Genomic prediction, populations, machine learning

AX06072022

Alencar Xavier
Quantitative Geneticist at Corteva Biostatistics
Adjunct professor at Purdue University

Overview

- Brief intro
- Application of genomic prediction in a breeding program
  1. Trait selection
  2. Practical considerations for application
  3. Training population theory
- Genomic enhanced cross prediction in breeding
- Utilizing machine learning to increase selection accuracy and efficiency

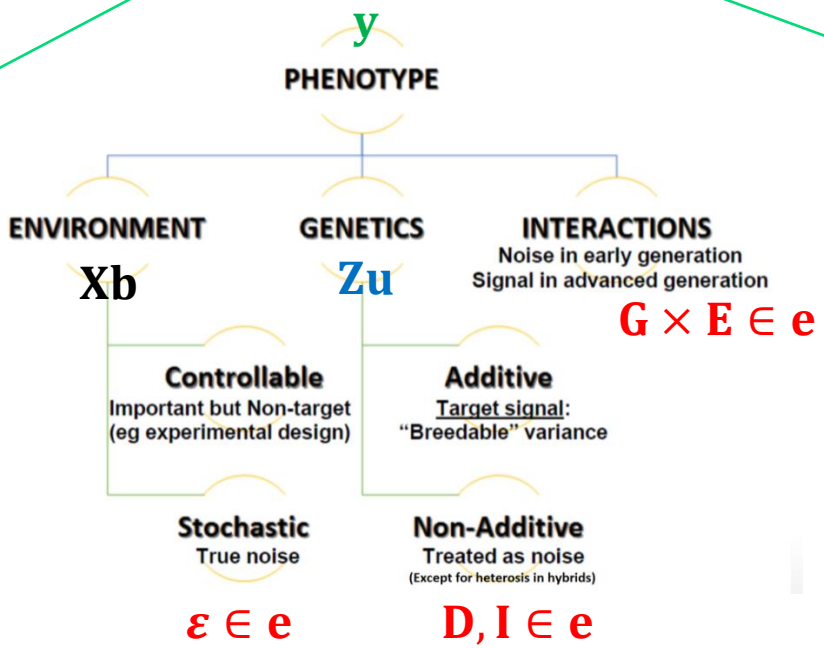
Alencar.Xavier@Corteva.com
Corteva Agrisciences, Latin America and Africa Breeding Analyst

- Brief intro to mixed models
- Application of genomic prediction in a breeding program
  1. Trait selection
  2. Practical considerations for application
  3. Training population theory
- Genomic enhanced cross prediction in breeding
- Utilizing machine learning to increase selection accuracy and efficiency

Alencar.Xavier@Corteva.com
Corteva Agrisciences, Latin America and Africa Breeding Analyst

A simple model

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$$



Model notation

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$$

$$\begin{cases} \mathbf{y} \sim N(\mathbf{Xb}, \mathbf{V}) \\ \mathbf{y} \sim N(\mathbf{Xb}, \mathbf{ZGZ}' + \mathbf{R}) \end{cases}$$

$$\mathbf{u} \sim N(0, \mathbf{G}) \qquad \mathbf{e} \sim N(0, \mathbf{R})$$

where

$$\mathbf{G} = \mathbf{A}\sigma_a^2 \qquad \mathbf{R} = \mathbf{I}\sigma_e^2$$

$$\mathrm{cov}(\mathbf{Zu}, \mathbf{e}) = \mathbf{0}$$

n = number of observations
p = number of parameters
q = number of individuals

y = vector of observations (n)
X = design matrix of fixed effects (n × p)
b = vector of fixed effect coefficients (p)
Z = incidence matrix of random effects (n × q)
u = vec. of random effects − genetics values (q)
e = vector of residuals (n)
σ_a^2 = random effect variance (1)
σ_e^2 = residual variance (1)
R = residual variance matrix (n × n)
G = genetic variance matrix (q × q)
A = relationship matrix (q × q)
$\lambda = \sigma_e^2 : \sigma_a^2$ = regularization parameter (1)

What the variance-covariances mean?

$$\mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}; \qquad \mathbf{G} = \mathbf{A}\sigma_a^2; \qquad \mathbf{R} = \mathbf{I}\sigma_e^2;$$

- The covariance between i^{th} and j^{th} genotypes is:

$$g_{ij} = a_{ij}\sigma_a^2$$

Without relationship $a_{ij} = 0$

- The covariance between i^{th} and j^{th} observations is:

$$v_{ij} = z_i A z_j' \sigma_a^2 + r_{ij}\sigma_e^2$$

Key metrics

- **Heritability** (plot level): $\mathbf{H_p = V_y^{-1} V_G}$

Heritability on balanced populations without relationship:

$$\frac{\sigma_a^2}{\sigma_a^2 + n^{-1}\sigma_e^2}$$

- **Heritability** (entry level): $\mathbf{H_e = V_u V_{\hat{u}}^{-1} = G\left(G - C^{22}\right)^{-1} = \left(I - C^{22}G^{-1}\right)^{-1}}$

- **Accuracy**: $\mathbf{a} = \mathrm{cor}(u, \hat{u}) = \frac{\mathbf{cov}(\hat{u},u)}{\sqrt{\mathbf{var}(\hat{u})\mathbf{var}(u)}} = \sqrt{\frac{GZ'V^{-1}ZG}{G}}$

Accuracy on (observed) balanced population without relationship:

$$\sqrt{\frac{\sigma_a^2}{\sigma_a^2 + n^{-1}\sigma_e^2}}$$

- **Reliability**: $\mathbf{r} = \sqrt{\mathrm{diag}(\mathbf{H_e})}$

Reliability of observed individuals from population without relationship:

$$\sqrt{\frac{\sigma_a^2}{\sigma_a^2 + n_i^{-1}\sigma_e^2}}$$

Key metrics

- Heritability
  - Direct measure of genetic control
  - Assess statistical models, experimental designs
- Accuracy
  - Check how well we can predict something
  - Optimize TPE/TPG, experimental designs, training sets
  - Response to selection ($R \propto i \times r_{g,\hat{g}} \times \sigma_a$)
- Reliability
  - Direct measure of confidence
  - Deregression = Unshkring BLUPs for GWAS and multistage analysis
    - Mitigate Bulmer effect (changes in relative ranking)

Alencar.Xavier@Corteva.com
Corteva Agrisciences, Latin America and Africa Breeding Analyst

How are the parameters estimated?

