Combination of genotype, pedigree, local and foreign information

J. Vandenplas1, F.G. Colinet1, N. Gengler1,*

1 University of Liège, Gembloux Agro-Bio Tech, Belgium
2 National Fund for Scientific Research, Brussels, Belgium

Introduction

✓ For genomic prediction
  - To get reliable GEBV use of multiple sources needed
✓ Traditional traits (i.e., in dairy cattle)
  - Local genomic evaluations ➡ MACE
✓ Novel traits (e.g., milk quality, feed efficiency, methane)
  - Combining different sources of often locally sparse phenotypic data even more needed

Aims

✓ To develop strategies to use simultaneously:
  - Genotypes,
  - Pedigree,
  - Local phenotypes and,
  - Foreign phenotypes
✓ By combining and adapting:
  - Single-step genomic evaluations (ssGBLUP) and,
  - Bayesian procedure to integrate external information*


Aims

✓ To develop strategies to use simultaneously:
  - Genotypes,
  - Pedigree,
  - Phenotypes from multiple sources

ssGBAYES
ssGBLUP

- Single-step genomic evaluation (ssGBLUP)
  - Allows direct combination of genomic, pedigree and phenotypic information
- However: current limitation
  - Only available local phenotypic records can be used

ssGBLUP

- Single-step genomic evaluation (ssGBLUP)
  - Allows direct combination of genomic, pedigree and phenotypic information
- However: current limitation
  - Only available local phenotypic records can be used

In opposition to multi-step methods (e.g., use of local-EBV and MACE-EBV in prediction equation step)

Modified MME

\[
\begin{bmatrix}
X'X & X'Z \\
Z'X & Z'Z + G^{-1}
\end{bmatrix}
\begin{bmatrix}
\hat{\beta}_M \\
\hat{u}_l
\end{bmatrix} =
\begin{bmatrix}
X'Y_L \\
Z'Y_L
\end{bmatrix}
\]

- \( G^{-1} = H^{-1} \otimes G_s^{-1} \): inverse of combined genomic-pedigree based (co)variances matrix
- \( Y_L \): vector of local phenotypes
- \( \hat{\beta}_M \): vector of estimated local fixed effects
- \( \hat{u}_l \): vector of estimated local GEBV
- \( p(u_l) = MVN(0, G^{-1}) \)
**Alternative**

- As alternative to introduce phenotypes
  - Also allowing for foreign information
- However avoiding double counting
  - E.g. MACE-EBV contain our phenotypes
- Also avoiding deregression
  - Potential source of trouble
- Be simple and flexible
  - Allowing to extend to multiple sources

**Bayesian priors**

- Assumption: *a priori* known information on $\hat{u}_{i}^{*}$
  - $y_{i}$ replaced by $\hat{u}_{i}$ and $D_{i}$
  - Source of phenotypic information
    - Vector of local EBV: $\hat{u}_{i}$ = “external” information
    - Prediction error (co)variances matrix: $D_{i}$
- Allows use of simplified models
  - *E.g.*, Test-day model $\rightarrow$ lactation EBV

**Concept of “Hybrid” ssGBLUP**

$$
\begin{bmatrix}
X'R^{-1}X & X'R^{-1}Z \\
Z'R^{-1}X & Z'R^{-1}Z + G^{-1}
\end{bmatrix}
\begin{bmatrix}
\hat{\beta}_{i} \\
\hat{u}_{i}^{*}
\end{bmatrix}
= 
\begin{bmatrix}
X'R^{-1}y_{i} \\
Z'R^{-1}y_{i}
\end{bmatrix}
$$

$$(G^{-1} + \Lambda_{i})\hat{u}_{i}^{*} = D_{i}^{-1}\hat{u}_{i}$$

$\rightarrow$ More details later

**“Hybrid” ssGBLUP $\Rightarrow$ ssGBAYES**

- Using only local information as source
  - $X'R^{-1}X$ $X'R^{-1}Z$ $\begin{bmatrix}
\hat{\beta}_{i} \\
\hat{u}_{i}^{*}
\end{bmatrix}
= 
\begin{bmatrix}
X'R^{-1}y_{i} \\
Z'R^{-1}y_{i}
\end{bmatrix}
$$

$$(G^{-1} + \Lambda_{i})\hat{u}_{i}^{*} = D_{i}^{-1}\hat{u}_{i}$$

Least square part of LHS of hypothetical BLUP generating local information

RHS of hypothetical BLUP generating local information

*“external” = outside of original system of MME*
Bayesian priors

✓ Assumption: a priori known information on $\hat{u}_l^*$
  - Extended to 2 sources of phenotypic information
    - Vector of local EBV: $\hat{u}_l$ (1st "external" information)
    - Vector of foreign EBV: $\hat{u}_f$ (2nd "external" information)
    - Prediction error (co)variances matrices: $D_L$, $D_F$

✓ Issue: only available for some animals
  - $\hat{u}_l$, $\hat{u}_f$, $D_L$, and $D_F$: (partially) unknown

* "external" = outside of original system of MME

Methods

✓ For both sources: estimation of $\hat{u}_i$ (i=L,F)
  - Available
    - EBV of some animals (so-called "external" $\hat{u}_i$ )
  - "Internal" animals: prediction of EBV ($\hat{u}_i$)
    $\mathbf{P}(\hat{u}_i | \hat{u}_i) = \mathcal{MVN}(\mathbf{G}_u \hat{u}_i | \mathbf{G}_u^\top (\mathbf{G}_i)^{-1})$
    $\mathbf{G} = \mathbf{A} \otimes \mathbf{G}_u$: genetic (co)variances matrix
  - $\hat{u}_i = [\hat{u}_i \hat{u}_i]$

