

Rapid changes of genetic parameters under genomic selection and challenges of estimating genetic parameters

Ignacy Misztal



UNIVERSITY OF
GEORGIA

Estimating changing parameters in pigs (Hidalgo et al., 2019)

- Growth ($h^2=0.3$) and fertility ($h^2=0.07$)
- 40k genotypes over 10 years
- 400k animals

- Data slices for 3 years
 - Elimination of noncontributing genotypes
 - gibbs3f90

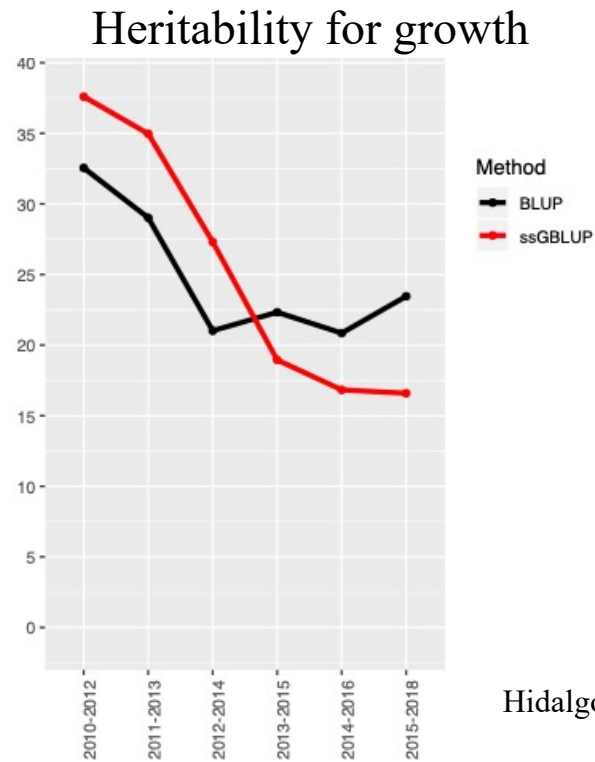
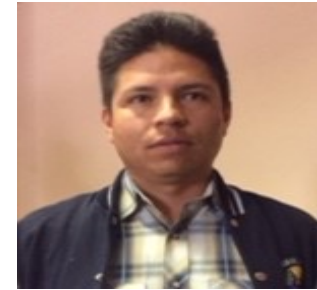
ANIMAL GENETICS AND GENOMICS

Changes in genetic parameters for fitness and growth traits in pigs under genomic selection

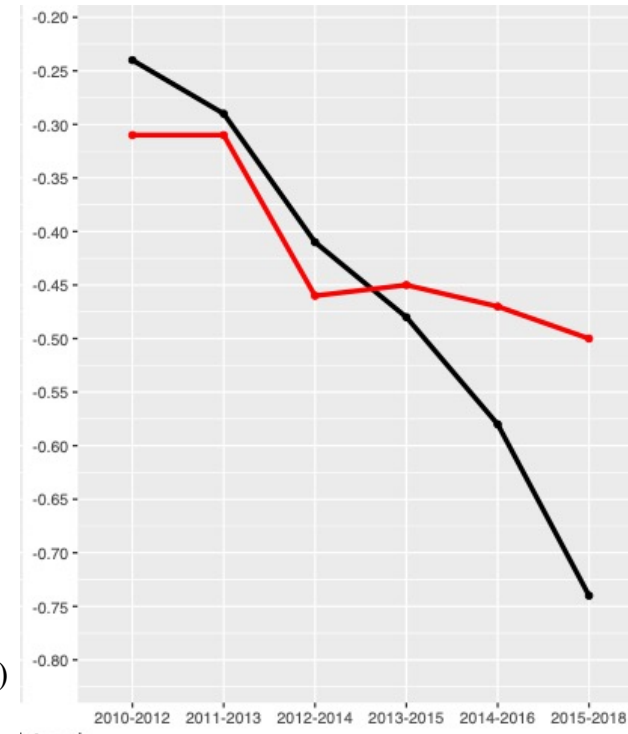
Jorge Hidalgo,^{†,1} Shogo Tsuruta,[†] Daniela Lourenco,[†] Yutaka Masuda,[†] Yijian Huang,[‡] Kent A. Gray,[‡] and Ignacy Misztal[†]

