



**10 Y 11 DE MAYO 2023**  
METROPOLITANO, ROSARIO, ARGENTINA

Nutrición en la Agricultura Digital: Hacia una  
**Agronomía + Cuantitativa:**  
¿Cómo analizamos los datos?

**ADRIAN CORRENDO**

Carlos Hernandez  
Josefina Lacasa, &  
Ignacio Ciampitti



Ciampitti Lab

**KANSAS STATE**  
**UNIVERSITY**



# Los datos, ¿son poder?

...Hay que usarlos

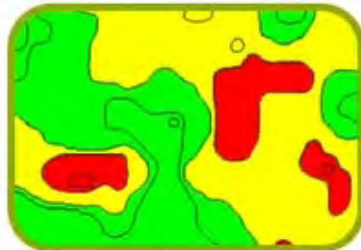
## #1 ¿Qué hacemos con los datos?



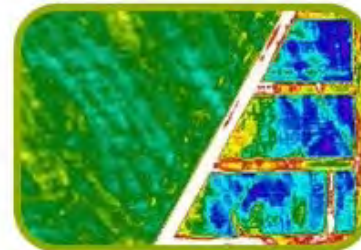
**Agronómicos**



**Maquinaria**



**Prescripción**



**Sensores**



**Producción**

*Fuente: J. Fulton et al. (2019), Data Literacy*

# DESAFIOS

## 1 - EDUCACIÓN

Curricula limitada en “ciencia de los datos aplicada”



## 2 - REPRODUCIBILIDAD

Aun limitada capacidad de repetir analisis y resultados

*NO se comparten las DATOS, los codigos TAMPOCO*



## 3 - ACCESSIBILIDAD

Todavía NO somos capaces de traducir, lo suficiente, la ciencia en herramientas de decision



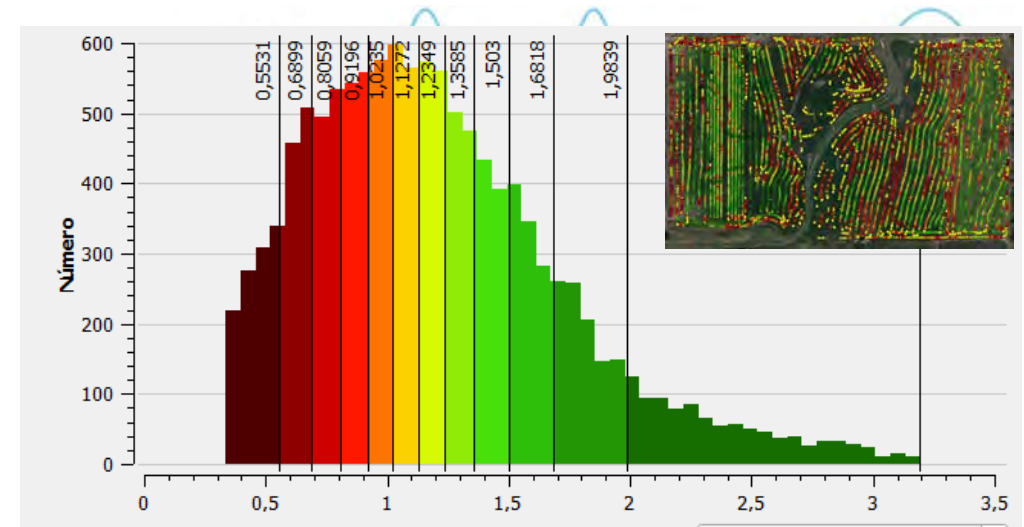
# ALGUNOS MITOS

Y VICIOS que arrastramos

## 1 - DISTRIBUCION NORMAL

Lo requería el ANOVA tradicional

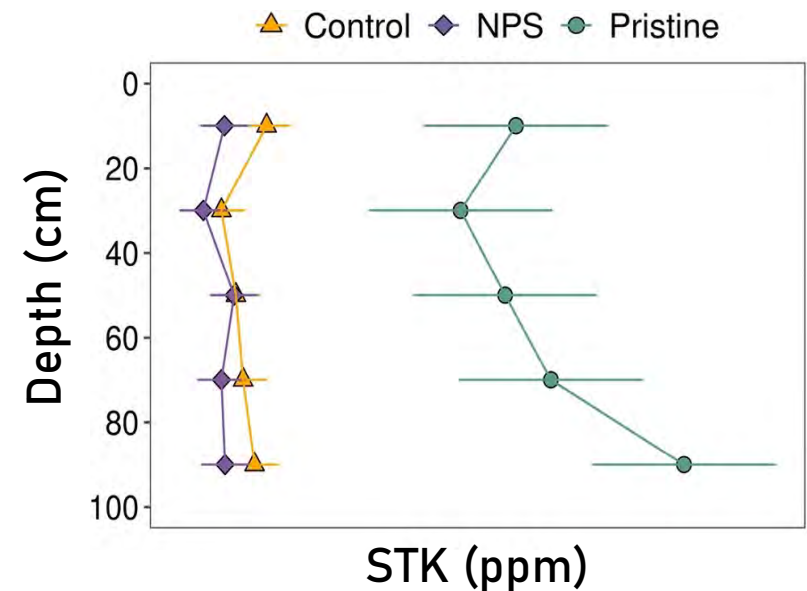
Muchas +opciones (modelos generalizados)



