



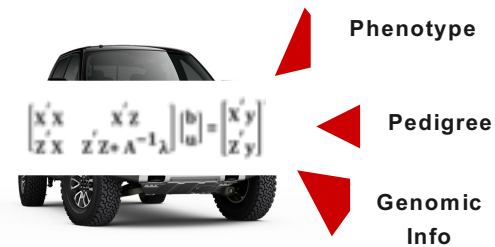
Developments in single-step for beef cattle genomic evaluation in the US

Daniela Lourenco

S. Tsuruta, I. Pocrnic, A. Legarra, B.O. Fragomeni, Y. Masuda,
I. Aguilar, S. Miller, D. Moser, I. Misztal

BIF - 06/22/2018

Single-step

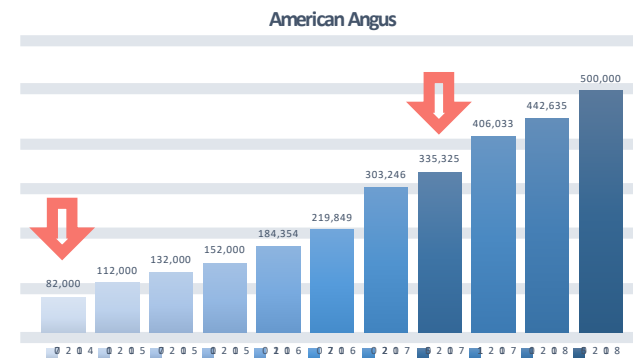


Developments in single-step

- Validation
 - Categorical traits and maternal effect
- Large-scale genomic evaluation
- Indirect prediction with APY
- Recent projects by UGA group

3

Number of genotyped animals



4

Ability to predict future performance

2014

- 8M animals in pedigree
- 6M BW and WW
- 3.4M PWG
- 52k genotyped animals
- 18.7k born in 2013

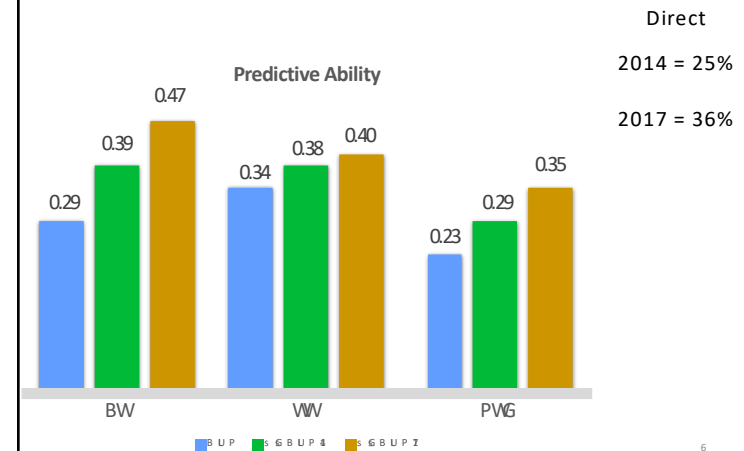
2017

- 10M animals in pedigree
- 8M BW and WW
- 4.2M PWG
- 335k genotyped animals
- 18.7k born in 2016

Predictive ability = $\text{COR}(Y_{\text{adj}}, \text{GEBV})$

5

Growth traits



Validation for Calving Ease

Feb/2017

- 9M animals in pedigree
- 8M BW
- 1.5M CE
- 303k genotyped animals
- 6.6k born in 2016

Predictive ability = $\text{COR}(Y_{\text{adj}}, \text{GEBV})$

7

Calving Ease is categorical (binary)!

- Phenotypes are 1 and 2
- Adjusted phenotypes?
- EBV and GEBV are in “continuous scale”
- Lourenco et al. (2015)
- EBV = 0.12 vs. GEBV = 0.13

Predictive ability = $\text{COR}(Y_{\text{adj}}, \text{GEBV})$



8

Maternal effect goes back!