[†]Department of Animal and Dairy Science, University of Georgia, Athens, GA 30602, [‡]Smithfield Premium Pigs, Smithfield, NC 27870

Changes in (co)variances in pigs due to genomic selection



Genetic correlation with reproduction



Hidalgo et al. (2019)

Heritability halved, antagonistic correlations $-0.3 \rightarrow -0.5$

Why parameters change

- Bulmer effect
- Response in unselected traits when other trait(s) strongly selected
- Epistatic effects
- Change in trait definition:
 - Management
 - Changing resource allocation

Cases of changing correlations - resource allocation

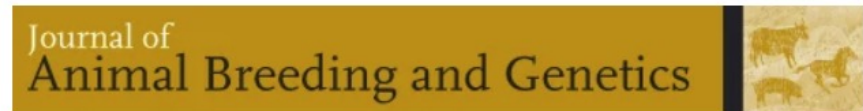
- Milk and dairy form
 - Old times: fat cows lose milk by getting fat
 - New time: Cows need fat as body reserves during negative energy balance
- Production and fertility
 - When production low, fertility OK
 - When production very high, energy redirected from fertility
 - $\text{fertility} = \text{innate_fertility} - \alpha \text{ production}$

Old times for parameter estimation

- Pick a subset
- Estimate variance components
- Use a fact that mixed-model equations sparse

- AIREML
 - 100k-1M animals
 - Cubic costs with multiple traits, crashes if too many
 - Often fast convergence to final value

- Gibbs sampler
 - Over 1 M animals
 - Linear costs with traits
 - Not sure when to finish samples



[Full Access](#)