## 2 - HOMOGENEIDAD DE VARIANZA

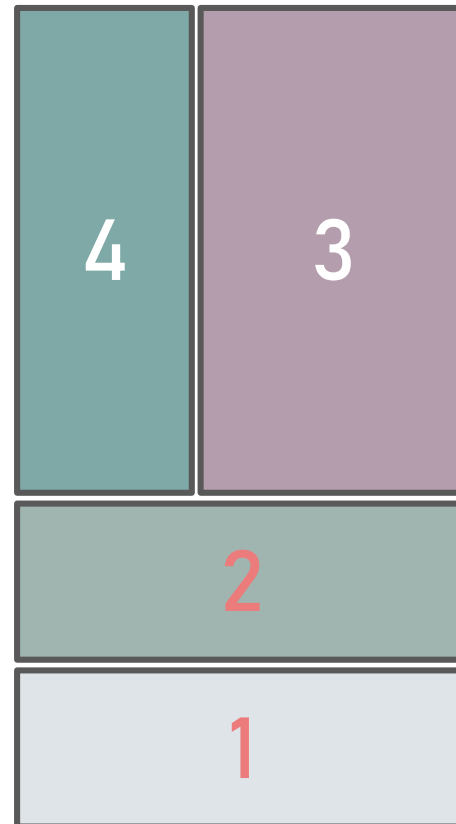
Lo requería el ANOVA tradicional

Podemos especificar varianza heterogenea en los modelos



# DISEÑO EN BLOQUES

*Lacasa et al. (2023) MS Tesis*



**MALA ESPECIFICACIÓN?**

ALEATORIZACIÓN inadecuada

**Bloques INCOMPLETOS ?**

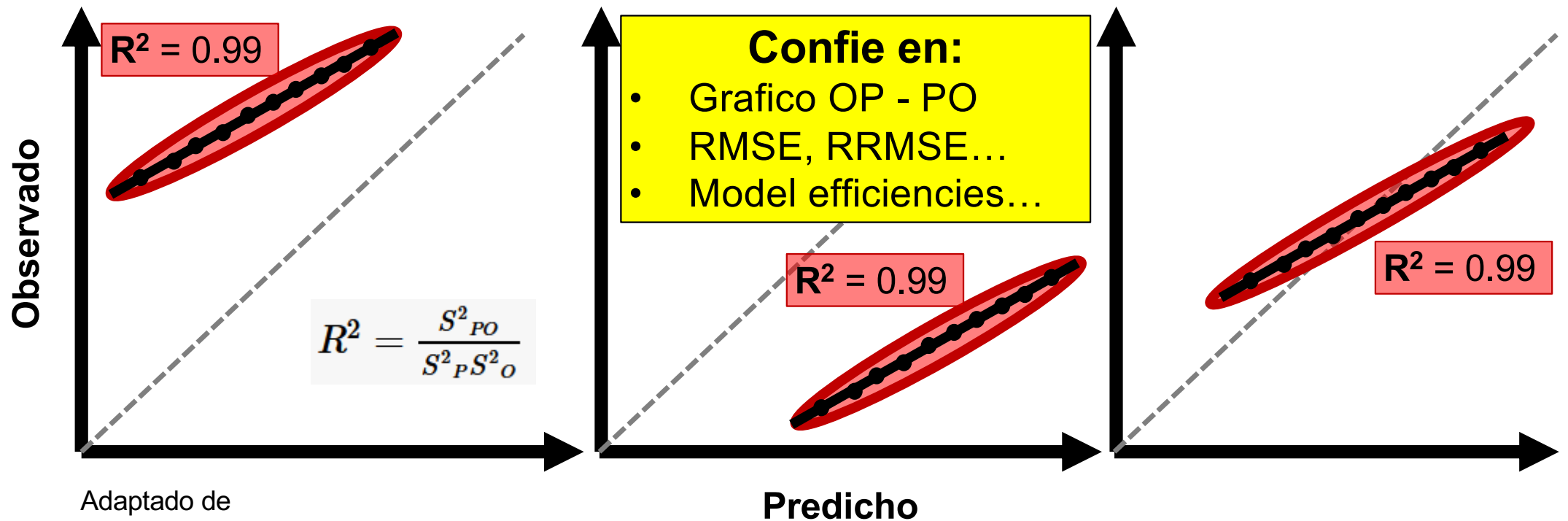
Analisis espacial

Estadística Bayesiana +  
Machine Learning (TGP)

# OBSESIÓN POR EL R<sup>2</sup>

PARA EL CASO PREDICHO-OBSERVADO

**NO DICE NADA** DE LA EXACTITUD (accuracy)



Adaptado de  
Lin et al. (1989), Correndo et al. (2021, 2022)

# OBSESIÓN POR EL **NUMERO** EXACTO



**VALOR VERDADERO**

**CURVA VERDADERA**

**NIVEL CRITICO**

**DOSIS OPTIMA**



**PROBABILIDAD**

**INCERTIDUMBRE**

**CURVA + PROBABLE**

**RANGOS CRITICOS**

**P-value  $\leq 0.05$**



**ENFOQUE EN EL  
ERROR**

# Que se discute en



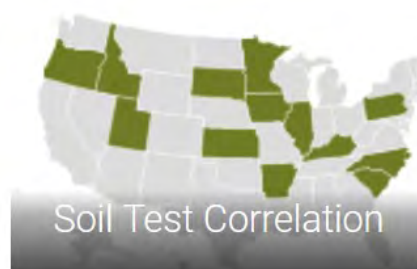
# ?

**+100 especialistas**

- + Transparencia en diagnóstico con análisis de suelo
- Interpretación clara y consistente
- Base de datos → Decision Support Tool: P, K, pH



<https://soiltestfrst.org/>





# Preguntas principales

→ Rendimiento **Relativo**... a qué?



→ RR objetivo? 90%, 95%, 99%...variable?

→ ¿**Umbr**ales criticos o **Rangos** criticos?

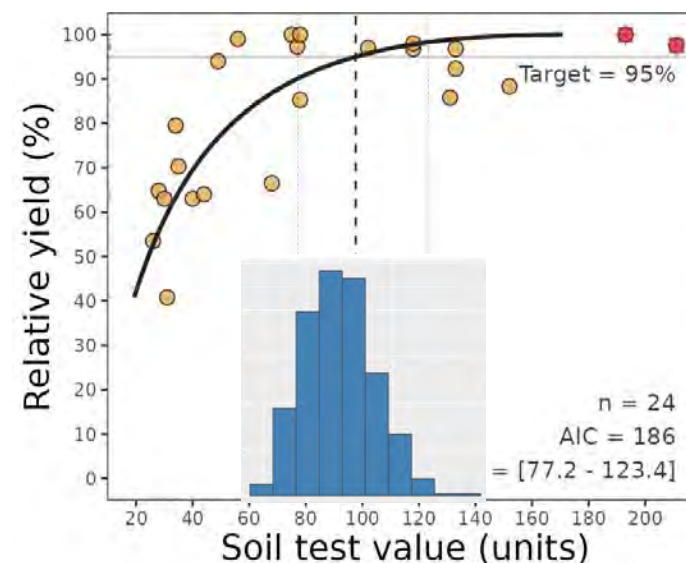
→ **Incertidumbre** (Bootstrapping -remuestreo-)

→ ¿**Probabilidad** de respuesta ?

