Genomic evaluation validation test proposed by Interbull is necessary but not sufficient because it does not check the correct genetic trend

JM. Astruc * , G.Baloche, F.Barillet, A.Legarra

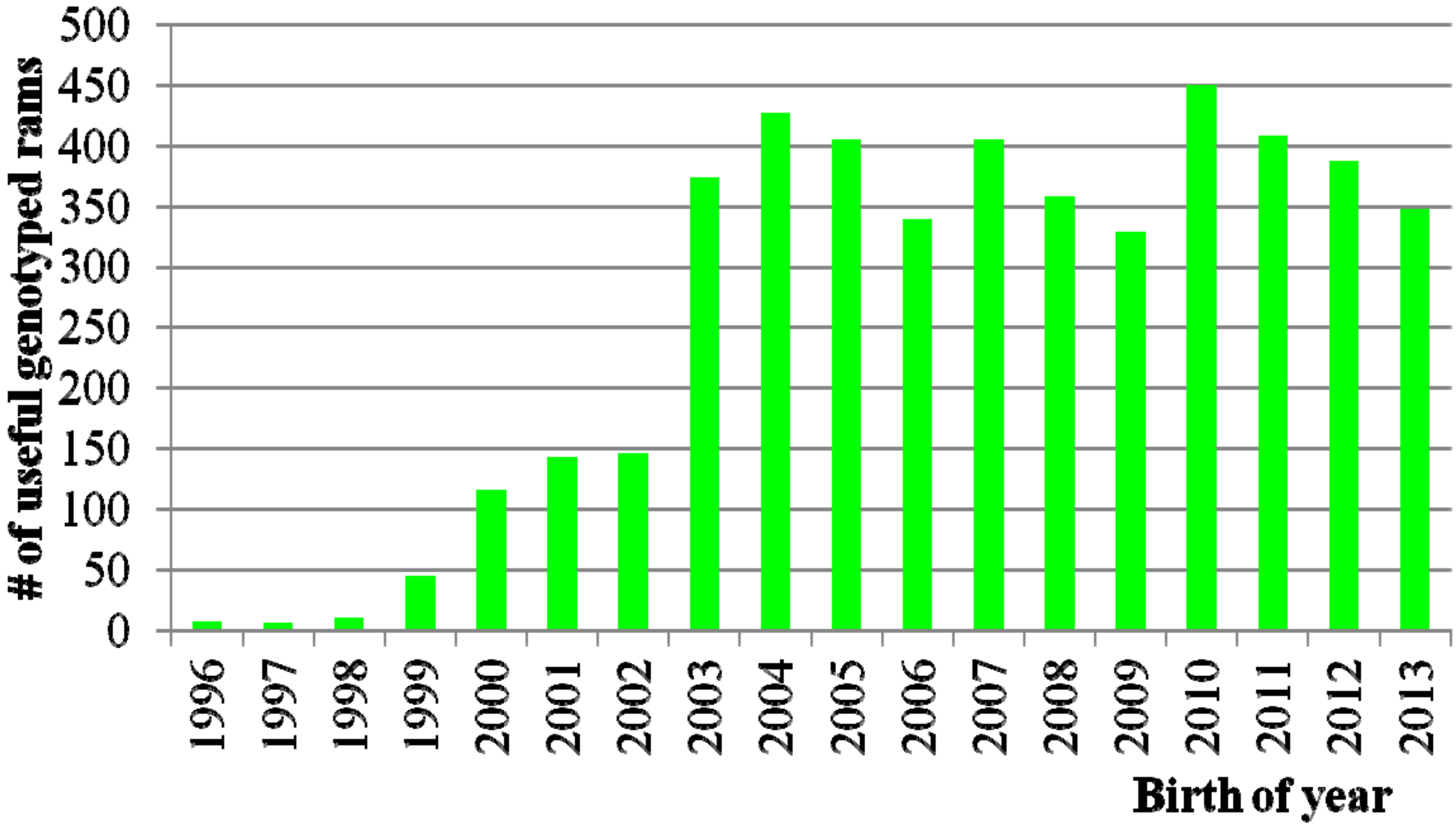
* *French Livestock Institute - Toulouse, France*

Session S2 Dairy Sheep and Goats,

Berlin, Germany, 23 May 2014

Towards genomic selection in French dairy sheep (cf. Barillet et al., ICAR, 2014)

Reference population in Lacaune (rams phenotyped et genotyped)



2015
GS

Towards genomic selection in French dairy sheep (cf. Barillet et al., ICAR, 2014)