- Henderson's equation ($Cg = r$)

$$\begin{bmatrix} X'R^{-1}X & Z'R^{-1}X \\ X'R^{-1}Z & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}$$

- We know (data): $x = \{y, X, Z, A\}$
- We want (parameters): $\theta = \{b, u, \sigma_a^2, \sigma_e^2\}$
- Parameter estimation based on Gaussian likelihood: $L(x|\theta)$

Review lecture on variance component estimations
https://rpubs.com/alenxav/varComp

$$\hat{\sigma}_u^2 = \frac{\hat{\mathbf{u}}'\mathbf{A}^{-1}\hat{\mathbf{u}} + \mathbf{A}^{-1}\mathbf{C}^{22}}{q}$$

$$\hat{\sigma}_e^2 = \frac{\mathbf{y}'\mathbf{e}}{n - r_X}$$

Example from **Cunningham & Henderson 1968**

DATA AND INCIDENCE MATRICES

y	μ	a_1	a_2	b_1	b_2	b_3
3	1	1	0	1	0	0
2	1	1	0	0	1	0
3	1	1	0	0	0	1
2	1	1	0	1	0	0
3	1	1	0	0	1	0
5	1	1	0	0	1	0
6	1	1	0	0	1	0
7	1	1	0	0	1	0
2	1	0	1	1	0	0
8	1	0	1	0	1	0
4	1	0	1	0	0	1
3	1	0	1	1	0	0
8	1	0	1	0	1	0
4	1	0	1	0	0	1
9	1	0	1	0	1	0
3	1	0	1	0	0	1
2	1	0	1	0	0	1
5	1	0	1	0	0	1

$$\mathbf{y} = \mathbf{Xa} + \mathbf{Zb} + \mathbf{e}$$

The least squares equations (ignoring $\boldsymbol{\mu}$) are

$$\left[\begin{array}{cc|ccc} 8 & 0 & 2 & 5 & 1 \\ 0 & 10 & 2 & 3 & 5 \\ \hline 2 & 2 & 4 & 0 & 0 \\ 5 & 3 & 0 & 8 & 0 \\ 1 & 5 & 0 & 0 & 6 \end{array}\right]\left[\begin{array}{c} a_1 \\ a_2 \\ \hline b_1 \\ b_2 \\ b_3 \end{array}\right] = \left[\begin{array}{c} 31 \\ 48 \\ \hline 10 \\ 48 \\ 21 \end{array}\right]$$

In algebraic terms, these equations are

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} \end{bmatrix}\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \end{bmatrix}$$

$$\lambda = 0.5721$$

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \mathbf{I}k \end{bmatrix}\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \end{bmatrix}$$

$$\left[\begin{array}{cc|ccc} 8 & 0 & 2 & 5 & 1 \\ 0 & 10 & 2 & 3 & 5 \\ \hline 2 & 2 & 4.5721 & 0 & 0 \\ 5 & 3 & 0 & 8.5721 & 0 \\ 1 & 5 & 0 & 0 & 6.5721 \end{array}\right]\left[\begin{array}{c} a_1 \\ a_2 \\ \hline b_1 \\ b_2 \\ b_3 \end{array}\right] = \left[\begin{array}{c} 31 \\ 48 \\ \hline 10 \\ 48 \\ 21 \end{array}\right]$$

```
> solve(C, g)
 [1] 2.9371 4.8684 -1.2272 2.1826 -0.9554
```

a1 a2 b1 b2 b3

Example from **Robinson 1991**

Data

Herd	Sire	Yield
1	A	110
1	D	100
2	B	110
2	D	100
2	D	100
3	C	110
3	C	110
3	D	100
3	D	100

Design matrices

$$X = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \quad Z = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Solving

$$\left(\begin{array}{ccc|cccc} 2 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 3 & 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 4 & 0 & 0 & 2 & 2 \\ \hline 1 & 0 & 0 & 11 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 11 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 12 & 0 \\ 1 & 2 & 2 & 0 & 0 & 0 & 15 \end{array}\right) \begin{pmatrix} \hat{h}_1 \\ \hat{h}_2 \\ \hat{h}_3 \\ \hat{s}_A \\ \hat{s}_B \\ \hat{s}_C \\ \hat{s}_D \end{pmatrix} = \begin{pmatrix} 210 \\ 310 \\ 420 \\ 110 \\ 110 \\ 220 \\ 500 \end{pmatrix}$$

MME

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{Z}'\mathbf{X} \\ \mathbf{X}'\mathbf{Z} & \mathbf{Z}'\mathbf{Z} + \lambda\mathbf{K}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$

$$(\lambda = \sigma_e^2/\sigma_a^2)$$

which has solution

$$(1.4) \qquad \hat{\beta} = (105.64, 104.28, 105.46)^T,$$
$$\hat{u} = (0.40, 0.52, 0.76, -1.67)^T.$$

- Brief intro to mixed models
- Application of genomic prediction in a breeding program
  1. Trait selection
  2. Practical considerations for application
  3. Training population theory
- Genomic enhanced cross prediction in breeding
- Utilizing machine learning to increase selection accuracy and efficiency

Alencar.Xavier@Corteva.com
Corteva Agrisciences, Latin America and Africa Breeding Analyst

Trait selection

- Breeding objective?

  - Set of traits of interest (**TOI**)

  bred into a

  - Target population of genotypes (**TPG**)

  for a given

  - Target population of environments (**TPE**)

TPE/TPG

- Target population of environments (TPE)
  - Influences accuracies via GxE correlation
  - Which environments should I be able to predict?

- Target population of genotypes (TPG)
  - Influences accuracies via genetic relationship
  - Which genetics should I be able to predict?

From QTLs to Adaptation Landscapes: Using Genotype-To-Phenotype Models to Characterize G×E Over Time

Daniela Bustos-Korts[1], Marcos Malosetti[1], Karine Chenu[2], Scott Chapman[3,4], Martin P. Boer[1], Bangyou Zheng[3] and Fred A. van Eeuwijk[1]*

What Should Students in Plant Breeding Know About the Statistical Aspects of Genotype × Environment Interactions?