→ ¿Qué **modelo** utilizar?

→ ALCC; Cate-Nelson; Regresiones → Se pueden combinar?

**Maximo Rendimiento** (Pearce et al., 2022)

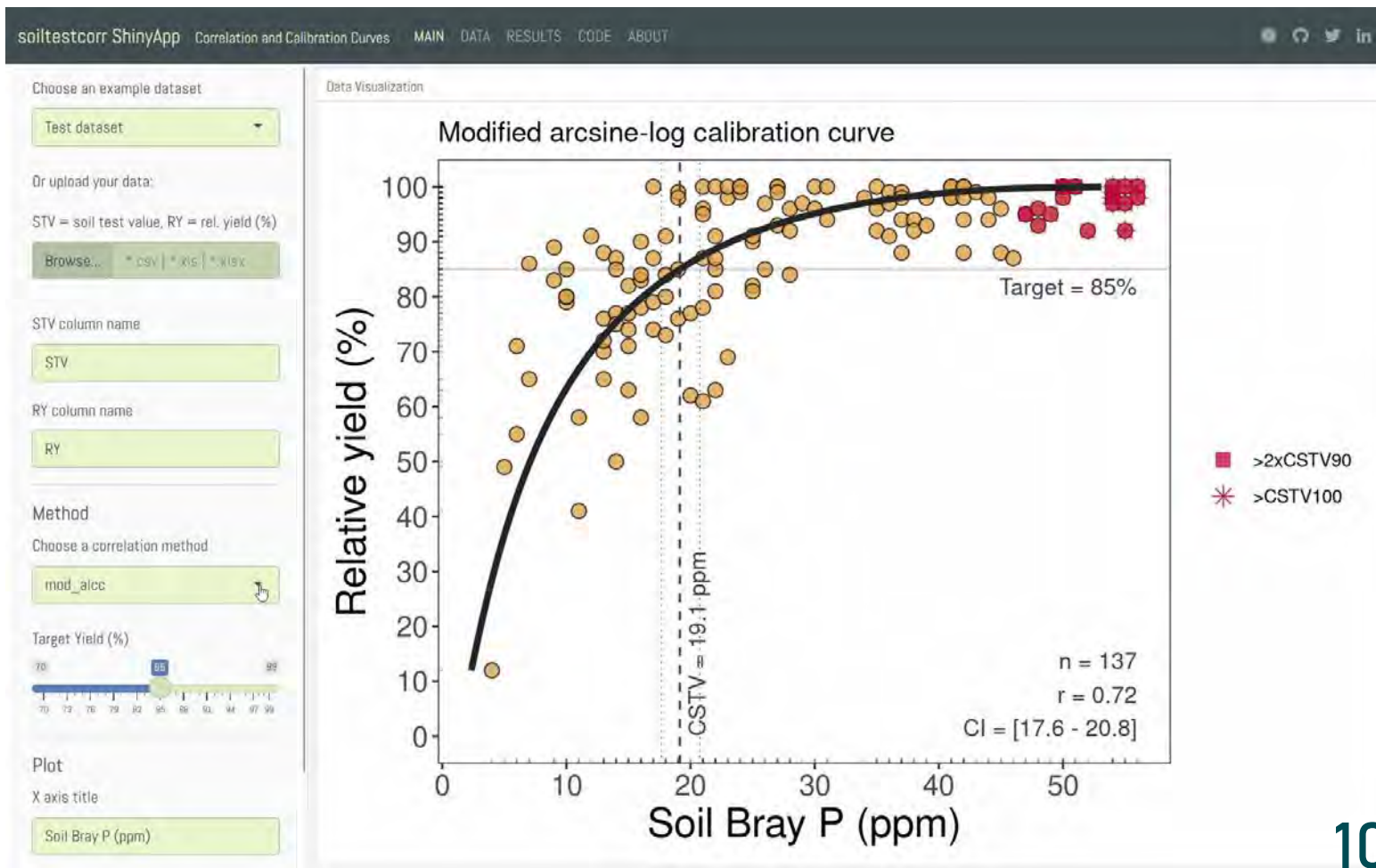


# Software + ShinyApp

Correndo et al. (2023) SoftwareX



[bit.ly/3t4vmtf](https://bit.ly/3t4vmtf)



