

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^{1,2}, Thuy Nguyen³, and Gert Nieuwhof³

¹GHPC CONSULTING AND SERVICES PTY. LTD., Australia

²Centre for Quantitative Genetics and Genomics, Aarhus University, Denmark

³DataGene Limited, Australia

30/05/2022



Why integration

Bull XYZ

Country A

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

Why integration

Bull XYZ

```
graph TD; BullXYZ[Bull XYZ] --> CountryA[Country A]; BullXYZ --> CountryB[Country B];
```

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

Bull XYZ

Country A

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

unhappy breeder

genetic progress

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

information

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

- happy breeders
- genetic progress

informa

Country B

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

Evaluation model

multiple-trait random-regression test day model

$$\begin{pmatrix} y_{1,t} \\ \vdots \\ y_{3,3} \end{pmatrix} = \begin{pmatrix} 0 & 0 & X_3 \otimes I_3 \\ \vdots & \vdots & \vdots \\ 0 & 0 & X_3 \otimes I_3 \end{pmatrix} \begin{pmatrix} b_{3,3} \\ \vdots \\ b_{3,3} \end{pmatrix} + \begin{pmatrix} ([f_0(K_1), \dots, f_4(K_1)] \otimes I_3)(1_1 \otimes I_{15}) & 0 \\ 0 & \vdots \\ 0 & 0 ([f_0(K_3), \dots, f_4(K_3)] \otimes I_3)(1_3 \otimes I_{15}) \end{pmatrix} \begin{pmatrix} c_{1,t,f_0} \\ \vdots \\ c_{3,3,f_4} \end{pmatrix} + \begin{pmatrix} ([f_0(K_1), \dots, f_2(K_1)] \otimes I_3)(Z_1 \otimes I_9) & 0 \\ 0 & \vdots \\ 0 & 0 ([f_0(K_3), \dots, f_2(K_3)] \otimes I_3)(Z_3 \otimes I_9) \end{pmatrix} \begin{pmatrix} u_{1,t,f_0} \\ \vdots \\ u_{3,3,f_2} \end{pmatrix} + \begin{pmatrix} ([f_0(K_1), \dots, f_2(K_1)] \otimes I_3)(Z_1 \otimes I_9) & 0 \\ 0 & \vdots \\ 0 & 0 ([f_0(K_3), \dots, f_2(K_3)] \otimes I_3)(Z_3 \otimes I_9) \end{pmatrix} \begin{pmatrix} q_{1,t,f_0} \\ \vdots \\ q_{3,3,f_2} \end{pmatrix} + \begin{pmatrix} e_{1,t} \\ \vdots \\ e_{3,3} \end{pmatrix}$$

Evaluation 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

$$\begin{pmatrix} y_{1,t} \\ \vdots \\ y_{3,3} \end{pmatrix} = \begin{pmatrix} 0 & 0 & X_3 \otimes I_3 \\ \vdots & \vdots & \vdots \\ 0 & 0 & X_3 \otimes I_3 \end{pmatrix} \begin{pmatrix} b_{3,3} \\ \vdots \\ b_{3,3} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix} \begin{pmatrix} C_{1,t,f_0} \\ \vdots \\ C_{3,3,f_4} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix} \begin{pmatrix} U_{1,t,f_0} \\ \vdots \\ U_{3,3,f_2} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix} \begin{pmatrix} q_{1,t,f_0} \\ \vdots \\ q_{3,3,f_2} \end{pmatrix} + \begin{pmatrix} e_{1,t} \\ \vdots \\ e_{3,3} \end{pmatrix}$$

Evaluation 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

$$[u_{1,t,f_0}, \dots, u_{3,3,f_2}] \sim N(0, \Gamma \otimes \Sigma_g)$$

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

- H → single step H matrix
- Q → genetic group regression matrix
- Θ → $I \times 0.65$

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:
 - $rel_{itb} - rel_{domestic} > 0.01$

Integration data set

Abulls

- 466 individuals
- information provided to Interbull
- 212 genotyped
- eligibility for integration:
 - $rel_{itb} - rel_{domestic} > 0.01$

Bbulls

- 15,597 individuals
- no information provided to Interbull
- 116 genotyped
- eligibility for integration:
 - all

Integration data set

Abulls

- 466 individuals
- information provided
- 212 genotyped
- eligibility for integration:
 - $rel_{itb} - rel_{domestic} > 0.01$

Bbulls

- 15 597 individuals
- information provided
- 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:

- $\mathcal{R}_{27} \rightarrow \mathcal{R}_9$
 - $\Sigma_g^* = \Omega \Sigma_g \Omega'$
 - $\Sigma_g \rightarrow 27 \times 27$ genetic co-variance matrix
 - $\Omega \rightarrow 9 \times 27$ block matrix of Legendre polynomial coefficients
- $\mathcal{R}_3 \rightarrow \mathcal{R}_9$
 - $u_{i,:}^* = Ku_{i,:}$
 - $r_{i,:}^* = Kr_{i,:}$
 - $p_{i,:}^* = (1 - r_{i,:}^*) \odot \text{diag}(\Sigma_g^*)$
 - $u_{i,:}$ → vector of breeding values of animal i
 - $r_{i,:}$ → vector of reliabilities of animal i
 - $p_{i,:}$ → 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 $\mathcal{R} > 1$
 - maybe $>50k$ iterations (parameterization, Σ 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
 - $r_{ITB} > r_{SENT}$
 - $$\begin{pmatrix} \sigma_{ITB}^2 & \sigma_{SENT}^2 \\ \sigma_{SENT}^2 & \sigma_{SENT}^2 \end{pmatrix}$$
- expectation
 - $r_{blended} \approx r_{ITB}$
 - $U_{blended} \approx U_{ITB}$
- evaluation parameters
 - $cor(:blended, :ITB)$
 - $lm(:blended \sim :ITB)$

Method evaluation

ABulls

- expectation driver
 - $r_{ITB} > r_{SENT}$
 - $\begin{pmatrix} \sigma_{ITB}^2 & \sigma_{SENT}^2 \\ \sigma_{SENT}^2 & \sigma_{SENT}^2 \end{pmatrix}$
- expectation
 - $r_{blended} \approx r_{ITB}$
 - $U_{blended} \approx U_{ITB}$
- evaluation parameters
 - $cor(:,blended, :ITB)$
 - $lm(:,blended \sim :ITB)$

BBulls

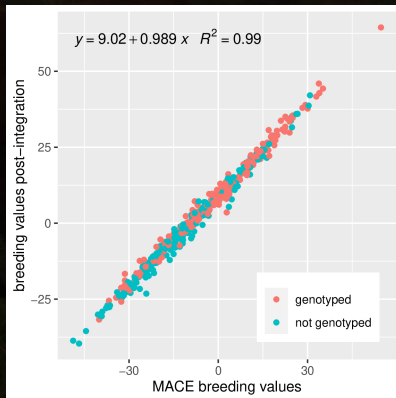
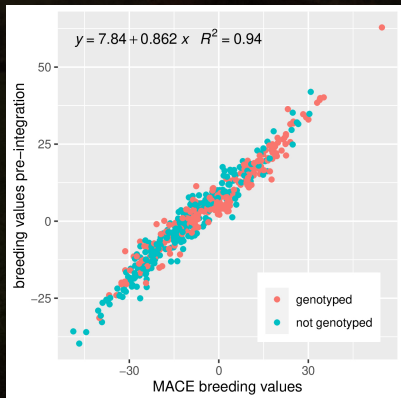
- expectation driver
 - $\begin{pmatrix} \sigma_{ITB}^2 & 0 \\ 0 & \sigma_{DOM}^2 \end{pmatrix}$
- expectation
 - $r_{blended} \geq \max(r_{ITB}, r_{DOM})$
- evaluation parameters
 - ? \rightarrow nice plots

Results

- very similar for all three traits
- only shown for protein

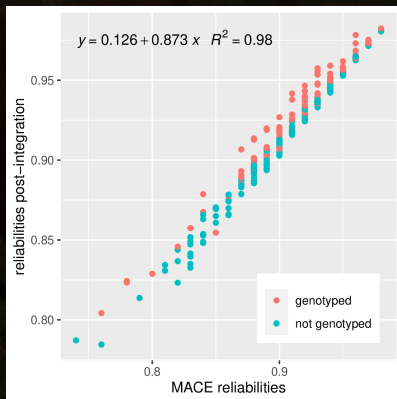
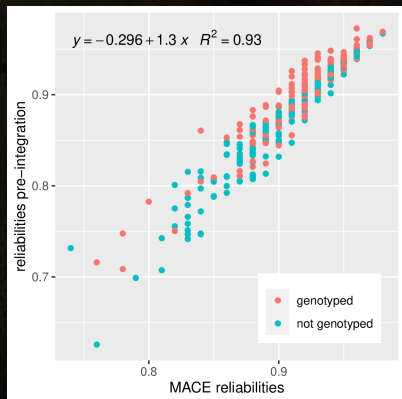
Results: **ABulls**