Fred A. van Eeuwijk,* Daniela V. Bustos-Korts, and Marcos Malosetti

An Equation to Predict the Accuracy of Genomic Values by Combining Data from Multiple Traits, Populations, or Environments FREE

Yvonne C J Wientjes ✉, Piter Bijma, Roel F Veerkamp, Mario P L Calus

Genetics, Volume 202, Issue 2, 1 February 2016, Pages 799–823,
https://doi.org/10.1534/genetics.115.183269

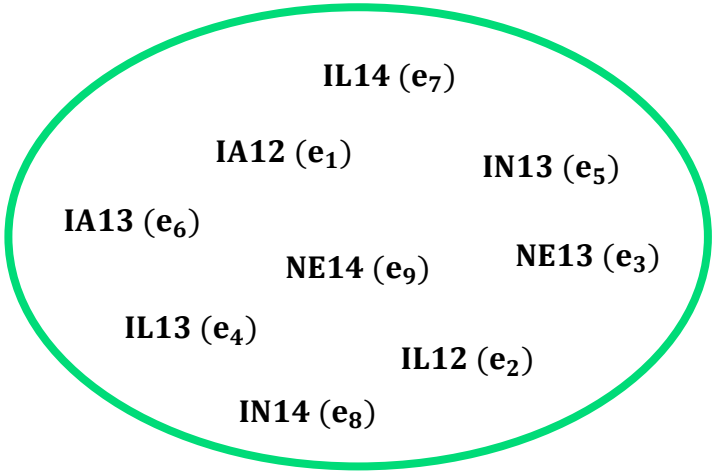
Multiple environments

- Any given breeding trial happens in each environment that is sample of many environments:

$$e_i \in E$$

That is:

$$\mathbf{TPE\ (E) =}$$



$$\begin{bmatrix} y_{e_i} \\ y_{e_j} \\ g_E \end{bmatrix} = \begin{bmatrix} \sigma^2_{g(e_i)} + \sigma^2_{\epsilon(e_i)} & \sigma_{g(e_i,e_j)} & \sigma_{g(e_i,E)} \\ \sigma_{g(e_j,e_i)} & \sigma^2_{g(e_j)} + \sigma^2_{\epsilon(e_j)} & \sigma_{g(e_j,E)} \\ \sigma_{g(E,e_i)} & \sigma_{g(E,e_j)} & \sigma^2_{g(E)} \end{bmatrix}$$

Multiple traits and environments

$$y = \{y_1, y_2, \ldots, y_k\}$$

With multiple traits, the relation among traits is modeled

$$V(u) = A \otimes \Sigma_a = \begin{bmatrix} A\sigma^2_{a_1} & A\sigma_{a_{12}} \\ A\sigma_{a_{21}} & A\sigma^2_{a_2} \end{bmatrix}$$

$$V(e) = I \otimes \Sigma_e = \begin{bmatrix} I\sigma^2_{e_1} & I\sigma_{e_1e_2} \\ I\sigma_{e_2e_1} & I\sigma^2_{e_2} \end{bmatrix}$$

Alencar.Xavier@Corteva.com
Corteva Agrisciences, Latin America and Africa Breeding Analyst

Practical example – Estimate covariances

Consider a dataset with multiple traits (columns) and individuals (rows)

```
> head(Y,10)
         TRAIT1 TRAIT2 TRAIT3 TRAIT4 TRAIT5
GENO001      NA   9.24   8.91  10.11   9.78
GENO002   13.26  11.03  13.85     NA  10.89
GENO003   12.30  10.69  12.24     NA  10.11
GENO004   10.53  10.55   9.05   7.39     NA
GENO005   10.80  11.25     NA   9.35  12.66
GENO006   10.26  10.73     NA  12.16  10.62
GENO007    9.60   8.94   8.46   6.63  10.01
GENO008      NA   9.58  10.08   9.07  12.49
GENO009   10.40     NA     NA   9.26   7.65
GENO010   12.17  10.83   9.89  11.14  12.05
```

Fit a model using phenotypes (Y) and genotypes (X)

```
> require(bWGR)
> fit = mrr(Y,X)
> round(fit$h2,2)
[1] 0.38 0.48 0.71 0.63 0.60
> round(fit$GC,2)
     [,1] [,2] [,3] [,4] [,5]
[1,] 1.00 0.76 0.70 0.64 0.62
[2,] 0.76 1.00 0.56 0.65 0.39
[3,] 0.70 0.56 1.00 0.71 0.23
[4,] 0.64 0.65 0.71 1.00 0.24
[5,] 0.62 0.39 0.23 0.24 1.00
```

Genomic heritability

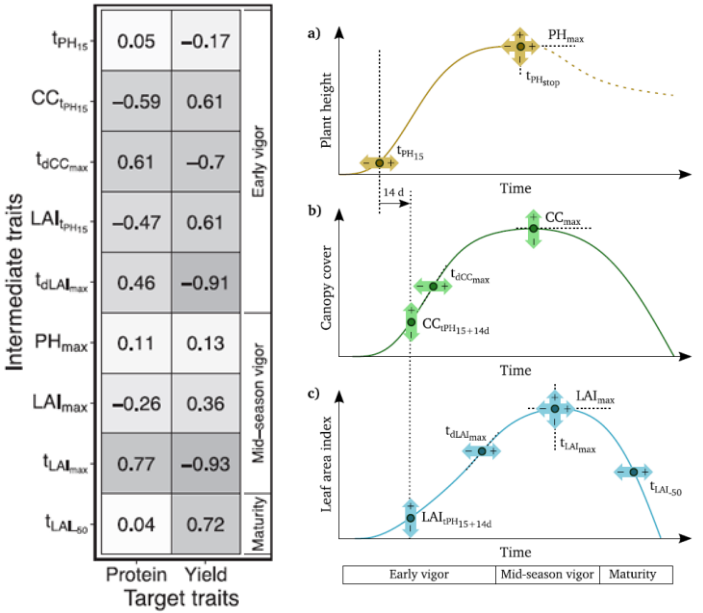
Genetic correlations

+ Genomic breeding values to make selections
+ Marker effects to predict new individuals
+ Variance components to create selection indices

High-throughput phenotyping

Singh A.K. et al. (2021) High-Throughput Phenotyping in Soybean. https://doi-org.ezproxy.lib.purdue.edu/10.1007/978-3-030-73734-4_7

Correlation(HTP, main trait)



Roth et al. (2022) High-throughput field phenotyping of soybean: Spotting an ideotype. https://doi.org/10.1016/j.rse.2021.112797

