

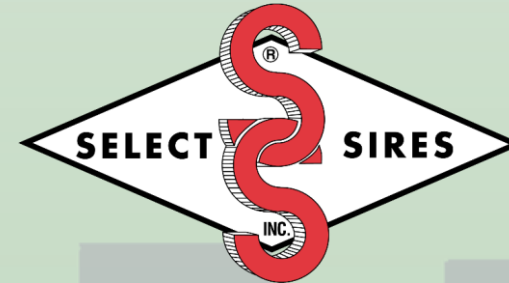


# Thank you to the following sponsors for their support of the Interbull Annual Meeting

**Silver**



**Bronze**





# Evaluations: New Methodologies

Chair: P. VanRaden

Masuda, Y.

Single-step GBLUP including more than 2 million genotypes with missing pedigrees for production traits in US Holstein

Lawlor, T.

Modelling different forms of selection for linear type traits in a single-step GBLUP analysis.

Mrode, R.

The application of several genomic models for the analysis of small holder dairy cattle data.

Jiang, J.

A scalable Bayesian mixed model approach for GWAS and genomic prediction

Liu, Z.

Genomic prediction of health traits using a mixed bull and cow reference population for German Holsteins

Mota, R.R.

Strategy to stabilize GEBV estimation under a quickly evolving mixed sire and cow based reference population in the single-step evaluation system of the Walloon Region of Belgium

# Single-step GBLUP including more than 2 million genotypes with missing pedigrees for production traits in US Holstein

Y. Masuda<sup>1</sup>, S. Tsuruta<sup>1</sup>, E Nicolazzi<sup>2</sup>, and I. Misztal<sup>1</sup>

1 University of Georgia, USA

2 Council on Dairy Cattle Breeding (CDCB), USA

2019 Interbull Meeting, June 22, Cincinnati, OH

# Background

- Genomic prediction with single-step GBLUP (ssGBLUP)
- Required: compatibility in scale among relationship matrices

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}$$

- Issue: missing pedigrees with many genotypes
  - Missing elements in  $\mathbf{A}^{-1}$  and  $\mathbf{A}_{22}^{-1}$
  - Compensation by unknown parent groups (UPG) or metafounders (MF)
- Several models for UPG in  $\mathbf{H}^{-1}$

# Missing parents in ssGBLUP

- Classical UPG

- $\mathbf{H}^* = \mathbf{A}^* + \begin{bmatrix} 0 & 0 & 0 \\ 0 & \mathbf{G}^{-1} - \omega \mathbf{A}_{22}^{-1} & 0 \\ 0 & 0 & 0 \end{bmatrix}$

- Complete UPG

- $\mathbf{H}^* = \mathbf{A}^* + \begin{bmatrix} 0 & 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -\mathbf{G}^{-1} \mathbf{Q}_2 \\ 0 & -\mathbf{Q}'_2 \mathbf{G}^{-1} & \mathbf{Q}'_2 \mathbf{G}^{-1} \mathbf{Q}_2 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -\mathbf{A}_{22}^{-1} \mathbf{Q}_2 \\ 0 & -\mathbf{Q}'_2 \mathbf{A}_{22}^{-1} & \mathbf{Q}'_2 \mathbf{A}_{22}^{-1} \mathbf{Q}_2 \end{bmatrix}$

- Pedigree UPG (complete except for  $\mathbf{Q}'\mathbf{G}^{-1}\mathbf{Q}$ )

- $\mathbf{H}^* = \mathbf{A}^* + \begin{bmatrix} 0 & 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} & 0 \\ 0 & 0 & 0 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -\mathbf{A}_{22}^{-1} \mathbf{Q}_2 \\ 0 & -\mathbf{Q}'_2 \mathbf{A}_{22}^{-1} & \mathbf{Q}'_2 \mathbf{A}_{22}^{-1} \mathbf{Q}_2 \end{bmatrix}$

It turns out that this formula is similar to  $\mathbf{H}^{\Gamma^{-1}}$  with metafounders.

