Accuracy or Accuracy?

Daniela Lourenco

BLUPF90 TEAM – 11/2022

- We should look at accuracy!!!

Prediction (validation) Accuracy

Accuracy
Reliability
Prediction Accuracy
Prediction Reliability
Predictive Ability (predictability)

Theoretical Accuracy

Accuracy
Reliability
BIF Accuracy

- **Theoretical Accuracy**

  - Precision or stability of EBV

  - How much EBV changes when more data is added (risk)

  - SE of EBV

  - $Acc = \sqrt{1 - \dfrac{PEV}{\sigma_u^2}}$

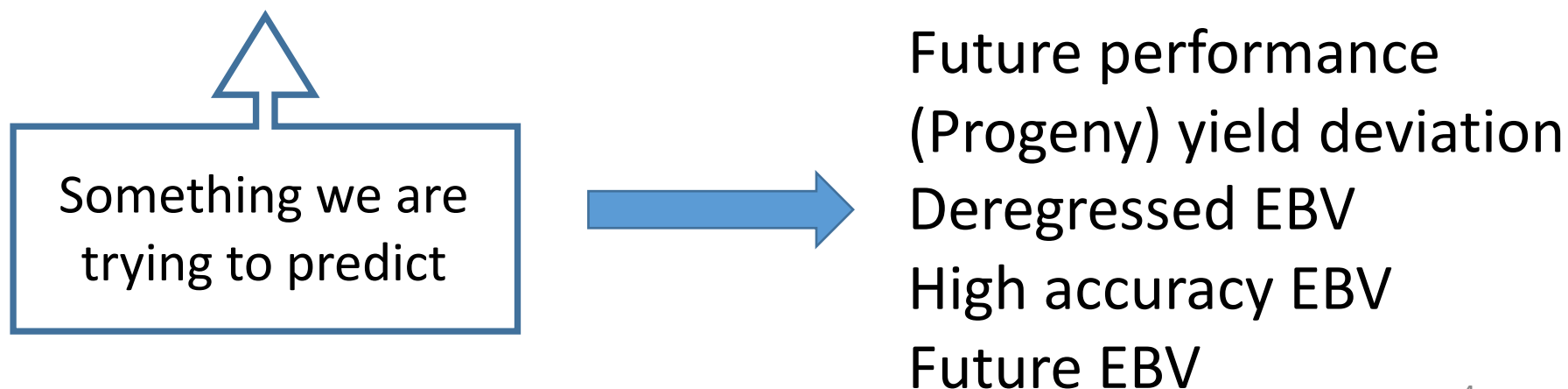 - $Acc\_inb = \sqrt{1 - \dfrac{PEV}{\sigma_u^2(1+f)}}$

