Model R^2's in single-step evaluation for udder depth in US Holsteins with different number of genotyped animals and use of external information from Interbull

Tom Lawlor Holstein Association USA

Shogo Tsuruta, Daniela Lourenco, Breno Fragomeni, Yutaka Masuda, Ignacy Misztal University of Georgia

Ignacio Aguilar INIA, Uruguay

U.S. Registered Holsteins
FOR MAXIMUM PROFIT

THE UNIVERSITY OF GEORGIA

2014 Data

- US Holsteins
- 10,067,745 Linear Trait scores on Udder Depth
- 9,561,998 animals
- Model R^2 from Validation bulls

  No daughters in 2010, daughters in 2014

Category of animals	Number of genotyped animals
US Proven bulls	14,447
All Proven bulls	17,310
Proven bulls + cows with records	50,165
Proven bulls + cows with short pedigrees	386,579
All genotyped animals (2014)	569,404

Value of utilizing all genotypes

Genotypes Included	Number of genotyped animals included in analysis		R^2
Proven bulls	17,310	No APY	0.42
Proven bulls *plus* Cows with records	50,165	No APY	0.43
All genotypes	569,404	APY	0.42 b_1=0.98

Refinements to the analysis

- $\mathbf{G}^{-1}$ replaced with $\mathbf{G}_{APY}^{-1}$ from Algorithm for Parent and Young
  - Selection of CORE animals.

- Inbreeding considered as a part of $\boldsymbol{A}^{-1}$

- Account for different segments of the breed.
  - North America versus Europe, Well vs. Poorly recorded (Long vs. Short pedigrees), highly selected vs. not

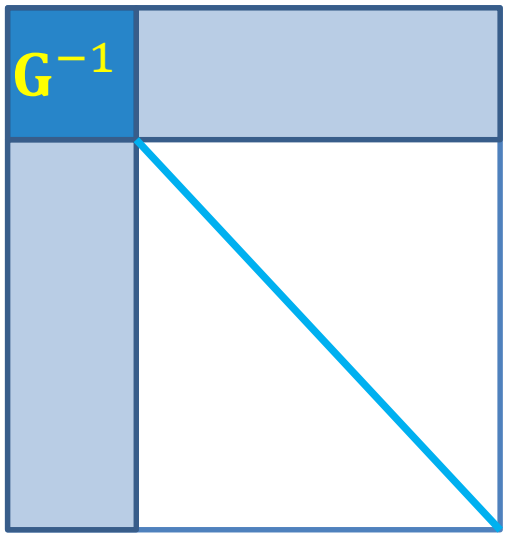
- Including Mace into the ssGBLUP

Inversion of G by APY algorithm

Invert matrix of 1,000,000 animals

Regular $\mathbf{G}^{-1}$

Invert matrix of 15,000 animals



$\mathbf{G}^{-1}$ by APY

1. Define 2 groups: “core” and “non-core”
2. Invert $\mathbf{G}$ of core animals only
3. Calculate APY $\mathbf{G}^{-1}$ using a recursive equation

- Misztal et al. (2014) and Fragomeni et al. (2015)

Inversion of G by APY
Algorithm for Proven and Young
Algorithm for CORE and Non-CORE

- Why does it work?
- How many core animals should you use?
- How should the core animals be selected?

Why does the APY algorithm work.