# Previous studies

- Koivula et al. (2017)
  - Complete UPG: Reasonable results
- Bradford et al. (2019)
  - Complete UPG: Low accuracy and high bias
  - Metafounders: Accurate and unbiased for young animals
- Masuda et al. (2018, 2019)
  - Complete UPG: Low accuracy in predictions
  - Pedigree UPG: Reasonable accuracy and inflation
- Not clear how  $\mathbf{H}^*$  is justified in theory.
  - Also, no tests on ssGBLUP including UPG with > 2M genotypes.

# Objectives

- To derive a reasonable inverse of the relationship-matrix ( $\mathbf{H}^{-1}$ ) with UPG or MF in ssGBLUP
- To implement the inverse in a genetic-evaluation software to handle millions of genotypes
- To validate the genetic trends and the predictability of young-bull predictions for production traits in US Holstein

# Complete UPG

- Misztal et al. (2013)

$$\mathbf{H}^* = \left( \text{var} \begin{bmatrix} \mathbf{u}_1^* \\ \mathbf{u}_2^* \\ \mathbf{g} \end{bmatrix} \right)^{-1} = \mathbf{A}^* + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} & -(\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1})\mathbf{Q}_2 \\ \mathbf{0} & -\mathbf{Q}_2'(\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}) & \mathbf{Q}_2'(\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1})\mathbf{Q}_2 \end{bmatrix}$$

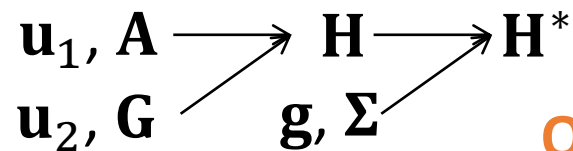
- Original derivation: QP-transformation of MME

- Formal derivation as Quaas (1988):

- $\mathbf{u}^* | \mathbf{g} \sim N(\mathbf{Q}\mathbf{g}, \mathbf{H})$  and  $\mathbf{g} \sim N(\mathbf{0}, \mathbf{\Sigma})$

- Joint density:  $p(\mathbf{u}^*, \mathbf{g}) \propto \exp\left([\mathbf{u}^{*'} \quad \mathbf{g}'] \mathbf{H}^* \begin{bmatrix} \mathbf{u}^* \\ \mathbf{g} \end{bmatrix}\right)$ , then  $\mathbf{\Sigma} \rightarrow 0$

- Updating process:



**Question:** Why should we apply UPG for genomic relationships?

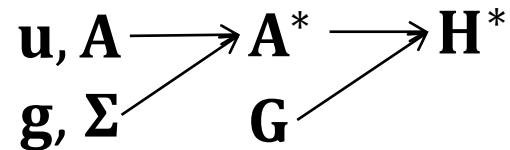


# Pedigree UPG

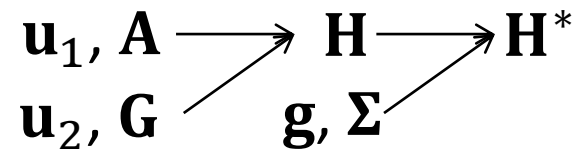
- Bradford et al. (2019) and Masuda et al. (2018, 2019)

$$\mathbf{H}^* = \left( \text{var} \begin{bmatrix} \mathbf{u}_1^* \\ \mathbf{u}_2^* \\ \mathbf{g} \end{bmatrix} \right)^{-1} = \mathbf{A}^* + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} & -(-\mathbf{A}_{22}^{-1})\mathbf{Q}_2 \\ \mathbf{0} & -\mathbf{Q}_2'(-\mathbf{A}_{22}^{-1}) & \mathbf{Q}_2'(-\mathbf{A}_{22}^{-1})\mathbf{Q}_2 \end{bmatrix}$$

- Original idea: decreasing the contribution of  $\mathbf{G}^{-1}$  to UPG
- Formal derivation:
  - I skip it because of the time limit.
  - Updating process:



Complete UPG:



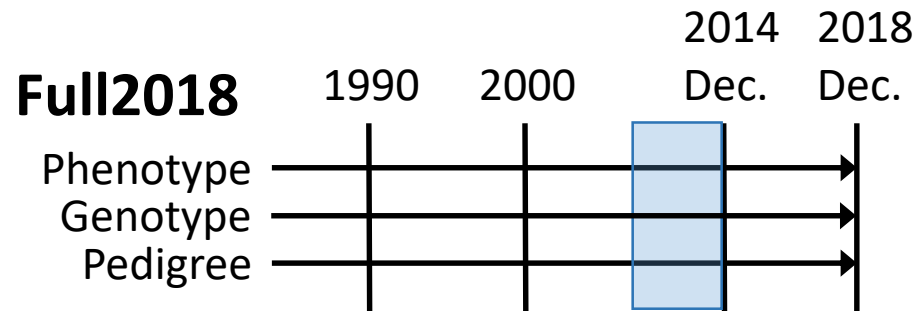
# Tested $\mathbf{A}^*$ and $\mathbf{H}^*$

Model	Inverse matrix	Abbr.
Pedigree BLUP with the standard UPG	$\mathbf{A}^* = \begin{bmatrix} \mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{Q} \\ -\mathbf{Q}'\mathbf{A} & \mathbf{Q}'\mathbf{A}^{-1}\mathbf{Q} \end{bmatrix}$	BLUP
ssGBLUP with the standard UPG	$\mathbf{H}^* = \mathbf{A}^* + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - 0.9\mathbf{A}_{22}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}$	ssGBLUP classicUPG
ssGBLUP with complete UPG	$\mathbf{H}^* = \mathbf{A}^* + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} & -(\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1})\mathbf{Q}_2 \\ \mathbf{0} & -\mathbf{Q}'_2(\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}) & \mathbf{Q}'_2(\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1})\mathbf{Q}_2 \end{bmatrix}$	ssGBLUP compUPG
ssGBLUP excluding genomic UPG	$\mathbf{H}^* = \mathbf{A}^* + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} & -(-\mathbf{A}_{22}^{-1})\mathbf{Q}_2 \\ \mathbf{0} & -\mathbf{Q}'_2(-\mathbf{A}_{22}^{-1}) & \mathbf{Q}'_2(-\mathbf{A}_{22}^{-1})\mathbf{Q}_2 \end{bmatrix}$	ssGBLUP pedUPG

# Full data in Holstein

	Description	Number of records/animals
Phenotype	<b>Protein yield</b> (305-d basis) for US Holstein cows recorded between Jan. 1990 and Dec. 2018	61,229,782
Pedigree	Animals born in Dec. 2018 or earlier (3-gen. back from phenotyped cows) <b>16 UPGs: sex by 4-yr group</b>	35,857,897
Genotype	Animals born in Dec. 2018 or earlier (79,294 markers) <b>15%</b> with missing sire and/or dam	2,334,951

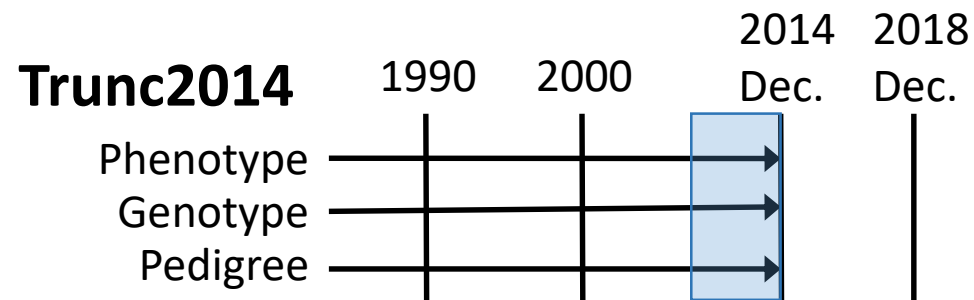
# Validation study



For benchmarks  
(DYD/GPTA2018)

