Single-step genomic evaluation of crossbreed dairy cattle in the US

Ignacy Misztal, D. Lourenco, S. Tsuruta, M. Bermann
University of Georgia
Alberto Cesarani University of Sassari
A. Legarra INRA
E.L. Nicolazzi CDCB
P. M. VanRaden USDA
Introduction

- Genomic evaluation in US by multistep method
- Evaluation by crossbreds by SNP effects from purebreds
  - Accuracy slightly higher than parent average
- Crossbred genotypes not used for purebred evaluation
- Move to single-step models
  - Avoid preselection bias, allow more complex models, simplify pipelines
Goals

- Implement single-step evaluation for crossbreds

Questions:

- Are reliabilities for crossbreds higher than based on parent average?
- Are PTA for purebreds negatively affected by crossbred data?
- Is computing time reasonable?
Single step in dairy at UGA

• Original single-step paper (2010)

• ..........many papers – inflation .................

• Holsteins – current CDCB data (2020)
  • High R² and low inflation
  • Data truncation before 2000 not affecting young animals

• 5 dairy breeds with 4 M genotypes (2021)
  • Similar accuracy for single- and multi-breed with choice of core animals
### Data

Phenotypes from 2000 recorded in Holstein, Jersey, and their crosses

<table>
<thead>
<tr>
<th>Category</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phenotypes</td>
<td>47,417,185</td>
</tr>
<tr>
<td>Cows with records</td>
<td>20,367,132</td>
</tr>
<tr>
<td>Tot. animals</td>
<td>27,111,201</td>
</tr>
<tr>
<td>Genotypes</td>
<td>1,424,863</td>
</tr>
</tbody>
</table>
Materials and Methods I

- Milk (MY), fat (FY), and protein (PY) 305-d yields recorded from January 2000 (Cesarani et al., 2021)

- Three-trait repeatability model
  \[ y = Xb + Z_h h + Z_a Q_a g_a + Z_a a + Z_p p + e \]
  
  **Factors:**
  - herd × sire
  - herd × management
  - age × parity
  - inbreeding
  - heterosis
Materials and Methods I

- Milk (MY), fat (FY), and protein (PY) 305-d yields recorded from January 2000 (Cesarani et al., 2021)

- Three-trait repeatability model

\[ y = Xb + Z_h h + Z_a Q_a g_a + Z_a a + Z_p p + e \]

- Complete data: phenotypes up to August 2021
- Reduced data: phenotypes up to August 2017
Materials and Methods I

• Milk (MY), fat (FY), and protein (PY) 305-d yields recorded from January 2000 (Cesarani et al., 2021)

• Three-trait repeatability model

\[ y = Xb + Z_h h + Z_a Q_a g_a + Z_a a + Z_p p + e \]

herd × sire

herd × management, age × parity, inbreeding, heterosis

• **Complete data**: phenotypes up to August 2021

• **Reduced data**: phenotypes up to August 2017

• UPG (8 groups per breed) in \( A^{-1} \) and \( A_{22}^{-1} \): difference by breed, YOB, and sex
Validation cows = genotyped females with no phenotypes in the reduced dataset

- **HO pure** = HO animals with both sire and dam HO
- **JE pure** = JE animals with both sire and dam JE
- **cross** = HO or JE animals with at least one parent of the opposite breed
- **cross F1** = cross animals with 100 heterosis

Validation method = predictivity based on adjusted phenotypes (correlation and $b_1$)

$\text{pred}=\text{corr}(\text{PTA, y-"fixed"})$
Predictive abilities for cows
Predictive abilities for cows

HO pure (N=688,985)
Predictive abilities for cows

**HO pure (N=688,985)**

- Milk: BLUP = 0.3, sGBLUP = 0.55
- Fat: BLUP = 0.35, sGBLUP = 0.56
- Protein: BLUP = 0.34, sGBLUP = 0.53

**JE pure (N=119,743)**

- Milk: BLUP = 0.31, sGBLUP = 0.52
- Fat: BLUP = 0.26, sGBLUP = 0.46
- Protein: BLUP = 0.32, sGBLUP = 0.51
Predictive abilities for cows

**HO pure (N=688,985)**

- Milk: 0.31, BLUP; 0.55, ssGBLUP
- Fat: 0.35, BLUP; 0.56, ssGBLUP
- Protein: 0.34, BLUP; 0.53, ssGBLUP

**JE pure (N=119,743)**

- Milk: 0.31, BLUP; 0.52, ssGBLUP
- Fat: 0.26, BLUP; 0.46, ssGBLUP
- Protein: 0.32, BLUP; 0.51, ssGBLUP

**cross (N=3,2353)**

- Milk: 0.46, BLUP; 0.67, ssGBLUP
- Fat: 0.31, BLUP; 0.51, ssGBLUP
- Protein: 0.38, BLUP; 0.59, ssGBLUP
Predictive abilities for cows

**HO pure (N=688,985)**

- Milk: 0.3, 0.55, 0.31
- Fat: 0.35, 0.52, 0.31
- Protein: 0.34, 0.56, 0.31

**JE pure (N=119,743)**

- Milk: 0.31, 0.52, 0.31
- Fat: 0.26, 0.46, 0.31
- Protein: 0.32, 0.51, 0.31

**Cross (N=3,2353)**

- Milk: 0.46, 0.67, 0.46
- Fat: 0.51, 0.51, 0.51
- Protein: 0.59, 0.59, 0.59

**Cross F1 (N=1,378)**

- Milk: 0.46, 0.67, 0.46
- Fat: 0.31, 0.51, 0.31
- Protein: 0.38, 0.59, 0.38

**Single breed analyzes**

- Milk: 0.3, 0.55, 0.31
- Fat: 0.35, 0.52, 0.31
- Protein: 0.34, 0.56, 0.31
b1 values

HO pure (N=688,985)

<table>
<thead>
<tr>
<th></th>
<th>BLUP</th>
<th>ssGBLUP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milk</td>
<td>0.83</td>
<td>1.04</td>
</tr>
<tr>
<td>Fat</td>
<td>0.9</td>
<td>0.98</td>
</tr>
<tr>
<td>Protein</td>
<td>0.88</td>
<td>1.01</td>
</tr>
</tbody>
</table>
b1 values

HO pure (N=688,985)

JE pure (N=119,743)
b1 values

HO pure (N=688,985)

JE pure (N=119,743)

cross (N=3,235)
Predictivity and accuracy

\[ \text{accuracy} = \frac{\text{pred}}{h} = \frac{\text{corr}(\text{PTA}, y - \text{"fixed"})}{h} \]

For Holstein milk: \( \text{pred} = 0.55, \text{acc} \approx 0.90, \Rightarrow h^2 \approx 0.22 \)

For crossbreds: \( \text{pred} = 0.67, \text{assume } h^2 \approx 0.22 \Rightarrow \text{acc} = 1.10 \quad ?????? \)

Breed type (F1, F2, reciprocal...) ignored
hard to do from existing data
<table>
<thead>
<tr>
<th>Step</th>
<th>Rounds</th>
<th>Sec / round</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>preGSf90</td>
<td></td>
<td></td>
<td>~ 10 h</td>
</tr>
<tr>
<td>BLUP</td>
<td>671</td>
<td>88.98</td>
<td>~ 17 h</td>
</tr>
<tr>
<td>ssGBLUP</td>
<td>459</td>
<td>134.06</td>
<td>~ 17 h</td>
</tr>
</tbody>
</table>
Conclusions

• Predictivity for crossbreds higher than expected
  • Results superior if used for management

• Purebred evaluations unaffected by crossbred data

• Computations feasible
Acknowledgements