Genomic evaluation

Method of computation : **Single Step GBLUP**
(from Misztal package BLUPF90)

From ewes performances

Entire population included in evaluation

Output **GEBV**

Objectives of the study

Compare ability of SS-GBLUP (GEBV) and pedigree-based BLUP (PA) to predict genetic merit

Propose an experience feedback on our approach and our questioning

In this study : computation tested on 7 traits :

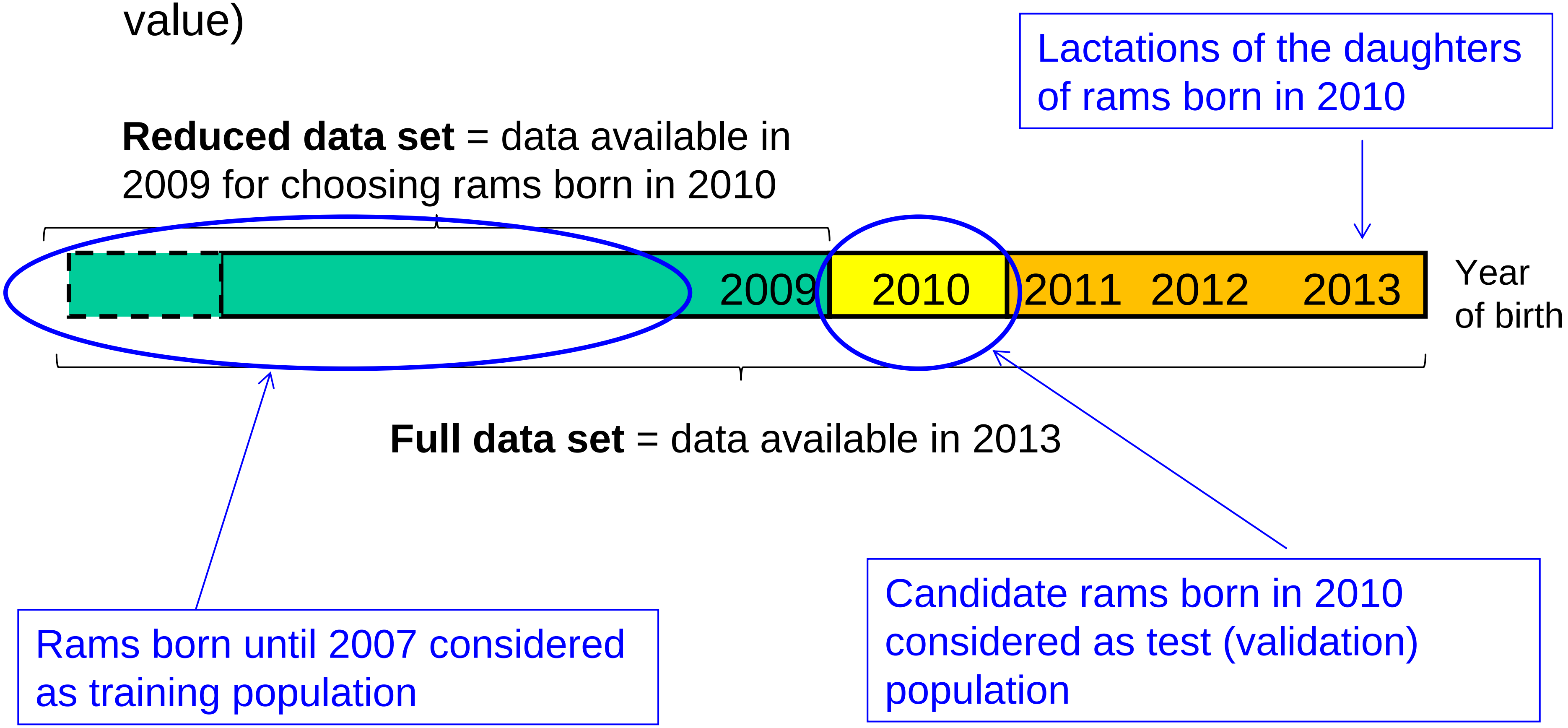
- → milk yield (MY),
- → fat content (FC),
- → protein content (PC)
- → somatic cell count
- → 3 udder morphology traits

} Results on these traits presented here

Method

Use Interbull validation test for genomic evaluation
(Mäntysaari et al., 2010)

→ Compare both PA and GEBV of one cohort of males at the moment of their choice with DYD (considered as true breeding value)



