Estimation of heritabilities and genetic correlations in very large datasets using predictivities within and across traits

Ignacy Misztal, Shogo Tsuruta, Zuleica Trujano, Mary Kate Hollifield, Gopal Gowane, Daniela Lourenco

UNIVERSITY OF GEORGIA

Challenge of parameter estimation in genomic era

Potential negative effects of genomic selection

Ignacy Misztal[1] and Daniela Lourenco

Department of Animal and Dairy Science, University of Georgia, Athens, GA 30602, USA

[1] Corresponding author: ignacy@uga.edu

- Possibly rapidly changing parameters
- Need estimates using complete data including genomic
- Computing issues with REML and Bayesian methods
- Can we estimate parameters without size restriction, generation by generation?

Realized and theoretical accuracies

Realized accuracy

$$acc = corr\left(y - Xb, \hat{u}\right)/h$$

Legarra et al. (2008)

y-Xb - adjusted phenotype

$\hat{u}$ - breeding value obtained without that phenotype

h^2 - heritability

Theoretical accuracy

$$acc = \sqrt{\frac{Nh^2}{Nh^2 + M_e}}$$

Daetwyler et al. (2008)

N – number of genotyped animals with phenotypes

M_e – number of independent chromosome segments

Me ≈ 5k (chickens, pigs), 10k (beef), 15k (Holsteins)

Pocrnic et al. (2017)

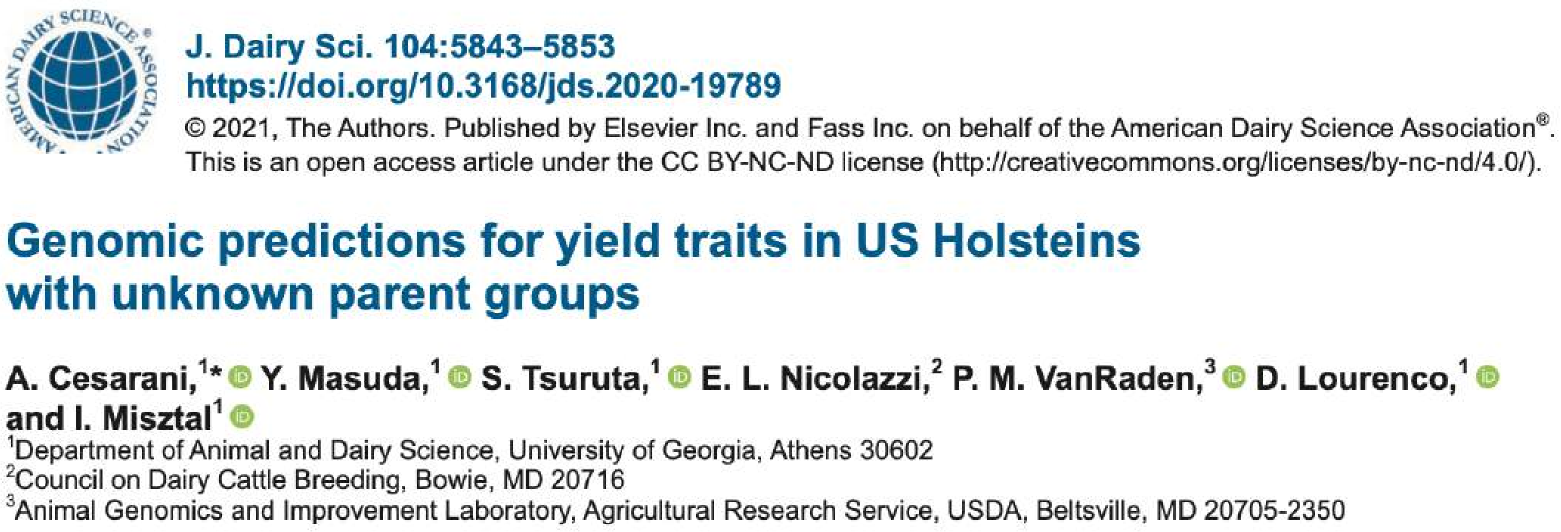
Formula for estimating heritability

$$\widehat{h^2} = \frac{c^2 + \sqrt{c^4 + 4c^2 M_e / N_{ref}}}{2} \mp \frac{3c}{\sqrt{N_{val}}}$$

$$c = corr(y - Xb, \hat{u})$$

N_{ref} – animals in reference population
Me – effective chromosome segments, ~5k in pigs and chicken, ~15k in cattle
N_{val} – number of animals in validation population

