Integration of MACE breeding values into the multiple-trait single-step random regression test-day genetic evaluation for yield traits of Australian red breeds

Interbull meeting, Montreal, 2022

Vinzent Boerner\textsuperscript{1,2}, Thuy Nguyen\textsuperscript{3}, and Gert Nieuwhof\textsuperscript{3}

\textsuperscript{1}GHPC CONSULTING AND SERVICES PTY. LTD., Australia
\textsuperscript{2}Centre for Quantitative Genetics and Genomics, Aarhus University, Denmark
\textsuperscript{3}DataGene Limited, Australia

30/05/2022
Why integration

Country A
- many off-spring (high-performing)
- additional off-spring
  - high parent-average
  - high selection probability

Bull XYZ
Why integration

**Country A**
- many off-spring (high-performing)
- additional off-spring
  - high parent-average
  - high selection probability

**Country B**
- few/no off-spring (still high-performing)
- maybe genotyped
- additional/first off-spring
  - inferior parent-average
  - low selection probability
Why integration

Bull XYZ

Country A
- many off-spring (high-performing)
- additional off-spring
  - high parent-average
  - high selection probability

Unhappy breeder

Country B
- few/no off-spring (still high-performing)
- maybe genotyped
- additional/first off-spring
  - inferior parent-average
  - low selection probability
Why integration

**Country A**
- many off-spring (high-performing)
- additional off-spring
  - high parent-average
  - high selection probability

**Country B**
- few/no off-spring (still high-performing)
- maybe genotyped
- additional/first off-spring
  - inferior parent-average
  - low selection probability

Unhappy breeder

Genetic progress

Bull XYZ
Why integration

**Country A**
- many off-spring (high-performing)
- additional off-spring
  - high parent-average
  - high selection probability

**Country B**
- few/no off-spring (still high-performing)
- maybe genotyped
- additional/first off-spring
  - inferior parent-average
  - low selection probability
Why integration

Bull XYZ

Country A
- many off-spring (high-performing)
- additional off-spring
  - high parent-average
  - high selection probability
- enhanced information
- happy breeders
- genetic progress

Country B
- few/no off-spring (still high-performing)
  - maybe genotyped
- additional/first off-spring
  - inferior parent-average
  - low selection probability
multiple-trait random-regression test day model
multiple-trait random-regression test day model

- Nine traits → three traits in three lactations
  - Milk
  - Fat
  - Protein
- 27 genetic effects
  - Legendre polynomials 0, 1 and 2
multiple-trait random-regression test day model

- nine traits $\rightarrow$ three traits in three lactations
  - milk
  - fat
  - protein
- 27 genetic effects
  - Legendre polynomials 0, 1 and 2

\[
\begin{pmatrix}
  y_{l,t} \\
  \vdots \\
  y_{3,3}
\end{pmatrix} =
\begin{pmatrix}
  X_l \otimes I_3 \\
  \qquad \qquad 0
\end{pmatrix}
\begin{pmatrix}
  b_{l,t} \\
  \qquad \qquad 0
\end{pmatrix} +
\begin{pmatrix}
  f_0(K_l), \ldots, f_4(K_l) \otimes I_3 \\
  \qquad \quad \vdots \\
  f_0(K_3), \ldots, f_2(K_3) \otimes I_3
\end{pmatrix}
\begin{pmatrix}
  c_{l,t} \\
  \qquad \qquad c_{3,3}
\end{pmatrix}
\flushright
[\begin{pmatrix}
  u_{l,t,f_0}, \ldots, u_{3,3,f_2}
\end{pmatrix}] \sim N(0, \Gamma \otimes \Sigma_g)
\]

\[
\Gamma = \begin{pmatrix}
  \Theta & \Theta Q' \\
  Q\Theta & Q\Theta Q' + H
\end{pmatrix}
\]

- $H \rightarrow$ single step $H$ matrix
- $Q \rightarrow$ genetic group regression matrix
- $\Theta \rightarrow I \times 0.65$
Evaluation data set

- 10,000,614 observations across three traits and three lactations
  - Milk yield in litre
  - Fat yield in kg
  - Protein yield in kg
- Pedigree
  - 975,532 individuals
  - 73 phantom parents
- Genotypes
  - 8,191 animals
  - Various platforms
  - Imputed to common set of 73,000
Integration data set

Abulls

- 466 individuals
- Information provided to Interbull
- 212 genotyped
- Eligibility for integration:
  - \( \text{rel}_{\text{itb}} - \text{rel}_{\text{domestic}} > 0.01 \)
Integration data set

