Trends of predictive breeding

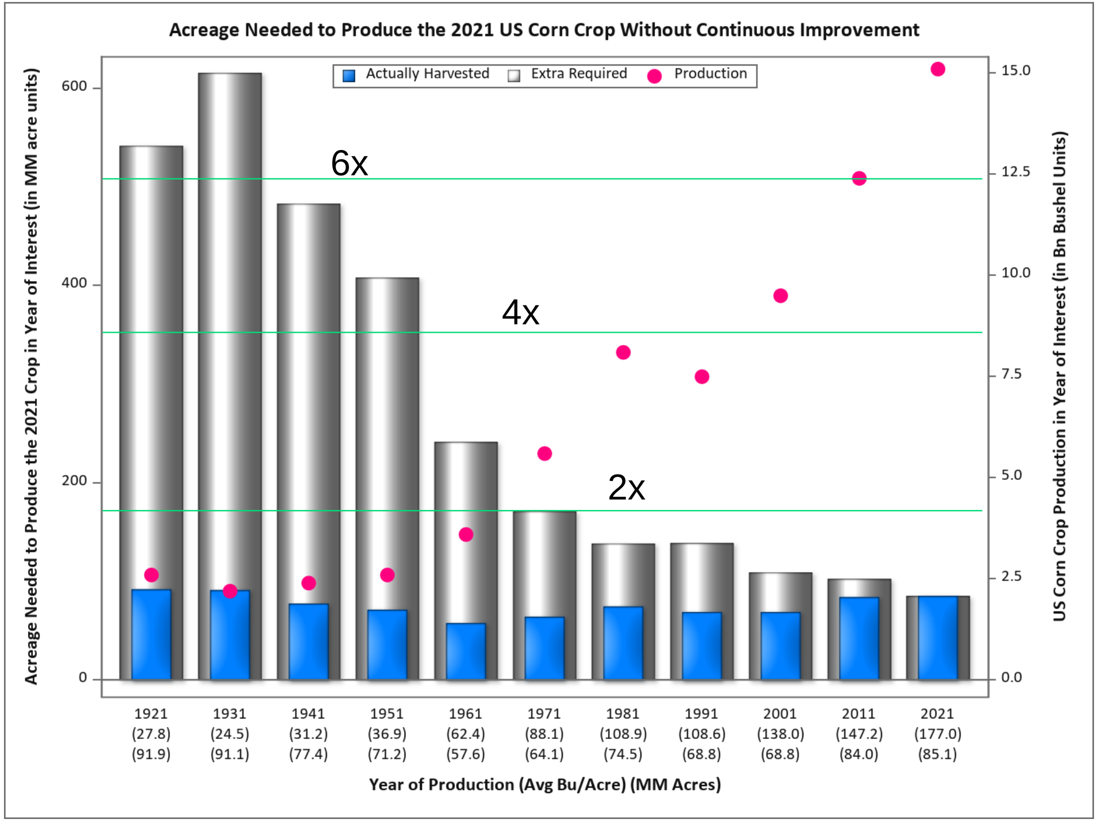
Changes in the plant breeding landscape driven by more, better and different data

Alencar Xavier
Breeding Analyst at Corteva Agrisciences
Adjunct professor at Purdue University

12/08/2023

Contributions from Radu Totir, Frank Technow,
David Habier, David Bubeck, Abelardo de la Vega

What are some implications of continuous Corn Improvement?



