INRA SCIENCE & IMPACT
SelGen

La Région Occitanie Pyrénées - Méditerranée

ardi
Research, Development and Innovation in sheep

Interreg POCTEFA
UNIÓN EUROPEA
UNION EUROPÉENNE

INRA
SCIENCE & IMPACT

Behavior of method LR to benchmark genetic evaluations

Fernando Macedo & Andrés Legarra
GenPhySE, INRA.
fernando.macedo@inra.fr
EAAP, August 2018, Dubrovnik, Croatia.

Summary

- Why another method to benchmark genetic evaluations?
- Method LR
- Simulations
- First results
- Conclusions

Why another method to benchmark genetic evaluations?

- In genomic evaluations cross validation is the most used tool for benchmarking

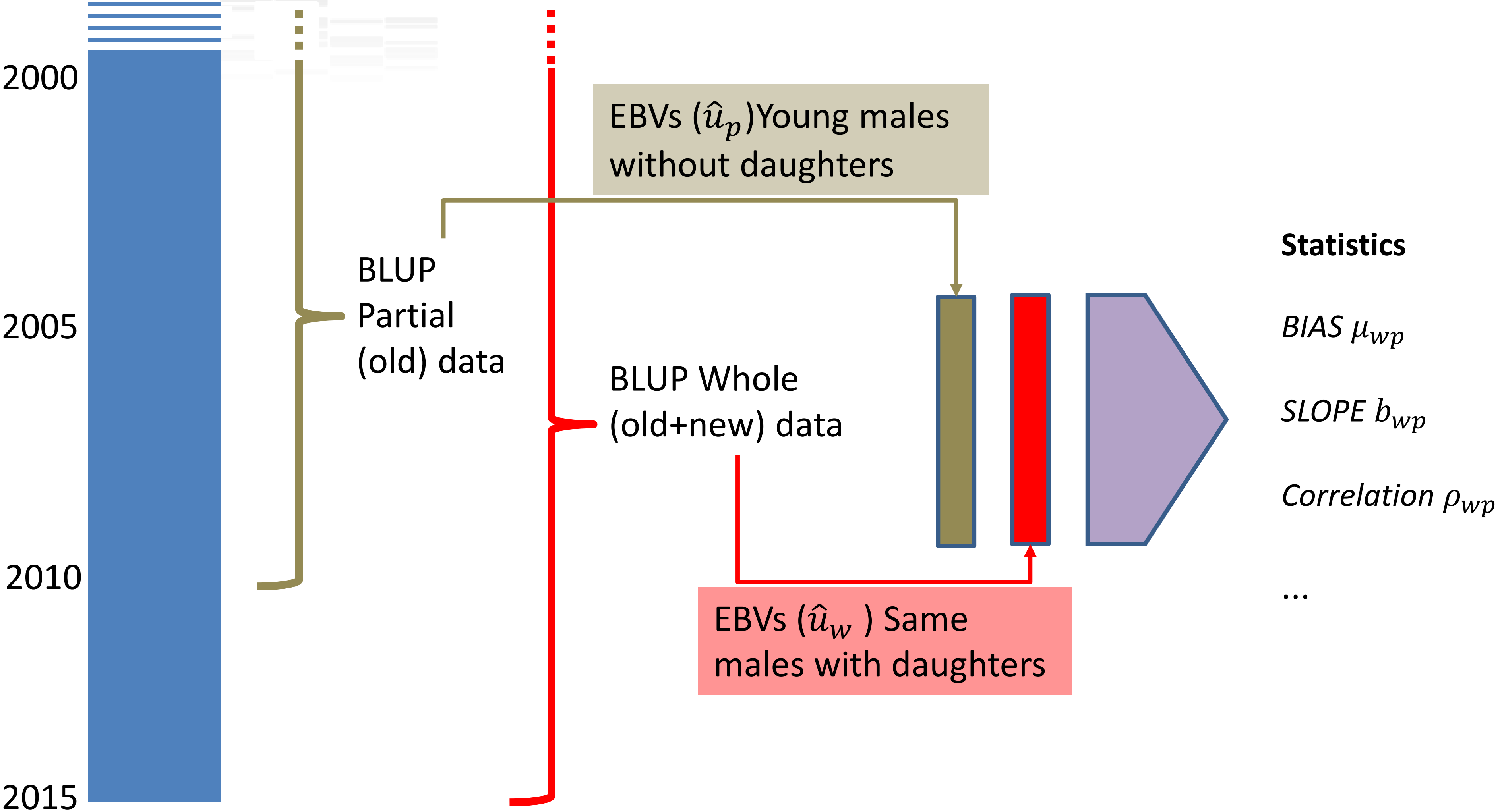
- All golden standard have problems:

