GENOMIC SELECTION AND GENOTYPE BY ENVIRONMENT INTERACTION IN A WHEAT BREEDING PROGRAM INCLUDING ENVIRONMENTAL INFORMATION

THE UNIVERSITY of WISCONSIN MADISON

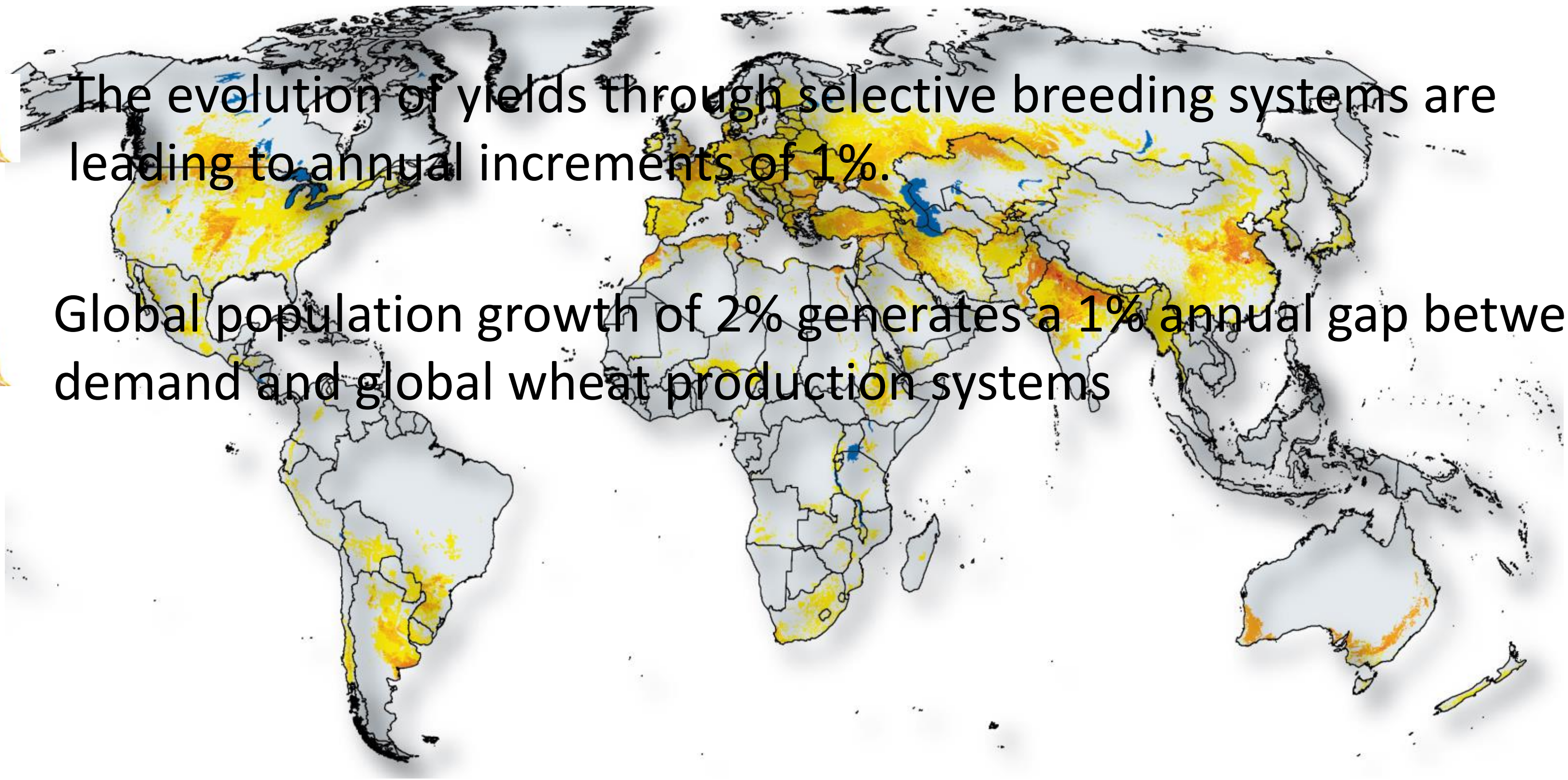
Pablo González Barrios (pgonzalez6@wisc.edu)
Dr. Gutiérrez Lab
Agronomy Department – University of Wisconsin - Madison

TEXAS A&M PLANT BREEDING SYMPOSIUM 2017

- Wheat is the most widely grown crop in the world and provides 20% of the daily protein and of the food calories for 4.5 billion people

- The evolution of yields through selective breeding systems are leading to annual increments of 1%.