Reliable computing in estimation of variance components

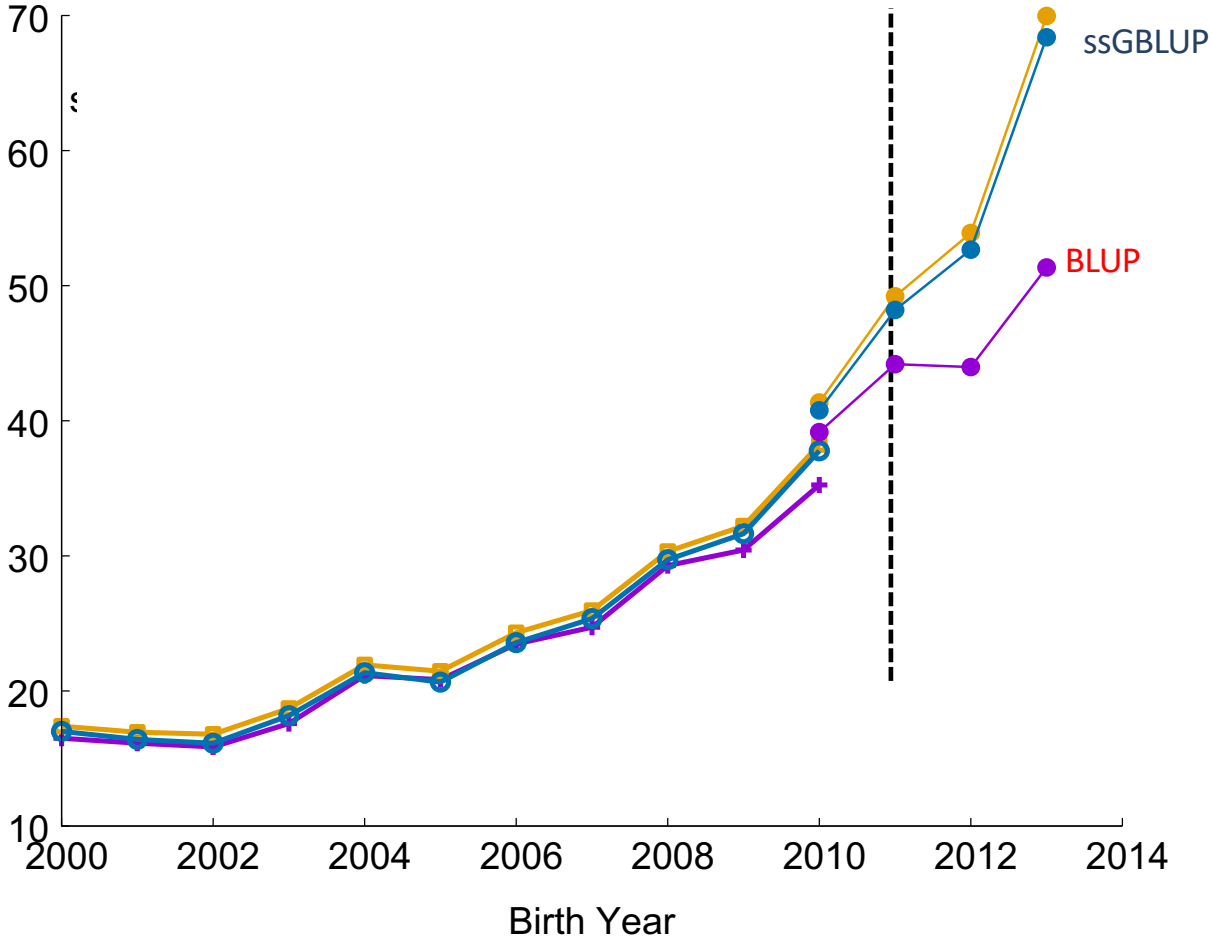
[Correction\(s\) for this article](#) ▾

I. Misztal

What changed with genomic selection?

- Possible biases due to genomic preselection
 - Evaluation unbiased if all data used for selection included in analyses (Henderson, 1984).
- Long or impossible computing
- Accelerated “aging” of the additive model

Genetic trend for genotyped bulls (Masuda et al., 2019)



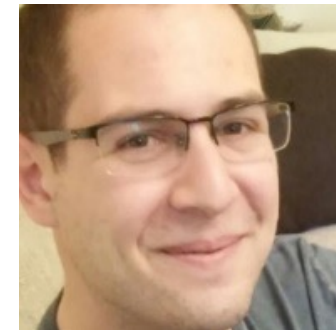
Computing issues

- Mixed models stored as sparse matrix
 - If 4 byte integer, number of elements $< 2^{31} = 2 \text{ G}$ elements
- Equations no longer sparse – dense blocks
 - Limit of about 50k genotypes for 1 trait, 30k for two traits
- REML: Sparse matrix factorization/inversion inefficient with dense blocks
- YAMS (Masuda et al., 2015) – sparse matrix package that recognizes dense blocks
 - Up to 100 time faster than older FSPAK
 - About 10 hrs (AIREMLF90) for 200k animal (15k genotyped) 4 trait model

When parameter estimation biased?

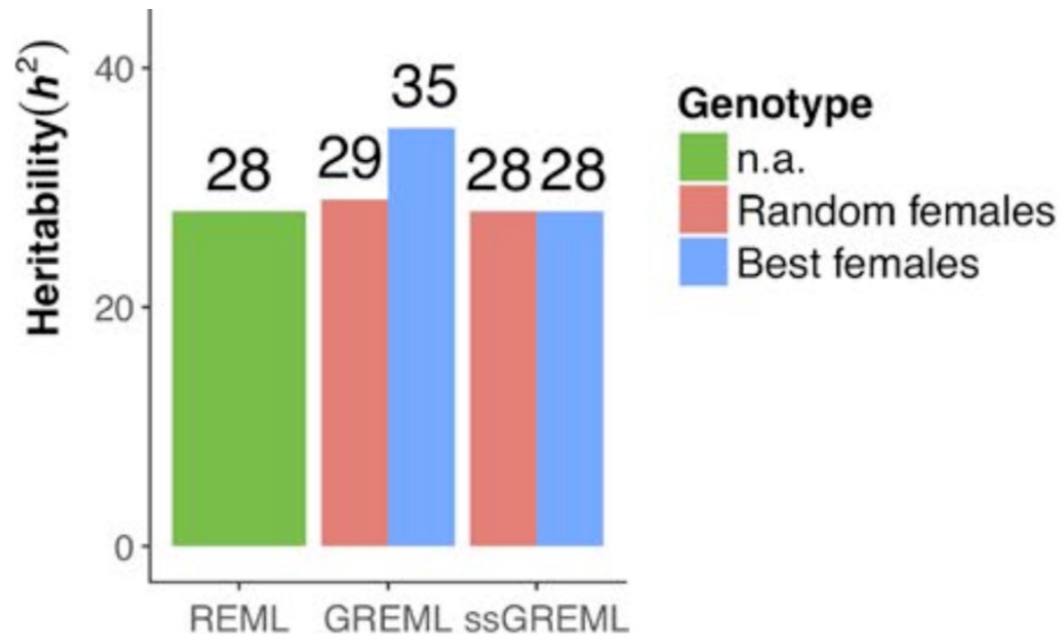
Simulation by Cesarani et al. (2018)