Lactations of the daughters of rams born in 2010

Reduced data set = data available in 2009 for choosing rams born in 2010

Full data set = data available in 2013

Rams born until 2007 considered as training population

Candidate rams born in 2010 considered as test (validation) population

Method

Validation on candidate rams

Regression of $\mathrm{DYD}_{\mathrm{full2013}}$ to $\mathrm{GEBV}_{\mathrm{reduced2009}}$ or $\mathrm{PA}_{\mathrm{reduced2009}}$

$$\mathrm{DYD}_{\mathrm{full2013}} = b_0 + b_1 * \mathrm{GEBV}_{\mathrm{reduced2009}}$$

$$\mathrm{DYD}_{\mathrm{full2013}} = b_0 + b_1 * \mathrm{PA}_{\mathrm{reduced2009}}$$

$$R^2_{\mathrm{Interbull}} = R^2\,(1+k/\mathrm{EDC}) \qquad k=(4-h^2)/h^2$$

Expected prediction bias (b_1) ~ 1

Comparison of $R^2_{\mathrm{Interbull}}$ and b_1 between GEBV and PA

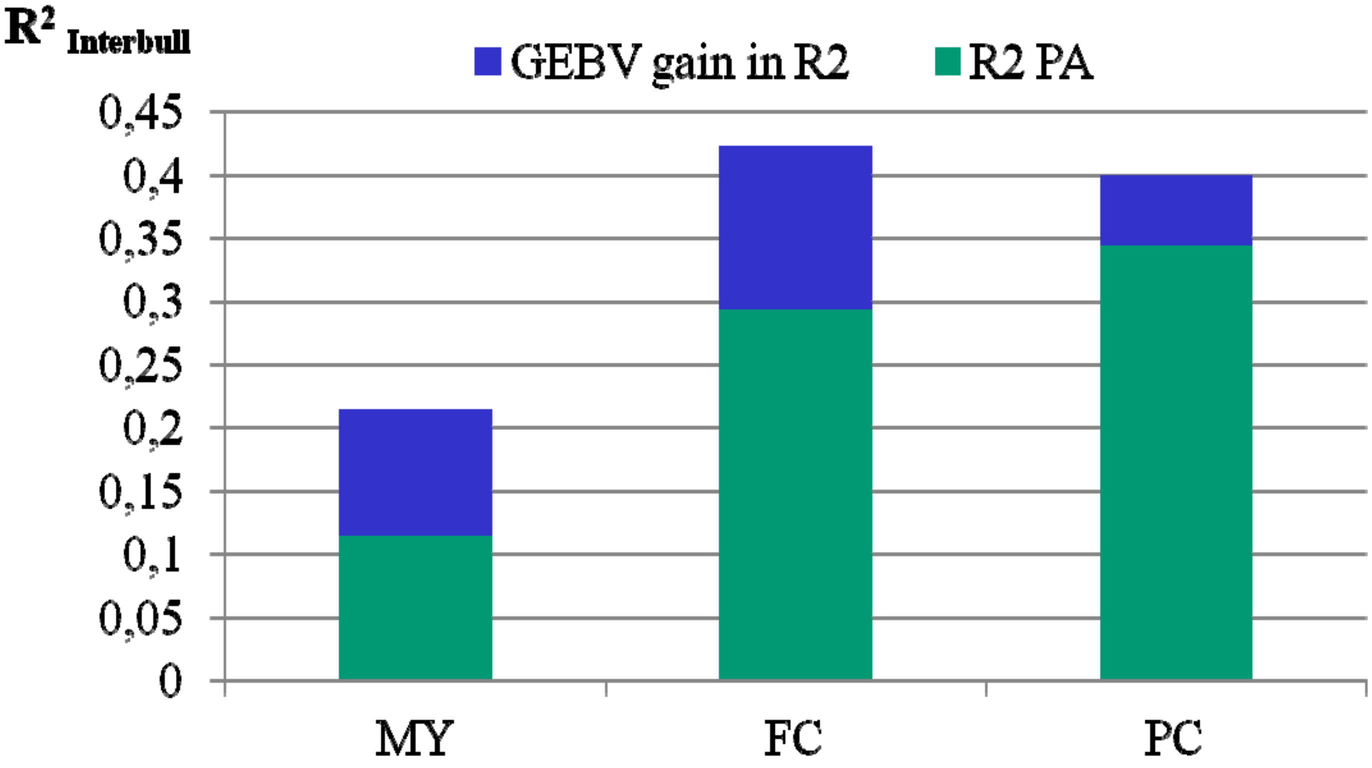
Available data : genotypes and phenotypes