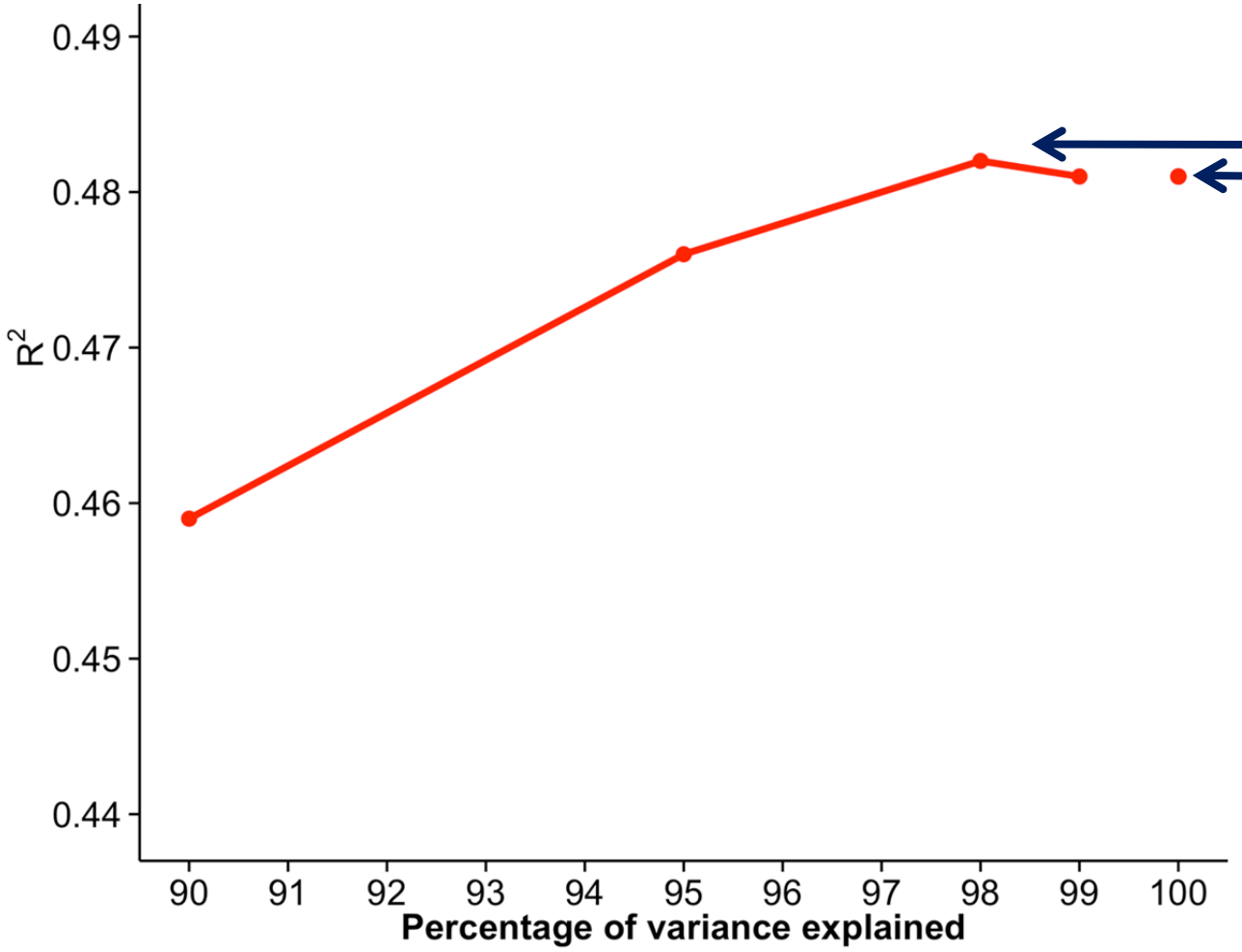
- To maximize the accuracy of the genomic predictions.

  The optimal number of CORE animals is limited in size because

- **The limited rank of the Genomic Relationship Matrix.**
- Number of eigenvalues to explain “most” of the genetic variation.
- **A function of the number of independent chromosome segments.**
- And subsequently the Effective Population Size.