DYD with different (G)PTA by  
VanRaden and Wiggans (1991)

DYD-BLUP, DYD-pedUPG,  
PTA-BLUP, and GPTA-pedUPG



For GPTA using  
ssGBLUP  
(GPTA2014)

Validation Bulls: Genotyped  
young bulls with no tested  
daughters in 2014 but with at  
least 50 tested daughters in  
2018 (N=2,315)

$$DYD/GPTA2018 = b_1 \times GPTA2014 + b_0$$

- $R^2$  : validation reliability
- Slope ( $b_1$ ): Inflation of prediction

# Model

- Same model as the official evaluation
  - Fixed effects: management, age\*parity, inbreeding, and heterosis
  - Random effects: herd\*sire interaction, additive genetic effect, permanent environmental effect, and residual effect
- APY for genomic relationships
  - 15,000 core animals (randomly chosen)
- UPG: sex by year-group
  - Full data (16): -1986, -1990, 1994, -1998, -2002, -2006, -2010, and 2011-
  - Truncated (14): -1986, -1990, 1994, -1998, -2002, -2006, and 2007-
- Genetic base: phenotyped cows born in 2005

# Solving MME in this study

	OpenMP-based solver	MPI-based solver
Parallelism	OpenMP	MPI and OpenMP
CPU-cores used	6	8
Iteration on data	Data and pedigree files	APY G-inverse files
Genotypes (core animals)	2.3 M (15K)	2.3 M (15K)
Total memory usage	> 267 GB	> 17 GB
Wall-clock time per round	35 s	39 s
WC time for 600 rounds	5.8 h	6.5 h

The software development is still going on especially for efficiency.

# R<sup>2</sup> and b1: DYD/GPTA2018 on GPTA2014

	<b>R2</b>		<b>b1</b>	
<b>Benchmark</b>	<b>BLUP</b>	<b>pedUPG</b>	<b>BLUP</b>	<b>pedUPG</b>
DYD2018-BLUP	<b>0.34</b>	0.67	<b>0.43</b>	0.79
DYD2018-pedUPG	0.33	<b>0.77</b>	0.42	<b>0.85</b>

For validation bulls with at least 50 daughters (N=2315)

