

UNIVERSITY OF GEORGIA
College of Agricultural & Environmental Sciences

How to validate genomic predictions?

Comparison between genomic and non-genomic models

Daniela Lourenco
BLUPF90 TEAM – 02/2023

1

UNIVERSITY OF GEORGIA

How to validate genomic predictions?

- We should look at accuracy!!!

Prediction (validation) Accuracy

- Accuracy
- Reliability
- Prediction Accuracy
- Prediction Reliability
- Predictive Ability (predictability)

Theoretical Accuracy

- Accuracy
- Reliability
- BIF Accuracy

2

UNIVERSITY OF GEORGIA

Understanding accuracy

- Theoretical Accuracy**
 - Precision or stability of EBV
 - How much EBV changes when more data is added (risk)
 - SE of EBV
 - $Acc = \sqrt{1 - \frac{PEV}{\sigma_u^2}}$
 - $Acc_{inb} = \sqrt{1 - \frac{PEV}{\sigma_u^2(1+F)}}$
 - Individual
 - Model-based

- Prediction Accuracy**
 - $\rho = COR(u, \hat{u})$
 - $\Delta G = i \rho \sigma_u / L$ (response to selection)
 - Potential ΔG of a breeding scheme
 - Very popular after genomics
 - Cross-validation
 - Population

3

UNIVERSITY OF GEORGIA

Theoretical Accuracy of EBV

Theoretical Accuracy Increases as an more information is added

Age (years)	Theoretical Accuracy
0	0.0
0.5	0.4
1.5	0.6
3.5	0.8

4

Accuracy and possible EBV change


BW EBV of 1.8	ACC	Possible Change
Bull A	.05	2.49
Bull B	.80	.53

Possible change = interval of 95% = $EBV \pm 1.96 \times SE$

Modified from: Dan Moser's GA talk - 2016

5

Accuracy and possible EPD change



Bull A, Acc. = .05
1.8 BW EBV \pm 2.49
-0.69 to 4.29

Bull B, Acc. = .80
1.8 BW EBV \pm .53
1.27 to 2.33

Which EBV will change most?

Which Bull has more reliable EBV?

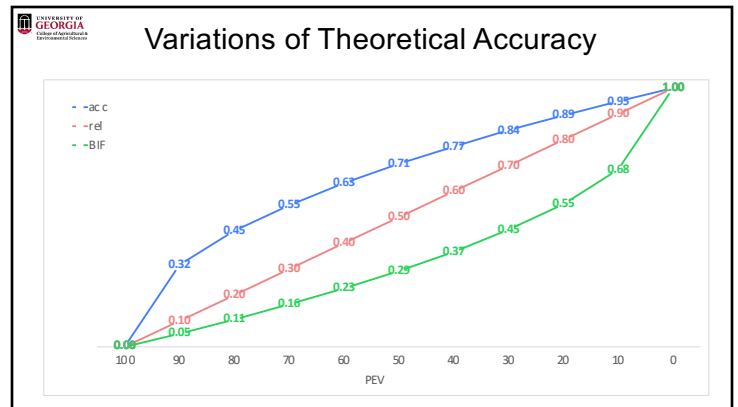
Modified from: Dan Moser's GA talk - 2016

6

Variations of Theoretical Accuracy

- Several: $Accuracy = \sqrt{1 - \frac{PEV}{\sigma_u^2(1+F)}}$ → Henderson (1975) Derivations under selection or not
- Beef cattle: $BIF Accuracy = 1 - \frac{PEV}{\sigma_u^2(1+F)}$ → Lower values Approaches 1 very slowly Willham (~1985)
- Dairy cattle: $reliability = 1 - \frac{PEV}{\sigma_u^2(1+F)}$ → Lower values Approaches 1 more slowly Fraction of σ_u^2 accounted for by EBV VanRaden et al. (~1989)

7



8

How to validate: Prediction accuracy

- Reflects the correlation between true and estimated breeding value
 - $accuracy = COR(u, \hat{u})$
- Do we have true breeding values in real populations?
 - $accuracy = COR(benchmark, \hat{u})$

Something we are trying to predict

→

Future performance
 (Progeny) yield deviation
 Deregressed EBV
 High accuracy EBV
 Future EBV

9

Prediction accuracy

Training

Validation

Adapted from Wójcik et al. (2015)

10

Prediction Accuracy

- Complete data (used to compute the benchmark)
- Reduced data (used to compute GEBV and EBV)

11

Which benchmark to use?

- $accuracy = COR(benchmark, \hat{u})$

Validation animals	Trait measured on	Benchmark
Dairy bulls	progeny	daughter yield deviation / deregressed proof

12

DYD or DEBV as benchmark

Reduced data:

Pedigree ———— Dec. 2013 ———— Dec. 2017

Genotype - - - - -

Phenotype ————

- Remove 4 to 5 years of data
- Bulls with no daughter records in the reduced data
- Bulls have at least 10 daughters in the complete data
- At least 100 bulls

• Benchmark: Deregressed EBV (based on VanRaden et al., 2009)

$$DEBV_{complete} = \frac{EBV_{complete} - PA_{complete}}{R_{complete}} + PA_{complete}$$

$$DEBV_{complete} = b_0 + b_1(G)EBV_{reduced}$$

R^2 = prediction reliability
 b_0 = bias
 b_1 = dispersion

13

DYD or DEBV as benchmark

- Are bias and dispersion also important?

Adapted from Legarra and Reverter (2017)

14

Adjusted phenotypes as benchmark

Reduced data:

Pedigree ———— 2018 ———— 2019

Genotypes - - - - -

Phenotypes ————

- Remove 1 or 2 years of data
- Validation animals with own phenotypes in the complete data
- Phenotypes adjusted for fixed effects (complete data)

Predictivity or predictive ability of (G)EBV = $Cor(Y_{adj}, (G)EBV_{reduced})$

$$Accuracy = \frac{Predictivity}{\sqrt{h^2}}$$

$$Y_{adj} = b_0 + b_1(G)EBV_{reduced}$$

b_0 = bias
 b_1 = dispersion

15

A new validation method

- LR Method
 - Linear Regression metrics
 - Legarra & Reverter (2018; GSE)
- Reduced (partial) and Complete (whole) data
- Validation animals have phenotypes in the *complete data* but not in the *reduced data*
- Benchmark: complete (G)EBV
- Compares EBV with EBV and GEBV with GEBV
 - Same scale

16

LR validation

- Accuracy

$$Accuracy = \frac{cov(\hat{u}_c, \hat{u}_r)}{\sqrt{(1+F-2f)\sigma_{u_{c\&os}}^2}} \quad \text{or} \quad Accuracy = \frac{cov(\hat{u}_c, \hat{u}_r)}{\sqrt{(1-F)\sigma_u^2}}$$

- Dispersion $\hat{u}_c = b_0 + b_1 \hat{u}_r$

- Bias $\mu_{cr} = \bar{\hat{u}}_r - \bar{\hat{u}}_c$

- Consistency between subsequent evaluations $cov_{cr} = cov(\hat{u}_c, \hat{u}_r)$

- Estimator of the ratio of accuracies using the "reduced" or the "complete" data