    - Pre-corrected phenotypes may be not well corrected
    - Daughter Yield Deviations are not always available or might be inaccurate
    - Some traits (like maternal effects) don't have direct observation related to animals

- Need simple general tools for varied situations in animal breeding systems

- Legarra & Reverter (2017) proposed a new method based on comparisons of EBV from partial (old) data vs whole (old+new) data.
    - Does not require "true" breeding values
    - Does not require pre-corrected phenotypes

Method LR



BLUP Partial (old) data

EBVs ($\hat{u}_p$)Young males without daughters

BLUP Whole (old+new) data

EBVs ($\hat{u}_w$) Same males with daughters

Statistics

BIAS μ_{wp}

SLOPE b_{wp}

Correlation ρ_{wp}

…

Method LR: Estimators.

Bias

$\mu_{wp} = \overline{\widehat{u_p}} - \overline{\widehat{u_w}}$. Expected value of 0 in absence of bias.

Slope of the regression EBVw on EBVp

$b_{w,p} = \frac{cov(\widehat{u_p},\widehat{u_w})}{var(\widehat{u_p})}$. With a value of 1 in unbiased procedure.

Correlation between EBVp and EBVw.

Direct estimator of **relative increase of accuracy** from partial to whole.

$\rho_{p,w} = \frac{cov(\widehat{u_w},\widehat{u_p})}{\sqrt{var(\widehat{u_w})var(\widehat{u_p})}}$. The expected value is $E(\rho_{p,w}) \approx \frac{acc_p}{acc_w}$.

Objective

Testing the estimators of bias, slope and accuracy using simulated selection schemes in several scenarios:

1. The genetic evaluation model is the correct one.
2. The genetic evaluation model is wrong.

Simulation details

Dairy sheep like scheme.
Simulation was performed with QMSim software (version 1.10) (Sargolzaei & Schenkel, 2009).

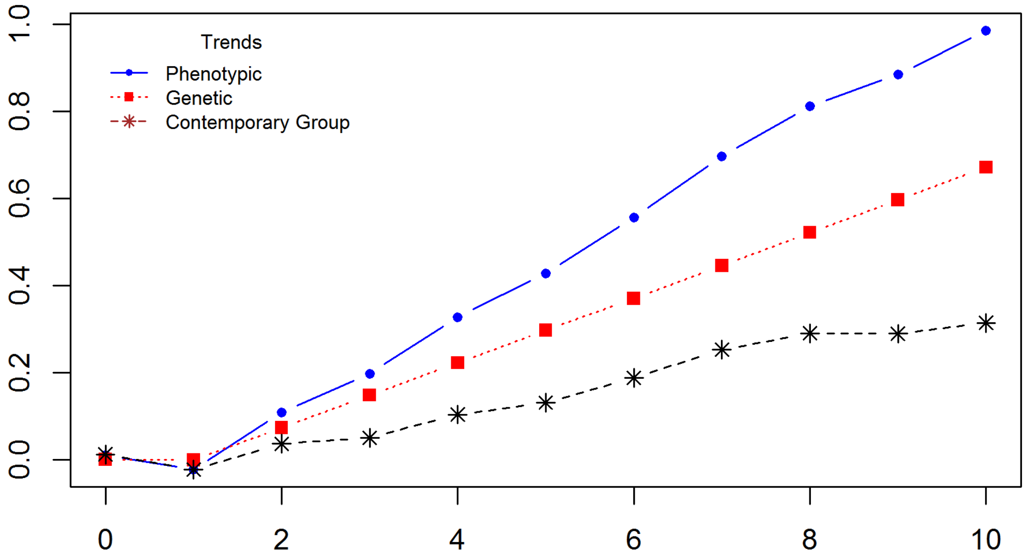
Parameters used:

- h^2 simulated: 0.05, 0.10, 0.25, 0.50
- 20 replicates for each h^2
- Records only in females
- 10 generations
- Total animals in each replicate around 500,000
- Selection by higher EBV's

The genetic evaluation model is wrong

2 Strategies:

1. Contemporary groups with phenotypic trend
    - Around 90 CG/Generation with about 500 animals each.
    - Simulated: True effect of CG as random with time trend
    - Estimated: in BLUP as fixed effect



2. Using different h^2 in blup evaluations to those used for simulation (results not shown)
    - e.g. simulated h^2=0.10 and evaluation h^2=0.05

Analyzed Datasets

Within each replicate:
e.g.
At the end of generation 5: estimate EBV of young males (without progeny) $\hat{u}_p$
At the end of generation 6: estimate EBV of the same males (with progeny) $\hat{u}_w$
Compute statistics:

Bias

$$\mu_{wp} = \overline{\widehat{u_p}} - \overline{\widehat{u_w}}$$

Slope

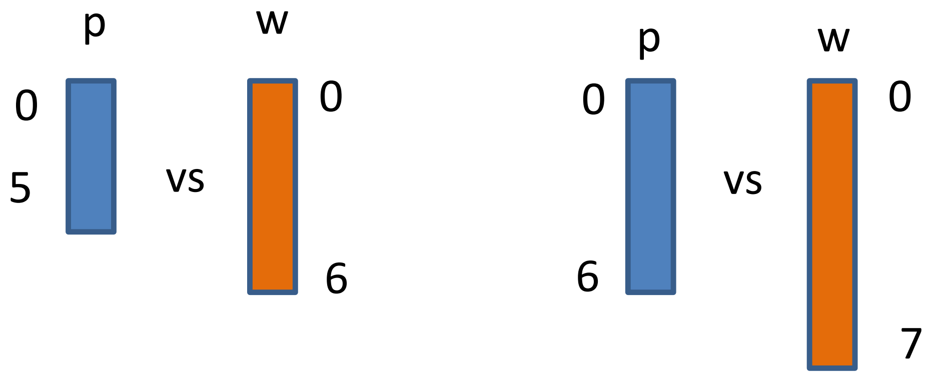
$$b_{w,p} = \frac{cov(\widehat{u_p},\widehat{u_w})}{var(\widehat{u_p})}$$

Correlation

$$\rho_{p,w} = \frac{cov(\widehat{u_w},\widehat{u_p})}{\sqrt{var(\widehat{u_w})var(\widehat{u_p})}}$$

In this work we estimate the statistics for generations 5 to 9:

5 vs 6
6 vs 7
7 vs 8
8 vs 9
9 vs 10



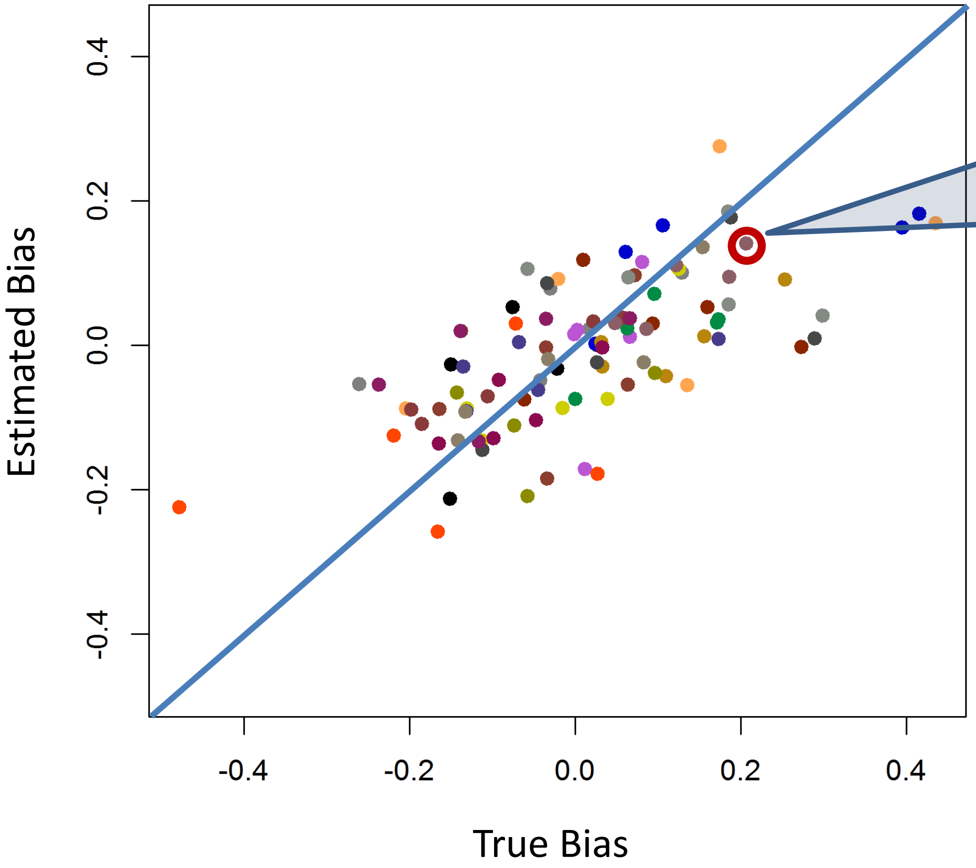
First Results

Using the correct evaluation model

BIAS

$$Estimated \Rightarrow \mu_{w,p} = \widehat{\overline{u_p}} - \widehat{\overline{u_w}}$$