- Global population growth of 2% generates a 1% annual gap between demand and global wheat production systems

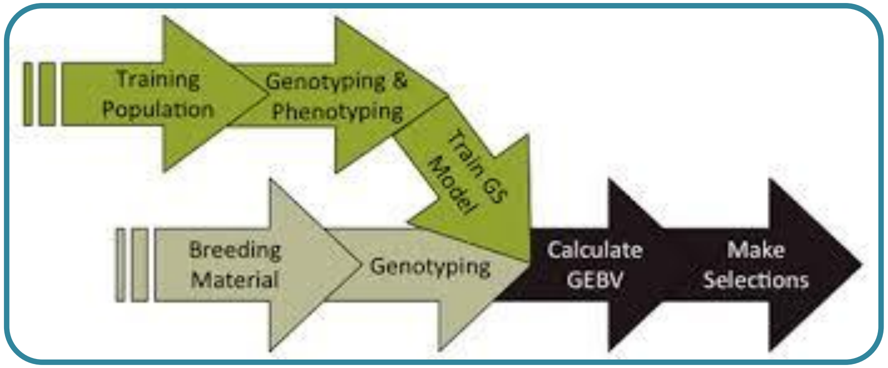


Area harvested Wheat

Production Wheat

Source: FAOSTAT (Feb 12, 2017)

Genomic Selection



- Genomic selection emerged in the last years as the most efficient and economical tool in comparison to other plant breeding methods to achieve this objectives
- Strategy for selecting individuals by predicting estimated breeding values (GEBVs)
- Using phenotypic and genotypic data from a training population to fit a statistical model.
- It allows to calculate a "phenotype" value through the only genotypic information and the trained statistical model