Heritabilitv for milk in Holsteins



J. Dairy Sci. 104:5843–5853
https://doi.org/10.3168/jds.2020-19789
© 2021, The Authors. Published by Elsevier Inc. and Fass Inc. on behalf of the American Dairy Science Association®.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Genomic predictions for yield traits in US Holsteins with unknown parent groups

A. Cesarani,[1]* Y. Masuda,[1] S. Tsuruta,[1] E. L. Nicolazzi,[2] P. M. VanRaden,[3] D. Lourenco,[1] and I. Misztal[1]
[1]Department of Animal and Dairy Science, University of Georgia, Athens 30602
[2]Council on Dairy Cattle Breeding, Bowie, MD 20716
[3]Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705-2350

Reference: 580k   Validation 381k

Starting $h^2 = 0.35$ Me=15k chromosome segments

Predictivity = 0.55

Calculated $h^2 = 0.33$

Formula for genetic correlations

$corr\left(y_i - Xb_i, \widehat{u_j}\right)$ Predictivity of trait i by trait j

$$corr_{ij} = \frac{corr\left(y_i - Xb_i, \widehat{u_j}\right)}{h_i \; acc_j} \mp \frac{1}{h_i \; acc_j \sqrt{N_{val}}}$$

Under correct model: $corr_{ij} = corr_{ji}$

Procedure

- Select reference population, number of genotyped N > 10,000
- Select validation population, number of genotyped $N_{val} > 5000$
- Estimate GEBV with phenotypes of reference population and genotypes of both populations; treat traits uncorrelated
- Calculate predictivities within and across traits
- Recalculate heritabilities $\widehat{h^2} = \frac{c^2+\sqrt{c^4+4c^2 M_e/N}}{2}$
- Recalculate accuracies $acc = corr(y - Xb, \hat{u})/h$
- Calculate genetic correlations $corr_{ij} = \frac{corr(y_i - Xb_i, \widehat{u_j})}{h_i \; acc_j} \pm \frac{1}{h_i \; acc_j \sqrt{N}}$

Data simulation

- y_i^j, u_i^j - phenotype and breeding values of trait i in j-th generation

- Two uncorrelated traits
  - “production” with $h^2 = 0.4$
  - “base fitness” with $h^2 = 0.1$

