Risks of forecasting in Animal Breeding

- Our models are an oversimplification of reality: good in the short run, bad in the long run
- Spelman et al:
  - New Zealand Animal Evaluation Limited (NZAEL), […] first included genomic information in the national evaluation in 2009 […]
  - The first two crops of DNA-proven sires, used in 2008 and 2009 […] the initial GEBVs of these sires were found to be over-estimated […] as a gesture of appreciation to the early adopters of genomic evaluation, LIC credited the $5 premium that the farmers paid
- We need tools to rank, understand and quantify the behavior of prediction models in an “animal breeding” context
- The need for these tools has dramatically increased with genomic selection, that takes riskier decisions

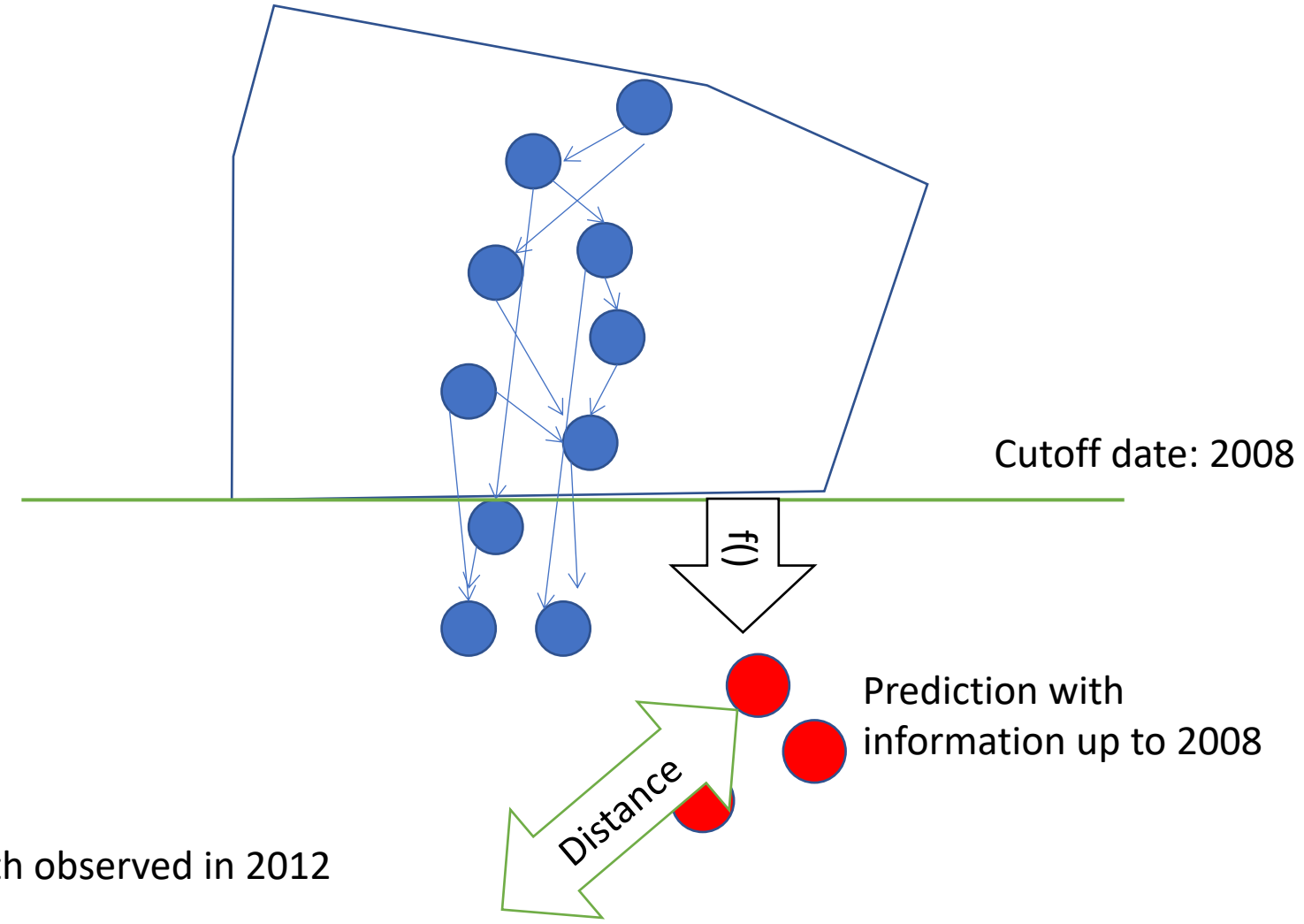
Why do we need cross-validation?

- Classic statistical methods to compare models (AIC…) only inform about global fit and parsimony
- All records are equal, but some records are more equal than others

Cross-validation in a nutshell

- Split data into “training” and “validation”,
- using a model and data in “training” predict “data” in “validation”
- Measure distance from predicted to observed data

Example: Forward cross-validation



Cross-validation in a nutshell

- Split data into “training” and “validation”, using a model predict “data” in “validation”
- Measure distance from predicted to observed data
- **Problems:**
  - Which distance
  - How to split
  - How we interpret the results
- In Animal Breeding Data Is Selected and has a time trend
- The machine learning people does not have the problem of selected data

Which kind of cross-validation should we use?