# DON-Maiz

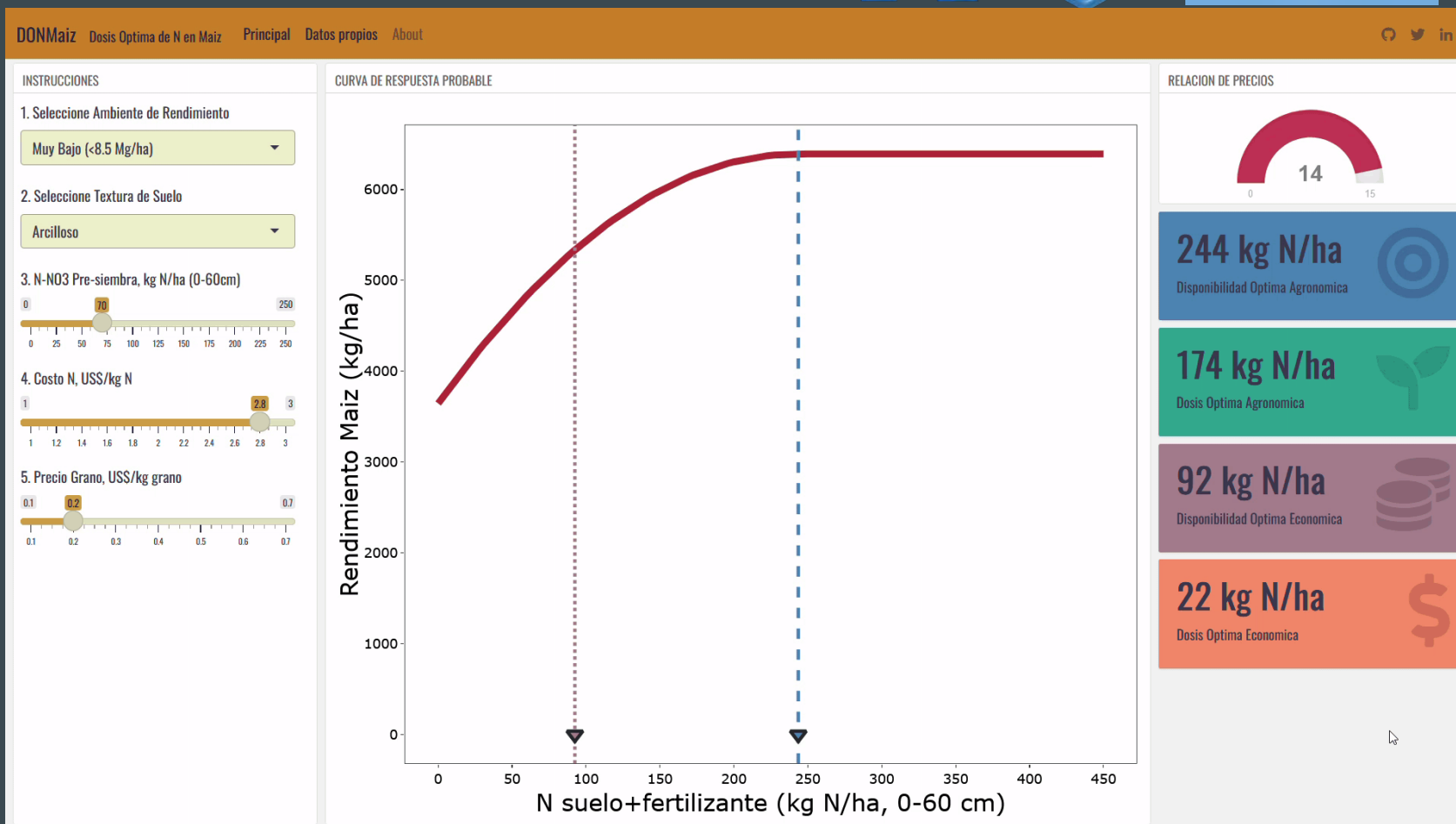


Explore  
Resultados

A su ritmo



[bit.ly/3TZCrHc](https://bit.ly/3TZCrHc)



# EXPLICAR o PREDECIR?

## MODELOS EXPLICATIVOS

*Objetivo: “comprender” lo que ya pasó.*

Estudiar un sistema a través de “componentes”

Limitados ante interacciones complejas

**Mecanísticos** → PERFECTO!

Machine Learning → puede ser! (Arboles)

NO requiere ENTRENAR - VALIDAR





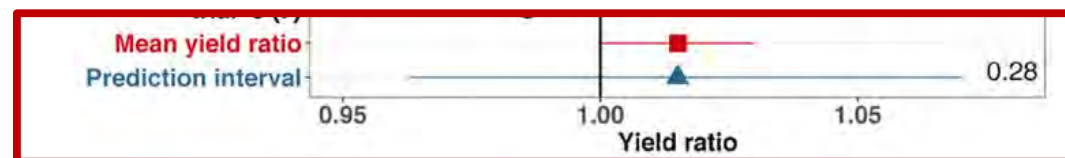
European Journal of Agronomy

Volume 120, October 2020, 126127



Going beyond mean effect size:  
Presenting prediction intervals for on-farm network trial analyses

[Anabelle Laurent](#)<sup>a</sup>  , [Fernando Miguez](#)<sup>a</sup>, [Peter Kyveryga](#)<sup>b</sup>,  
[David Makowski](#)<sup>c</sup>



# EXPLICAR o PREDECIR?

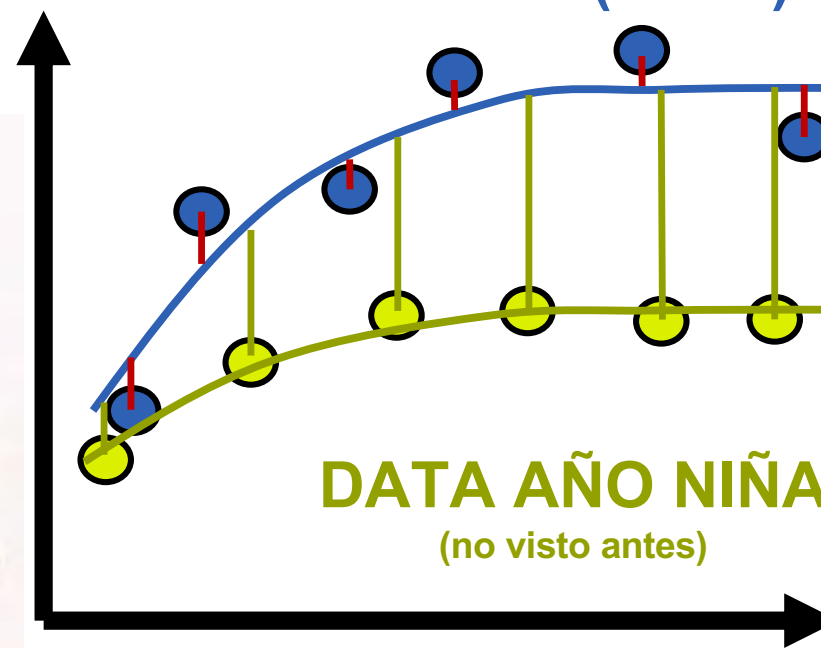
## MODELOS PREDICTIVOS

MAMÁ DICE QUE LA VIDA ES COMO UNA  
CAJA DE BOMBONES, NUNCA SABES  
EL QUE TE VA A TOCAR.