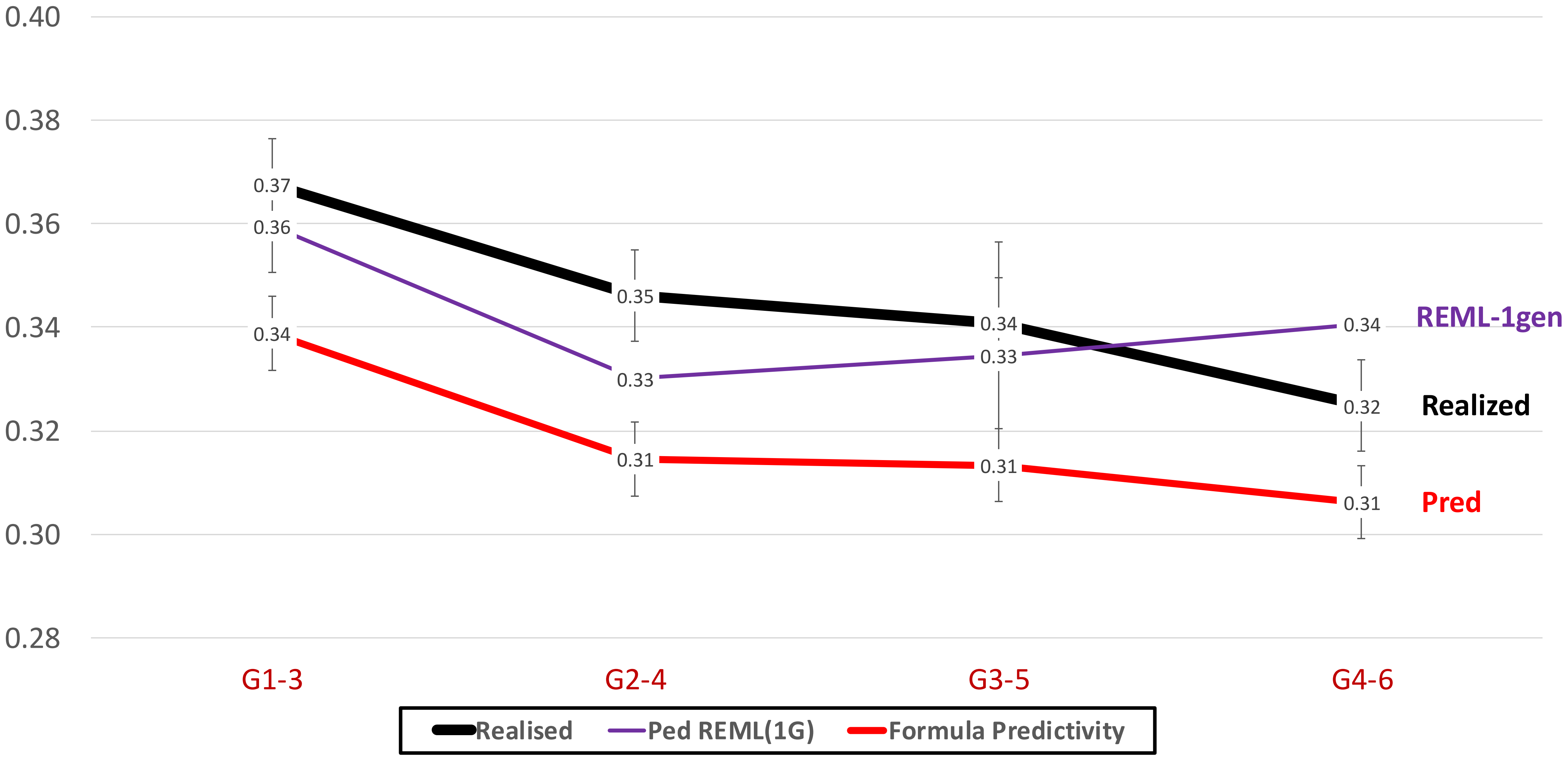
- Evolving fitness trait $u_3^j = \alpha_j \left(u_2^j - \beta u_1^j \overline{(u_1^j - u_1^0)} \right)$
  - α_j - scaling factor so that $var\left(u_3^j\right) = const$
  - β - chosen to change genetic correlation between traits 1 and 3 by about -0.1 each generation.

Data simulation (2)