3,550 genotyped rams with 39,483 SNPs (from the 54k Illumina beadchip)

Phenotypes extracted from official French dairy sheep evaluation

- Full data from 2013 evaluation (4,733,295 lactations)
- Reduced data from 2009 evaluation (4,125,703 lactations)
- DYD from regular evaluation weighted by EDC (equivalent daughters contribution)
- Ewes performances
- 450 candidate rams born in 2010 (validation population)
- 2,415 genotyped learning rams (born 1998-2007)

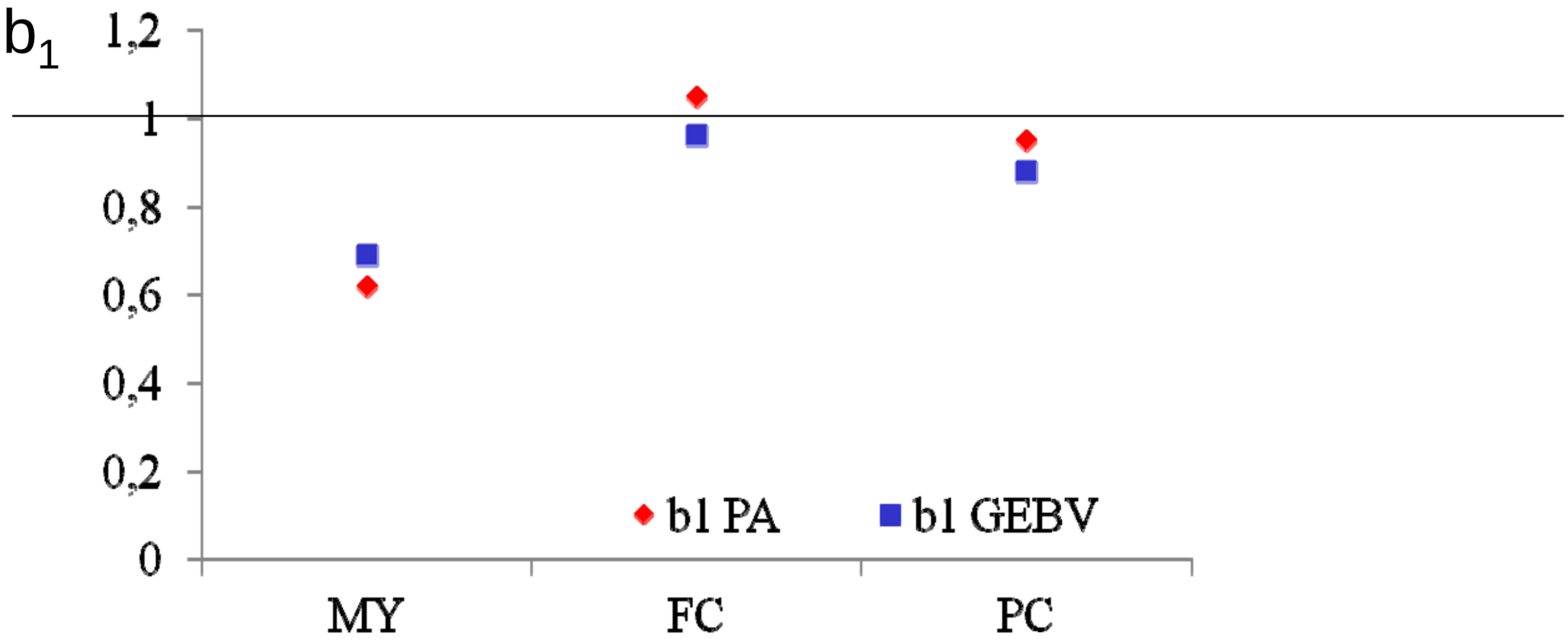
GEBV reliability *versus* PA reliability



GEBV outperformed PA for all traits

Astruc et al - 39th ICAR Session, Berlin, Germany, 2014

GEBV b_1 *versus* PA b_1



Slope b_1 near to 1 for fat and protein content both for PA and genomic prediction : OK.

Inflation of predictions for milk yield (b_1<1) in both estimations.
Slightly better with GEBV.

Glance on genetic trend

Rams : genetic trend from 1990 to 2007

Trait	Expected genetic trend *	Genetic trend observed with GEBV
MY	5.26 liters	2.71 liters
FC	0.20 g/l	0.16 g/l
PC	0.17 g/l	0.12 g/l

* From official conventional genetic evaluation

- Strong underestimation of genetic trend for milk yield
- Significant underestimation of genetic trend for fat and protein content

What possible explanation ?