GxE interaction

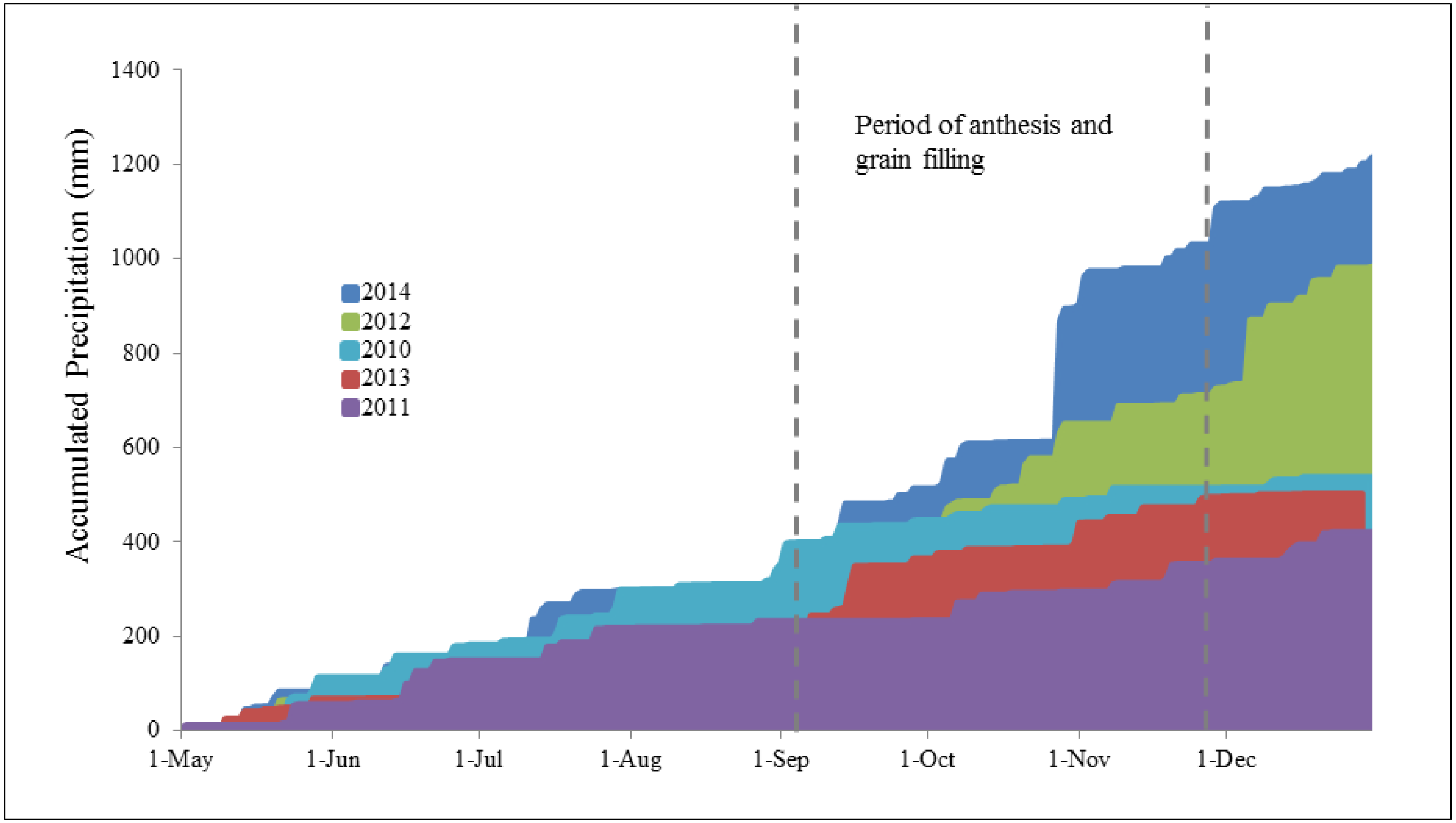
- Genotype by environment interaction (GEI) is the change in the relative performance of a character measured in two or more genotypes which are measured in two or more environments *(Bowman. 1972).*

- In plant breeding the most important GEI occurs when a change of ranking of the genotypes in different environment (crossover).

- GS studies have included GEI information by performing overall predictions across environments *(Heffner et al.. 2011; Resende et al.. 2011)*. within environments *(Burgueño et al.. 2012; Dawson et al.. 2013; Heslot et al.. 2014)*. or groups of environments or using marker-by-environment predictions (*Jarquín et al.. 2014; Lopez-Cruz et al.. 2015*).

- It is unclear the best alternative to incorporate environmental information in GS models that exploit GEI.

MEGA-ENVIRONTMENTS



Lado *et al*. (2016). Crop Sci. 56:2165-2179.

OBJECTIVES

General

- To evaluate different strategies to model GEI by incorporating selected environmental covariates into prediction models.

Specifics

- To quantify and analyze GEI patterns

- To model genotype by environment interaction with different levels of information from genotypes and environments

- To evaluate different strategies to predict new environments including environmental covariates information

GBLUP Genomic Predictions

$$y = X\beta + Zu + \varepsilon$$

GBLUP$_{(gxe)}$ + EC:

$y_{(nx1)}$: vector of mean yield in each environment (n = number of genotypes (N) by environment (k): Nxk).

$X_{(nx1)}$: is the associated design matrix of length n

β: vector of fixed terms (Environmental Covariates)

$u_{(nx1)}$: genotype by environment predictors.

$u \sim N(0.\ \sigma^2_G\ G_{(NxN)} \quad \rho_{(kxk)})$.

G realized additive relationship matrix

$\rho_{(kxk)}$ correlation matrix among environments

$Z_{(nxn)}$ incidence matrix

ε residual errors vector. $\varepsilon \sim N(0.\ \sigma^2\ R_{(NxN)})$.

R heterogeneity in mean estimate precision.

R heterogeneity in mean estimate precision.