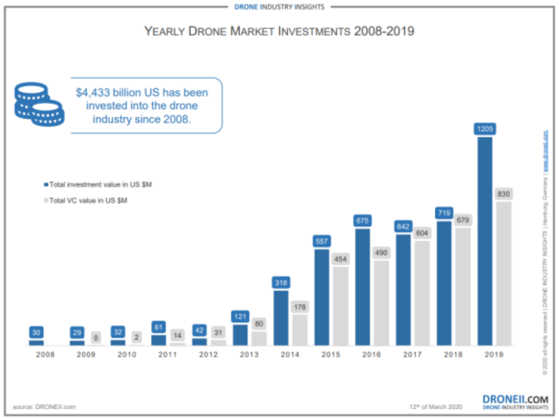
***Based on 2021 USDA NASS data**

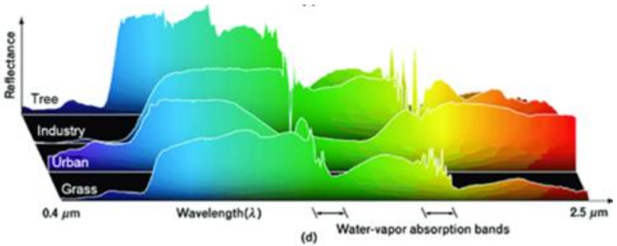
Source: Totir 2021, ASTA

CORTEVA agriscience

Alencar.Xavier@Corteva.com
Quantitative Geneticist, Breeding Analyst LAAF

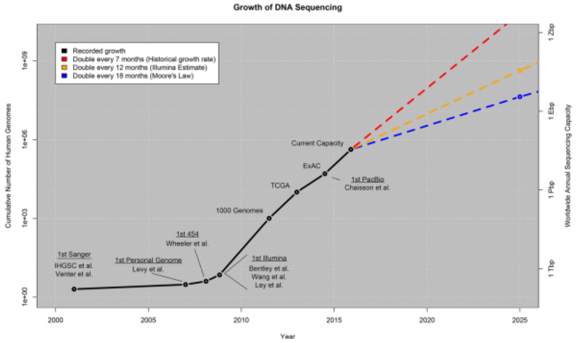
More Pheno



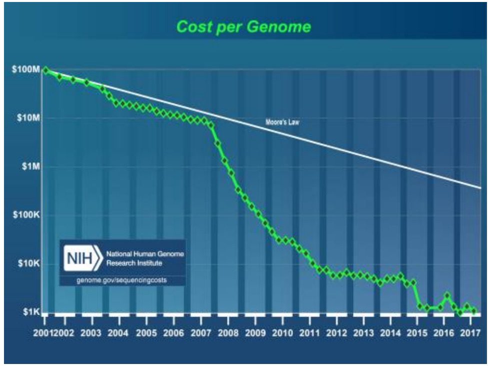


https://www.mdpi.com/2076-3417/12/5/2570

More Geno



The Cost of Sequencing a Human Genome. NIH.
https://www.genome.gov/27565109/the-cost-of-sequencing-a-human-genome/



Stephens, Z. D.et al. (2015). Big data: astronomical or genomical? *PLoS biology*, *13*(7), e1002195.

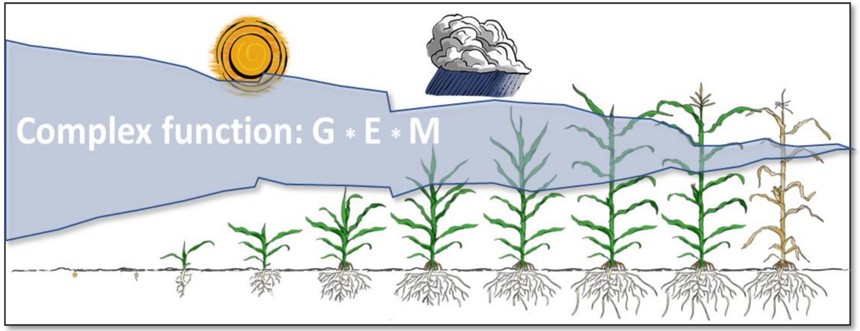
More Env

- UC Merced GridMET
- NWS NOAA
- NASA GISS, NASA power
- Harmonized SoilDB
- USDA SSURGO