- Index

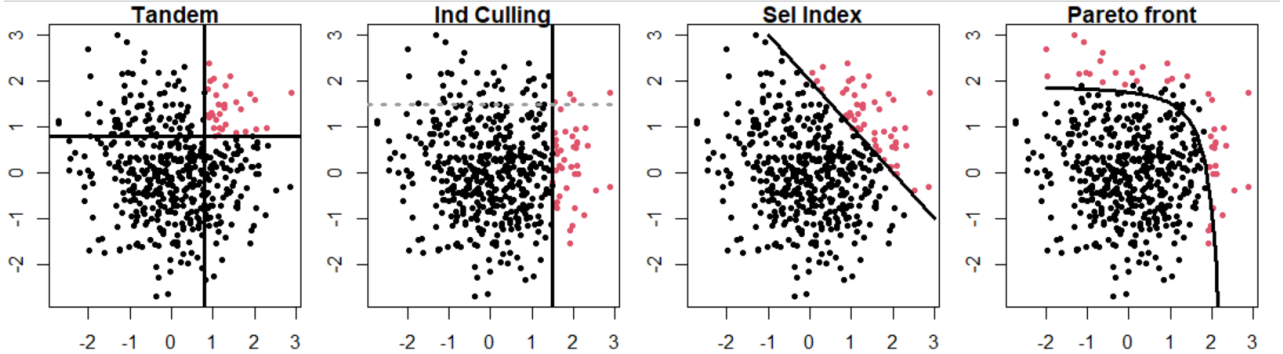
$$\mathbf{I} = \mathbf{Uw} = u_1 w_1 + u_2 w_2 + \cdots + u_K w_K$$

- Optimal selection index (Smith-Hazel)
  - $\boldsymbol{w} = \mathbf{G}\mathbf{V}^{-1}\alpha = \boldsymbol{\Sigma}_a(\boldsymbol{\Sigma}_a + \boldsymbol{\Sigma}_e)^{-1}\boldsymbol{\alpha}$
  - $\boldsymbol{\alpha} = economic\ value$

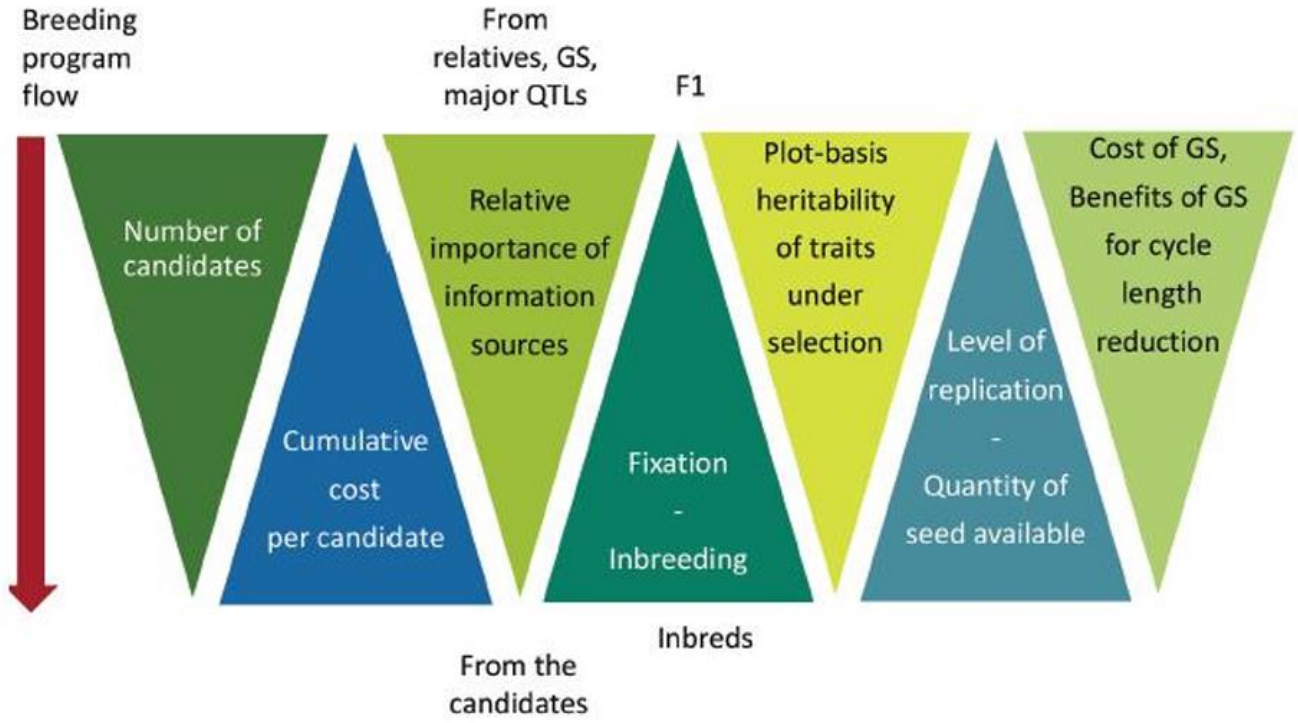
- Estimated from multi-variate models:
  - $\boldsymbol{\Sigma}_a$ = genetic covariance among traits (k × k)
  - $\boldsymbol{\Sigma}_e$ = residual covariance among traits (k × k)

Alternatives to selection index?

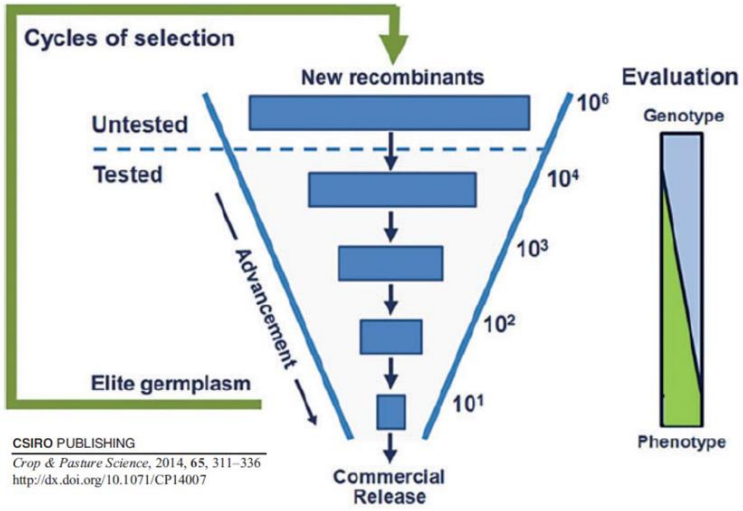
- Tandem selection
- Independent culling
- Multi-objective selection