  - Individual

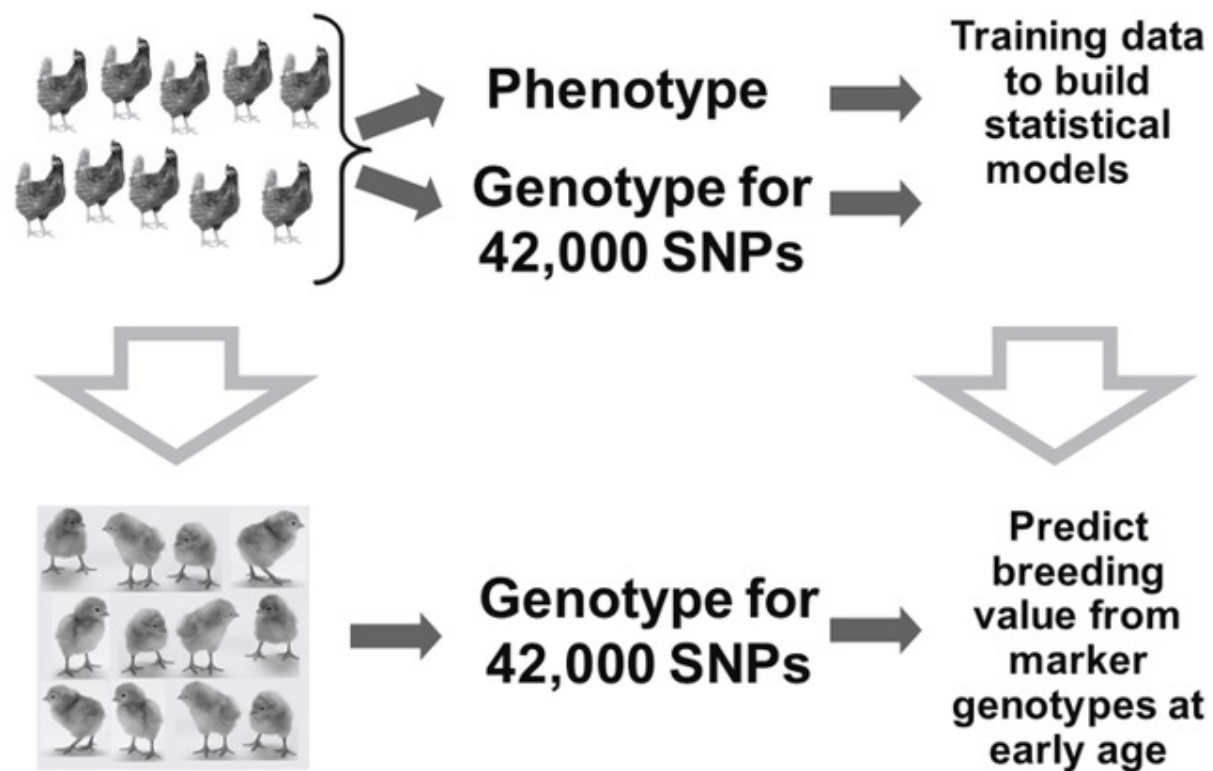 - Model-based

- **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

- Reflects the correlation between true and estimated breeding value

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

- Do we have true breeding values in real populations?
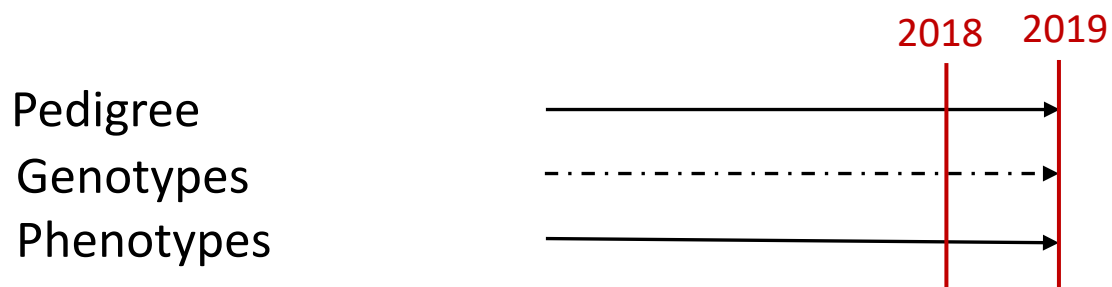
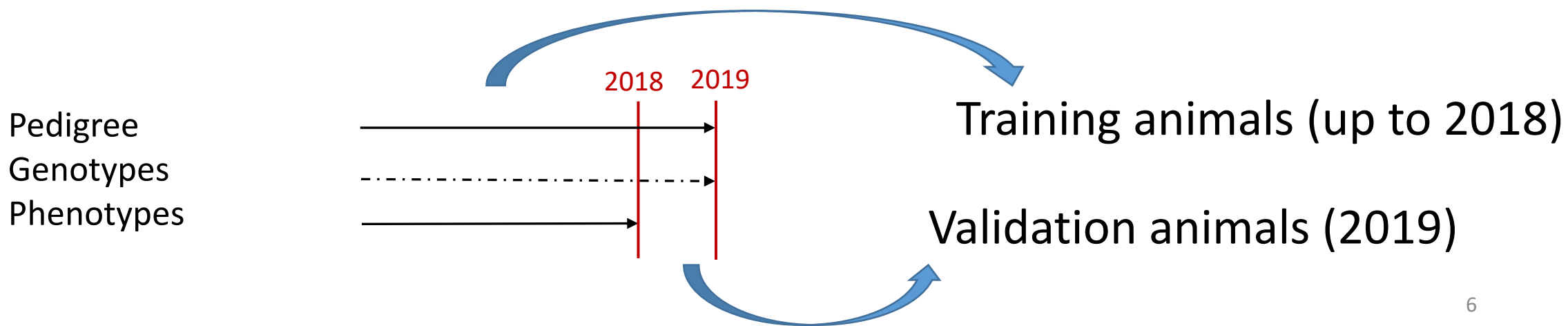 - $accuracy = COR(benchmark, \hat{u})$