What difference of model between conventional and genomic evaluation ?

One possibility :
in GBLUP, unknown parent group not taken into account

Taking into account Unknown Parent Groups (UPG)

Favorable situation of missing pedigree in French Lacaune

All rams' sires and dams are known (100%);
More than 99% of the ewes' dams are known;
Roughly 95% of the ewes' sires are known.

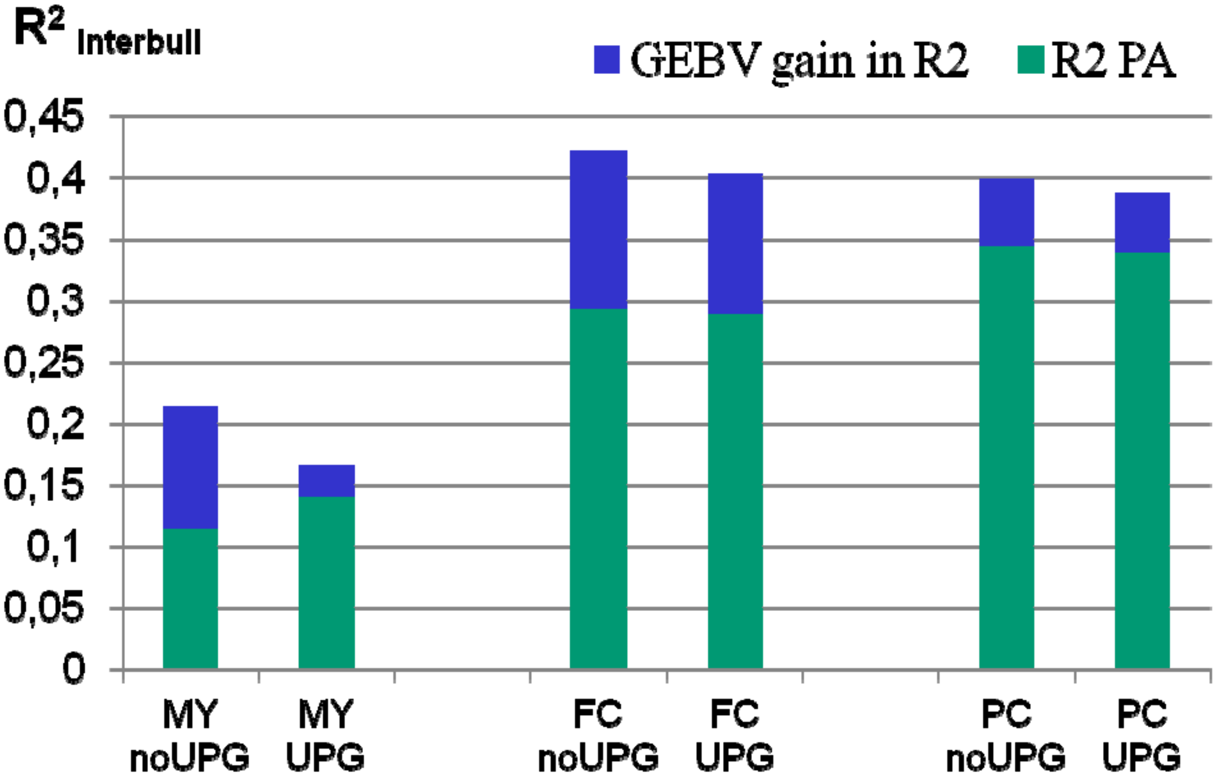
→ UPG were not taken into account in a first time.

Add unknown parent groups in the SSGBLUP model
(by explicitly adding covariates in the model with the contribution of each group – Misztal et al, 2013)

22 groups depending on the birth year of the animal (whatever male or female)