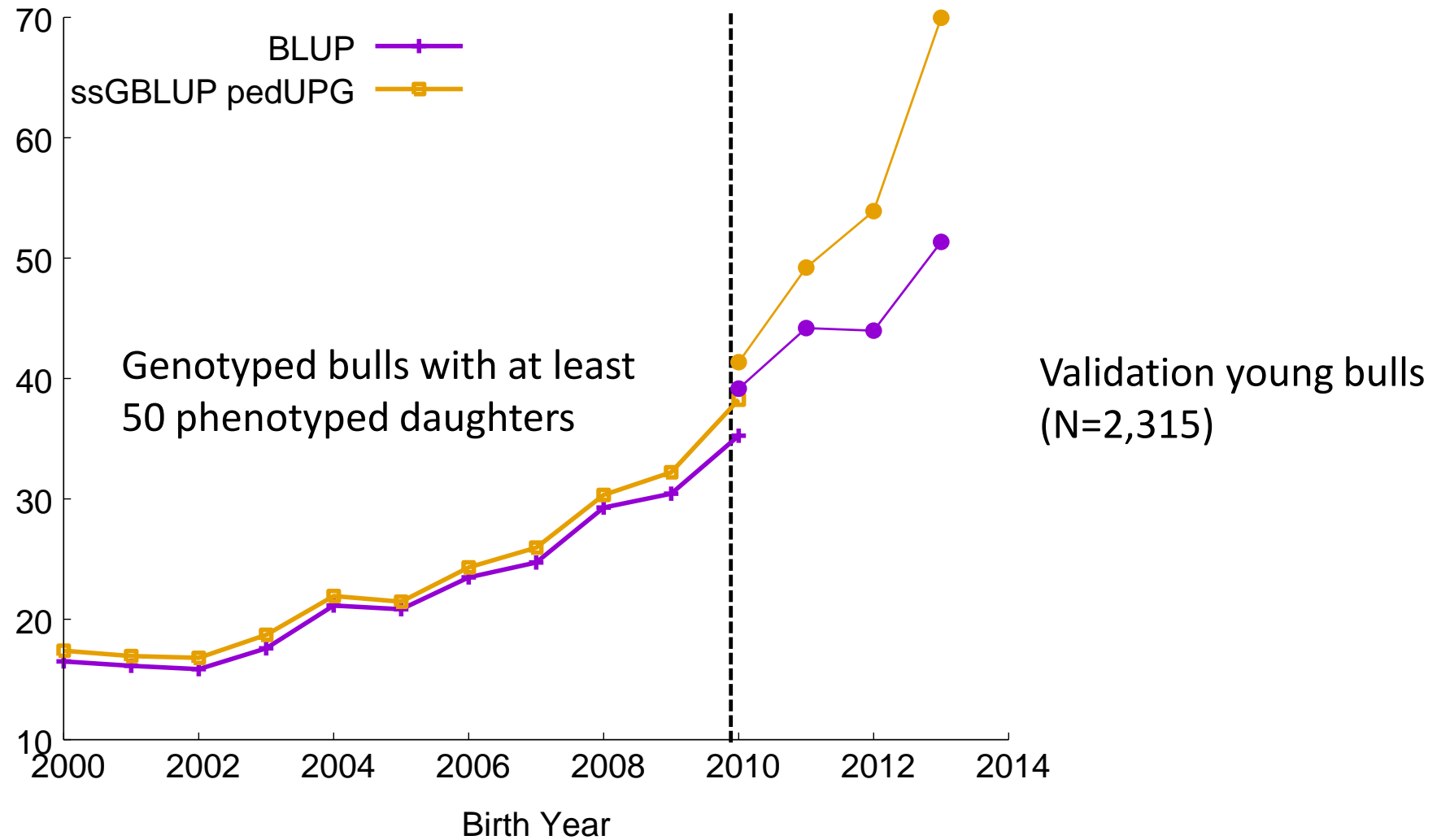
# R<sup>2</sup> and b1: DYD/GPTA2018 on GPTA2014

	<b>R2</b>		<b>b1</b>	
<b>Benchmark</b>	<b>BLUP</b>	<b>pedUPG</b>	<b>BLUP</b>	<b>pedUPG</b>
DYD2018-BLUP	<b>0.34</b>	0.67	<b>0.43</b>	0.79
DYD2018-pedUPG	0.33	<b>0.77</b>	0.42	<b>0.85</b>
PTA2018-BLUP	<b>0.38</b>	0.68	<b>0.47</b>	0.83
GPTA2018-pedUPG	0.34	<b>0.82</b>	0.44	<b>0.90</b>