**Abulls**
- 466 individuals
- Information provided to Interbull
- 212 genotyped
- Eligibility for integration: \( r_{itb} - r_{domestic} > 0.01 \)

**Bbulls**
- 15,597 individuals
- No information provided to Interbull
- 116 genotyped
- Eligibility for integration: all
Integration data set

**Abulls**
- 466 individuals
- 212 genotyped
- Eligibility for integration:
  - \( rel_{itb} - rel_{domestic} > 0.01 \)

**Bbulls**
- 15,597 individuals
- 116 genotyped
- Eligibility for integration:
  - All

**Dimension problem:**
27 genetic effects vs 3 Interbull breeding values
Integration methodology

Global approach: pseudo data point with data-point specific residual variance as the single tuning parameter

Overcome dimension discrepancy:

- $\mathbb{R}_{27} \rightarrow \mathbb{R}_{9}$
  - $\Sigma_g^* = \Omega \Sigma_g \Omega'$
    - $\Sigma_g \rightarrow 27 \times 27$ genetic co-variance matrix
    - $\Omega \rightarrow 9 \times 27$ block matrix of Legendre polynomial coefficients
- $\mathbb{R}_{3} \rightarrow \mathbb{R}_{9}$
  - $u_{i,*} = Ku_{i,:}$
  - $r_{i,*} = Kr_{i,:}$
  - $p_{i,*} = (1 - r_{i,*}) \odot \text{diag}(\Sigma_g^*)$
    - $u_{i,:} \rightarrow$ vector of breeding values of animal $i$
    - $r_{i,:} \rightarrow$ vector of reliabilities of animal $i$
    - $p_{i,:} \rightarrow$ vector of prediction error variances of animal $i$
    - "." sent to Interbull(SENT) or received from Interbull(ITB)
    - $K \rightarrow 9 \times 3$ matrix of diagonal blocks
Integration methodology

Global approach: pseudo data point with data-point specific residual variance as the single tuning parameter

Derive a pseudo data point:

- $\text{diag}((D_{i,:} + \Pi_i)^{-1}) \equiv p_{i,:}$
  - iterative procedure if $\mathbb{R} > 1$
  - maybe >50k iterations (parameterization, $\Sigma$ radius, consistency of $p_i$)
- $D_{i,:} \approx R^{-1}$
- $\Pi_i$?
  - $C_{i,i} - C_{i,:}C_{i,:}^{-1}C_{i,:} \rightarrow$ exact but infeasible
  - $\Sigma_g^{-1} \rightarrow$ usually sufficient
Integration methodology

Global approach: pseudo data point with data-point specific residual variance as the single tuning parameter

Derive a pseudo data point:

**Abulls**
- $D_{i,R} = D_{i,ITB} - D_{i,SENT}$
- $y_i^* = D_{i,R}^{-1}((D_{i,ITB} + \Sigma_a^{*-1})u_{i,ITB}^* - (D_{i,SENT} + \Sigma_a^{*-1})u_{i,SENT}^*)$

**Bbulls**
- $y_i^* = D_{i,ITB}^{-1}((D_{i,ITB} + \Sigma_a^{*-1})u_{i,ITB}^*)$
Integration methodology

Global approach: pseudo data point with data-point specific residual variance as the single tuning parameter

Tweak MIX99 input:
- add $y^*$ to data
- add $D_{i,R}^{-1}(ABulls)$ or $D_{i,ITB}^{-1}(BBulls)$ to residual variance file
- set “DIM” for $y^*$ to 306
- add extra level associated to $y^*$ to all fixed classification effects
- add row 306 to regression table file $\rightarrow \int_0^{305} L_i$
Method evaluation

**ABulls**

- **expectation driver**
  - \( r_{ITB} > r_{SENT} \)
  - \( \begin{pmatrix} \sigma^2_{ITB} & \sigma^2_{SENT} \\ \sigma^2_{SENT} & \sigma^2_{SENT} \end{pmatrix} \)

- **expectation**
  - \( r_{blended} \approx r_{ITB} \)
  - \( u_{blended} \approx u_{ITB} \)

- **evaluation parameters**
  - \( \text{cor}(::blended, ::ITB) \)
  - \( \text{lm}(::blended \sim ::ITB) \)
**Method evaluation**