FORREST GUMP



EXPLICATIVO (NIÑO)



Modelo

Combinado

Que reconozca  
el tipo de año

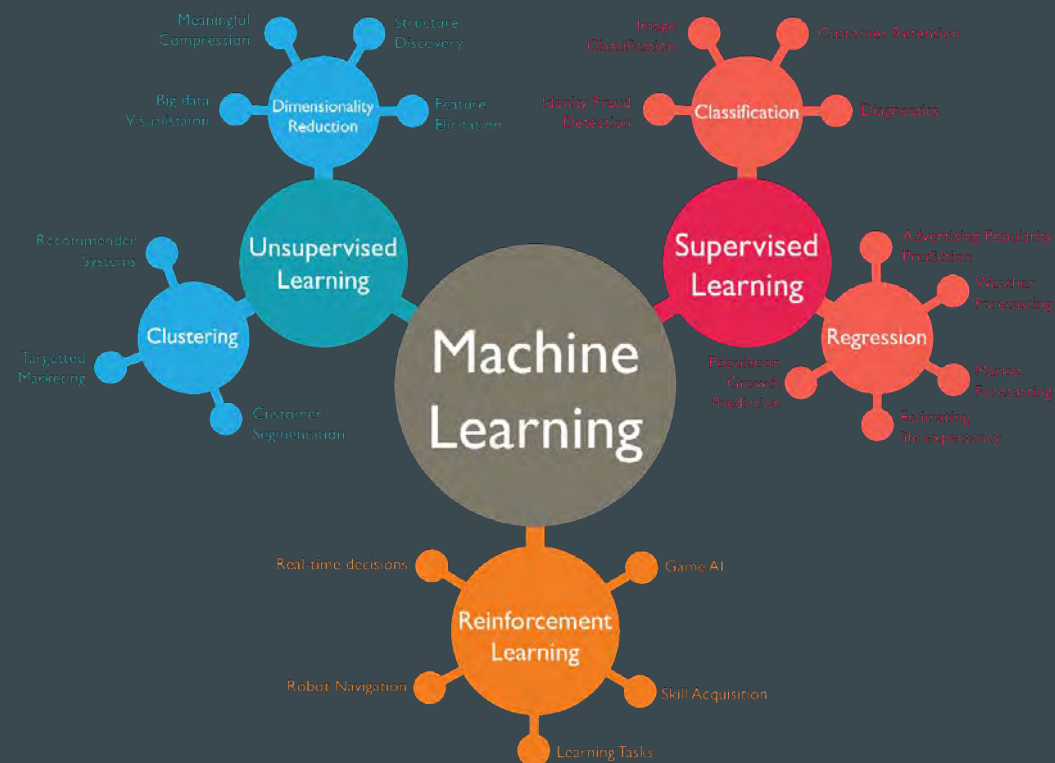
# MACHINE LEARNING

AI

## Centrado en “PREDECIR”

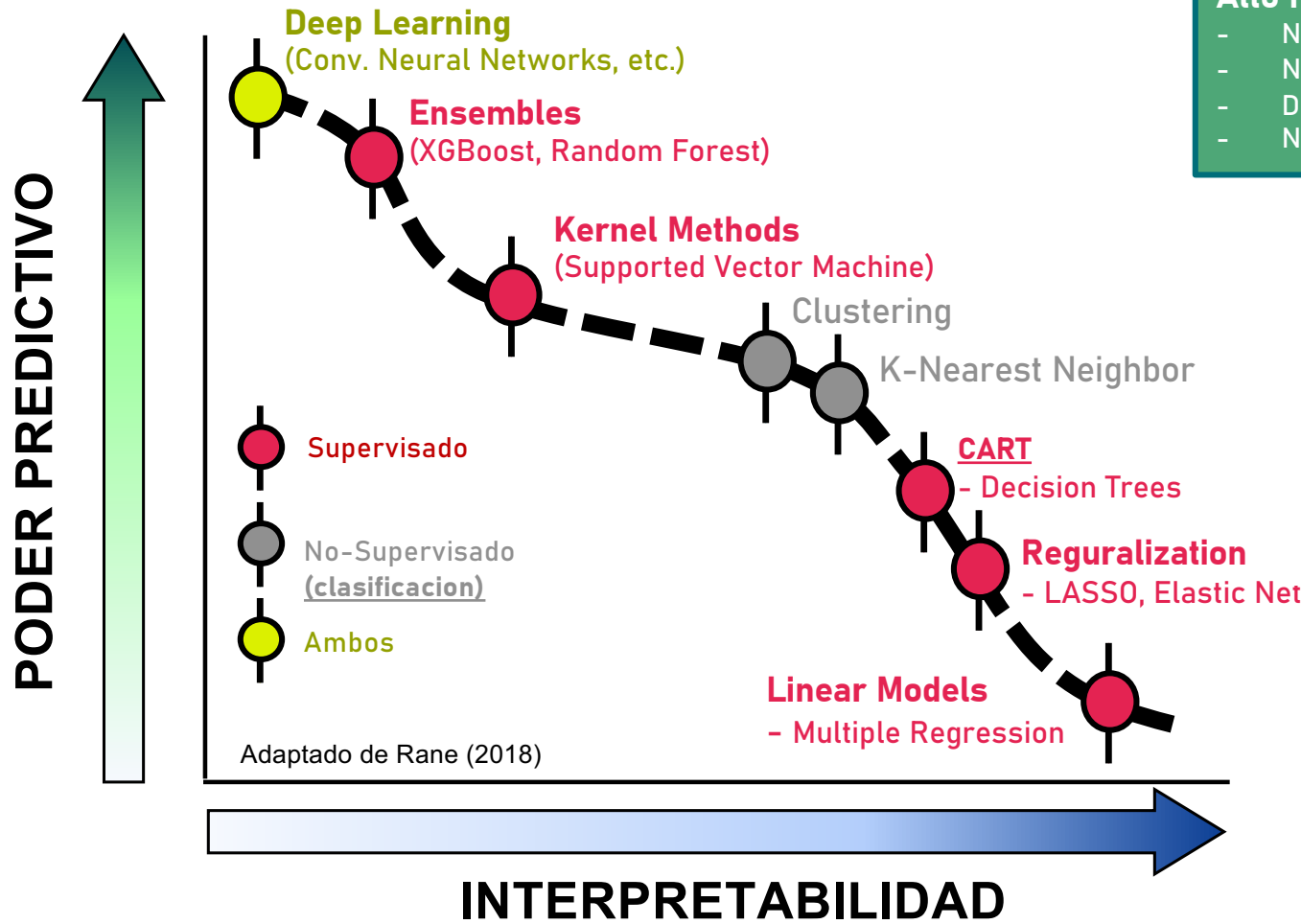
- Son modelos **MULTIVARIABLES**
  - Explorar interacciones complejas
    - No SON MAGICOS!
  - Requieren:
    - **Mucha data** (de calidad)
    - **Variabilidad** (para aprender)
- El **ALGORITMO**...
  - Ingeniería & Selección de variables
- La **VALIDACIÓN** es crucial!
  - Training – Testing (o validation)