- Random k-fold, leave-one-out, k-means cluster for crossvalidation:
  - Easy to implement
  - Produce s.e. of the estimates
  - Not very useful for genetic improvement
- Random k-fold: you predict individuals from contemporaries (or even offspring)
- Leave-one-out: overfit (there's always a close sib with information), and are we interested in predicting well one individual?
- K-means cluster: it measures whether we can predict *far* animals

Which kind of cross-validation should we use?

For selection:

- Is my genetic evaluation leading me to maximization of genetic progress?
- We want the method that best predicts future performance
- Forward cross-validation (or retrospective analysis)
  - (Interbull tests)
  - Cut data at date *t*
  - Could we have predicted at time *t* the data that was actually observed after *t*?

Metrics

- Theory of quantitative genetics suggest using Metrics from linear regression:

- Bias: $b_0 = E(u - \hat{u})$
- Slope: $b_1 = \frac{Cov(u,\hat{u})}{Var(\hat{u})}$
- Accuracy: $r = \frac{Cov(u,\hat{u})}{\sqrt{Var(u)Var(\hat{u})}}$
- $MSE = b_0^2 + \sigma_u^2(1 + r^2/b_1^2 - 2r^2/b_1)$

- Similar, but not exactly the same, to Interbull tests
- Why are these relevant? Genetic progress !!

Metrics are statistics

- Bias: $b_0 = E(u - \hat{u})$
- Slope: $b_1 = \frac{Cov(u,\hat{u})}{Var(\hat{u})}$
- Accuracy: $r = \frac{Cov(u,\hat{u})}{\sqrt{Var(u)Var(\hat{u})}}$

- They depend on the data: they have distributions
  - is $r = 0.52$ different from $r = 0.53$?
- There are ways to assess these distributions (see paper)
  - parametric (Fisher's transform, Hotelling-Williams) and
  - sampling-based (bootstrap, jacknife)

How should we cross-evaluate?

- We can see $\hat{u}$ (EBVs), we can't see u (TBVs)
- Dairy breeders pretend they have pseudo-TBV after progeny testing (although this is sometimes not true for e.g. fertility)
- Most species and traits don't have this luxury
- So, often we have used precorrected phenotypes y^* or deregressed proofs and compare predictions with observations
- e.g. $r \approx cor(y^*, \hat{y})/h$ (Legarra et al. 2008)
- But this ignores the covariance structure in precorrected $y^*$and leads to paradoxes:
  - $r > 1$ (observed in chicken)
  - $r_{pedigree} > r_{genomic}$ (observed in dairy cattle for fertility)

Predictivity

- Precorrected data is obtained with the "whole" data set $\boldsymbol{y}^* = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}$
- For "new" records in "whole" but not in "partial"
  - $Var(\boldsymbol{y}^*_{new}) = \boldsymbol{R} + \boldsymbol{G} - \boldsymbol{X}\boldsymbol{C}^{\boldsymbol{\beta}\boldsymbol{\beta}}_{w}\boldsymbol{X}'$
- $b_{y^*_{new},p}$: slope of the regression $\boldsymbol{y}^*_{new}$ ~ $\widehat{\boldsymbol{u}}_p$: should yield 1
- Correlation $r\left(\boldsymbol{y}^*_{new}, \widehat{\boldsymbol{u}}_p\right)/h$ may be biased and not be a good estimator of accuracy
- Still, comparing models should be fine (better models give better $r\left(\boldsymbol{y}^*_{new}, \widehat{\boldsymbol{u}}_p\right)$

The proposed method $\mathcal{LR}$

- Run one evaluation with “whole” (w) data and get EBVs: $\hat{u}_w$
  - For instance, data from 1990 to 2017
- Run one evaluation with “partial” (p) data and get EBVs: $\hat{u}_p$
  - For instance, data from 1990 to 2010

The proposed method $\mathcal{LR}$

- Regression of EBVs obtained with "whole" (w) data on EBVs estimated with "partial" (p) data, $b_{w,p} = \frac{Cov(\hat{u}_w, \hat{u}_p)}{Var(\hat{u}_p)}$ should be 1
- The correlation of partial on whole $\rho_{p,w} = \frac{Cov(\hat{u}_p, \hat{u}_w)}{\sqrt{Var(\hat{u}_w)Var(\hat{u}_p)}}$ is a function of respective accuracies $E(\rho_{w,p}) = \sqrt{\frac{\mu_{acc_p^2}}{\mu_{acc_w^2}}}$

The proposed method $\mathcal{LR}$

Technical note: detection of bias in genetic predictions

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

J Anim Sci 1994. 72:34-37.