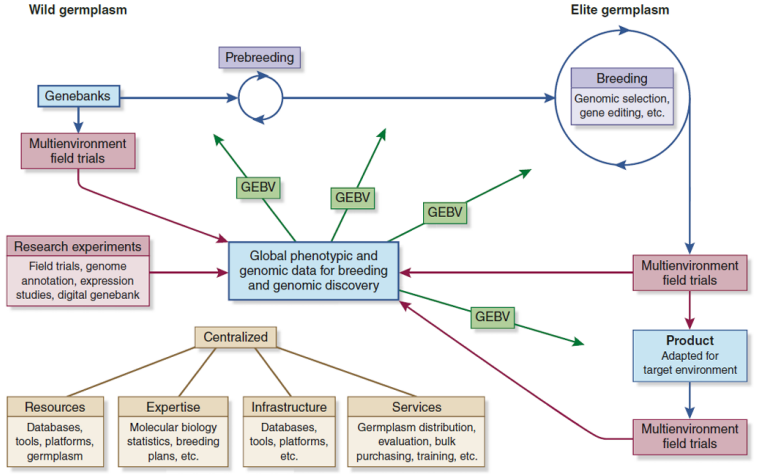


Practical considerations for application



Training population theory





Hickey et al. (2017) *Nature genetics* 49(9):1297

Training population theory

- Where does the data come from?
  - Breeding pipeline
  - Designed experiments
- Accuracy-based optimization (Wientjes et al 2016,Mangin et al 2019)
  - Accuracy is a function of
    1. Trait H2, genetic architecture
    2. Relationship between ES and PS

Accuracy

Accuracy is the square root of reliability ($a = \sqrt{r^2}$). Accuracy is generally defined as the correlation between estimated and true breeding values.

$$a = cor(u, \hat{u}) = \frac{cov(u, \hat{u})}{sd(u) \times sd(\hat{u})}$$

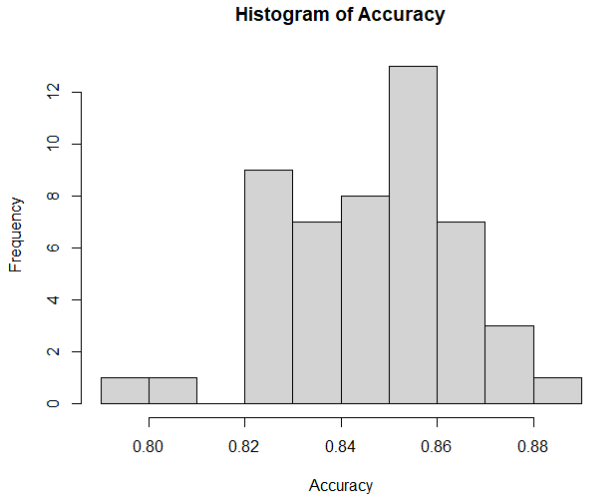
If the statistical model is the true model, $cov(u, \hat{u}) = cov(\hat{u}, \hat{u}) = \hat{\sigma}^2_u$

$$a_i = \sqrt{\frac{cov(u, \hat{u})^2}{var(\hat{u})var(u)}} = \sqrt{\frac{cov(u, \hat{u})}{var(u)}} = \sqrt{\frac{G_{iy}Z'V^{-1}ZG_{yi}}{G_{ii}}}$$

When all individuals are observed, the denominator G_{ii} cancels out with the nominator G_{iy}, yielding $a = \sqrt{Z'V^{-1}ZG}$.

Simple, practical example

```
> require(bWGR)
> data(tpod)
> Accuracy = EigenAcc(gen[1:100,],gen[101:150,])
> hist(Accuracy)
> mean(Accuracy)
[1] 0.8466465
```



- Brief intro to mixed models
- Application of genomic prediction in a breeding program
  1. Trait selection
  2. Practical considerations for application
  3. Training population theory
- Genomic enhanced cross prediction in breeding
- Utilizing machine learning to increase selection accuracy and efficiency

Alencar.Xavier@Corteva.com
Corteva Agrisciences, Latin America and Africa Breeding Analyst

Cross predictions

- Simulation?
  - Cases with trait introgression (TI) or marker assisted selection (MAS)
  - Cases with GWS not based on linear models
- Deterministic?
  - Marker effects (β) are known, and offspring genotypes ($X_{A\times B}$) are known

$$\mathrm{V}[X_{A\times B}\beta] = \beta' X' X \beta$$

$$= 2\sum_{j=1}^{J} p_j(1-p_j)\beta_j^2 = 0.5\sum_{j=1}^{J}\beta_j^2 =$$

half of the sum of b^2, using only markers segregating between the pair of genotypes

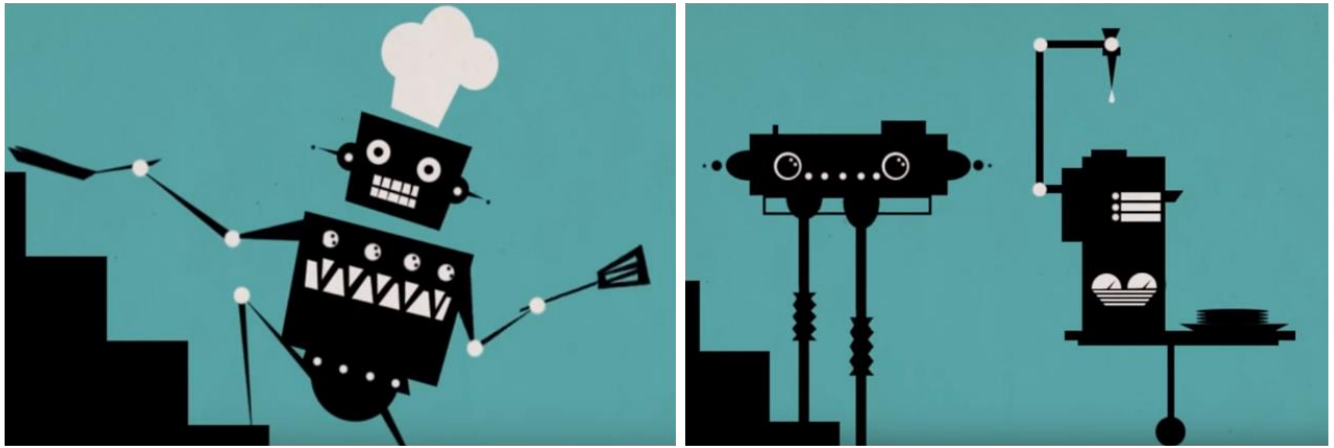
Alencar.Xavier@Corteva.com
Corteva Agrisciences, Latin America and Africa Breeding Analyst