Something we are trying to predict

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

Prediction accuracy

Adapted from Wolc et al. (2015)

Training

Validation

Prediction Accuracy

- Complete data (used to compute the benchmark)

2018 2019

Pedigree
Genotypes
Phenotypes

- Reduced data (used to compute GEBV and EBV)

2018 2019

Pedigree
Genotypes
Phenotypes

Training animals (up to 2018)

Validation animals (2019)

Which benchmark to use?

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

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

Reduced data:

Dec. 2013   Dec. 2017

Pedigree

Genotype

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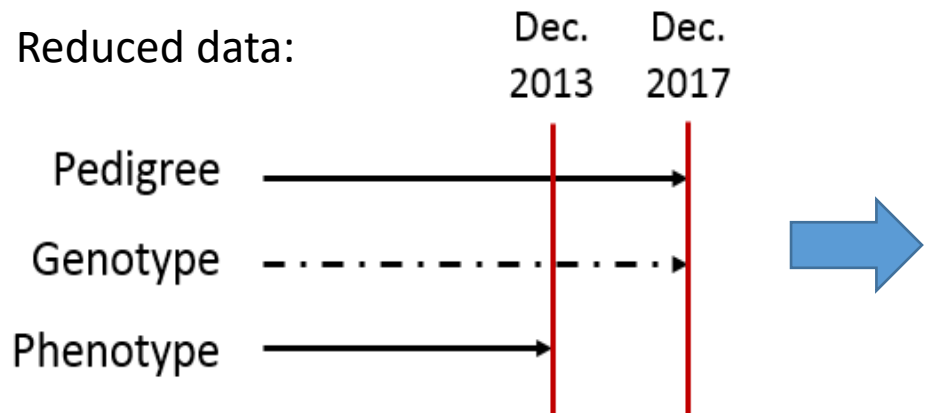
- Remove 4 to 5 years of data
- Bulls with no daughter records in the reduced data
- Bulls have at least 10 daughters in the complete data
- At least 100 bulls

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

$$\text{DEBV}_{\text{complete}} = \frac{\text{EBV}_{\text{complete}} - \text{PA}_{\text{complete}}}{\text{R}_{\text{complete}}} + \text{PA}_{\text{complete}}$$

$$\text{DEBV}_{\text{complete}} = \text{b}_0 + \text{b}_1(\text{G})\text{EBV}_{\text{reduced}}$$

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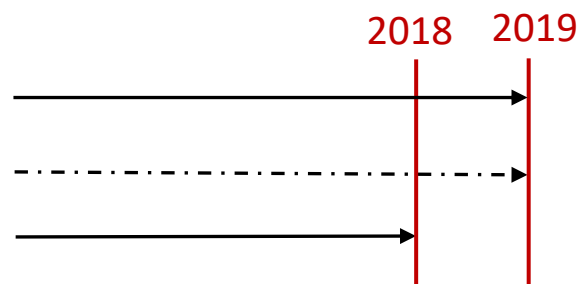b_1 = dispersion

- Are bias and dispersion also important?