Check of bias on succesive evaluations

- You run evaluations in 2009, 2010, 2011, 2012… with more data each year
- You want to check if evaluations are consistent to each other
- There is a theory for that

Technical note: detection of bias in genetic predictions

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

J Anim Sci 1994. 72:34-37.

  - "Whole" -> evaluation with all data (say 1990-2017)
  - "Partial" -> evaluation with data up to a given cut-off date (say 1990-2013)
- Changes from "whole" to "partial" are predictable
- This is different from MethodR were records are deleted at random

Bias, dispersion and accuracy of method X evaluations using method $\mathcal{LR}$ [updated version]

- Compute EBV's $\widehat{\boldsymbol{u}}_w$ for <u>all animals</u> with "whole" data and method X
- Create "partial" data:
  - Set values after cutoff date to missing (or draw at random if 1 generation)
- Compute EBVs $\widehat{\boldsymbol{u}}_p$ for <u>all animals</u> based on "partial" data and method X
- Compute statistics for "focal" individuals of interest (e.g. candidates to selection)
  - $b_0 = (\mathbf{1}'\widehat{\boldsymbol{u}}_w - \mathbf{1}'\widehat{\boldsymbol{u}}_p)/n$
  - $b_1 = \frac{\widehat{\boldsymbol{u}}_p'\widehat{\boldsymbol{u}}_w}{\widehat{\boldsymbol{u}}_p'\widehat{\boldsymbol{u}}_p}$
  - $r = \frac{\widehat{\boldsymbol{u}}_p'\widehat{\boldsymbol{u}}_w}{\sqrt{\widehat{\boldsymbol{u}}_w'\widehat{\boldsymbol{u}}_w\widehat{\boldsymbol{u}}_p'\widehat{\boldsymbol{u}}_p}}$

An even more updated version (1)

- $\mu_{wp} = \overline{\widehat{\boldsymbol{u}}_p} - \overline{\widehat{\boldsymbol{u}}_w}$, has an expected value of 0
- $b_{w,p} = \frac{cov(\widehat{\boldsymbol{u}}_w, \widehat{\boldsymbol{u}}_p)}{var(\widehat{\boldsymbol{u}}_p)}$ has an expectation, $E(b_{w,p}) = 1$
- $\rho_{p,w} = \frac{cov(\widehat{\boldsymbol{u}}_p, \widehat{\boldsymbol{u}}_w)}{\sqrt{var(\widehat{\boldsymbol{u}}_w) var(\widehat{\boldsymbol{u}}_p)}}$ has expected value $E(\rho_{w,p}) \approx \frac{acc_p}{acc_w}$

Ratio of accuracies, not accuracy per se

Another estimator of accuracy (2)

- $\rho^2_{Cov_{w,p}} = \frac{cov(\hat{\boldsymbol{u}}_w, \hat{\boldsymbol{u}}_p)}{(1+\bar{F}-2\bar{f})\sigma^2_{u,\infty}}$ has $E\left(\rho^2_{Cov_{w,p}}\right) \approx acc_p^2$.

Average inbreeding and relationships

Variance of individuals on the test (this is less than the genetic variance)

Yet another estimator of accuracy (3)

- $b_{p,w} = \frac{cov(\hat{\boldsymbol{u}}_w, \hat{\boldsymbol{u}}_p)}{var(\hat{\boldsymbol{u}}_w)}$ has $E(b_{p,w}) = \frac{acc_p^2}{acc_w^2}$

Ratio of accuracies, not accuracies

About bias

If both “whole” and “partial” are biased by θ_p^2 and θ_w^2

- $b_{w,p} = \frac{(\hat{\boldsymbol{u}}_p - \bar{\hat{\boldsymbol{u}}}_p)'(\hat{\boldsymbol{u}}_w - \bar{\hat{\boldsymbol{u}}}_w)}{(\hat{\boldsymbol{u}}_p - \bar{\hat{\boldsymbol{u}}}_p)'(\hat{\boldsymbol{u}}_p - \bar{\hat{\boldsymbol{u}}}_p)}$ has, $E(b_{w,p}) = \frac{\theta_w}{\theta_p}$

For instance, a value of $b_{w,p} < 1$ may be due to

- overdispersion of the “partial” but also
- underdispersion of the “whole” (or low estimate of genetic trend???).

Predictability

- Precorrected data is obtained with the "whole" data set $\boldsymbol{y}^* = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}$
- For "new" records in "whole" but not in "partial"
  - $Var(\boldsymbol{y}^*_{new}) = \boldsymbol{R} + \boldsymbol{G} - \boldsymbol{X}\boldsymbol{C}^{\boldsymbol{\beta}\boldsymbol{\beta}}_{w}\boldsymbol{X}'$
- $b_{y^*_{new},p}$: slope of the regression $\boldsymbol{y}^*_{new}$ ~ $\widehat{\boldsymbol{u}}_p$: should yield 1

- But correlation $r\left(\boldsymbol{y}^*_{new}, \widehat{\boldsymbol{u}}_p\right)$ may be biased