- Brief intro to mixed models
- Application of genomic prediction in a breeding program
  1. Trait selection
  2. Practical considerations for application
  3. Training population theory
- Genomic enhanced cross prediction in breeding
- Utilizing machine learning to increase selection accuracy and efficiency

Alencar.Xavier@Corteva.com
Corteva Agrisciences, Latin America and Africa Breeding Analyst

Why is machine learning good for?

Good for solving **single** **well-defined** **problem**



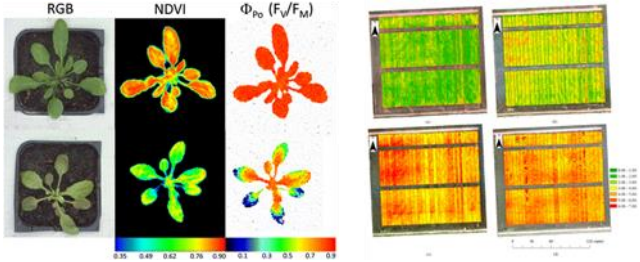
Source: https://www.youtube.com/watch?v=MPR3o6Hnf2g

Where has ML been *most* successful in PB?

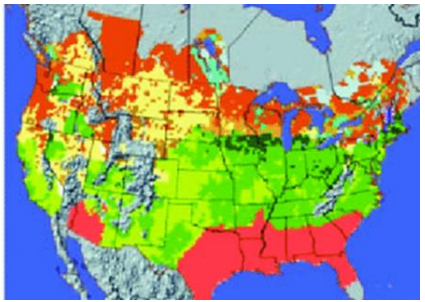
- HTP (e.g., drone, computer vision)
- Environmental classification (TPE'ing?)
- What about genomic prediction??

Machine Learning in GS (Xavier et al 2017,Xavier 2019,Xavier 2021)

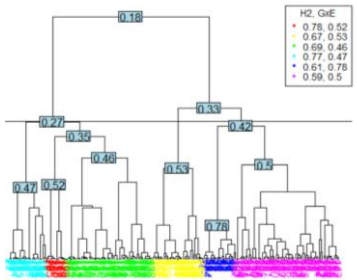
https://www.biomedcentral.com/collections/phenomics (left)
https://www.mdpi.com/2072-4292/11/17/2021 (right)

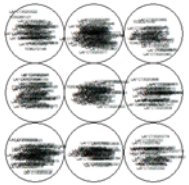


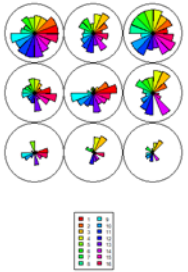
https://www.publish.csiro.au/cp/CP14007

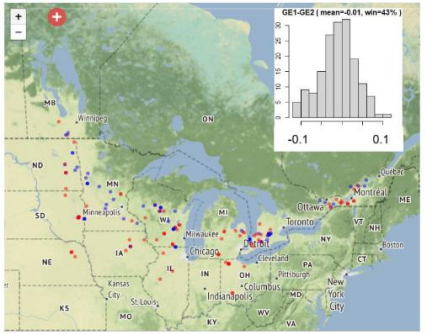


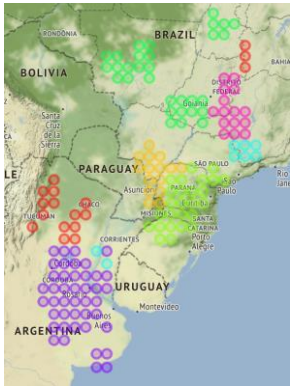
Machine learning to work on TPE and GxE patterns

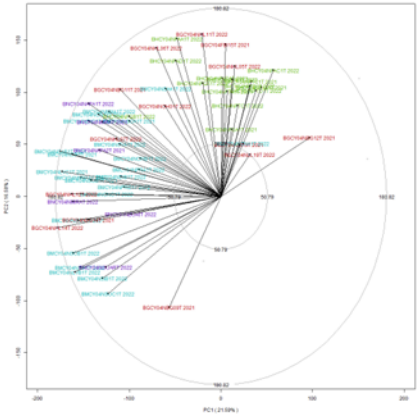


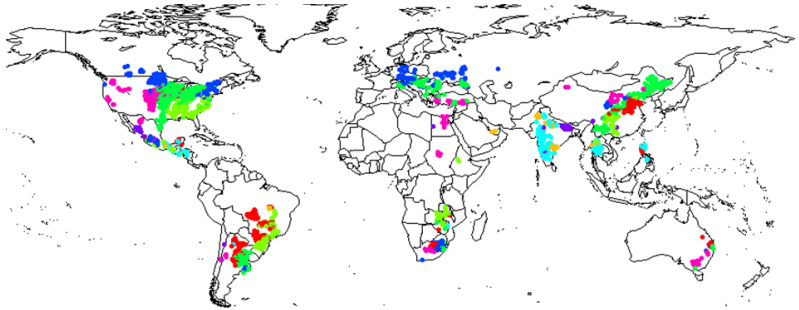












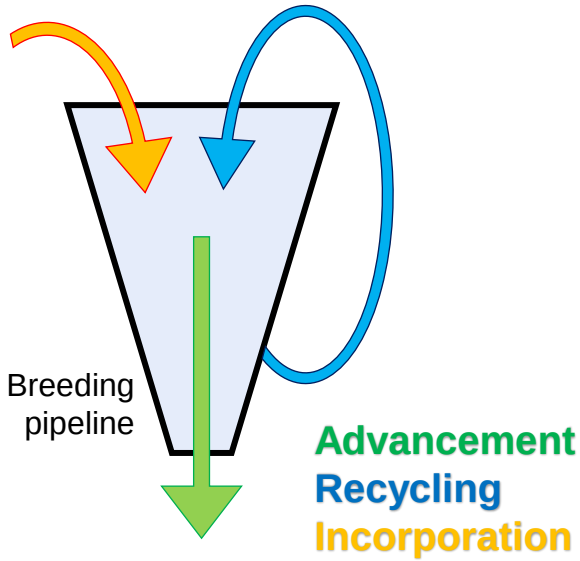
Alencar.Xavier@Corteva.com
Corteva Agrisciences, Latin America and Africa Breeding Analyst