## MACHINE LEARNING



# ¿QUE OBJETIVO PERSIGO?

# ¿NECESITO ENTENDER?



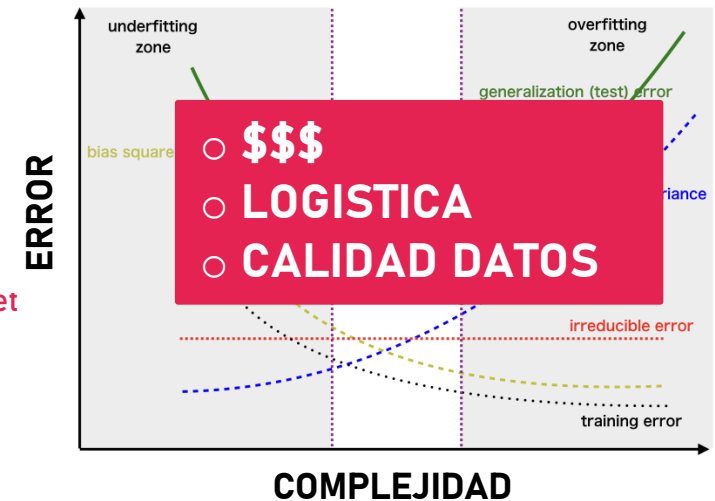
## Alto Poder Predictivo

- No lineales
- No estacionarios
- Dificil de computar
- Necesitan +datos

## Alta Interpretabilidad

- Lineales
- Estacionarios
- Facil de computar
- Limitada prediccion

# ¿CUAN COMPLEJO?

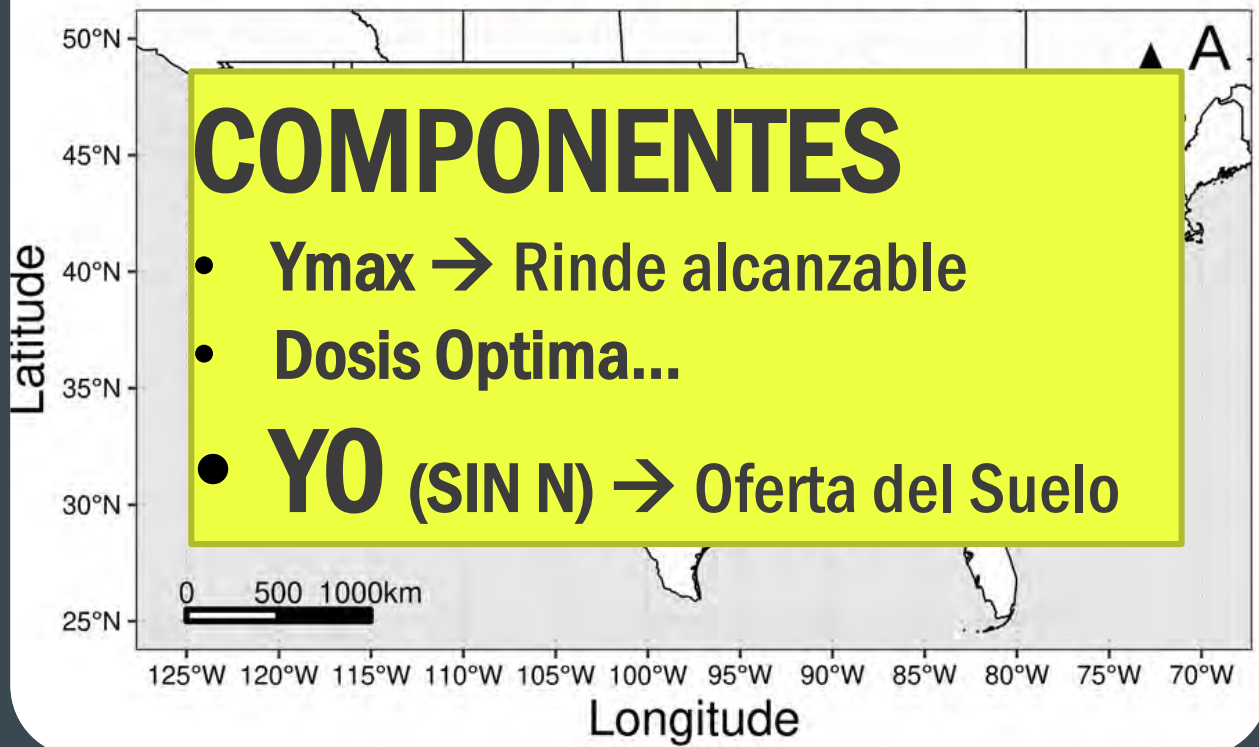


# MAIZ – Rendimiento y N

Correndo et al. (2021a, 2021b)



$Y_0(\text{Mg ha}^{-1})$  0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18



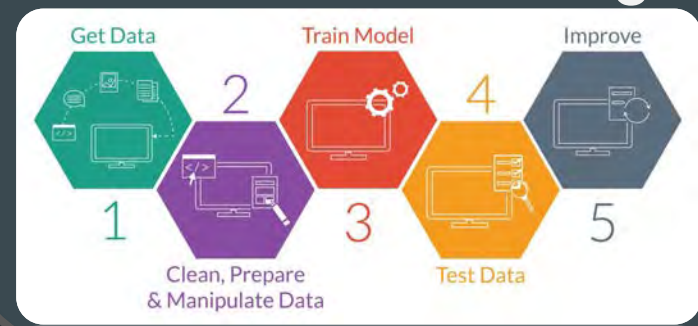
## COMPONENTES

- $Y_{max}$  → Rinde alcanzable
- Dosis Optima...
- $Y_0$  (SIN N) → Oferta del Suelo