- 3 of 3k animals each
- 30% animals in each generation genotyped
- REML, GREML (genotyped animals only) and ssGREML
- Selection or random mating
- Random or best animals genotyped
- 1 to 3 generations of data

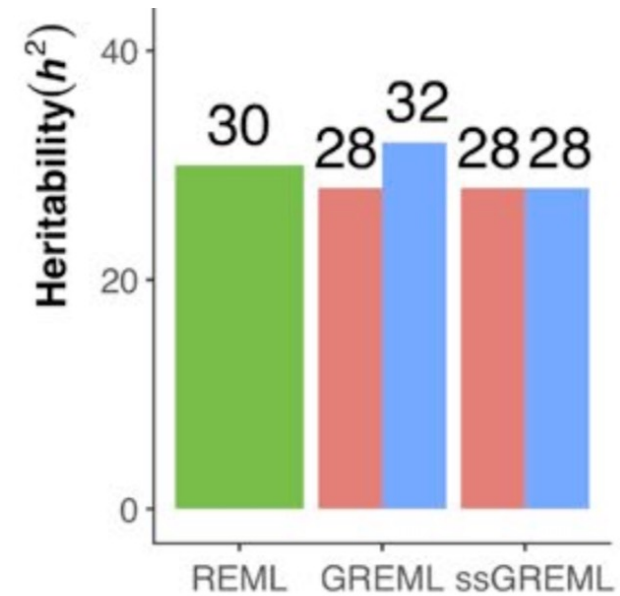


No selection ($h^2 = 30\%$)