Chasing the right signal

- Breeding value (**GEBV**)
  - *Pattern*: ADDITIVE GENETICS
  - *Method*: GBLUP, BayesABC, LASSO
  - *Suits*: **RECYCLING**, **ADVANCEMENT**

- Genomic value (**EGV**)
  - *Pattern*: ANY GENETICS
  - *Method*: RKHS, DNN, Random Forest
  - *Suits*: **ADVANCEMENT, PRODUCT PLACEMENT**

ML!

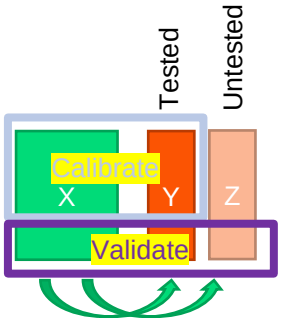


Testing machines for different scenarios of genomic prediction

	Genotype	Environment	Prediction Difficulty
CV00	New	New	*****
CV0	Observed	New	***
CV1	New	Observed	***
CV2	Observed	Observed	*

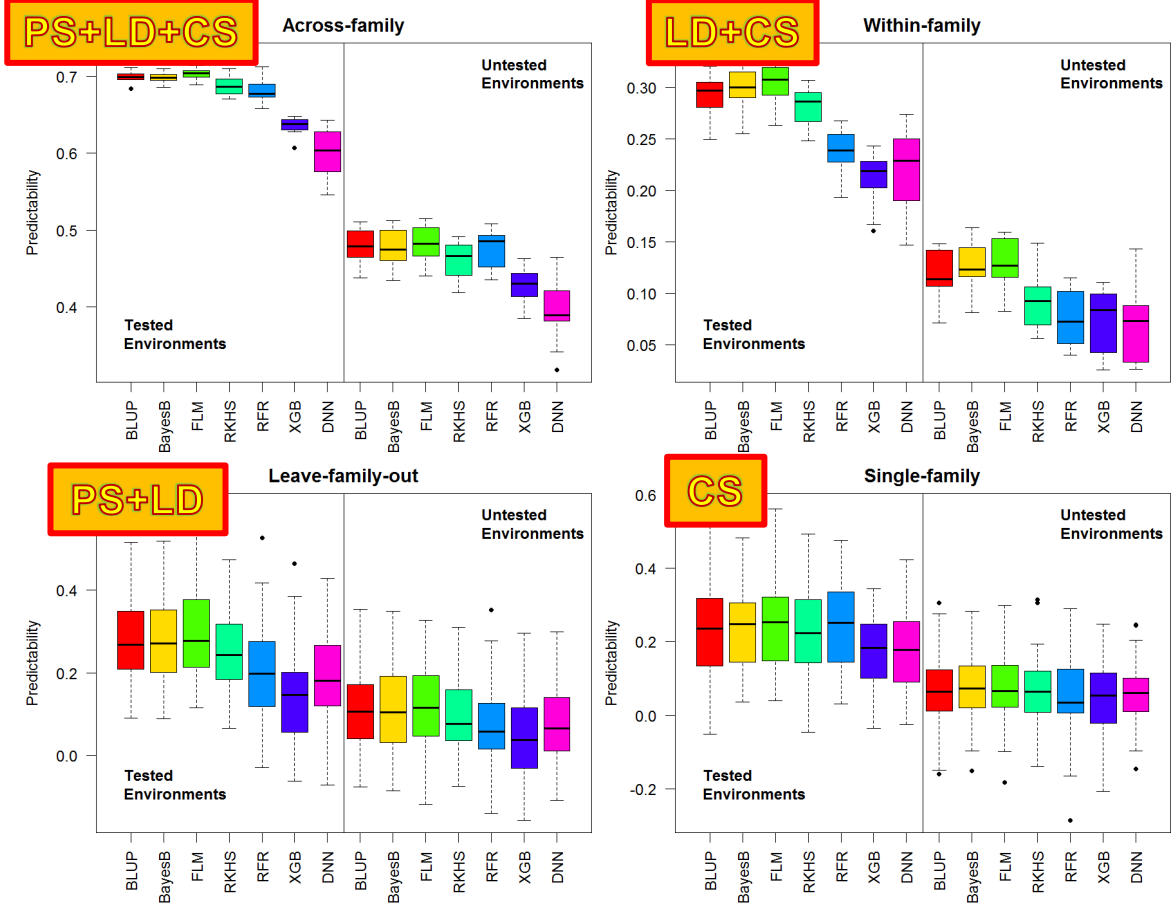
Adapted from Crossa et al. (2017) doi.org/10.1016/j.tplants.2017.08.011

CV scheme



Type of information captured by SNP

- Population structure (PS)
- Linkage disequilibrium (LD)
- Cosegregation / Haplotype (CS)

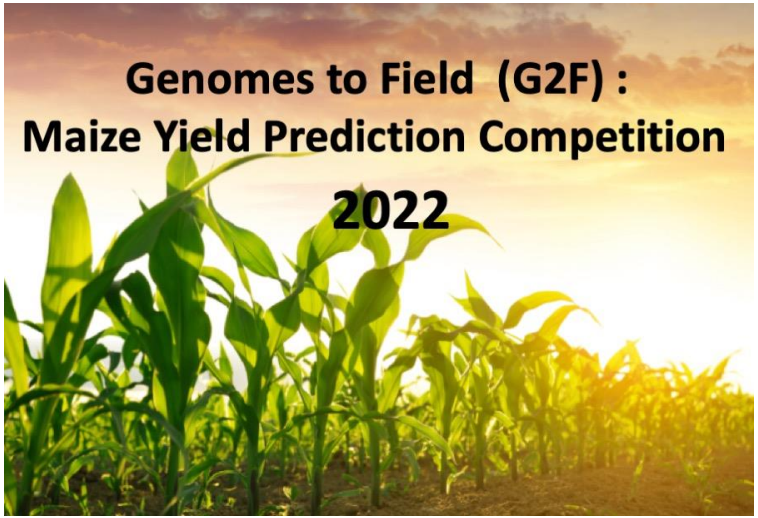


SoyNAM data
ES: 2012 (7 loc)
PS: 2013 (4 loc)
#Fam = 40
Genos = 5600
SNPs = 4300
Obs: 3k-5k obs/loc