$$True \Rightarrow \mu_{u,p} = \widehat{\overline{u_p}} - \bar{u}$$



True Bias = 0.19
Estimated Bias= 0.18

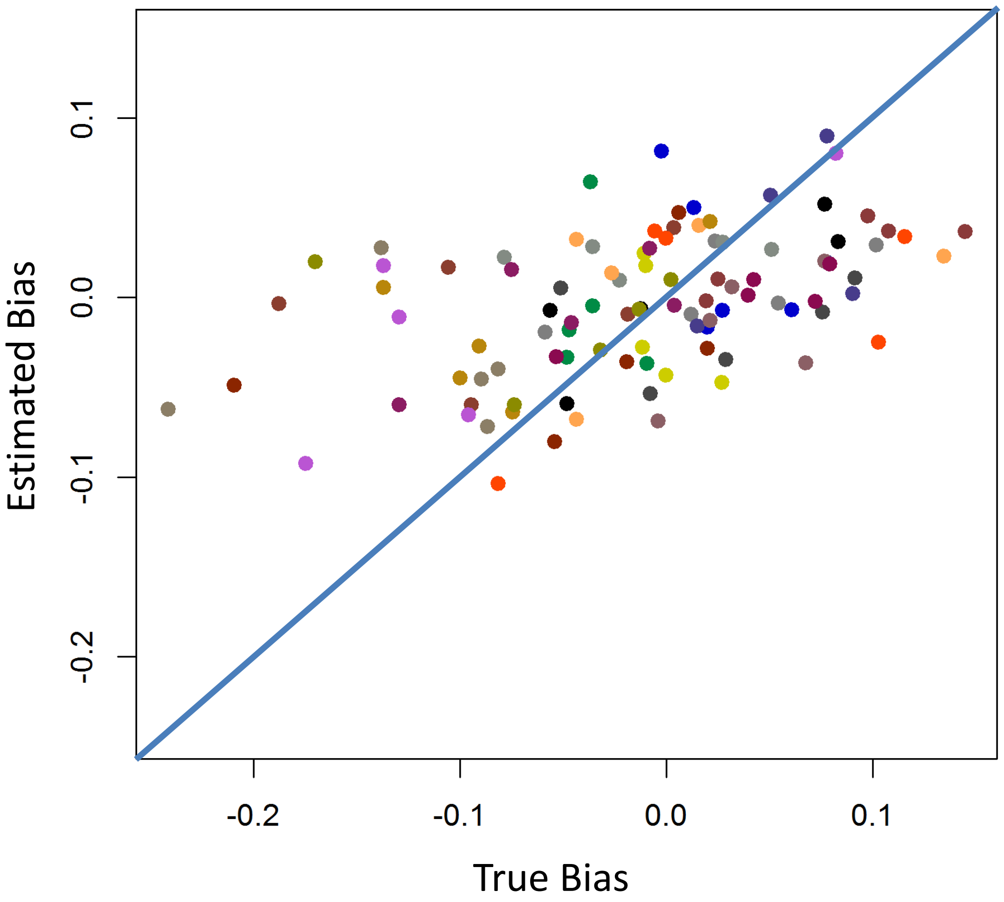
replicate=14
G9 (partial) vs G10 (whole)

h^2=0.5

BIAS

$Estimated\ Bias \Rightarrow\ \mu_{w,p} = \overline{\widehat{u_p}} - \overline{\widehat{u_w}}$

