Computing strategies for national dairy cattle evaluations

M. Bermann, D. Lourencoc, I. Misztal

06/2022
2022 ADSA National Meeting
Single-step genetic evaluations

• Move to single-step evaluations

• Less biased than multi-step methods

• Flexible for various models:
  • Repeatability
  • Multiple-trait
  • Maternal
  • Threshold
  • Random regression

• Software availability
Dairy cattle in the US

1.09 million animals genotyped in the last 12 months (10.1.20 to 9.30.21)

6 annual DHI summaries produced

4.08M cows in 11,691 herds enrolled in DHI test plans (2021)

4 selection indexes

5 production traits

21 health, fertility & calving traits

22 conformation traits

25 official genetic conditions & haplotypes

50 traits calculated by CDCB

Weekly genomic predictions for new genotyped animals

Monthly genomic evaluations

Triannual evaluations conventional, genomic & Interbull (in APR, AUG & DEC)

13M Lactation, Calving, Breeding and Health records added for each triannual evaluation

6.7M DHI records in CDCB health evaluations Holstein

885K DHI records in CDCB health evaluations Jersey

ACTIVITY REPORT OCT 2020/Sep 2021
Dairy cattle in the US
Single-step genetic evaluations

\[
\begin{pmatrix}
X'R^{-1}X & X'R^{-1}Z \\
Z'R^{-1}X & Z'R^{-1}Z + \sigma_g^{-2}H^{-1}
\end{pmatrix}
\begin{pmatrix}
\beta \\
\hat{u}
\end{pmatrix}
= 
\begin{pmatrix}
X'R^{-1}y \\
Z'R^{-1}y
\end{pmatrix}
\]

\[H^{-1} = A^{-1} + \begin{pmatrix}
0 & 0 \\
0 & G^{-1} - A_{22}^{-1}
\end{pmatrix}\]
Single-step with APY

\[ H^{-1} = A^{-1} + \begin{pmatrix} 0 & 0 \\ 0 & G^{-1} - A^{-1}_{22} \end{pmatrix} \quad \Rightarrow \quad H^{-1} = A^{-1} + \begin{pmatrix} 0 & 0 \\ 0 & G_{\text{APY}}^{-1} - A^{-1}_{22} \end{pmatrix} \]
From raw data to EBVs

- Read genotypes
- Quality control
- Set up single-step matrices
- Read data and pedigree
- Solve the MME
- Approximate accuracies

Genomic setup

Breeding value estimation
Challenges with more genotyped animals

1. Read and store genotypes
2. Set up APY blocks
3. Blending the GRM
4. Increased number of rounds for PCG
5. Calculating accuracy
### Genomic setup: read and store genotypes

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Int representation</th>
<th>4-byte integer</th>
<th>2-bit integer</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>2</td>
<td>0000 0000 0000 0000 0000 0000 0000 0010</td>
<td>10</td>
</tr>
<tr>
<td>Aa</td>
<td>1</td>
<td>0000 0000 0000 0000 0000 0000 0000 0001</td>
<td>01</td>
</tr>
<tr>
<td>aa</td>
<td>0</td>
<td>0000 0000 0000 0000 0000 0000 0000 0000</td>
<td>00</td>
</tr>
<tr>
<td>Missing</td>
<td>5</td>
<td>0000 0000 0000 0000 0000 0000 0000 0101</td>
<td>11</td>
</tr>
</tbody>
</table>

- 1 million 50k genotypes
- 200 gb -> 12.5 gb
- ~2-5 times faster
Genomic setup: APY blocks

- Dense matrices
- Optimized libraries
- Parallel computing

Implementation of genomic recursions in single-step genomic best linear unbiased predictor for US Holsteins with a large number of genotyped animals

Y. Masuda,∗ † I. Misztal,∗ S. Tsuruta,∗ A. Legarra,† I. Aguilar,‡ D. A. L. Lourenco,∗ B. O. Fragomeni,∗ and T. J. Lawlor§

∗Department of Animal and Dairy Science, University of Georgia, Athens 30602
†Institut National de la Recherche Agronomique, UMR1388 GenPhySE, 31326 Castanet Tolosan, France
‡Instituto Nacional de Investigación Agropecuaria, Canelones, Uruguay 90200
§Holstein Association USA Inc., Brattleboro, VT 05301
Genomic setup: creating $G_{\text{APY}}^{-1}$