- Phenotypes recorded in the progeny
- Genetic + environment of dam
- $\text{Cor}(Y_{i_adj}, \text{EBV}_{i_mat})?$
- $\text{Cor}(Y_{i_adj}, \text{EBV}_{mat_dam})?$
- $\text{Cor}(Y_{i_adj}, \text{EBV}_{total_maternal})?$



More robust way to validate in this situation

9

Validation

- LR Method
 - Linear Regression metrics
 - Legarra & Reverter (2017; 2018)



Technical Note: Detection of Bias in Genetic Predictions^{1,2}

A. Reverter, B. L. Golden, R. M. Bourdon, and J. S. Brinks

Department of Animal Sciences, Colorado State University, Fort Collins 80523

- Consistency between subsequent evaluations
- Partial and Whole evaluations
- Validation animals have no phenotypes in *Partial data* but do have phenotypes in *Whole data*
- Metrics

10

LR Validation

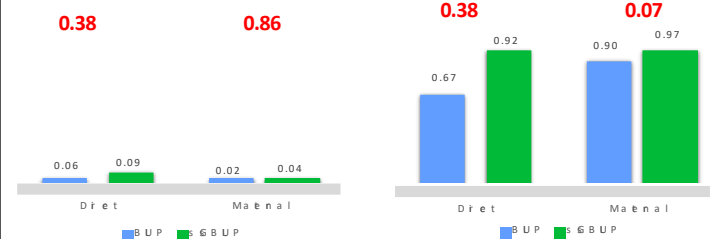
- Relative increase in accuracy EBV = $\rho_{EBV_p, EBV_w} = \text{COR}(EBV_p, EBV_w)$
- Relative increase in accuracy GEBV = $\rho_{GEBV_p, GEBV_w} = \text{COR}(GEBV_p, GEBV_w)$
- Gain in Accuracy = $\rho_{GEBV_p, GEBV_w} / \rho_{EBV_p, EBV_w}$
- Inflation of EBV = $EBV_w = b_0 + b_1 EBV_p$
- Inflation of GEBV = $GEBV_w = b_0 + b_1 GEBV_p$

11

Validation for CE – Predictivity

$\text{COR}(Y_{adj}, \hat{u})$

LR Method
 $\text{COR}(\hat{u}_p, \hat{u}_w)$

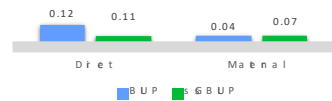


12

Validation for CE – b1 or inflation

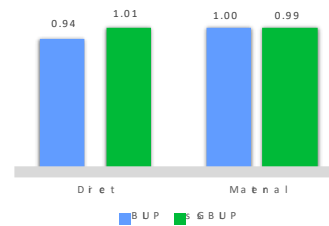
$$Y_{\text{adj}} = b_0 + b_1 \hat{u}$$

Y_{adj} and u not in the same scale!



LR Method

$$\hat{u}_w = b_0 + b_1 \hat{u}_p$$



13

LR Validation

- Compares EBV with EBV and GEBV with GEBV
 - Similar scale
- Seems to work for complex models and traits
 - Binary, low heritability, maternal models
- Still needs extensive tests (Macedo et al., 2018)
 - Extreme scenarios

14

Large-scale genomic evaluations

- Few organizations
- Methods available:
 - APY ssGBLUP (Misztal et al., 2014)
 - Indirect representations of \mathbf{G}
 - ssGBLUP with SNP effect and GEBV (Legarra & Ducroq, 2012)
 - SSBR or Super Hybrid Model (Fernando et al., 2016)
 - Sherman-Woodbury inversions
 - ssGTBLUP (Mantysaari et al., 2017)

$$\mathbf{G}^{-1} = \mathbf{I} - \left(\mathbf{Z} \left(\mathbf{Z}^T \mathbf{Z} + \mathbf{I} \right)^{-1} \mathbf{Z}^T \right)$$

15

Comparisons

APY ssGBLUP vs. ssGTBLUP

APY ssGBLUP vs. Super Hybrid Model

16

Large-scale genomic evaluations APY ssGBLUP

Single-step
before APY



Single-step
after APY



APY ssGBLUP

- APY is just an algorithm to construct \mathbf{G}^{-1} when inverting \mathbf{G} is **computationally not feasible**
- Based on the dimensionality of genomic information

$$\# \text{ genotyped animals} > \# \text{ SNP}$$

$$\mathbf{G} = \alpha \mathbf{G} + (1-\alpha) \mathbf{A} \mathbf{A}^T$$

VanRaden (2008)

- \mathbf{G} has a limited dimensionality

Independent blocks

Dependent blocks

- Independent blocks = Core (truck)
- Dependent blocks = Noncore (trailer)

18

How many core animals in APY?