$True\ Bias \Rightarrow\ \mu_{u,p} = \overline{\widehat{u_p}} - \bar{u}$



h^2=0.05

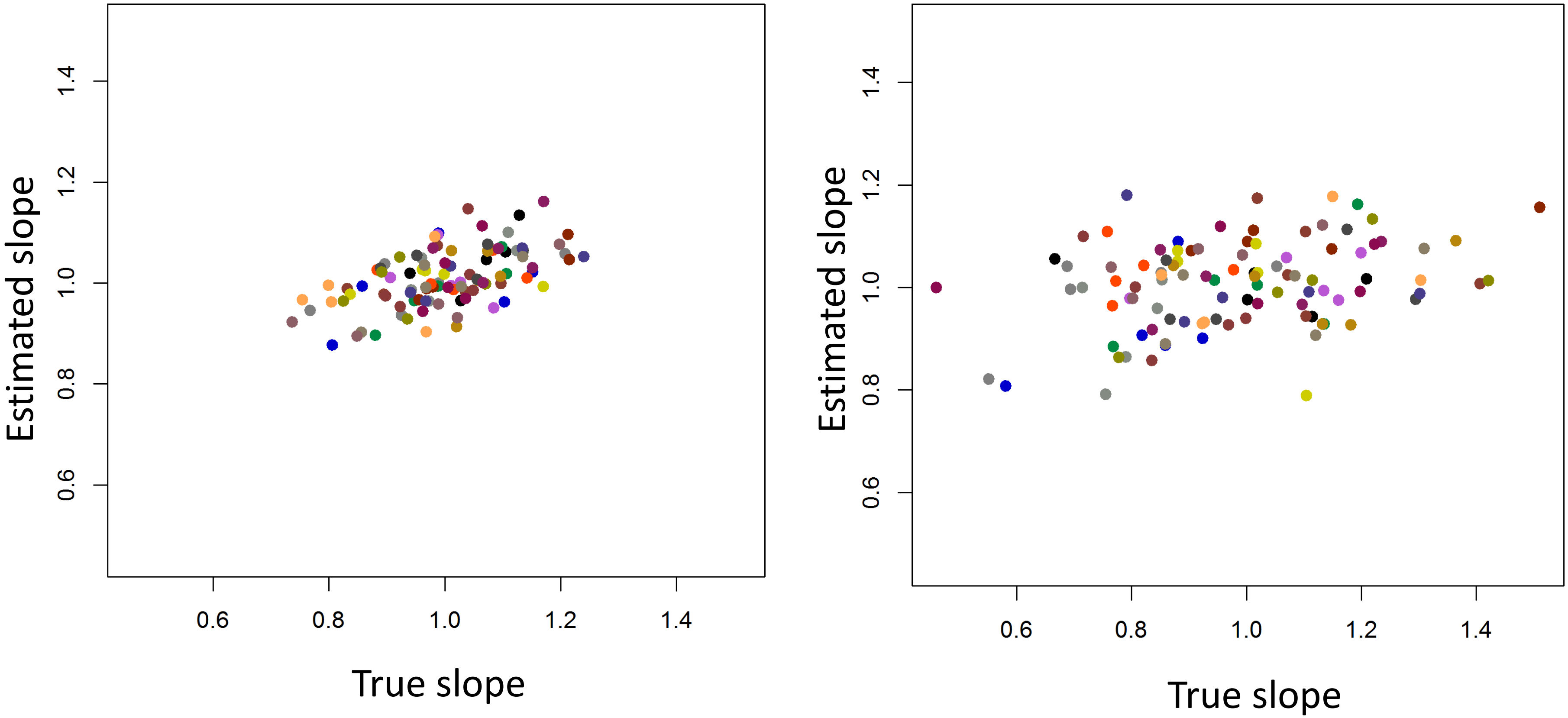
SLOPE $b_{w,p}$

Estimated $\Rightarrow b_{w,p} = \frac{cov(\widehat{u_p}, \widehat{u_w})}{var(\widehat{u_p})}$