Adapted from Legarra and Reverter (2017)
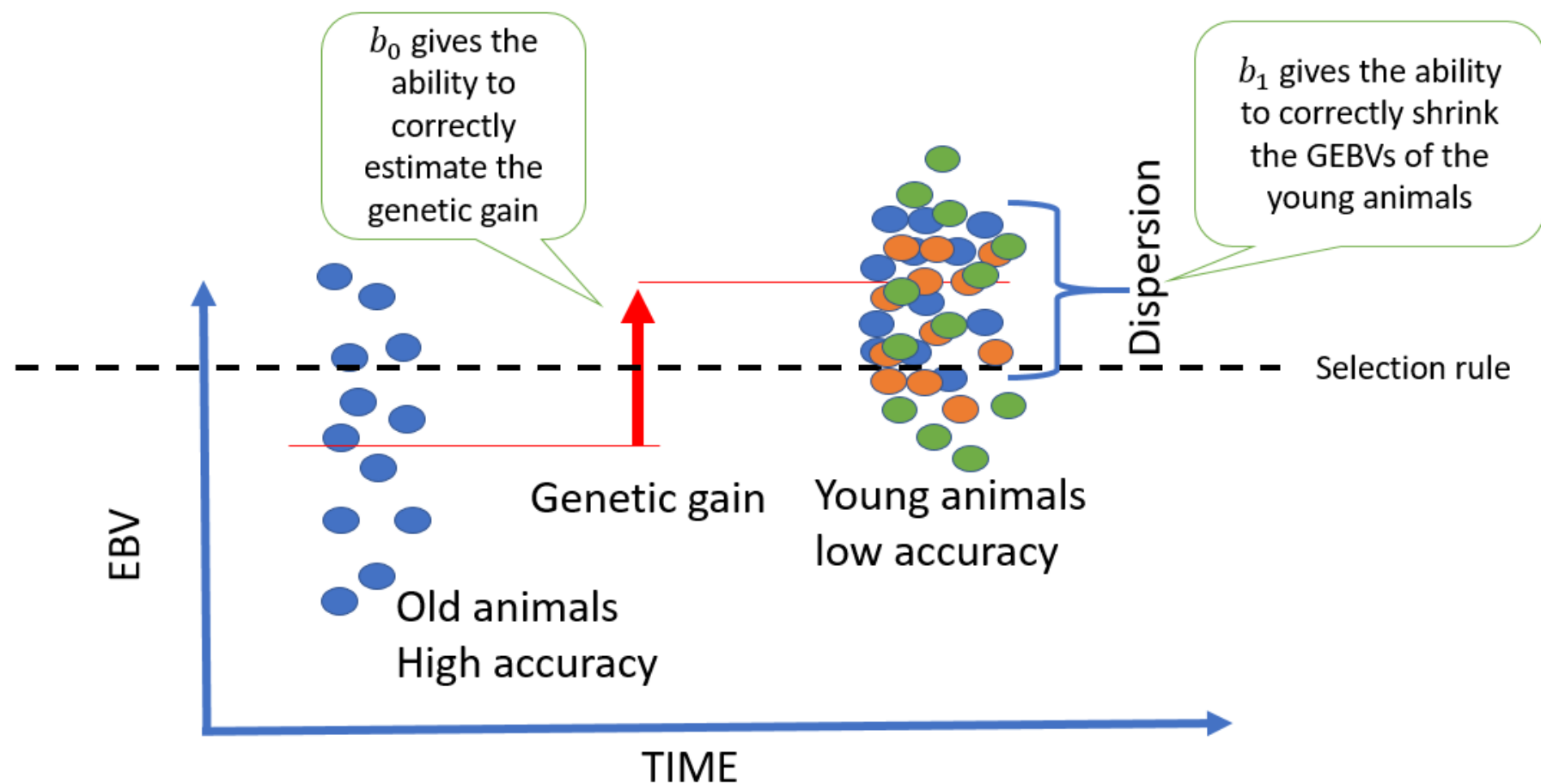
Reduced data:

Pedigree
Genotypes
Phenotypes

2018   2019

- 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

- 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

- Accuracy

$$Accuracy = \sqrt{\frac{cov(\widehat{\boldsymbol{u}}_c, \widehat{\boldsymbol{u}}_r)}{(1 + \bar{F} - 2\bar{f})\sigma_{u,\infty}^2}}$$
or
$$Accuracy = \sqrt{\frac{cov(\widehat{\boldsymbol{u}}_c, \widehat{\boldsymbol{u}}_r)}{(1 - \bar{F})\sigma_u^2}}$$

- Dispersion $\qquad \widehat{\boldsymbol{u}}_c = \mathrm{b}_0 + \mathrm{b}_1 \widehat{\boldsymbol{u}}_r$

- Bias $\qquad \mu_{cr} = \overline{\widehat{\boldsymbol{u}}_r} - \overline{\widehat{\boldsymbol{u}}_c}$

- Consistency between subsequent evaluations $\qquad cor_{c,r} = cor(\widehat{\boldsymbol{u}}_c, \widehat{\boldsymbol{u}}_r)$

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