One generation of data



Three generations of data

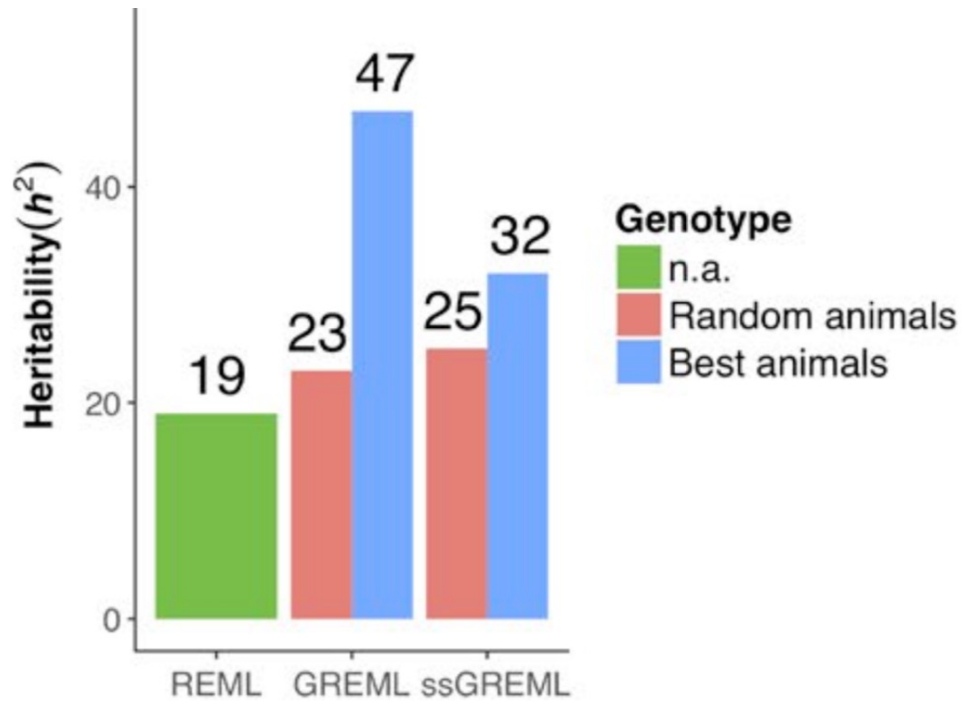


Mostly unbiased estimates without selection

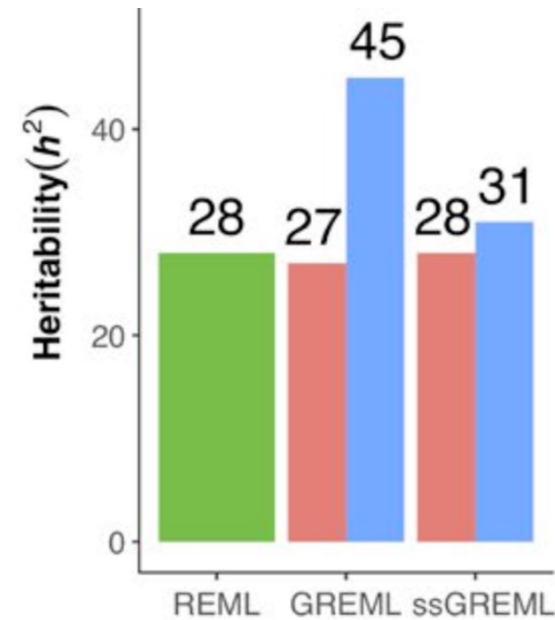
Some inflation with small data in GREML and selective genotyping

EBV selection ($h^2=30\%$)

One generation of data



Three generations of data



Heritability biased down in REML with small data
Heritability inflated in GREML and selective genotyping

Genetic variances in ssGREML and REML with selective genotyping (Wang et al., 2020)

Analysis model	Analysis genotyping strategy and proportion (%)	$\hat{\sigma}_a^2$ (SE)	$\hat{\sigma}_e^2$ (SE)	
H-AM ssGREML	Selective	10	13,358 (635) ^{ab}	22,358 (324) ^{ab}
		20	40,597 (729) ^{ab}	11,904 (243) ^{ab}
		30	55,051 (639) ^{ab}	8695 (136) ^{ab}
	Random	10	9265 (469)	24,231 (270) ^{ab}
		20	8967 (440)	24,382 (243)
		30	8873 (421)	24,461 (223)
A-AM GREML		11,475 (544) ^{ab}	23,148 (217) ⁵⁶	

Assumed $\sigma_a^2=9644$
 $\sigma_e^2=24,798$

Estimated heritability for broiler body weight:
 ~ 0.30 with pedigree
 ~ 0.80 with genotypes

Richter et al., 2022

Wang et al. Genet Sel Evol (2020) 52:31
<https://doi.org/10.1186/s12711-020-00550-w>



RESEARCH ARTICLE Open Access

Effect of genomic selection and genotyping strategy on estimation of variance components in animal models using different relationship matrices

Lei Wang^{1*}, Luc L. Janss¹, Per Madsen¹, John Henshall², Chiyong-Huoy Huang³, Danye Marois², Setegn Alemu¹, AC Sørensen¹ and Just Jensen¹

Recommendation for parameter estimation

- Any methods works if unselected data
 - Genomic info reduces SE
- If selection, use at least 2-3 generations
- REML may be OK with many generations
- Costs with ssGBLUP reduced if data pruned
 - Phenotypes for 2-3 generations
 - Pedigrees 1-2 generations behind phenotypes
 - Only contributing genotypes
- Beware of very low and very high estimates – data selection or poor model