# +1000 ensayos



## Machine Learning



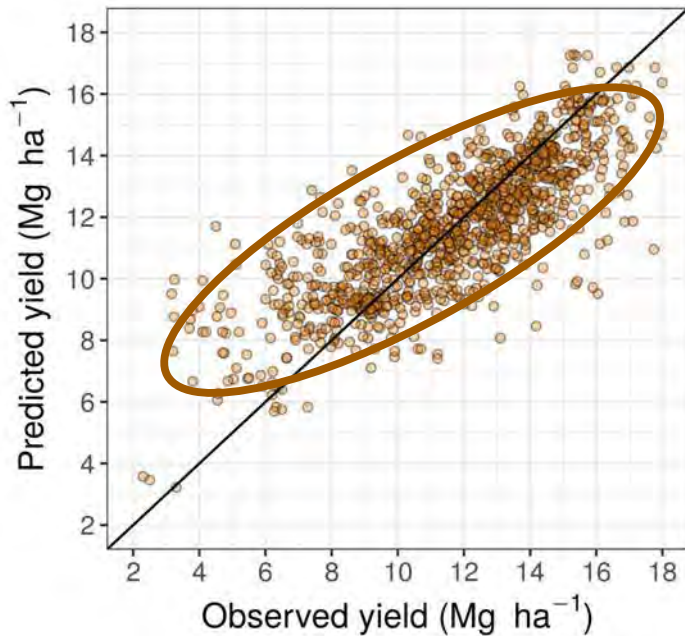
This Photo by Unknown Author is licensed under [CC BY-SA-NC](#)