  ➤ Correct propagation of information

Methods

✓ For both sources: estimation of $\hat{D}_i$ (i=L,F)
  - $\hat{D}_i = \mathbf{G}^{-1} + \Lambda_i$
  - $\Lambda_i$ = block diag($\Delta_j \mathbf{R}_y \Delta_j$); $j = 1,...,n$ animals
    - For external animals: $\Delta_j = \text{diag}(\sqrt{\mathbf{RE}}_k); k = 1,...,t$ traits
    - For internal animals: $\Delta_j = \mathbf{0}$

  ➤ All matrices $\Lambda_i$ depend only on contributions due to own records

Issue

✓ Non-Independence of information sources
  - E.g., Local information included in MACE-EBV

    $\mathbf{D}_1^{-1}\hat{u}_f = \mathbf{D}_1^{-1}\hat{u}_f - \mathbf{D}_1^{-1}\hat{u}_L$

    ➤ Estimation of external information free from local information
Methods

✓ Combination of genotype, pedigree, local and foreign information (ssGBayes)

\[
\left(G^{-1} + \Lambda_l + \Lambda_f - \Lambda_{lc}\right)\hat{u}_l^{**} = D_l^{-1}\hat{u}_l + D_{f}^{-1}\hat{u}_f - D_{lc}^{-1}\hat{u}_{lc}
\]

Least square part of LHS of hypothetical BLUP of local information

Least square part of LHS of hypothetical BLUP of foreign information

RHS of hypothetical BLUP of local information

RHS of hypothetical BLUP of foreign information free of local information

No double counting of contributions

Example: Walloon genomic evaluation

✓ ssGBAYES results for milk yield
  ❏ But already used for all INTERBULL traits

✓ 1,909 genotyped Holstein bulls and cows

✓ 16,234 animals (genotyped and ancestors)
  ❏ 12,046 animals with Walloon EBV (EBV\textsubscript{w})
  ❏ 1,981 bulls with MACE EBV (EBV\textsubscript{m})
  ❏ 601 bulls with Walloon EBV contributing to MACE (EBV\textsubscript{wc})

✓ Reliabilities (REL) for GEBV obtained through inversion of left-hand side

Results

✓ Averaged REL (SD) associated to EBV\textsubscript{w}, GEBV\textsubscript{w} and GEBV\textsubscript{w+m-wc} for genotyped bulls for milk yield

<table>
<thead>
<tr>
<th>Class of REL\textsubscript{w}</th>
<th>N</th>
<th>EBV\textsubscript{w}\textsuperscript{1} (SD)</th>
<th>No. of publishable bulls</th>
<th>GEBV\textsubscript{w}\textsuperscript{2} (SD)</th>
<th>No. of publishable bulls</th>
<th>GEBV\textsubscript{w+m-wc}\textsuperscript{3} (SD)</th>
<th>No. of publishable bulls</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 0.50</td>
<td>647</td>
<td>0.25 (0.32)</td>
<td>0</td>
<td>0.44 (0.09)</td>
<td>194</td>
<td>0.80 (0.06)</td>
<td>628</td>
</tr>
<tr>
<td>0.50 – 0.74</td>
<td>273</td>
<td>0.63 (0.07)</td>
<td>173</td>
<td>0.69 (0.06)</td>
<td>173</td>
<td>0.87 (0.05)</td>
<td>173</td>
</tr>
<tr>
<td>≥ 0.75</td>
<td>390</td>
<td>0.90 (0.07)</td>
<td>390</td>
<td>0.91 (0.06)</td>
<td>390</td>
<td>0.94 (0.04)</td>
<td>390</td>
</tr>
</tbody>
</table>

1 REL obtained from Walloon polygenic evaluation

2 REL obtained from Walloon genomic evaluation using only EBV\textsubscript{w}

3 REL obtained from Walloon genomic evaluation using EBV\textsubscript{w}, EBV\textsubscript{m} and EBV\textsubscript{wc}

Results

✓ Increase of REL for genotyped animals without MACE results and sired by genotyped bulls with MACE results
Conclusions

✓ Applied to Walloon dairy genomic evaluations
   - Bayesian approach integrates well MACE results into ssGBLUP
   - Recover large amount of phenotypic information
   - More accurate predictions for genotyped animals and their progeny
   - Correct propagation of all available information

Further implications ssGBayes

✓ No deregression
   - Direct use of EBV from multiple sources
✓ Applicable to multi-trait models
   - E.g., external information for correlated and/or predictor traits
✓ Open general framework, can be modified to accommodate latest genomic models, e.g.:
   - GWAS models based on ssGBLUP
   - SNP based single-step models

General notation

✓ Combining “s” sources of information

\[
(G^{-1} + \sum_{i=1,s} A_i) \hat{u}_c = \sum_{i=1,s} D_i^{-1} \hat{u}_i
\]

Least square part of LHS of hypothetical BLUP
RHS of hypothetical BLUP for information source i

→ Potential to improve current genomic prediction strategies

Thank you for your attention

✓ Acknowledgments for financial support
   - Service Public of Wallonia SPW – DGO3 through project D31-1274 DairySNP
   - National Fund for Scientific Research and Wallonie Brussels International for scientific stays

✓ Acknowledgments
   - Animal Breeding Association for providing data
   - CECI for computational resources
   - Animal and Dairy Science Department, University of Georgia, Athens, USA
   - Animal Science Department, University of Ljubljana, Slovenia

Thank you for your attention

Acknowledgments for financial support

- Service Public of Wallonia SPW – DGO3 through project D31-1274 DairySNP
- National Fund for Scientific Research and Wallonie Brussels International for scientific stays

Acknowledgments

- Animal Breeding Association for providing data
- CECI for computational resources
- Animal and Dairy Science Department, University of Georgia, Athens, USA
- Animal Science Department, University of Ljubljana, Slovenia