**ABulls**
- expectation driver
  - \( r_{ITB} > r_{SENT} \)
  - \( \begin{pmatrix} \sigma^2_{ITB} & \sigma^2_{SENT} \\ \sigma^2_{SENT} & \sigma^2_{SENT} \end{pmatrix} \)
- expectation
  - \( r_{blended} \approx r_{ITB} \)
  - \( u_{blended} \approx u_{ITB} \)
- evaluation parameters
  - \( \text{cor}(\text{blended}, \text{ITB}) \)
  - \( \text{lm}(\text{blended} \sim \text{ITB}) \)

**BBulls**
- expectation driver
  - \( \begin{pmatrix} \sigma^2_{ITB} & 0 \\ 0 & \sigma^2_{DOM} \end{pmatrix} \)
- expectation
  - \( r_{blended} \geq \max(r_{ITB}, r_{DOM}) \)
- evaluation parameters
  - \( ? \rightarrow \) nice plots
Results

- very similar for all three traits
- only shown for protein
Results: ABulls

breeding values

\[ y = 7.84 + 0.862 \times \quad R^2 = 0.94 \]

\[ y = 9.02 + 0.989 \times \quad R^2 = 0.99 \]

- genotyped
- not genotyped
Results: **ABulls**

**reliabilities**

\[ y = -0.296 + 1.3 x \quad R^2 = 0.93 \]

\[ y = 0.126 + 0.873 x \quad R^2 = 0.98 \]
Results: **BBulls**

breeding values

\[ y = 12.6 + 0.02x \quad R^2 = 0.01 \]

\[ y = 14.5 + 1.01x \quad R^2 = 0.98 \]
Results: **B Bulls**

**reliabilities**

- Pre-integration:
  \[ y = -0.195 + 0.399 \times R^2 = 0.05 \]

- Post-integration:
  \[ y = -0.0543 + 1.12 \times R^2 = 0.83 \]
Results: Population impact

**reliabilities**

<table>
<thead>
<tr>
<th>statistic</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>min</td>
<td>0</td>
</tr>
<tr>
<td>max</td>
<td>0.4118</td>
</tr>
<tr>
<td>no change</td>
<td>53.3%</td>
</tr>
<tr>
<td>&gt;0.01 increase</td>
<td>46.7%</td>
</tr>
<tr>
<td>&gt;0.05 increase</td>
<td>5.613%</td>
</tr>
<tr>
<td>&gt;0.1 increase</td>
<td>4.914%</td>
</tr>
<tr>
<td>&gt;0.25 increase</td>
<td>0.001876%</td>
</tr>
</tbody>
</table>

**breeding values**

<table>
<thead>
<tr>
<th>statistic</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>min</td>
<td>-30.8</td>
</tr>
<tr>
<td>max</td>
<td>27.79</td>
</tr>
<tr>
<td>&gt;1% change</td>
<td>84.17%</td>
</tr>
<tr>
<td>&gt;5% change</td>
<td>42.28%</td>
</tr>
<tr>
<td>&gt;10% change</td>
<td>28.6%</td>
</tr>
<tr>
<td>&gt;25% change</td>
<td>14.96%</td>
</tr>
</tbody>
</table>
Results: **Population impact**

### animal re-ranking

![Graph showing animal re-ranking with a correlation coefficient of ρ = 0.99. The graph displays the relationship between pre-integration rank and post-integration rank, with color indicating counts ranging from 5000 to 15000.](image-url)
Conclusions

- integration successful
- expectations met for ABulls and BBulls
- population wide impact of integrated data points
Conclusions

- integration successful
  - expectations met for ABulls and BBulls
  - population wide impact of integrated data points

- possible sources for evaluation bias
  - approximation of $C_{i,i} - C_{i,\neq i}C_{\neq i,\neq i}^{-1}C_{\neq i,i}$
  - $D_{i,:}$ ignores relationships between integration candidates
    - joint $D_{i,:} \rightarrow$ computational feasible?
  - possible negative impact on long-term genetic trend
    - assumption $\Sigma_{a,\text{ITB}} \equiv \Sigma_{a,\text{DOM}}$ may not hold
Conclusions

- integration successful
  - expectations met for ABulls and BBulls
  - population wide impact of integrated data points

- possible sources for evaluation bias
  - approximation of $C_{i,i} - C_{i,\neq i}C_{\neq i,\neq i}^{-1}C_{\neq i,i}$
  - $D_{i,:}$ ignores relationships between integration candidates
    - joint $D_{i,:} \rightarrow$ computational feasible?
- possible negative impact on long-term genetic trend
  - assumption $\Sigma_{a,ITB} \equiv \Sigma_{a,DOM}$ may not hold

- impact of genomics
  - sources of evaluation bias are equivalent to no-genomics
  - extent of evaluation bias $\uparrow$?!
THANK YOU!!!