81k

11k ~ 14k



77k

14k ~ 19k



75k

11k ~ 16k



16k

4k ~ 6k



23k

4k ~ 6k

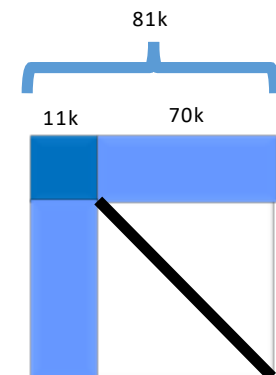
19

How many core animals in APY?



Core = 11k

noncore

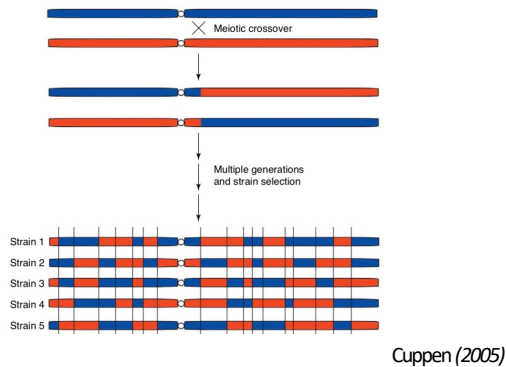


$$\text{Cor}(\text{GEBV}, \text{GEBV}_{\text{APY}}) > 0.99$$

20

Number of core animals in multibreed APY

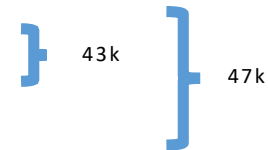
- More breeds = more segments



21

Number of core animals in multibreed APY

- Line 1 = 27k
- Line 2 = 16k
- F1 = 4k



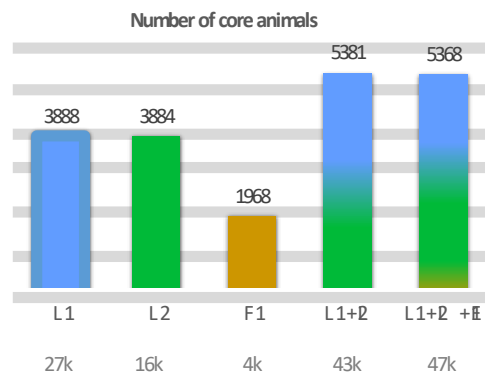
Poomic et al. (2018)



- Number of core animals
- Separate lines
- Both lines
- Both lines and F1

22

Number of core animals in multibreed APY



23

Who is using APY single-step?

- American Angus



- 500k genotyped animals
- 19k core
- all traits
- ~ 2 hour (G^{-1} APY)



Summer/2017

- Holsteins

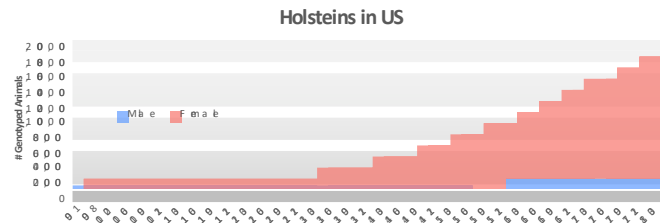


Fall/2016

- ~500k genotyped animals
- Wellness traits

24

APY with 2M genotyped animals?