True $\Rightarrow b_{u,p} = \frac{cov(\widehat{u_p}, u)}{var(\widehat{u_p})}$

h^2=0.50

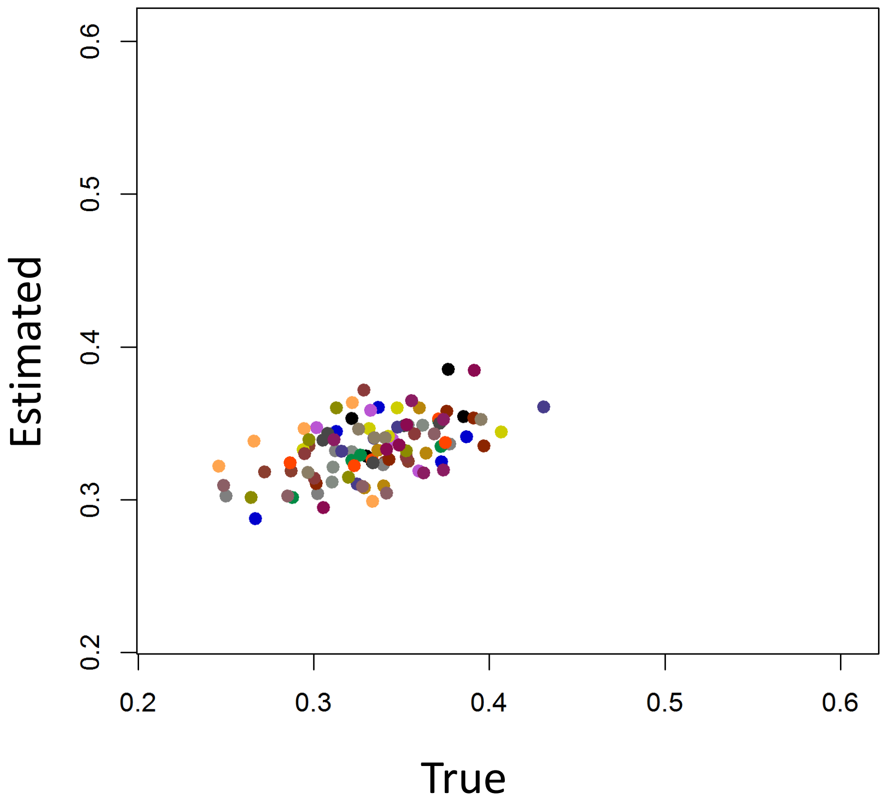
h^2=0.05



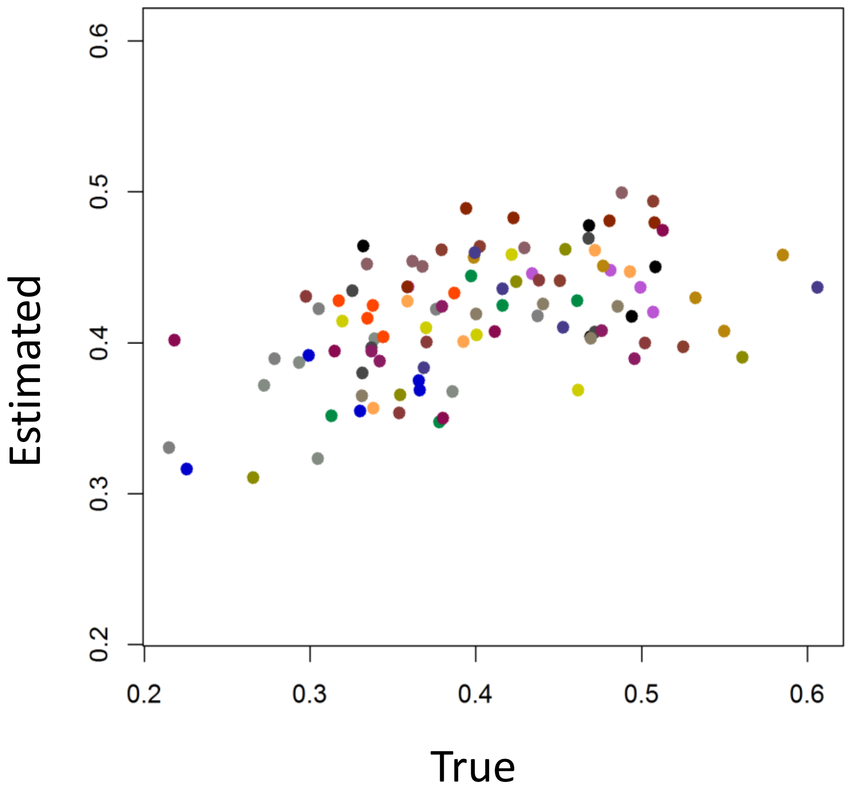
Relative accuracy gain

Estimated $\Rightarrow \rho_{p,w} = \frac{cov(\widehat{u_w},\widehat{u_p})}{\sqrt{var(\widehat{u_w})var(\widehat{u_p})}}$ True $\Rightarrow \frac{acc_p}{acc_w}$