. of

GBLUP$_{(M)}$

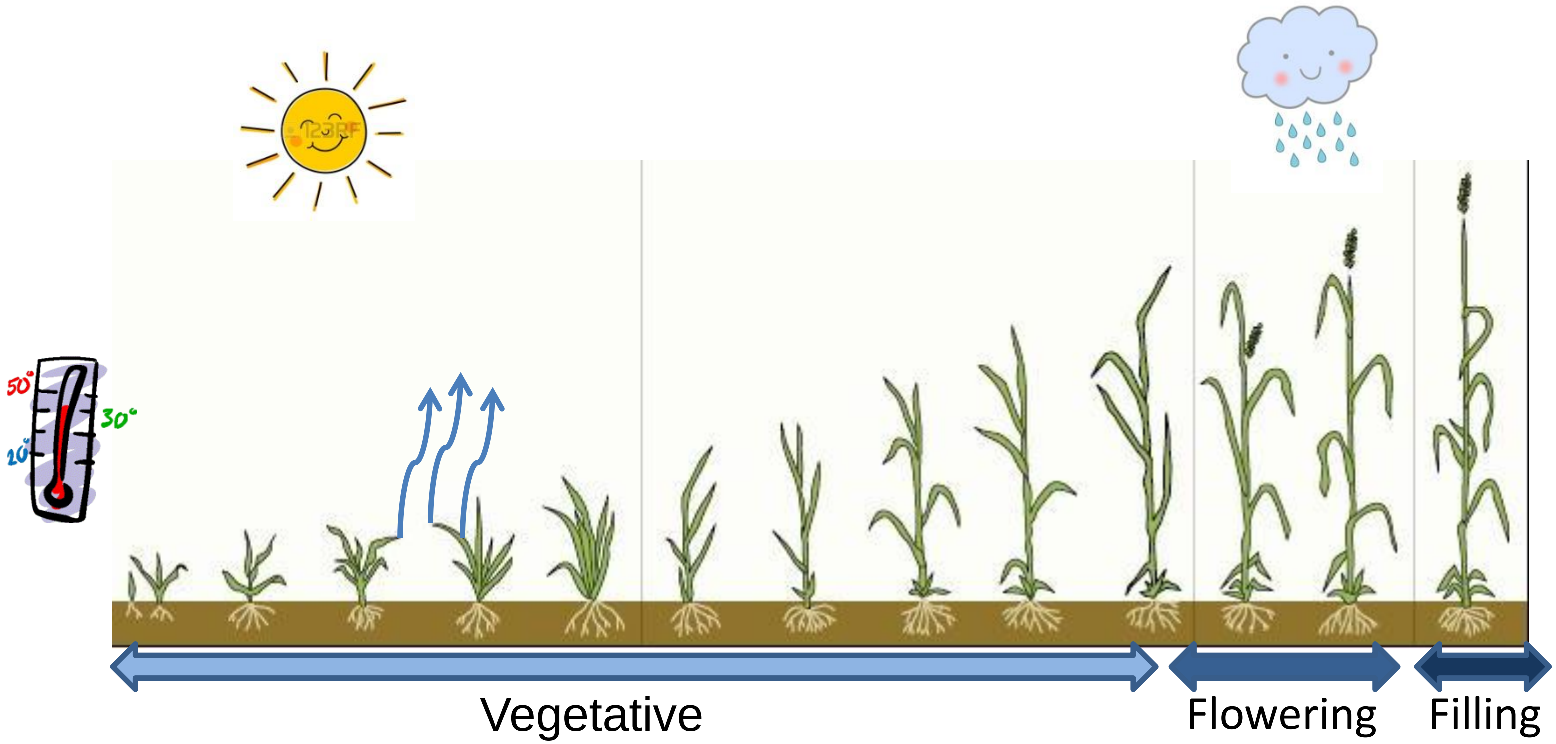
GBLUP$_{(GxE)}$

GBLUP$_{(GxE)}$ + EC

GBLUP$_{(GxE)}$ + Acov

Materials and Methods

Environmental covariates



Mean. min and max temp (°C)
Heliophany (h/dia)
Relative humidity (%)
Evapotranspiration(mm/day)
Accumulated rainfall (mm)

EC were calculated for each phenological stage

Selection of EC through Factor analysis

Materials and Methods

Phenotypic information

A total of 103 Elite inbred lines from the Uruguayan Wheat Breeding Program (UWBP)

One location / five years (2010-2014) / management (4 different sowing dates)= 19 environments

Adjusted means by comparing different experimental designs

Genotypic information

The lines were genotyped by genotyping by sequencing (GBS. *Elshire et al.. 2011*. modified by *Poland et al.. 2012* for wheat).