- Is it feasible?
 - $G^{-1} = 29 \text{ Tb}$
 - APY G^{-1} 14k core = 208 Gb



Yutaka Masuda

25

Additional features in ssGBLUP

- Single-step outputs GEBV
- We need SNP effect as well
- Commercial products
 - e.g. GeneMax for non-registered animals
 - Based on SNP effects

26

Indirect predictions in ssGBLUP

$$\begin{bmatrix} X'X & X'W \\ W'X & W'W+H^{-1}\lambda \end{bmatrix} \begin{bmatrix} h \\ t \end{bmatrix} = \begin{bmatrix} X'y \\ W'y \end{bmatrix}$$

$$\alpha = \lambda D Z' G^{-1} u$$

SNP effects GEBVs

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

G_{APY}^{-1}

Can we use G_{APY}^{-1} for SNP effect?

27

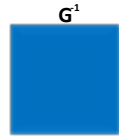
Dataset

- AAA
 - 8.2M animals in pedigree
 - 6.2M BW
 - 6.8M WW
 - 3.4M PWG
 - 81k genotyped
 - born 1977-2012: 66k
 - born 2013-2014: 15k
- Complete
 - Phenotypes up to 2012
 - Genotypes up to 2014 (81k)
- Reduced
 - Phenotypes up to 2012
 - Genotypes up to 2012 (66k)
- 3-trait with mat and mpe
 - Results for PWG

28

SNP effects in APY ssGBLUP

$$\mathbf{a}_g = \lambda \mathbf{DZ}' \mathbf{G}^{-1} \mathbf{u}$$



$$\mathbf{a}_{g_{APY}}^{-1} = \lambda \mathbf{DZ}' \mathbf{G}_{APY}^{-1} \mathbf{u}_{APY}$$



$$\mathbf{a}_{g_{cc}}^{-1} = \lambda \mathbf{DZ}' \mathbf{G}_{cc}^{-1} \mathbf{u}_{APY}$$



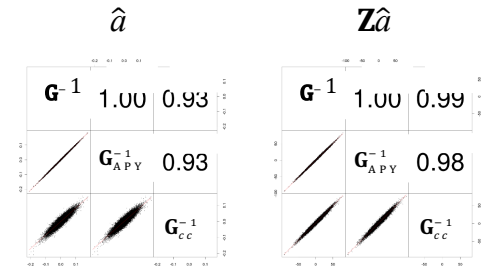
29

SNP effects in APY ssGBLUP

$$\mathbf{a}_g = \lambda \mathbf{DZ}' \mathbf{G}^{-1} \mathbf{u}$$

$$\mathbf{a}_{g_{APY}}^{-1} = \lambda \mathbf{DZ}' \mathbf{G}_{APY}^{-1} \mathbf{u}_{APY}$$

$$\mathbf{a}_{g_{cc}}^{-1} = \lambda \mathbf{DZ}' \mathbf{G}_{cc}^{-1} \mathbf{u}_{APY}$$



30

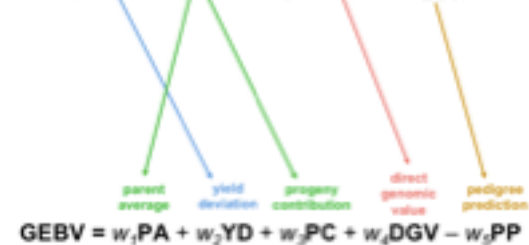
Additional features in ssGBLUP

- Interim evaluations
 - Indirect predictions
 - Quick evaluations between official runs
- Should be comparable to GEBV

31

Indirect predictions for young animals

$$\left\{ \mathbf{W}'\mathbf{W} + \alpha \mathbf{A}^{-1} + \alpha \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix} \right\} \hat{\mathbf{u}} = \mathbf{W}'\mathbf{y}$$



$$\text{GEBV}_y = w_1 \text{PA} + w_4 \text{DGV} - w_5 \text{PP}$$

$$\text{GEBV}_y \approx \text{DGV} = \mathbf{Z}\mathbf{a}$$

Lourenco et al., 2015

32

Problems with Indirect predictions

Genetic evaluation using single-step genomic best linear unbiased predictor in American Angus¹