More Computing

Alencar.Xavier@Corteva.com
Quantitative Geneticist, Breeding Analyst LAAF

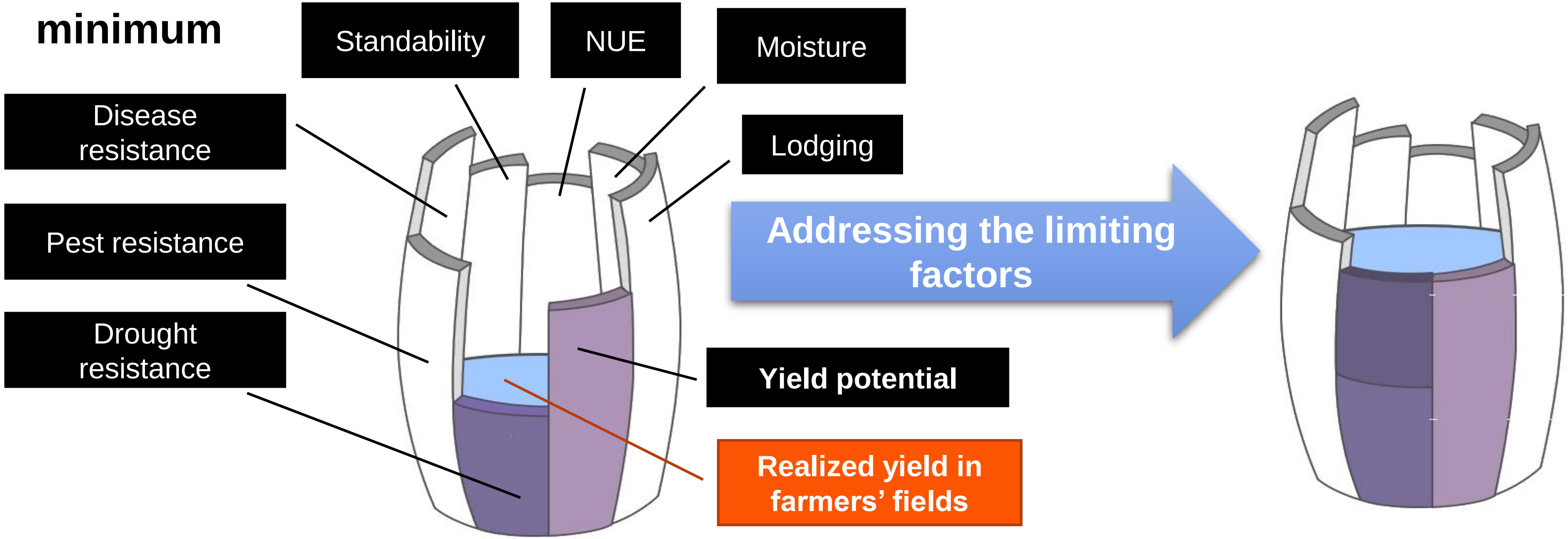
Challenges in Corn Improvement



Source: Totir 2021, ASTA

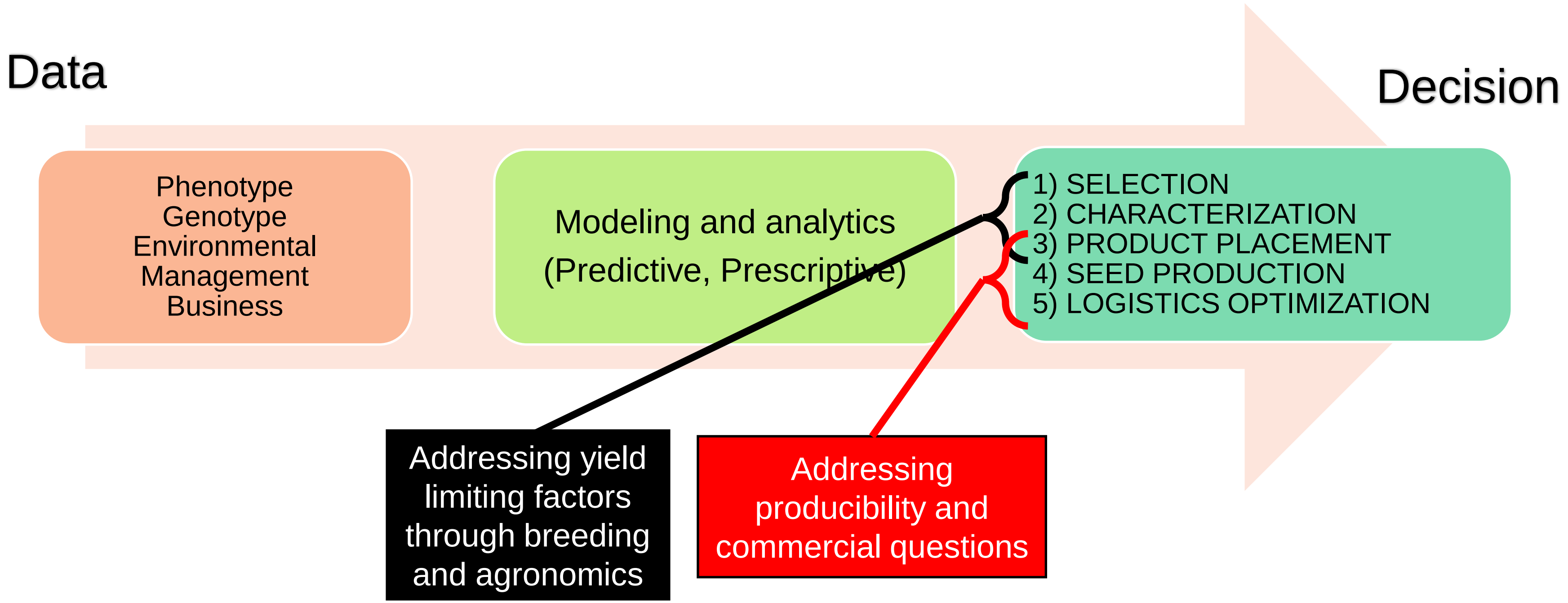
G = Genetics; E = Environment; M = Management