Optimum number of CORE animals

Model R^2 - Holsteins



Optimum $\mathbf{G}^{-1}$ by APY

regular G^{-1}

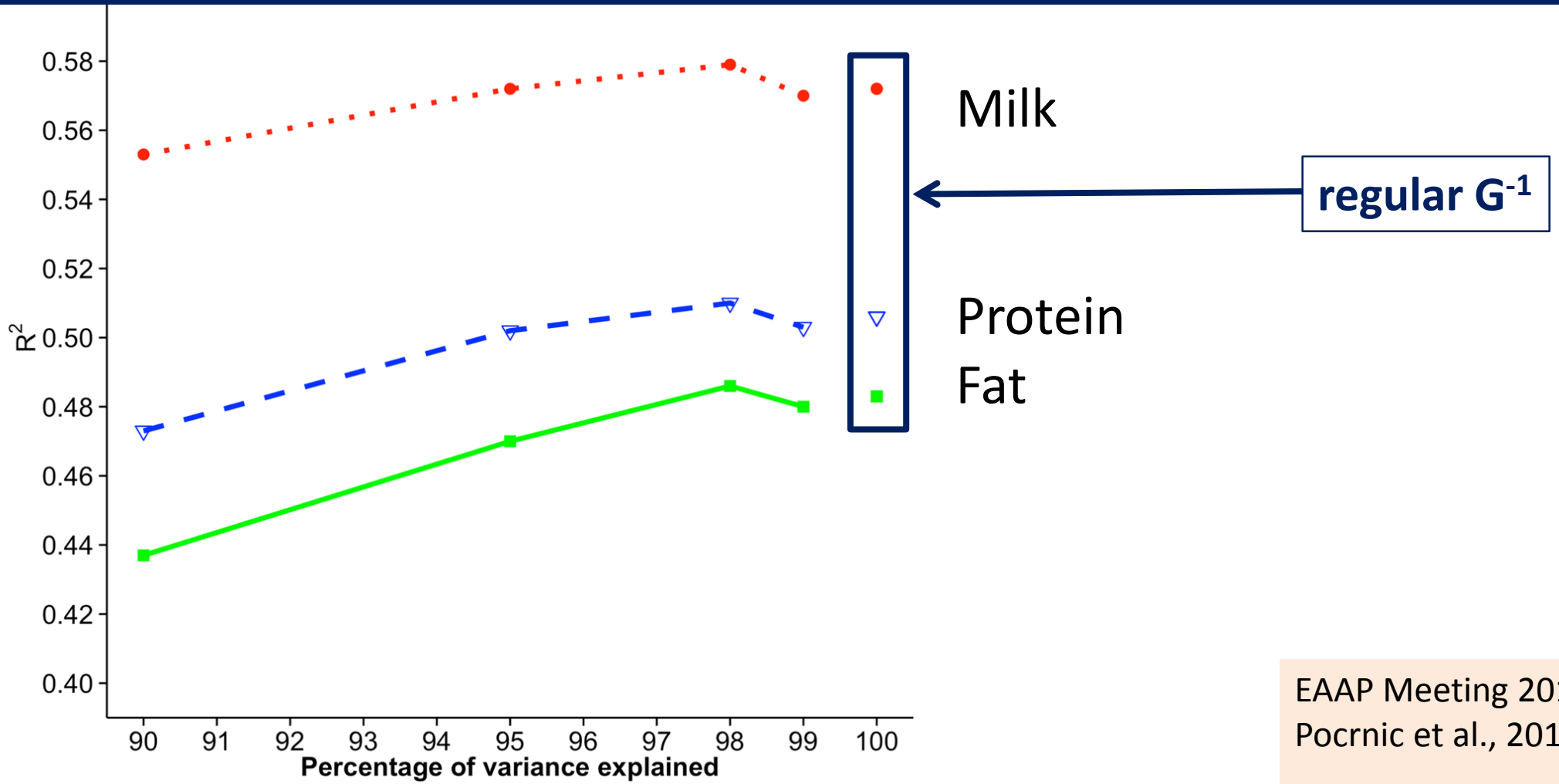
Core: 4500 8000 14,000 19,400

EAAP Meeting 2016
Pocrnic et al., 2016b

Overall Conformation

Optimum number of CORE animals

Model R^2 - Jerseys



Core: 3300 **6100** **11,500**

EAAP Meeting 2016
Pocrnic et al., 2016b

Production Traits

How should CORE animals be selected.

G^{-1} by APY

Description of Genotyped animals included in analysis	Number of Genotyped animals	Number of CORE animals	CORE animals chosen	R^2
Proven bulls plus cows with records	50,165	17,310 animals	All proven bulls	**0.43**
Proven bulls plus cows with records	50,165	17,310 animals	Chosen at random	**0.42**

With complete pedigree and a single genetic base --- choice of CORE is arbitrary. Bradford et al 2016

How should the CORE animals be selected?
Depends on what genotypes are included

Description of Genotyped animals included in analysis	Number of Genotyped animals	Number of CORE animals	CORE animals chosen	R^2
Proven bulls plus cows with short pedigrees	386,579	17,310 animals	All proven bulls	**0.37**
Proven bulls plus cows with short pedigrees	386,579	17,310 animals	Chosen at random	**0.42**

How CORE animals are selected is important.

Ostersen et al. 2016

- Random selection only ensures that some CORE animals are selected across generations

Better Way

- Chose animals from all generations
- **Include the genotyped parents with the most number of genotyped parents**
- Last generation – pick at random

Compatibility of sources of information

G and $\mathbf{A}_{22}$ should be compatible

Forni et. Al 2011; Vitezica et al., 2011, Christensen et al 2012

Degree of homozygosity should be similar between the two matrices

Degree of homozygosity in $\mathbf{A}^{-1}$ should MATCH the degree of homozygosity in $(\mathbf{G}^{-1} + \omega\mathbf{A}_{22}^{-1})$