breeding values



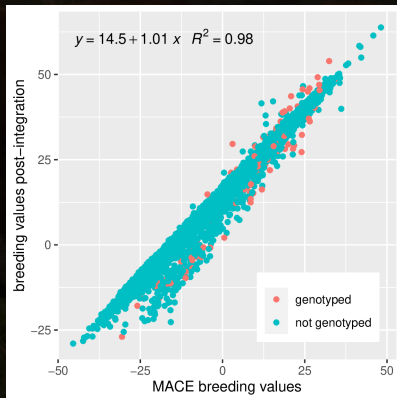
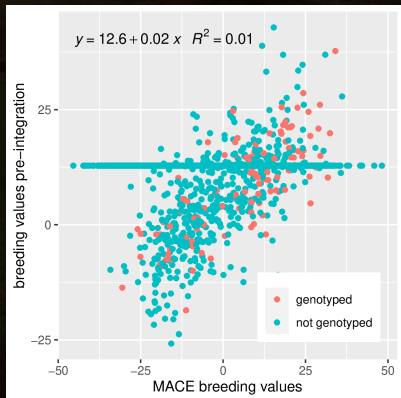
Results: **ABulls**

reliabilities



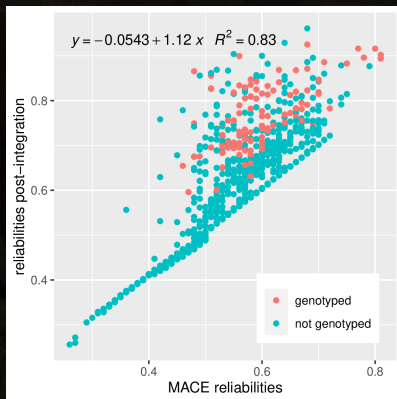
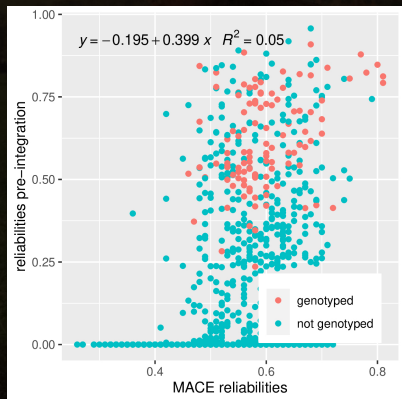
Results: *BBulls*

breeding values



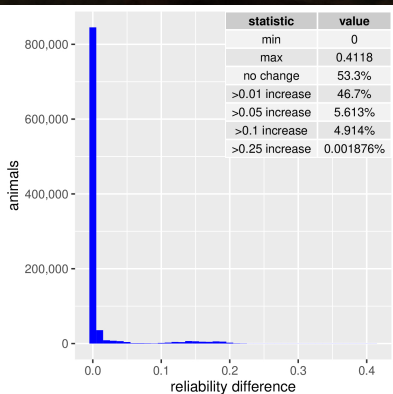
Results: *BBulls*

reliabilities

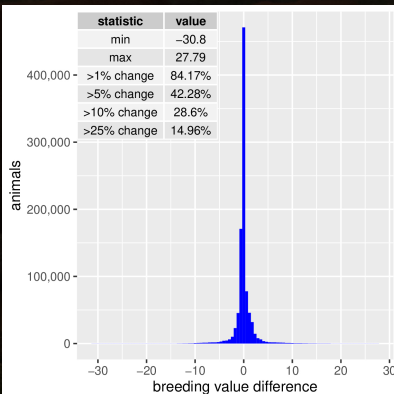


Results: Population impact

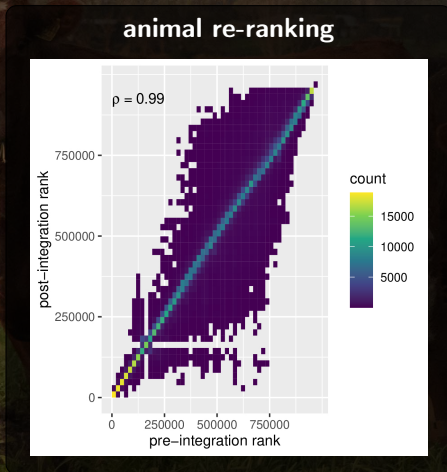
reliabilities



breeding values



Results: Population impact



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 $\Sigma_{a,ITB} \equiv \Sigma_{a,DOM}$ 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 $D_{i,:}$ → computational feasible?
- possible negative impact on long-term genetic trend
 - assumption $\sum_{a,ITB} \equiv \sum_{a,DOM}$ may not hold
- impact of genomics
 - sources of evaluation bias are equivalent to no-genomics
 - extent of evaluation bias ↗?!

THANK YOU!!!