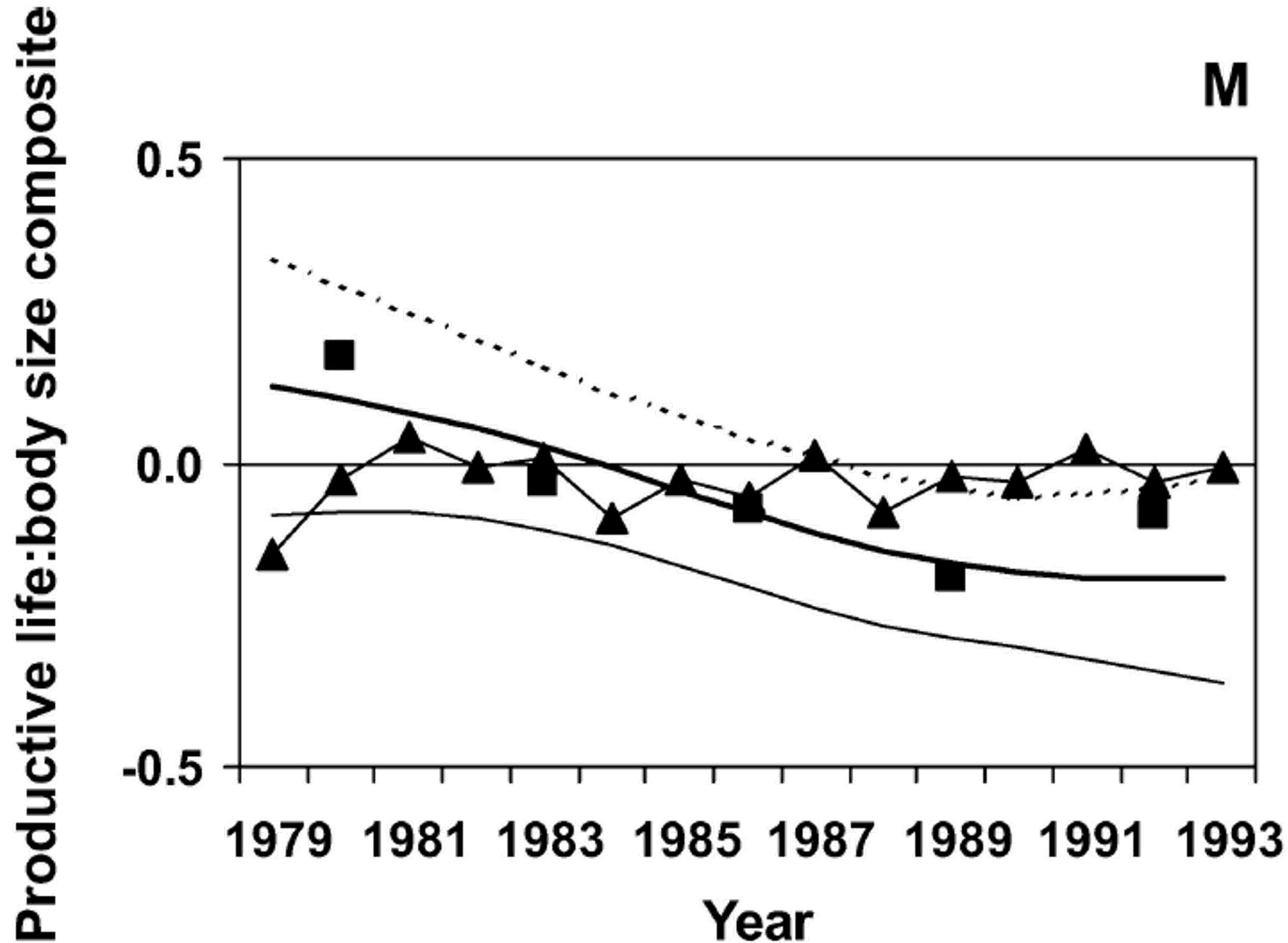
Packages for parameter estimation

- General based on relationships – extensive modeling capabilities
 - BLUPF90+
 - DMU
 - Wombat etc.
 - ASREML
 -
- Packages estimating SNP effects – usually limited modeling
 - GVCBLUP (UMN)
 - GS3 (INRA)
 - GENSEL (IASTATE)
 - Golden Helix
 -

How to estimate changes over time

- Random regression model on year of birth (Tsuruta et al., 2004)
 - Need to use all relevant genotypes
 - Multiple animal effects
 - Very expensive
- Computing by data slices
 - Small slices - less computing, more bias, parameters for base generation

Changing correlations between productive life and body size (Tsuruta et al., 2004)



J. Dairy Sci. 88:1156–1165

© American Dairy Science Association, 2005.