h^2=0.50

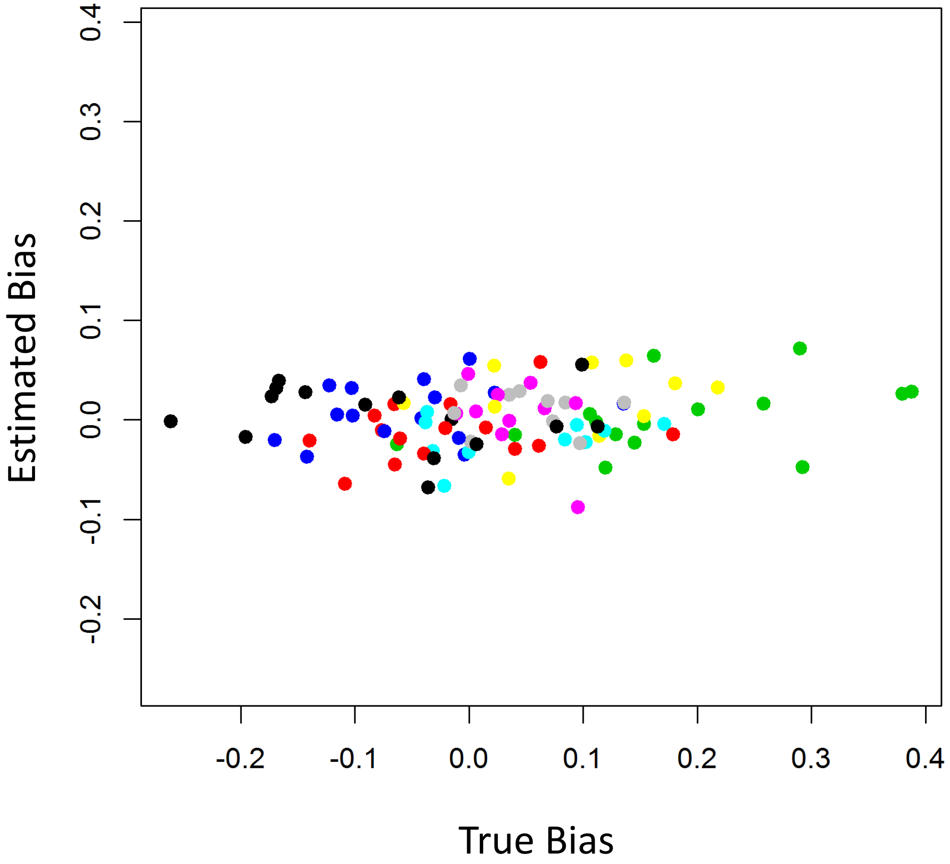


h^2=0.05



Using the wrong evaluation model

Wrong evaluation fitting CG as fixed when they have a time trend



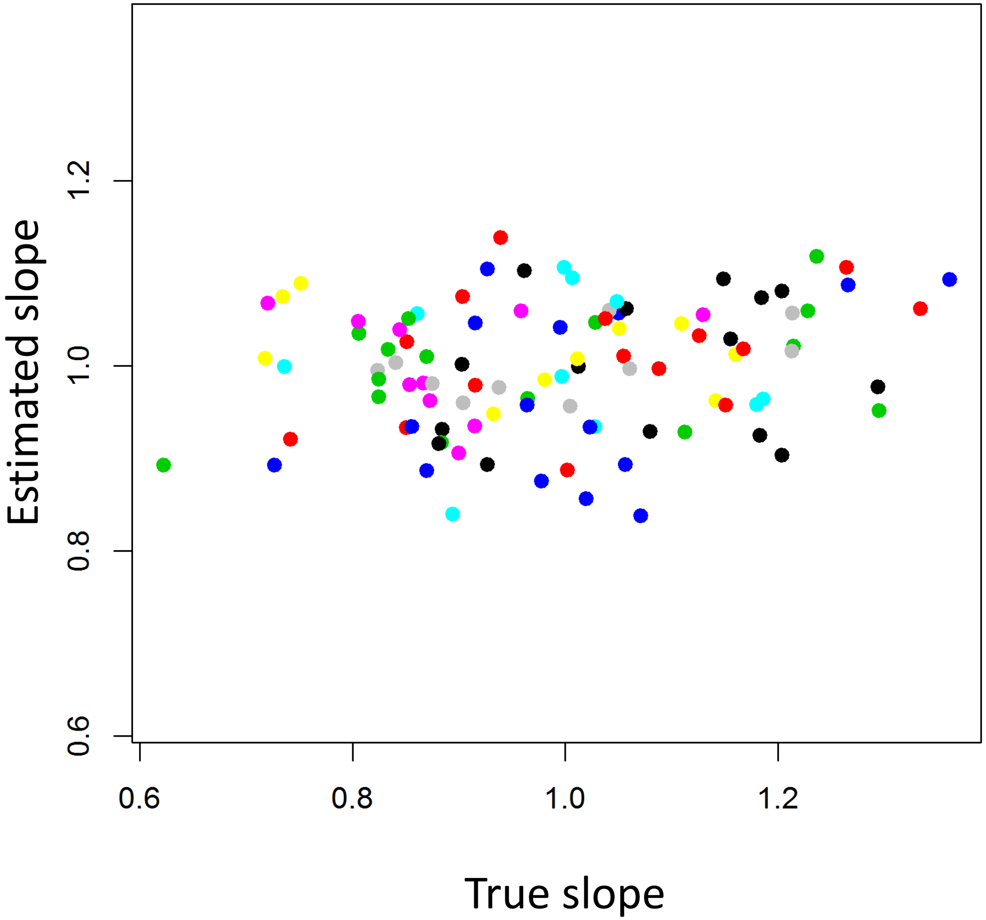
BIAS

$$Estimated \Rightarrow \mu_{w,p} = \overline{\widehat{u_p}} - \overline{\widehat{u_w}}$$