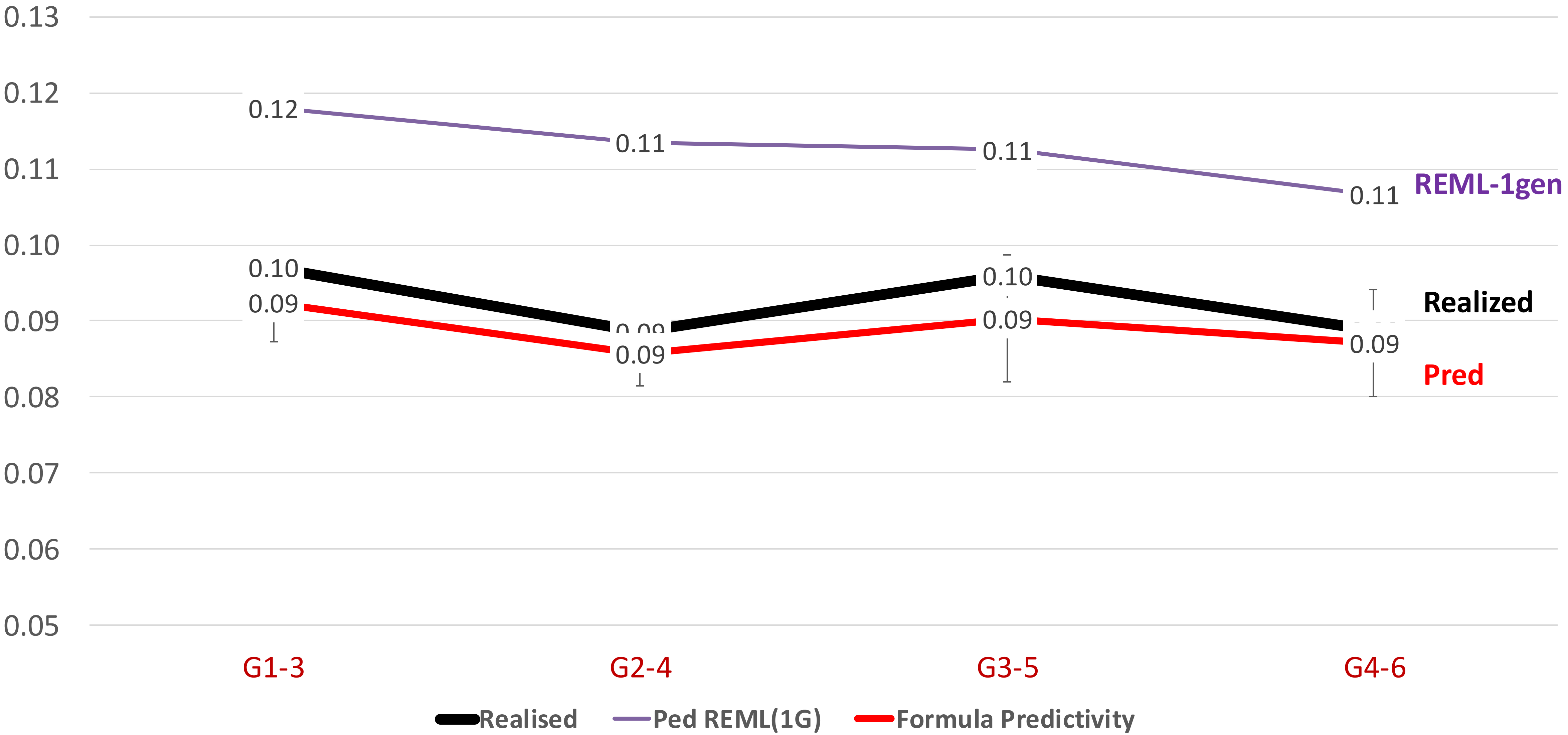
- Simulation by AlphaSim
  - 40k genotyped and phenotyped per generation
  - 6 generations
  - GBLUP selection in each generation
  - Effective population size 50

- Analyzes
  - Realized parameters for each generation
  - Pedigree REML
  - Parameters by predictivity, use 2 reference generations