Law of the minimum



CORTEVA agriscience
Alencar.Xavier@Corteva.com
Quantitative Geneticist, Breeding Analyst LAAF

End goal of GxExM characterization



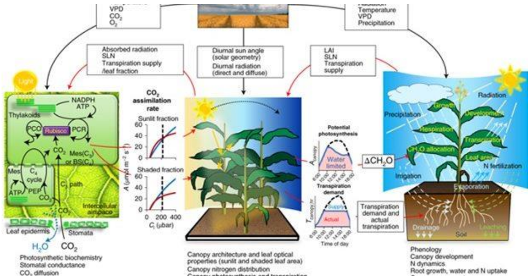
Alencar.Xavier@Corteva.com
Quantitative Geneticist, Breeding Analyst LAAF

Linear models

Crop models

Source: https://evolution.berkeley.edu/teach-resources/genes-environment-phenotype/

https://www.nature.com/articles/s41477-019-0398-8



Linear models

- **Phenotype** ↔ **Secondary phenotypes**
  - **Environment**
    - **Local control**
    - **Management**
    - **Soil & Weather**
  - **Genetics**
    - **Heritable**
    - **Non-heritable**
  - **Interactions**
    - **Residuals**
    - **GxE, GxM, ExM, GxExM**

General framework, trait-crop agnostic

Biological knowledge, trait-crop specificity

Crop models

- **Phenotype**
  - **Local control**
  - **Gen | Env**
    - **Underlying physiological traits**
      - **Heritable genetics**
      - **Per se genetics**
    - **Growth conditions**
      - **Management**
      - **Soil & Weather**
  - **Residuals**

Walking through the statistical black boxes of plant breeding

Alencar Xavier[1] · William M. Muir[2] · Bruce Craig[3] · Katy Martin Rainey[1]

Integrating Crop Growth Models with Whole Genome Prediction through Approximate Bayesian Computation

Frank Technow[1]*, Carlos D. Messina[2], L. Radu Totir[1], Mark Cooper[2]

Alencar.Xavier@Corteva.com
Quantitative Geneticist, Breeding Analyst LAAF

Linearly correlated phenotypes

Using machine learning to infer connections from data

$$\text{Phenotype (y)} = \text{Gen|Trials(g)} + \text{Plot level noise(e)}$$

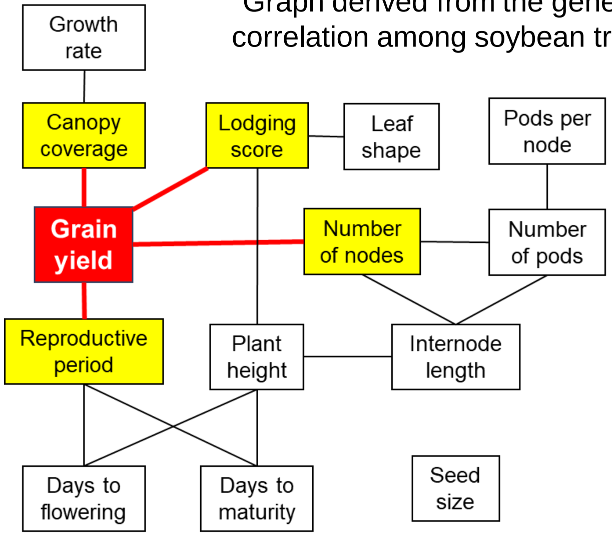
$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

$$\text{Variance} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \sigma^2_{g_1} & \sigma_{g_{12}} \\ \sigma_{g_{12}} & \sigma^2_{g_2} \end{bmatrix} + \begin{bmatrix} \sigma^2_{e_1} & \sigma_{e_{12}} \\ \sigma_{e_{12}} & \sigma^2_{e_2} \end{bmatrix}$$