We identified 81.999 SNPs.

Marker-data imputation was conducted using the realized relationship matrix through the multivariate normal expectation maximization method (MVN-EM) using *rrBLUP* package (*Endelman. 2012*) from R software (R Development Core Team. 2015).

Materials and Methods

Prediction strategies

TRN	TRN	TRN	TRN	TST
Year 1	Year 2	Year 3	Year 4	Year n

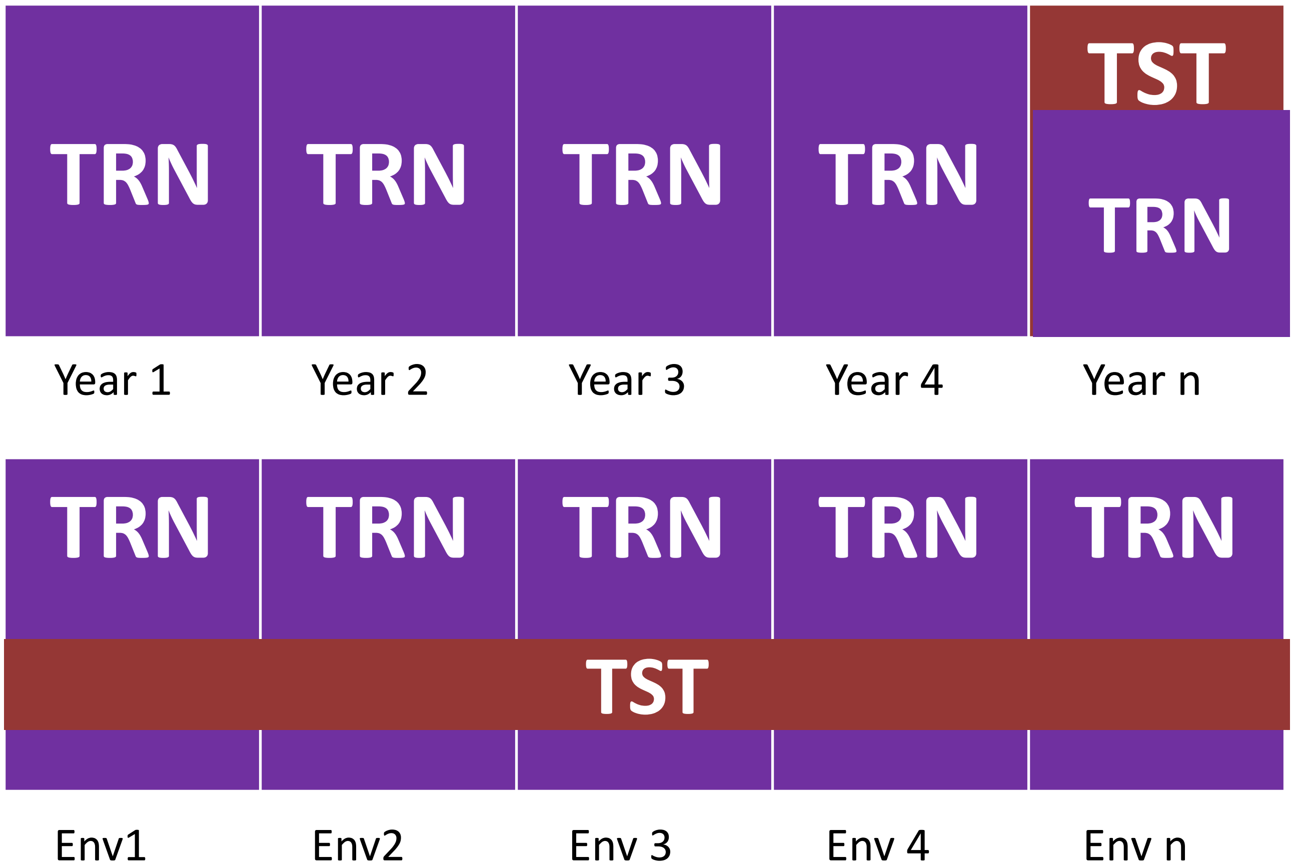
Know genotypes in unphenotyped years

TRN	TRN	TRN	TRN	TST
Env1	Env2	Env 3	Env 4	Env n

Know genotypes in unphenotyped environments

Materials and Methods

Prediction strategies



New genotypes in phenotyped environments

New genotypes in unphenotyped environments

RESULTS

Environmental covariates selection

Env. Covariate	2010	2011	2012	2013	2014
Mean temp VC	NS	NS	**	NS	NS
Mean temp ANT	NS	NS	**	**	NS
Mean temp GF	NS	NS	***	NS	NS
Minimum temp VC	NS	***	**	NS	NS
Minimum temp Ant	**	NS	**	**	NS
Minimum temp GF	NS	NS	NS	NS	NS
Maximum temp VC	**	**	**	NS	**
Maximum temp ANT	NS	NS	NS	**	NS
Maximum temp GF	NS	NS	NS	NS	NS
Accumulated rainfall VC	***	NS	NS	NS	**
Accumulated rainfall ANT	***	NS	**	NS	NS
Accumulated rainfall GF	NS	NS	NS	**	***
Heliophany VC	NS	***	***	**	***