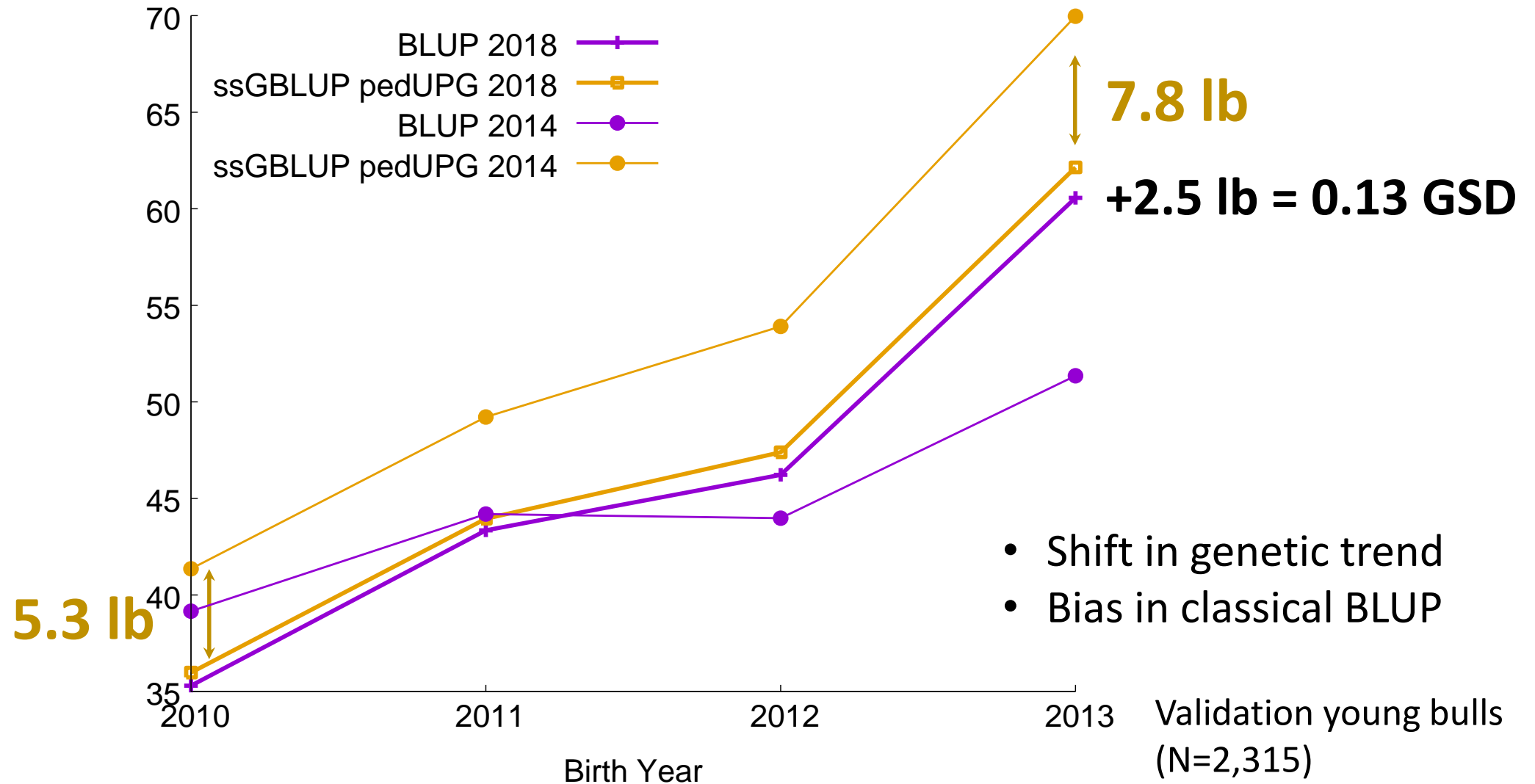
For validation bulls with at least 50 daughters (N=2315)