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

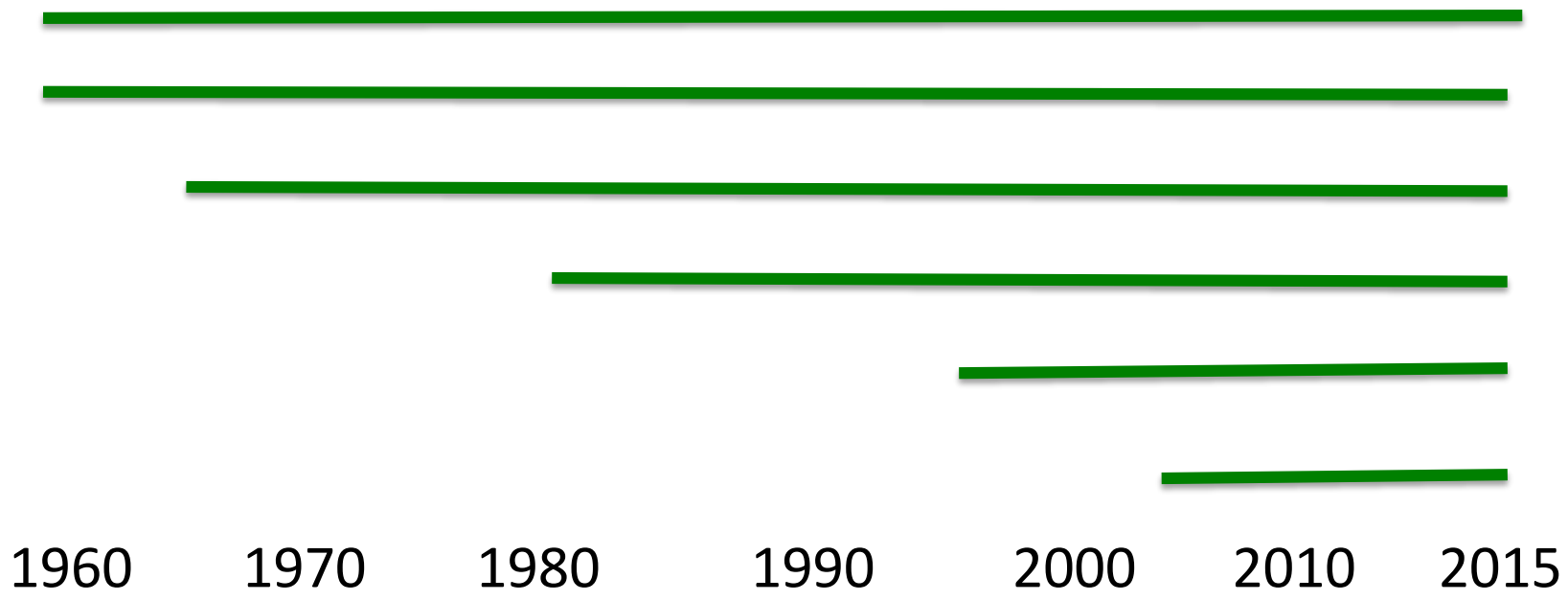
Model should account for differences in time span of Data and Pedigrees.

Including inbreeding in the calculation of $\mathbf{A}^{-1}$

	CORE	Adjustment to Pedigree Relationship matrix
When inbreeding was ignored in A^{-1}		optimal ω=0.70
When inbreeding included in A^{-1}	Random	optimal ω=0.98
When inbreeding included in A^{-1}	Proven	No adjustment ω= 1.00

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

Incomplete pedigrees can cause underestimation of inbreeding and relationships



Removing pedigrees and data prior to 1990 --- results in an increase in accuracy and a reduction in bias

Data	Number of genotyped animals included in analysis	R^2	b_1
All records & pedigrees	569,404	0.41	0.75
Only > 1990	569,404	0.42	0.98

G^{-1} by APY ***CORE =Proven*** **ω=0.98**

To do list ….
Approximate missing relationships
Metafounders *Legarra et al., 2016*

- Use genomic information to identify animals coming from different genetic bases (similar to Unknown Parent Groups).
- **Concept similar to VanRaden, 1992. Utilize average inbreeding of the contemporaries with known relationships**
- Calculate homozygosity relationships for within and across founder groups
- **Incorporate them into the model.**

To do list ….
Include external data from Interbull

- MACE $_{Now}$ = EBV from all countries combined
- **MACE $_{Needed}$ = DYD from all countries EXCLUDING domestic data**

- Literature on external information (e.g. Legarra et al., 2007; Vandenplas and Gengler, 2015)
- **Initial attempt -- Program BLUP90MBE (originally for multibreed beef)**

$$PTA_{NUSA}=(DE_{NUSA}+\alpha)^{-1}\,[(DE_{IB}+\alpha)PTA_{IB}-(DE_{USA}+\alpha)PTA_{USA}]$$

$$PTA_{NUSA,i}=\frac{(DE_{NUSA,i}+\alpha)PTA_{IB,i}-(DE_{USA,i}+\alpha)PTA_{USA,i}}{D_{NUSA,i}+\alpha}$$

Impact of including MACE data

Description of Genotyped animals included in analysis	Number of Genotyped animals	EXTERNAL Data	R^2
Proven bulls plus cows with records	50,165	none	**0.43**
Proven bulls plus cows with records	50,165	MACE	**0.49**

Neither model – adjusted to optimize R^2 or b_1

Conclusions

- $\mathbf{G}_{APY}^{-1}$ can handle a large number of genotypes.
- **CORE animals are now more clearly defined.**
- Need for Omega (**ω**), greatly reduced or eliminated.

- **To Do - account for multiple ancestral bases.**
  - **Include external data– with no double counting**