-2p_j, 1-2p_j, -2p_j, \text{ for genotype AA, AB, or BB of the } j\text{th SNP marker}), B \) is \(\left(\sum_{j=1}^m 2p_j[1 - p_j]D^{-1}\right)^{-1}\) and \(k\) is a parameter of residual polygenic variance.