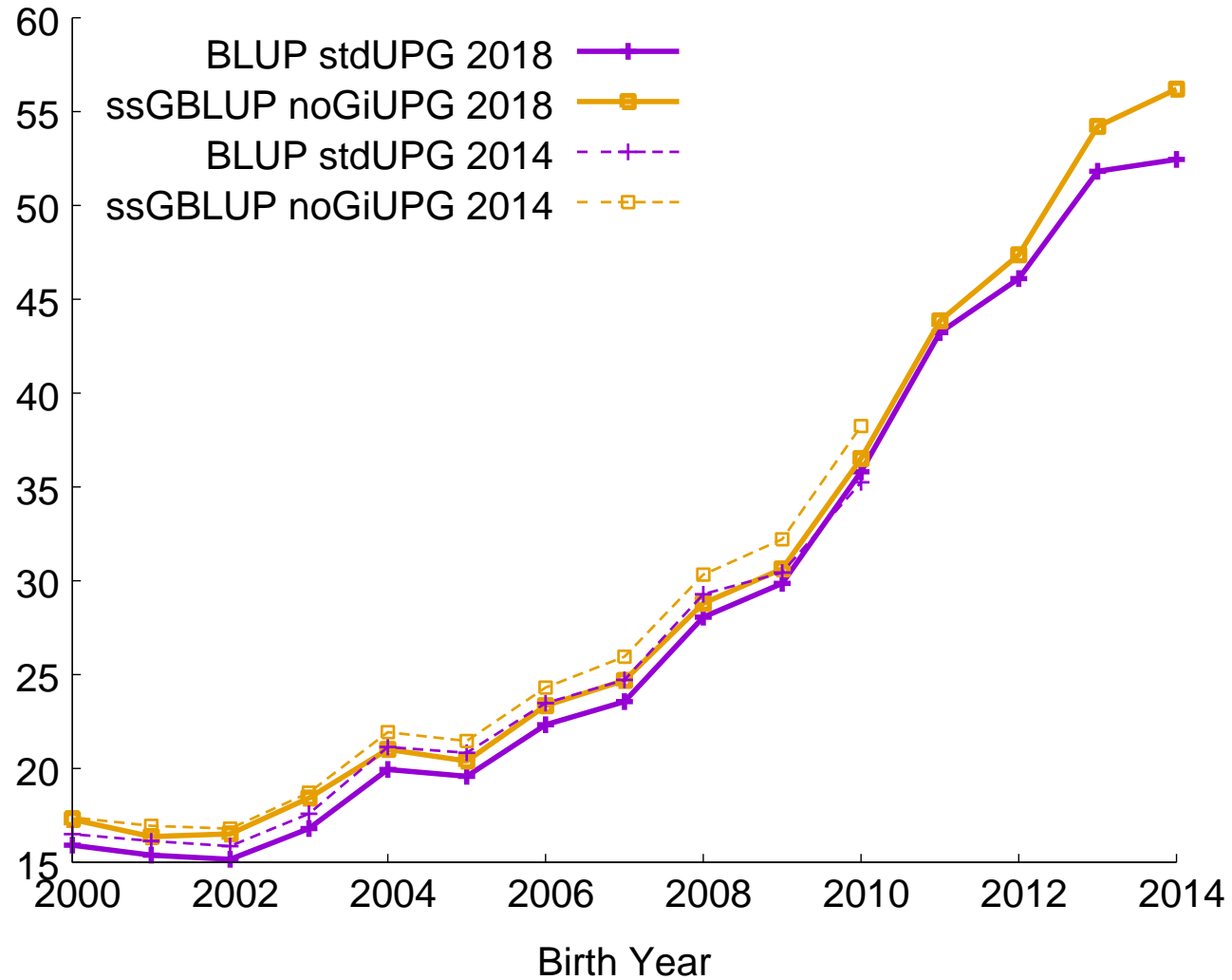
# Genetic trend for genotyped bulls in 2014