Changing Definition of Productive Life in US Holsteins: Effect on Genetic Correlations

S. Tsuruta,¹ I. Misztal,¹ and T. J. Lawlor²

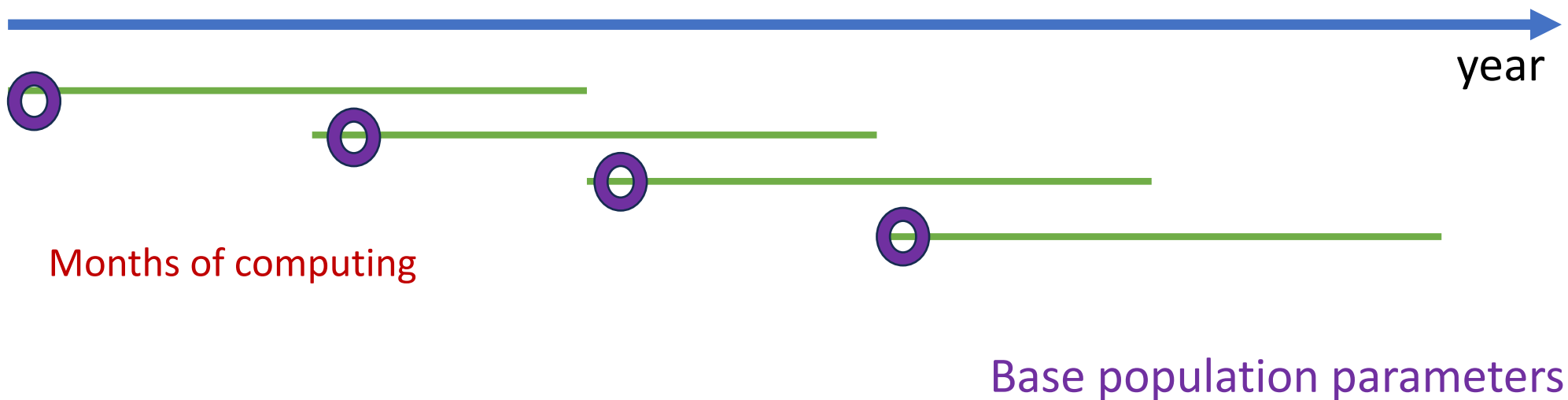
¹Animal and Dairy Science Department, University of Georgia, Athens 30602

²Holstein Association USA Inc., Brattleboro, VT 05301



Parameter estimates by time slices by traditional methods

- Data samples
- Few generations
- Base population parameters



Need new methods

- All data including genomic
- Parameters by last generation/ year
- Reasonable computing cost



Hours of computing

Last generation parameters

Realized and theoretical accuracies

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

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

$M_e \approx 5k$ (chickens, pigs), $10k$ (beef), $15k$ (Holsteins)
Pocrnic et al. (2017)

Pig data set

150k records on growth ($h^2=0.21$)

25k records on fitness ($h^2=0.05$)

53k genotyped animals



Hollifield et al., 2021

Theoretical accuracy

$$\sqrt{\frac{53k * \text{growth}}{53k * 0.21 + 5k}} = 0.83$$

$$\sqrt{\frac{25k * \text{fitness}}{25k * 0.05 + 5k}} = 0.44$$

Realized accuracy

0.82

0.41

Formulas for estimating heritability

$$\widehat{h^2}: \sqrt{\frac{Nh^2}{Nh^2 + M_e}} = \text{corr}(y - Xb, \hat{u})/h$$

$$\widehat{h^2} = \frac{c^2 + \sqrt{c^4 + 4c^2 M_e/N}}{2}, c = \text{corr}(y - Xb, \hat{u})$$

$$SE(\widehat{h^2}) \approx \frac{1}{\sqrt{N_{val}}} \left[c + \frac{2c^2 + \frac{4M_e}{N}}{\sqrt{c^2 + \frac{4M_e}{N}}} \right] \widehat{h^2} \approx \frac{3c}{\sqrt{N_{val}}}$$

N – # animals in reference N_{val} – number of animals in validation

Heritability 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. Misztal¹ 

¹Department of Animal and Dairy Science, University of Georgia, Athens 30602

²Council on Dairy Cattle Breeding, Bowie, MD 20716

³Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705-2350

# animals with phenotypes and genotypes	580k
# animals with validation	381k
Assumed M_e	15k
Initial h^2	0.35
Predictivity	0.55
Calculated h^2	0.33

How to estimate genetic correlations?