CORRELATED INFORMATION

Phenotypic network

Graph derived from the genetic correlation among soybean traits



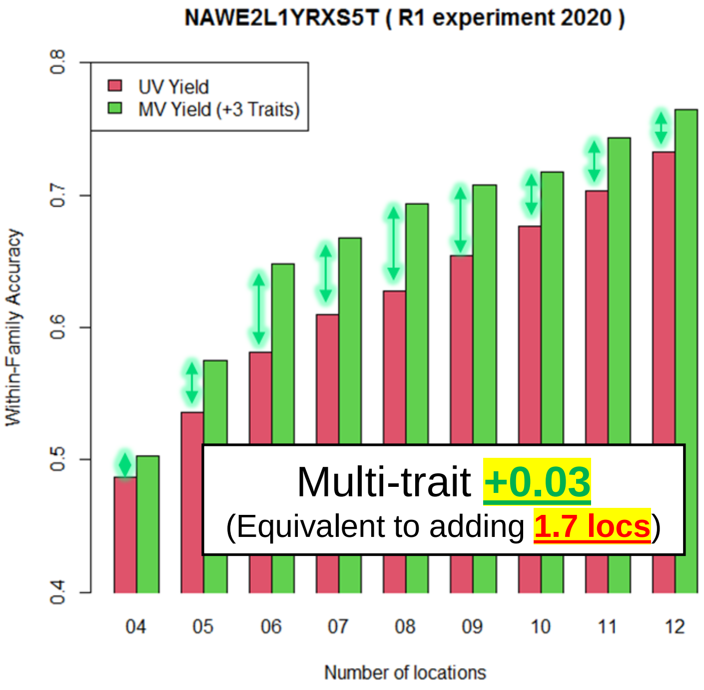
A new approach fits multivariate genomic prediction models efficiently

Alencar Xavier[1,2*†] and David Habier[1*†]

Using unsupervised learning techniques to assess interactions among complex traits in soybeans

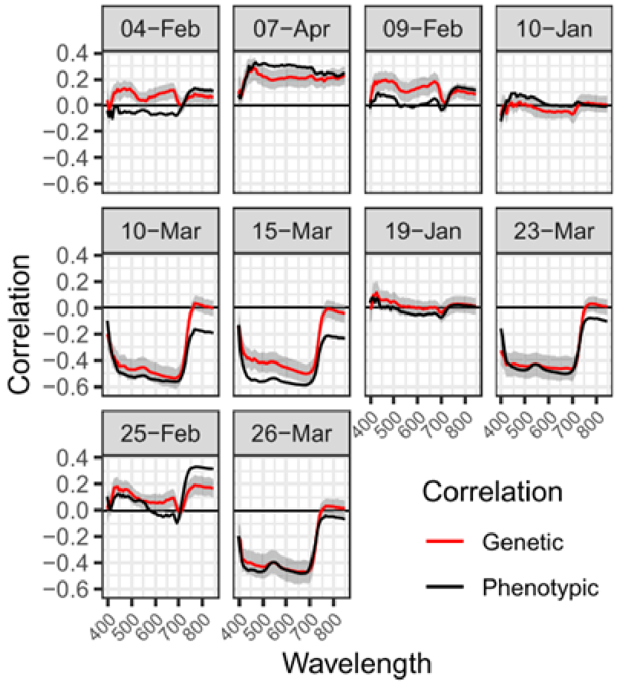
Alencar Xavier · Benjamin Hall · Shaun Casteel · William Muir · Katy Martin Rainey

Leveraging information from secondary traits



High-throughput phenotyping

Correlation(HTP, main trait) with ***thousands of traits***



Runcie et al. (2021) https://doi.org/10.1186/s13059-021-02416-w

CORTEVA agriscience
Alencar.Xavier@Corteva.com
Quantitative Geneticist, Breeding Analyst LAAF

"Breeding objective"

$f\,(\text{market segment, farming systems})$

- Set of traits of interest (**TOI**) bred into a

WHAT

Yield, moisture, relative maturity, disease resistance, stability, **trait package, producibility**

- Target population of genetics (**TPG**) for a given

WHO

Corn 111-121, corn 122-130
white corn 118-123

- Target population of environments (**TPE**) and management (**TPM**) practices

WHERE

HOW, WHEN

Drought, irrigation, early planting, varying levels of disease pressure, different soil types

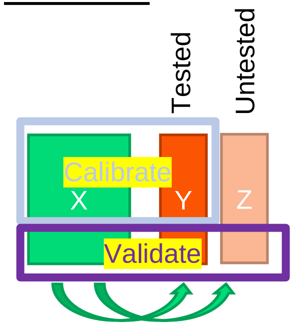
$$\rho_{\text{GxExM}} = \rho_{\text{TPG}} \times \rho_{\text{TPG}} \times \rho_{\text{TPM}}$$