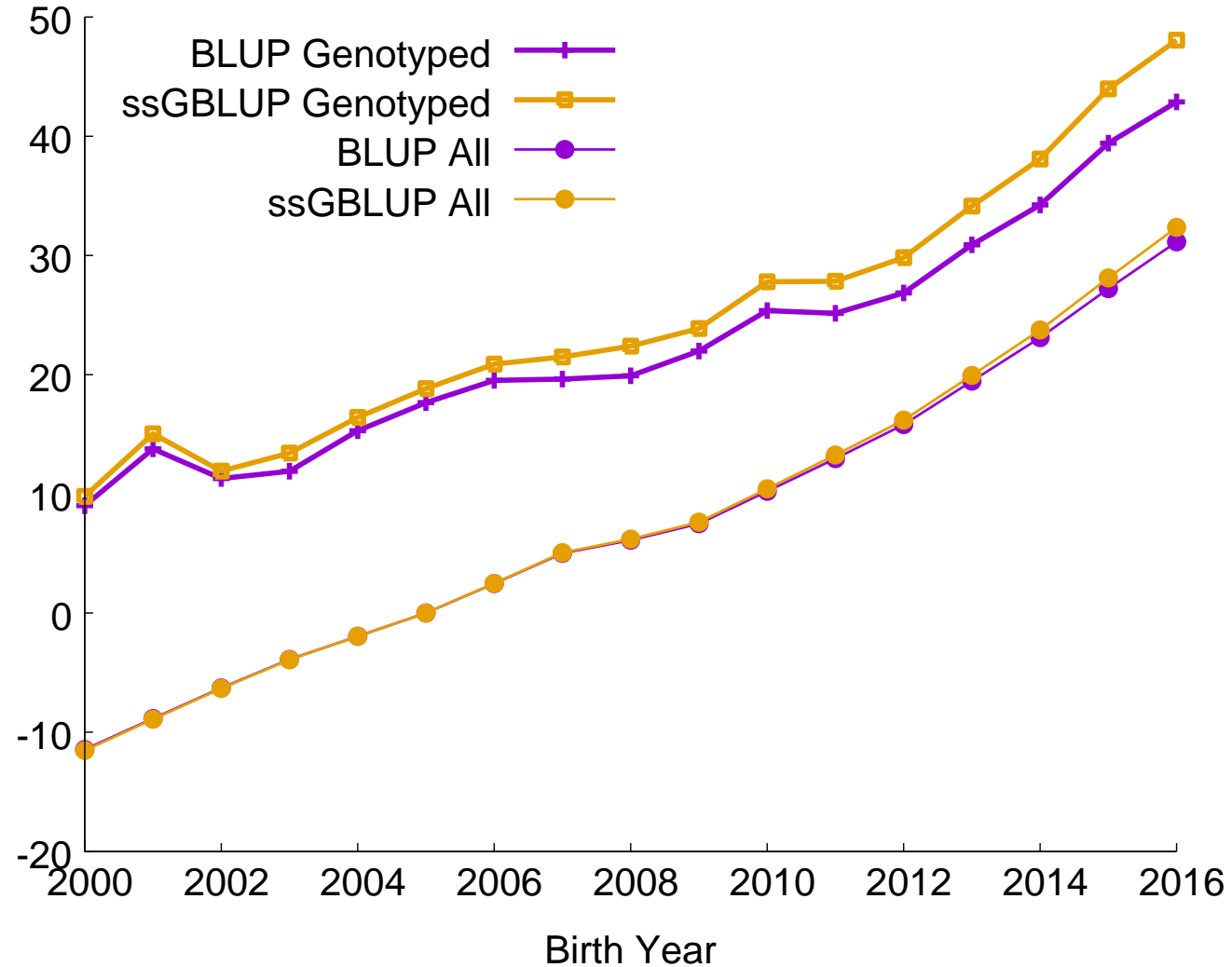
# Predicted and actual trends for valid. bulls



# Genetic trend for genotyped bulls in 2018



# All/genotyped cows with record(s) in 2018



# Summary

- An alternative  $\mathbf{H}^*$  with UPG for  $\mathbf{A}^{-1}$  and  $\mathbf{A}_{22}^{-1}$ , not for  $\mathbf{G}^{-1}$  is theoretically justified.
- The UPG model is reasonable in generic trends and predictability for young bulls.
- Single-step GBLUP with >2M genotypes is computationally feasible.
- This is a preliminary report. Additional research (with metafounders) is still in progress.

# Acknowledgement

- Council of Dairy Cattle Breeding (CDCB) for phenotype, genotype, and pedigree data.
- Holstein Association USA for financial support.
- Paul VanRaden (AGIL, USDA) for discussion on UPG in genetic evaluation
- Andres Legarra (INRA), Ignacio Aguilar (INIA), and Daniela Lourenco (UGA) for discussion on various topics in ssGBLUP