D. A. L. Lourenco,^{*2} S. Tsuruta,^{*} B. O. Fragomeni,^{*} Y. Masuda,^{*} I. Aguilar,[†]
A. Legarra,[‡] J. K. Bertrand,^{*} T. S. Amen,[§] L. Wang,[§] D. W. Moser,[§] and I. Misztal^{*}
© 2015 American Society of Animal Science. All rights reserved. J. Anim. Sci. 2015:93:2653–2662
doi:10.2527/jas2014-8836

$$\text{COR}(\text{GEBV}, \mathbf{Za}) > 0.99$$

$$\text{Avg}(\text{GEBV}) \approx 100 \quad \rightarrow \quad \text{Avg}(\mathbf{Za}) \approx 0$$

How to make \mathbf{Za} compatible to GEBV?

33

How to make $\mathbf{Z}\hat{a}$ compatible to $\widehat{\text{GEBV}}$?

Understanding genetic and genomic bases

- Base of BLUP: *founders of the pedigree*
- Base of SSGBLUP: modelled as a mean for genotyped
- $p(\mathbf{u}_g) = N(\mathbf{1}\mu, \mathbf{G})$
 $\mu = (\text{Pedigree base}) - (\text{Genomic base})$

Vitezica et al. (2011)

34

How to make $\mathbf{Z}\hat{a}$ compatible to $\widehat{\text{GEBV}}$?

1) Formula in Legarra (2017) $\mathbf{u}_{ip} = \mu + 0.95\mathbf{Za} + 0.05\mathbf{u}_{parents}$

2) Double fitting

a) fit a regression using genotyped animals in the evaluation

$$\text{GEBV}_{eval} = b_0 + b_1 \mathbf{Za}$$

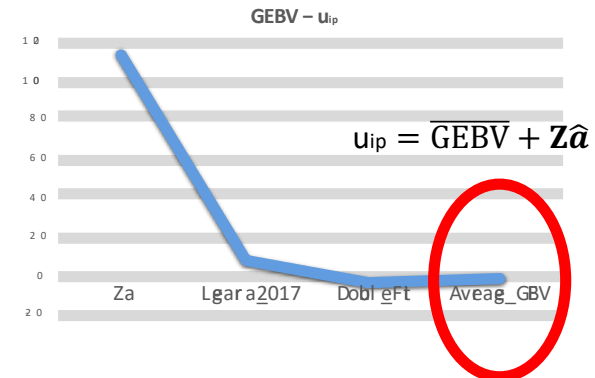
b) apply regression for indirectly predicted animals

$$\mathbf{u}_{ip} = b_0 + b_1 \mathbf{Za}$$

3) Add average GEBV $\mathbf{u}_{ip} = \overline{\text{GEBV}_{eval}} + \mathbf{Za}$

35

Bias of indirect predictions



Lourenco et al., 2018

36

APY ssGBLUP + Indirect Predictions

- Indirect predictions are unbiased after corrections
 - Average GEBV, double fitting or Legarra (2017)
 - Can be used as interim evaluation
- Indirect Predictions and SNP effects can be calculated
 - \mathbf{G}_{APY}^{-1} or core \mathbf{G}^{-1}
- Investigating with 500k genotyped animals for all traits

37

Under development at UGA

- PEV/PEC for SNP
 - Accuracy for Indirect Predictions
- Formulas to calculate SNP variance
- Multibreed evaluations
- Bias in genomic evaluations
- Dimensionality of genomic information
- Sequence data in ssGBLUP

38

Under development at UGA

- QCf90 (Masuda et al., 2018)
 - QC with bitwise operations
 - Works with raw or renumbered data
 - 570k genotyped for 61k SNP



Step	QCF90	PREGSF90
Computing time	917 sec.	2708 sec.
Memory usage	9 GB	257 GB

39

Take Home

- Single-step is the new standard for beef cattle genomic evaluation
- All industries are moving to single-step
- Under constant improvement
- Scientists: keep improving and developing methods for more accurate evaluation
- Producers: keep collecting data
 - Phenotypes, pedigree, genotypes

40