$$True \Rightarrow \mu_{u,p} = \overline{\widehat{u_p}} - \bar{u}$$

h^2=0.10

Wrong evaluation fitting CG as fixed when they have a time trend



SLOPE $b_{w,p}$

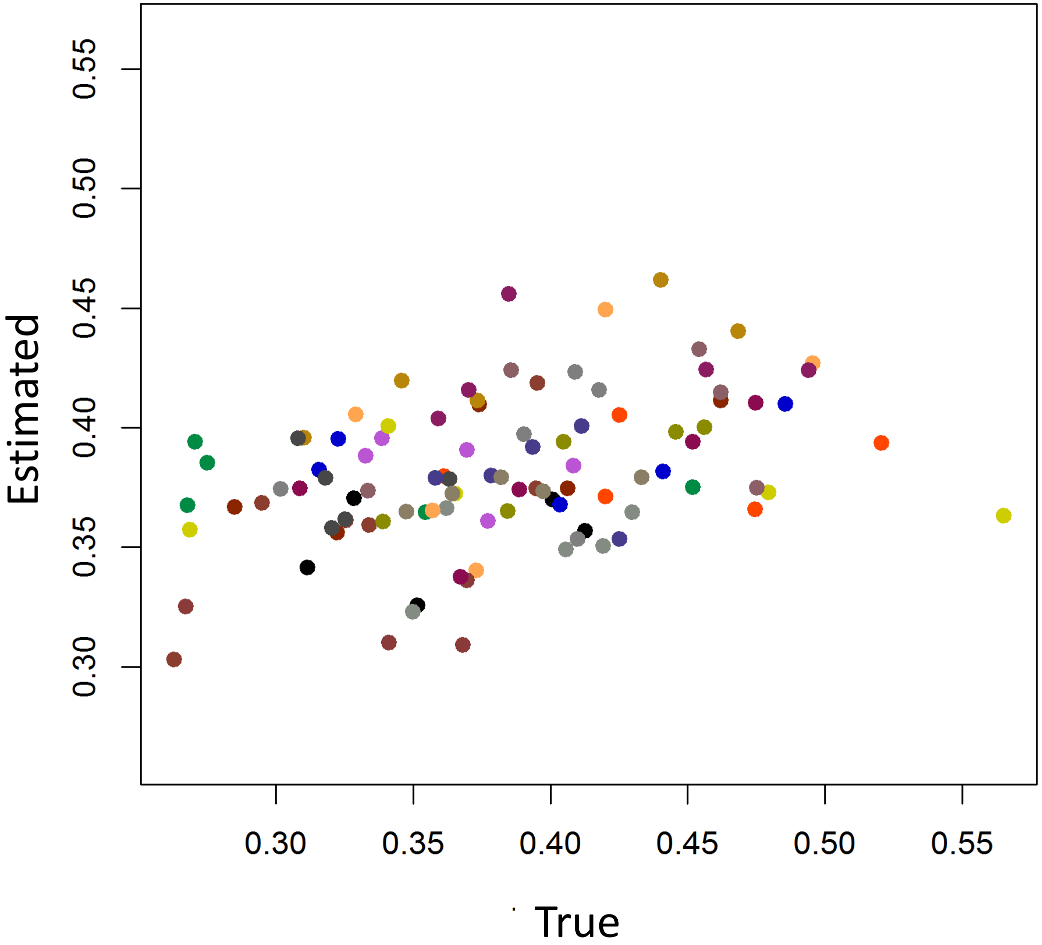
$$\text{Estimated} \Rightarrow b_{w,p} = \frac{cov(\widehat{u_p}, \widehat{u_w})}{var(\widehat{u_p})}$$

$$\text{True} \Rightarrow b_{u,p} = \frac{cov(\widehat{u_p}, u)}{var(\widehat{u_p})}$$

h^2=0.10

Wrong evaluation fitting CG as fixed when they have a time trend

Relative accuracy gain



Estimated $\Rightarrow \rho_{p,w} = \dfrac{cov(\widehat{u_w}, \widehat{u_p})}{\sqrt{var(\widehat{u_w})var(\widehat{u_p})}}$

True $\Rightarrow \dfrac{acc_p}{acc_w}$

h^2=0.10

Conclusions

- The proposed method LR estimates well bias, slope and accuracy when the model is in concordance with the reality.

But when the model has differences with the reality:

- With wrong model for contemporary groups
    - It is not possible to estimate bias or slope.
    - Accuracies can be estimated but not well

- With wrong heritabilities:
    - The bias could be under or over estimated
    - The slope is uninformative about the reality.
    - Accuracies can be estimated

Acknowledgements

Authors thank the following institutions that finance this research:

- **Poctefa Project ARDI**
- **La region Occitanie**
- **INRA - Metaprogram SelGen**

As well as to **Computing platform Bioinfo-Genotoul** for providing bioinformatics support.

Thank you for your attention!

Questions or comments?

INRA SCIENCE & IMPACT
SelGen

La Région Occitanie Pyrénées - Méditerranée

ardi
Research, Development and Innovation in sheep

Interreg POCTEFA
UNIÓN EUROPEA
UNION EUROPÉENNE

INRA SCIENCE & IMPACT