Predictivity for trait i

$$\text{corr}(y_i - Xb_i, \hat{u}_i) = \text{acc}_i h_i$$

What is predictivity from trait i to trait j?

$$\text{corr}(y_i - Xb_i, \hat{u}_j) = ?$$

.....
.....

$$\text{corr}(y_i - Xb_i, \hat{u}_j) = \text{acc}_j \text{corr}_{ij} h_i$$

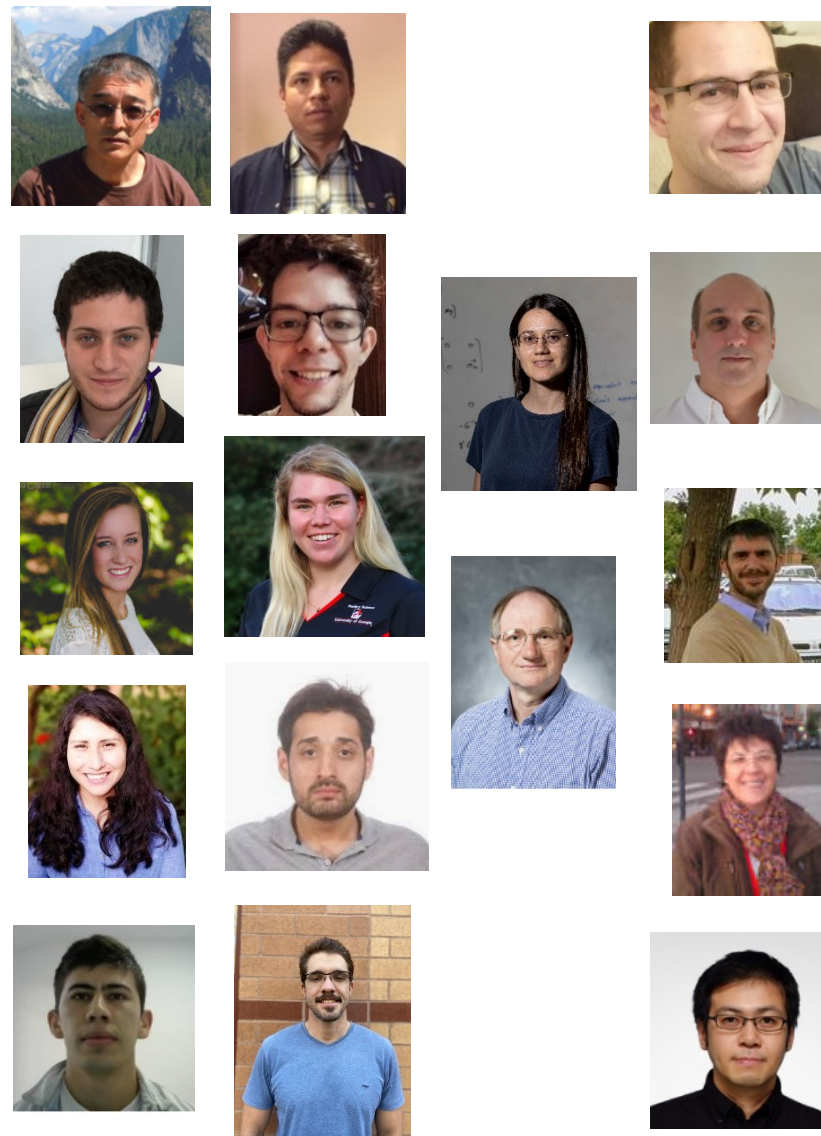
$$\text{corr}_{ij} = \frac{\text{corr}(y_i - Xb_i, \hat{u}_j)}{h_i \text{acc}_j}$$

$$SD(\text{corr}_{ij}) \approx \frac{1}{h_i \text{acc}_j \sqrt{N_{val}}}$$

Conclusions

- Possibly fast changes in genetic parameters under genomic selection
- Current estimates needed for selection index
- Estimation with genomic data by REML/Bayesian methods very expensive with more than 10-20k genotyped animals
- New methods needed to estimate parameters for last generation

UGA AB&G team



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