See genetic information theory (Habier et al 2007, Habier et al 2013)

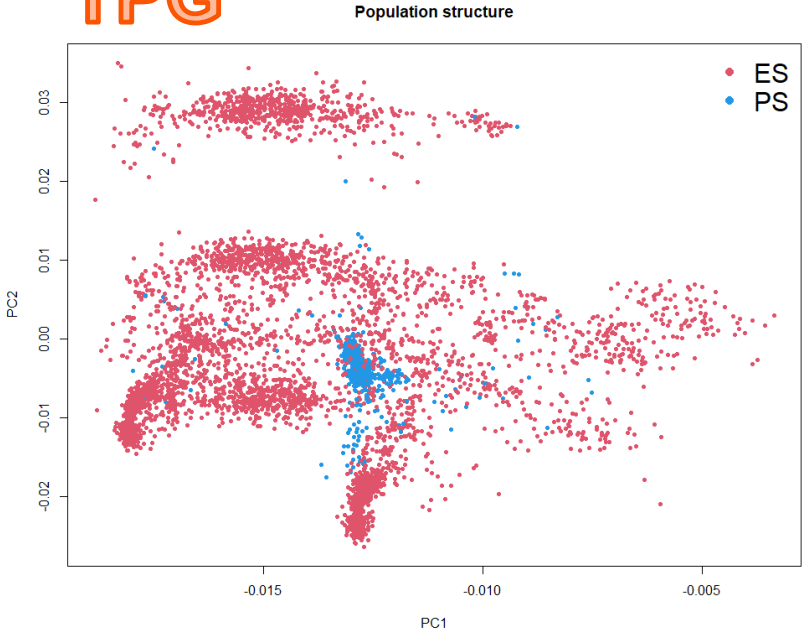
CORTEVA™ agriscience
Alencar.Xavier@Corteva.com
Corteva Agrisciences, Latin America and Africa Breeding Analyst

Case of study

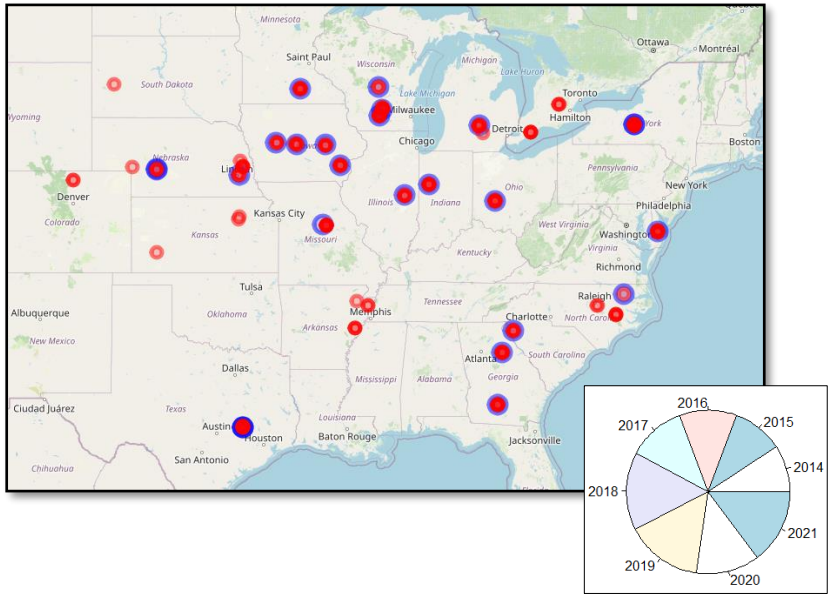


2022 G2F GxE prediction competition

TPG



TPE



Alencar.Xavier@Corteva.com
Corteva Agrisciences, Latin America and Africa Breeding Analyst

What to model?

$$y|E_i = \mu_i + g|E_i$$

$$\text{Phenotype @ i}^{\text{th}}\text{ Loc} = \text{i}^{\text{th}}\text{ Loc Mean} + \text{Genetic effect @ i}^{\text{th}}\text{ Loc}$$

- The winning approach:
  - Predict location means using mixed model and machine learning
  - Predict genetic performance with **selection index** based on **TPE/TPG**

Alencar.Xavier@Corteva.com
Corteva Agrisciences, Latin America and Africa Breeding Analyst

2022 G2F GxE prediction competition

Realized results

Team Name	Within RMSE
CLAC	2.329
igorkf	2.345
phenomaize	2.374
UCD_MegaLMM	2.387
CGM	2.391
breedingteam	2.398
Purdue	2.402
SmAL	2.425
ML_APT	2.472
MPB_Group	2.544

Ranking with alternative metrics

Team Name	Cor Within Loc
CLAC	0.357
CGM	0.353
MPB_Group	0.342
UCD_MegaLMM	0.338
SmAL	0.285
DeepCropVision	0.281
CropEnthusiast	0.279
AllModelsAreWrong	0.272
DataJanitors	0.256
supermanwasd	0.243

Team Name	Cor Across Loc
breedingteam	0.650
DataJanitors	0.644
CLAC	0.631
Purdue	0.631
UCD_MegaLMM	0.628
phenomaize	0.617
igorkf	0.600
CGM	0.587
SmAL	0.586
AllModelsAreWrong	0.575

Source: Jacob Washburn, Jose Ignacio Varela, Alencar Xavier

CORTEVA™ agriscience
Alencar.Xavier@Corteva.com
Corteva Agrisciences, Latin America and Africa Breeding Analyst

Thank you for your attention!

Final remarks:

1) The evaluation metric values locations means
2) ES-PS shared same locations environments
3) Machine learning can help with characterizing TPE/TPG

Questions??

Alencar Xavier

Alencar.Xavier@Corteva.com

- https://rpubs.com/alenxav/varComp
- https://alenxav.wixsite.com/home
- https://github.com/alenxav/Lectures