Heliophany ANT	***	**	NS	NS	***
Heliophany GF	NS	NS	NS	NS	NS
Evapotranspiration VC	NS	***	***	NS	***
Evapotranspiration ANT	NS	***	**	NS	***
Evapotranspiration GF	NS	NS	***	NS	***
Relative humidity VC	***	NS	NS	**	**
Relative humidity ANT	***	NS	**	NS	**
Relative humidity GF	NS	NS	***	NS	NS

Vegetative Period
Flowering
Grain fill

NS >0.05
** <0.05
***<0.001

Differences of more than 30 days between planting dates, affecting the

Most of the environmental covariates were significant in any of the years and /or phenological stage

RESULTS

Genomic selection model comparison

<u>Genotypes known in unphenotyped years</u>

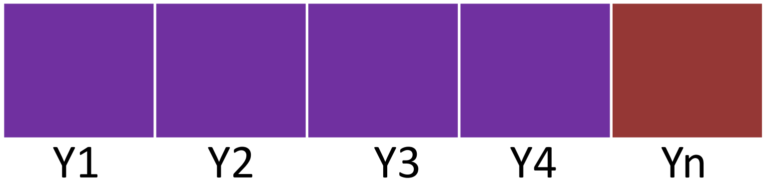


Table. Accuracy of genomic breeding values predictions for yield for different years and models.

Year	GBLUP(M)	GBLUP(GxE)	GBLUP(GxE) + EC	GBLUP(Acov)
2010	0.1063	0.5613	**0.5724**	0.0843
2011	0.2165	**0.4697**	0.4678	0.2485
2012	0.2954	0.4003	0.4011	**0.4395**
2013	0.4631	**0.4864**	0.4795	0.4396
2014	0.3101	0.2535	0.2663	**0.5868**

RESULTS

Known genotypes in unphenotyped environments

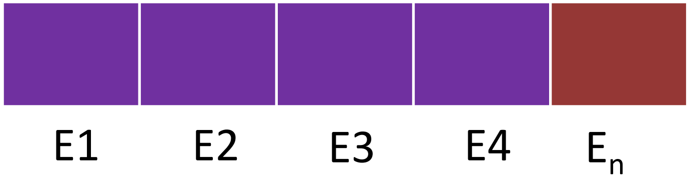


Table. Accuracy of genomic breeding values predictions for yield for different environments and models.