- Optimized libraries
- Parallel computing

$$G_{\text{pp}}^{-1} G_{\text{py}} G_{\text{yy}}^{-1} = G_{\text{pp}}^{-1} G_{\text{py}} G_{\text{yy}}^{-1} G_{\text{pp}}^{-1} + G_{\text{py}} G_{\text{pp}}^{-1}$$
Genomic setup: blending

- Make GRM invertible and add a polygenic effect
BV estimation: approximation of accuracies

- Combine different sources of information
  - Records
  - Pedigree
  - Genomics

Approximating Genomic Reliabilities for National Genomic Evaluation

Z. Liu¹, P. M. VanRaden², M.H. Lidauer³, M. P. Calus⁴, H. H. H. H. H. H. H. Jorjani³ and V. Ducrocq⁵

Efficient approximation of reliabilities for single-step genomic best linear unbiased predictor models with the Algorithm for Proven and Young

Matías Bemm, Daniela Lourenço, and Ignacy Misztal
Department of Animal and Dairy Science, University of Georgia, Athens, GA, USA
*Corresponding author: mbberrmann@uga.edu


## Multibreed genomic evaluation for production traits of dairy cattle in the United States using single-step genomic best linear unbiased predictor

A. Cesarani,†‡ D. Lourenço,† S. Tsuruta,† A. Legarra,‡ E. L. Nicolazzi,¶ P. M. VanRaden,* and I. Misztal†

†Department of Animal and Dairy Science, University of Georgia, Athens 30602
‡INRA, UMR1388 GenPhysE, Castanet-Tolosan, France, 31320
¶Council on Dairy Cattle Breeding, Bowie, MD 20716
*Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705

Table 7. Computational costs for all considered scenarios in terms of number of rounds and seconds per round needed to reach a convergence criterion of $10^{-15}$.

<table>
<thead>
<tr>
<th>Item</th>
<th>Rounds</th>
<th>s/round</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BLUP</td>
<td>ssGBLU</td>
<td>BLUP</td>
</tr>
<tr>
<td>AY</td>
<td>504</td>
<td>863</td>
<td>0.08</td>
</tr>
<tr>
<td>BS</td>
<td>364</td>
<td>867</td>
<td>0.18</td>
</tr>
<tr>
<td>GU</td>
<td>345</td>
<td>757</td>
<td>0.07</td>
</tr>
<tr>
<td>HO</td>
<td>457</td>
<td>473</td>
<td>21.25</td>
</tr>
<tr>
<td>JE</td>
<td>586</td>
<td>432</td>
<td>2.00</td>
</tr>
<tr>
<td>AY_BS_GU</td>
<td>592</td>
<td>1,534</td>
<td>3.01</td>
</tr>
<tr>
<td>AY_BS_GU_30k</td>
<td>1,581</td>
<td>1,529</td>
<td>2.22</td>
</tr>
<tr>
<td>AY_BS_GU_direct</td>
<td>1,581</td>
<td>1,529</td>
<td>2.22</td>
</tr>
<tr>
<td>5_BREEDS</td>
<td>643</td>
<td>1,142</td>
<td>27.01</td>
</tr>
<tr>
<td>5_BREEDS_45k</td>
<td>1,763</td>
<td>1,403</td>
<td>40.57</td>
</tr>
</tbody>
</table>

1AY = Ayrshire; BS = Brown Swiss; GU = Guernsey; HO = Holstein; JE = Jersey; 5_BREEDS = all 5 breeds together.
2Time to reach the convergence (i.e., only time for the preconditioning conjugate gradient with iteration on data). ssGBLU = single-step genomic BLUP.
3Computations were carried out on a Linux server (x86_64) equipped with Intel Xeon E5–2683 v4 2.10 GHz processor with 32 cores. Parallel threads used between 12 and 20 cores.
Challenges: more evaluations and information

• Convergence
  • Seconds per round
  • Code
    • Mathematical operations
  • Number of rounds
    • Preconditioning
    • Removing old data

• Complicated models
  • Multiple-trait
  • Threshold
  • Multi/cross breed

• New information

• Variance components
Conclusions

• Different computing strategies allow weekly single-step evaluations

• There is room for improving computational efficiency

• Computational efficiency takes time

• Anticipate next challenges and bottlenecks
Questions?

mbermann@uga.edu