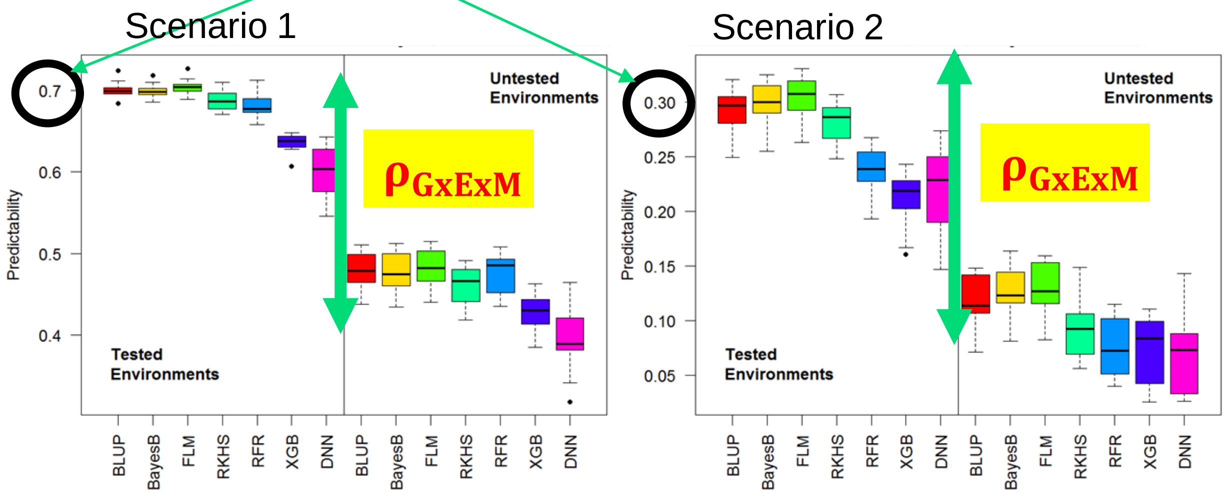
Model testing and validation schemes

$$\text{Prediction accuracy} \propto \sqrt{H^2} \times \rho_{GxExM}$$

Upper limit is determinate by the application $\cong \sqrt{H^2}$

CV scheme





Different prediction methods

Different prediction methods

Technical nuances of machine learning: implementation and validation of supervised methods for genomic prediction in plant breeding

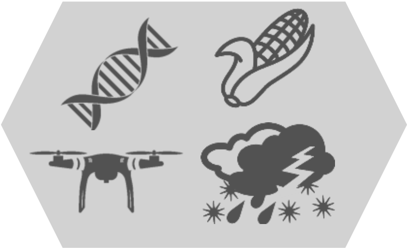
Alencar Xavier [1*]

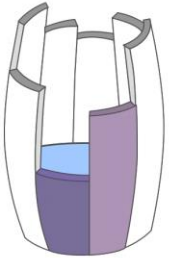
Alencar.Xavier@Corteva.com
Quantitative Geneticist, Breeding Analyst LAAF

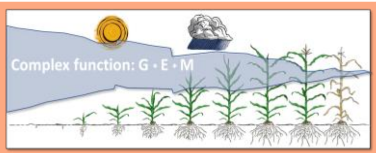
With <u>more data and better analytics</u>, breeding can <u>respond faster</u> the new <u>farming challenges and trends</u>

- New & better management
- Changing environment
- New pests and diseases

Predictive breeding
(TPE, TPG, TPM)

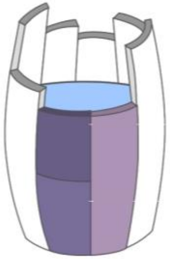






Crop models, linear models, machine learning

$$\text{Acc} \propto \mathbf{GxExM} \times \mathbf{h}^2$$



Performance

Producibility

Robustness

Alencar.Xavier@Corteva.com
Quantitative Geneticist, Breeding Analyst LAAF

Thank you for your attention!

Final remarks:

1) Plant breeding is changing with new data and modern analytics
2) Harnessing GxExM information benefits decision making, business and farmers
3) Modeling is contingent to the target genetics, environments and management

Questions??

Alencar Xavier

Alencar.Xavier@Corteva.com