Env	GBLUP(M)	GBLUP(GxE)	GBLUP(GxE) + EC	GBLUP(Acov)
2010LE1	0.2232	0.2582	0.2201	**0.273**
2010LE2	**0.3643**	0.2532	0.1962	0.1592
2010LE3	0.2571	0.2562	0.1983	**0.2812**
2010LE4	0.2438	**0.2853**	0.1787	0.162
2011LE1	0.3107	0.2218	0.2331	**0.3557**
2011LE2	0.2515	0.2849	0.7776	**0.2987**
2011LE3	0.2235	0.2879	**0.3146**	0.0723
2011LE4	0.2011	0.2621	0.1858	**0.3072**
2012LE1	**0.2913**	0.2431	0.2393	0.2395
2012LE2	0.2688	0.2540	0.240	**0.408**
2012LE3	0.2387	0.2651	0.297	**0.3668**
2012LE4	**0.2758**	0.2386	0.1818	0.2650
2013LE1	0.1224	0.2571	0.2544	**0.2690**
2013LE2	0.2421	0.2823	0.2645	**0.3562**
2013LE3	0.2752	0.2425	**0.2927**	0.1630
2013LE4	0.2509	0.2517	0.1825	**0.4643**
2014LE1	0.1279	0.2537	0.2597	**0.2721**
2014LE2	0.0803	**0.2535**	0.1722	0.737
2014LE3	0.0821	0.2537	0.2762	**0.2784**

RESULTS

New genotypes in phenotyped environments

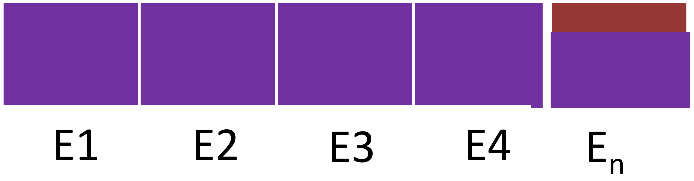


Table. Accuracy of genomic breeding values predictions for yield for different environments and models.

Env	GBLUP(M)	GBLUP(GxE)	GBLUP(GxE) + EC	GBLUP(Acov)
2010LE1	0.297	0.715	**0.720**	0.273
2010LE2	0.513	0.495	0.496	**0.592**
2010LE3	0.458	**0.906**	0.898	0.418
2010LE4	0.174	0.777	**0.787**	0.162
2011LE1	0.294	0.532	0.530	**0.557**
2011LE2	0.651	**0.778**	0.776	0.787
2011LE3	0.712	0.788	**0.795**	0.723
2011LE4	0.276	**0.875**	0.858	0.307
2012LE1	0.411	**0.654**	0.639	0.395
2012LE2	0.405	0.735	**0.740**	0.408
2012LE3	0.659	0.788	**0.797**	0.668
2012LE4	0.639	0.808	**0.818**	0.650
2013LE1	0.655	**0.571**	0.544	0.690
2013LE2	0.474	0.637	**0.645**	0.562
2013LE3	0.646	0.917	**0.927**	0.630
2013LE4	0.508	0.821	**0.825**	0.464
2014LE1	0.705	0.603	0.597	**0.721**
2014LE2	0.826	**0.759**	0.722	0.737
2014LE3	**0.795**	0.714	0.726	0.784

RESULTS

New genotypes in unphenotyped environments

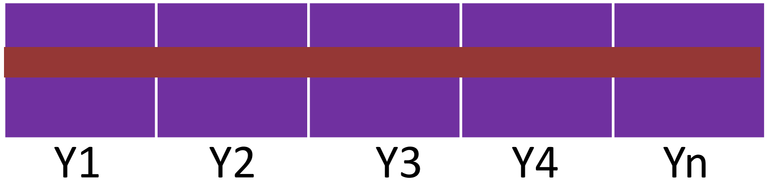


Table. Accuracy of genomic breeding values predictions for yield for different environments and models.

GBLUP	GBLUPgxe	GBLUPgxe+Cov	GBLUPcov
0.3122	0.4172	0.2763	0.4398

CONCLUSIONS

- The use of models that incorporate GEI information improved model prediction accuracies in most situations.

- The selection of environmental covariables by factor analysis was beneficial to determine covariates of greater effect on yield.

- The incorporation of environmental information within the prediction models showed better results through the use of environmental correlation matrices than through the use of fixed covariates in the model.

- Improvements in the systems of envirotyping and crop modeling could show positive advances in environmental characterization and improvement of genomic selection models.

Thank you!

Acknowledgment

INIA FPTA Project

Agronomy College - UdelaR