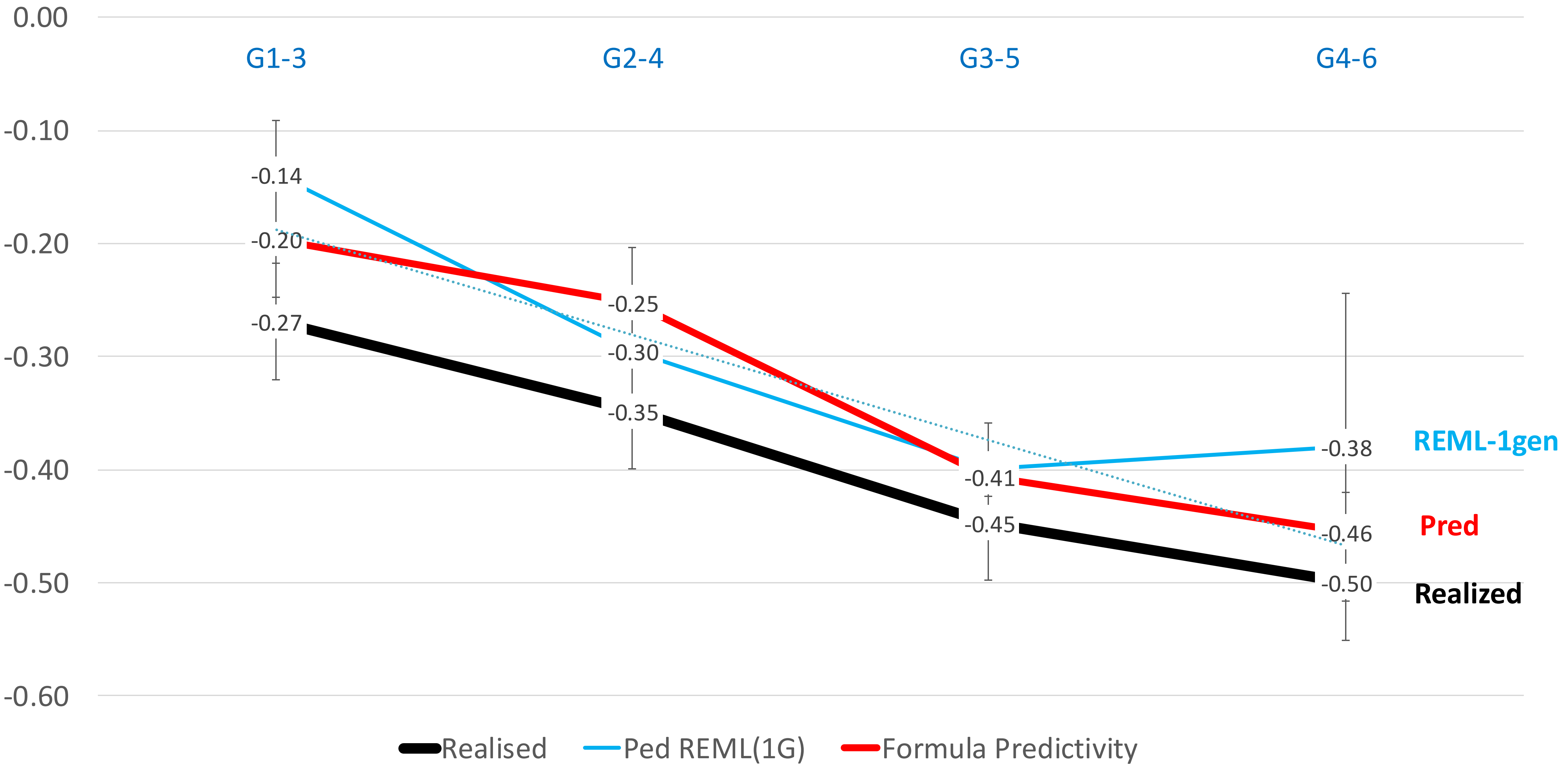
Heritability for "Production" Trait



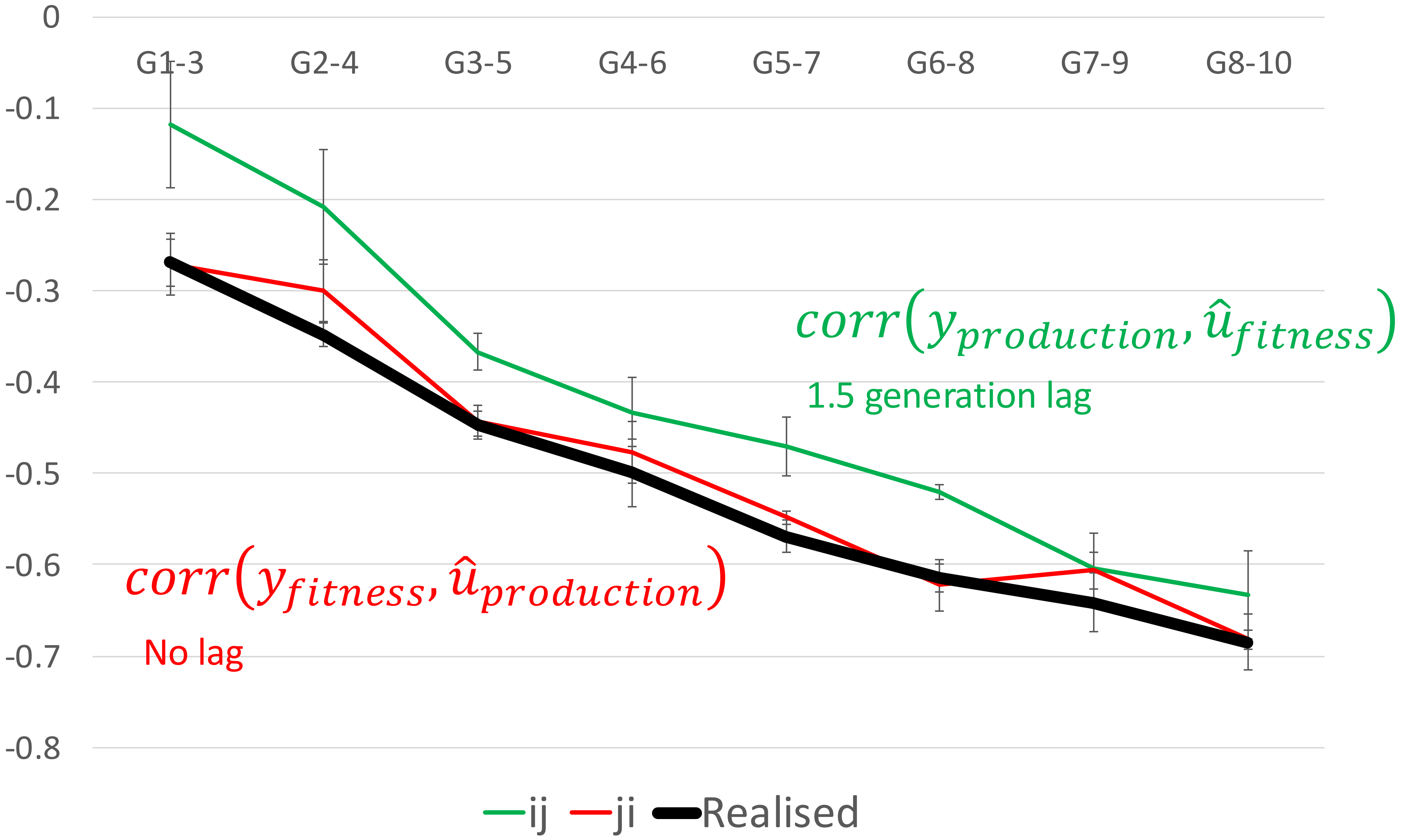
Heritability for “Fitness” Trait



Genetic correlations



Two predictivities: IJ and JI



$corr(y_{production}, \hat{u}_{fitness})$

1.5 generation lag

$corr(y_{fitness}, \hat{u}_{production})$

No lag

Comments

- Results with pedigree REML OK because of simple selection
- Order of genetic correlation important
- Predictivity sensitive to selected genotyping – no problem with current commercial data
- ssGBLUP: use number of both genotyped and phenotyped