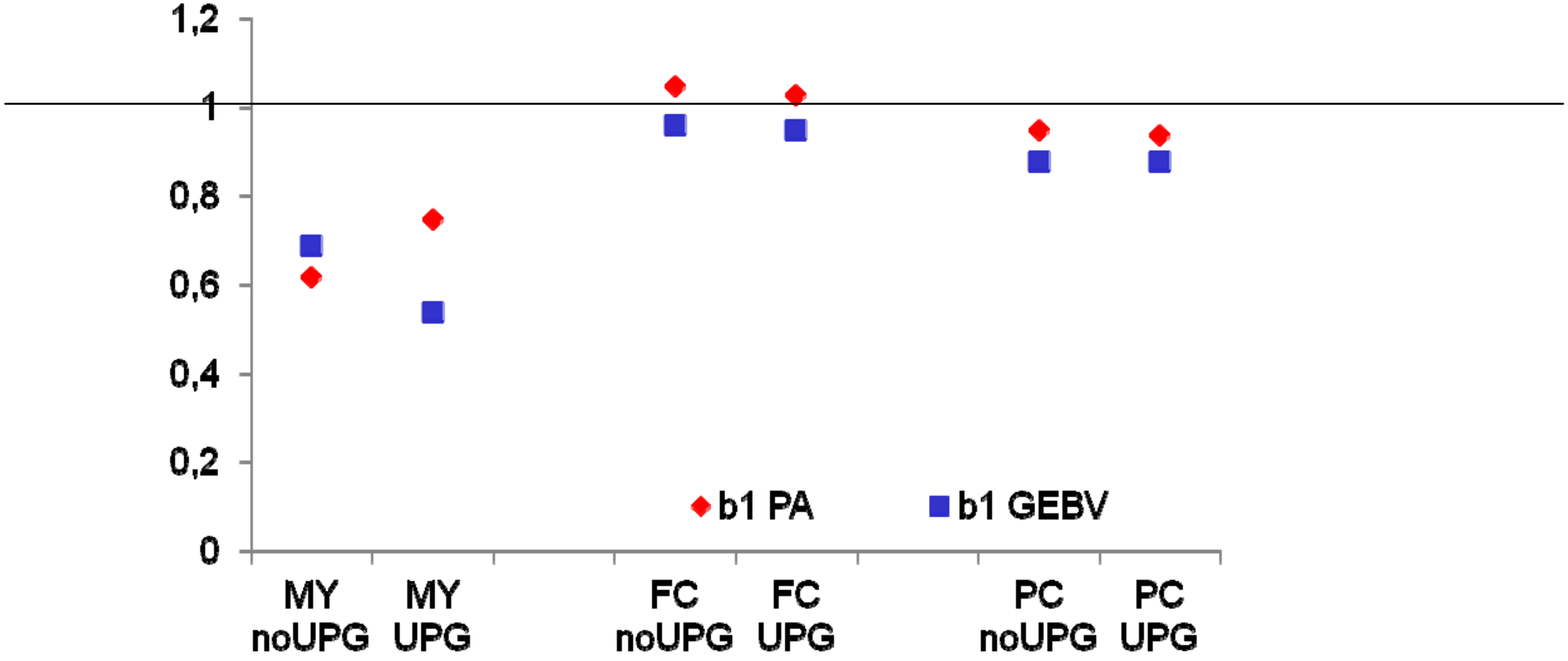
Same group for missing dam or missing sire.

What does UPG change regarding GEBV reliability *versus* PA reliability ?



Same improvement of GEBV over parental average for all traits. Although slightly different (lower) for milk yield

What does UPG change regarding GEBV b_1 *versus* PA b_1 ?



More or less the same pattern with or without UPG.
Slightly different for milk yield : with UPG, inflation of prediction greater in genomic situation

What does UPG change regarding genetic trend ?

Trait	**Expected genetic trend ***	**Genetic trend observed with GEBV without UPG**	**Genetic trend observed with GEBV with UPG**
MY	5.26 liters	2.71 liters	5.24 liters
FC	0.20 g/l	0.16 g/l	0.20 g/l
PC	0.17 g/l	0.12 g/l	0.16 g/l

* From official conventional genetic evaluation

- BUT : no more underestimation of genetic trend neither for milk yield nor for fat and protein content

Astruc et al - 39th ICAR Session, Berlin, Germany, 2014

Conclusion

Essential to check genetic trend when comparing genomic evaluation and conventional evaluation, in addition to the test on the bias and the reliability of both genomic and regular evaluation

Unknown parent groups to be considered in the model : play a key role, even when the rate of missing pedigree is low (as observed in French dairy Lacaune)

Remaining question, whatever genomic or conventional evaluation : modeling of milk yield must be improved.

THANK YOU FOR YOUR ATTENTION

Acknowledgements

Financial support provided by ANR, ApisGene, FUI, Midi-Pyrénées, Aveyron, Tarn, Rodez

Thanks to breeding organizations from Lacaune :
Upra Lacaune, Confédération Générale Roquefort, Ovitest

Thanks to I. Misztal and UGA team for utilization of GBLUP package