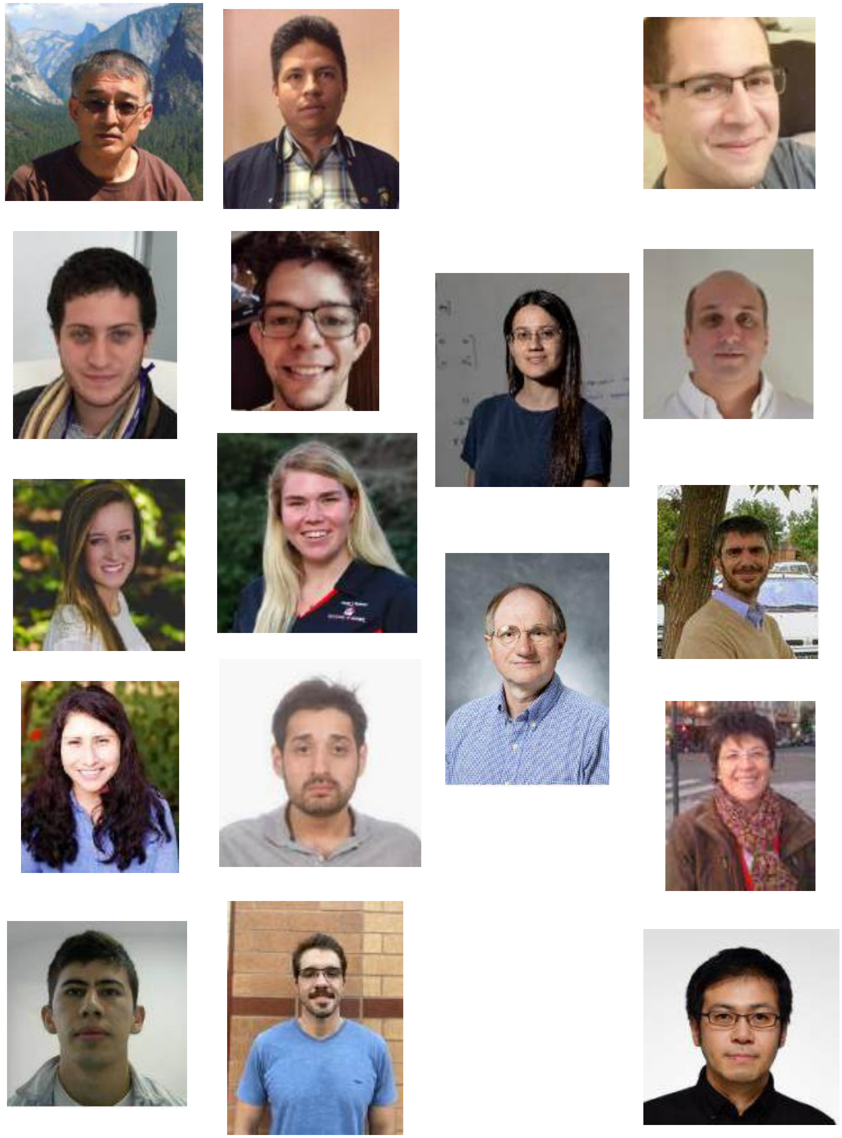
Conclusions

- Can estimate heritability from theoretical and realized accuracies
- Can estimate genetic correlations by predictivity across traits
- Any data size and data slice

- Applicable to models where predictivity is applicable

UNIVERSITY OF GEORGIA

UGA AB&G team



USDA

NIFA

PIC

ANGUS THE BUSINESS BREED
AMERICAN ANGUS ASSOCIATION
3201 Frederick Avenue • St. Joseph, MO 64506 • (816) 383-5100 • Fax (816) 233-9703 • E-mail: angus@angus.org

AGI

USDA United States Department of Agriculture
Agricultural Research Service
Warmwater Aquaculture Research Unit

USDA Agricultural Research Service
U.S. DEPARTMENT OF AGRICULTURE
Cool and Cold Water Aquaculture Research

zoetis

Smithfield®

Cobb®

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

DNA GENETICS
Greater gains. Improved performance.

ABCZ

Rural Development Administration
National Institute of Animal Science

WAGYU
AMERICAN WAGYU ASSOCIATION

BAYER Crop Science

U.S. Registered Holsteins
FOR MAXIMUM PROFIT
HOLSTEIN ASSOCIATION USA, INC.

THE MASCHHOFFS
Progressive Farming. Family Style.

ABS®

SEMEX®
Genetics for Life®

PEAK

PFHBiPM

PLEMDAT

Topigs Norsvin

methodgenetics

ANCP