# ALGORITMO

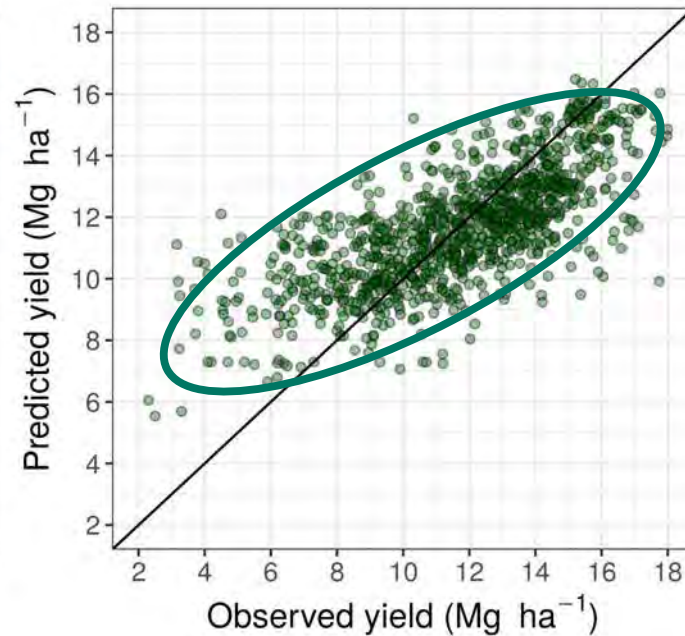
## PERFORMANCE Ymax

RMSE = 1.91 Mg ha<sup>-1</sup>



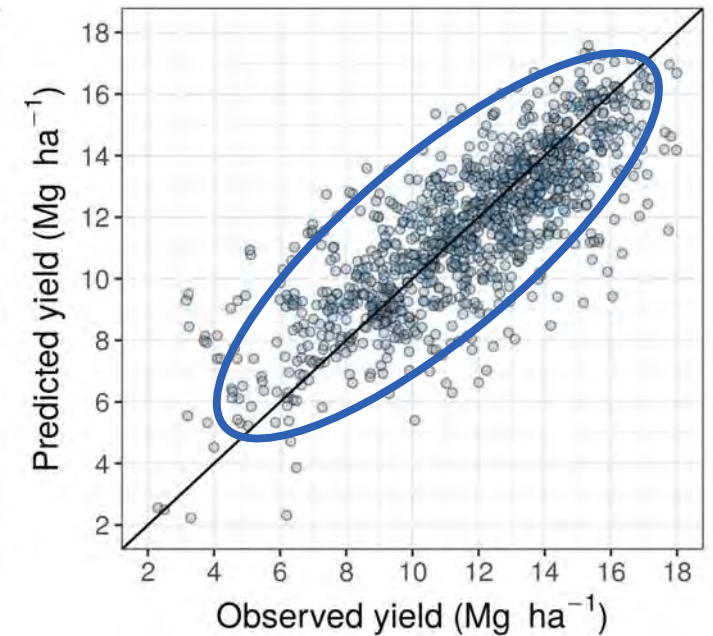
**randomforest()**

RMSE = 2.11 Mg ha<sup>-1</sup>



**cforest()**

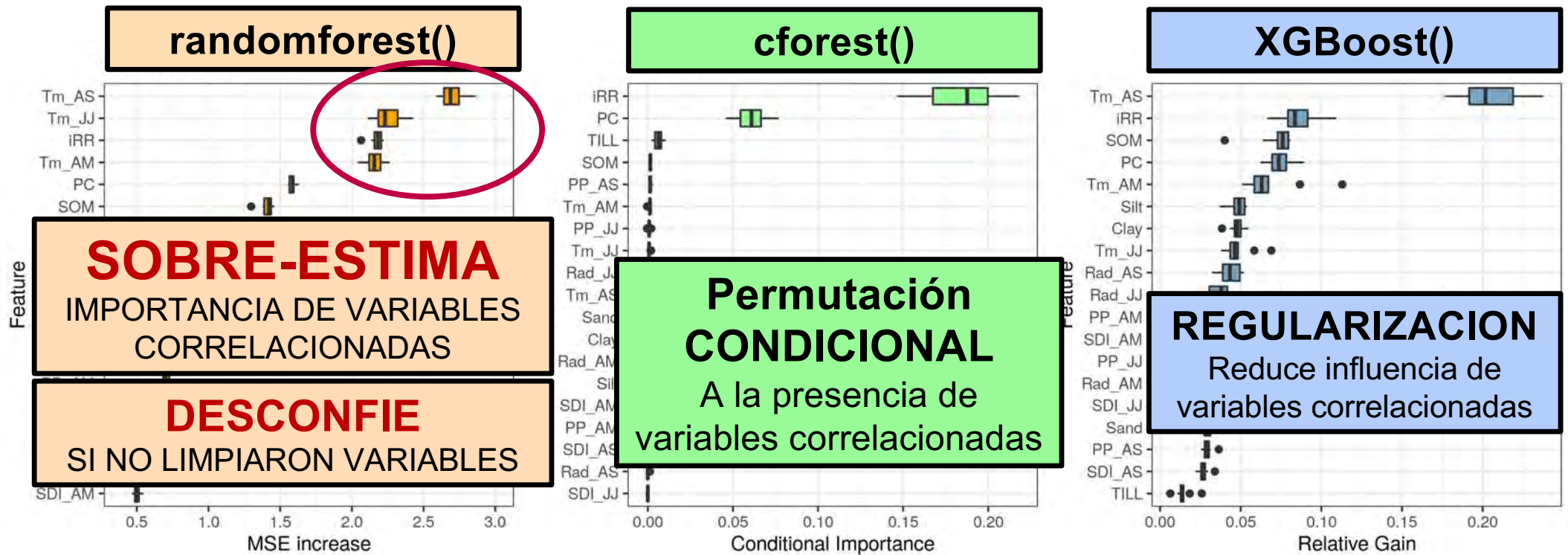
RMSE = 1.79 Mg ha<sup>-1</sup>



**XGBoost()**

# ALGORITMO

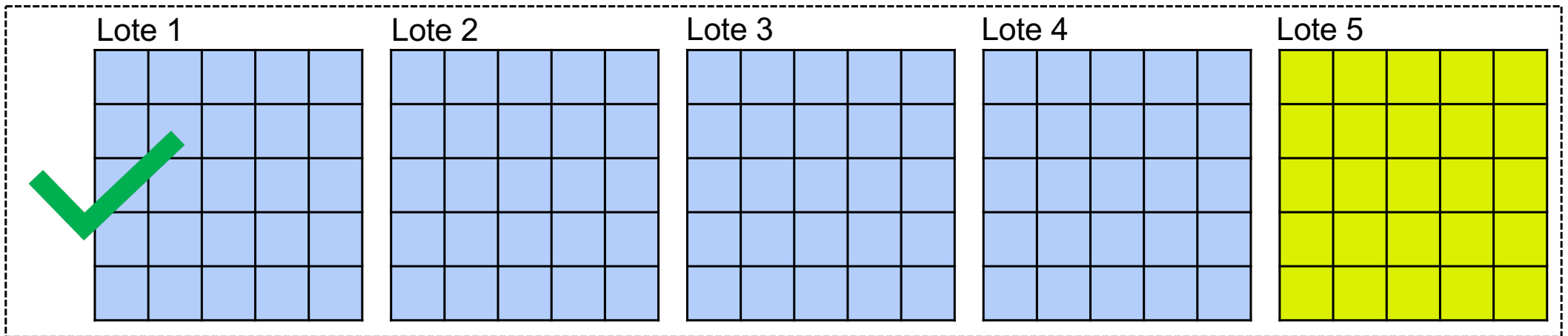
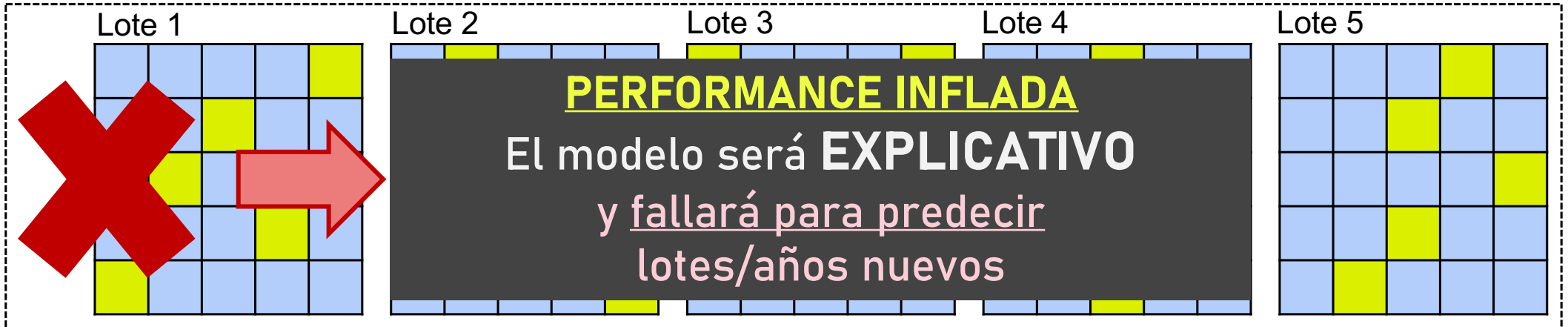
## IMPORTANCIA VARIABLES



Strobl et al. (2007)

<https://doi.org/10.1186/1471-2105-8-25>

# VALIDACIÓN → Usar Leave One Out (LOO)



# VALIDACIÓN

## DATOS

5 lotes/año con franjas de dosis de N para predecir rendimiento en maíz (suelo, NDVI, zona, etc.)

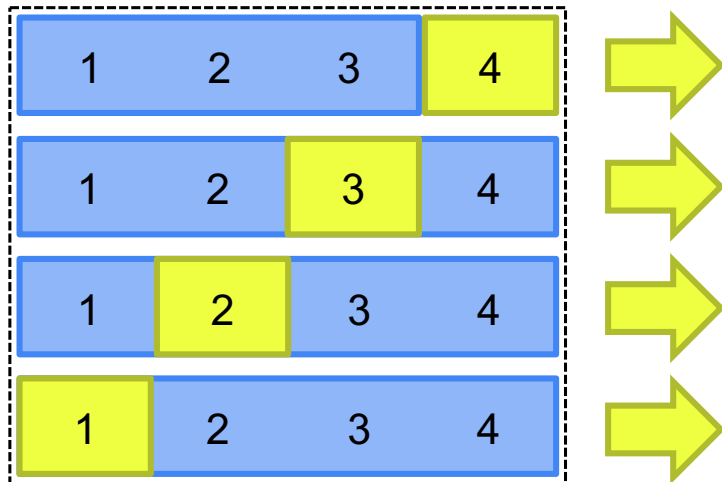
### EXTERNO, split 80:20



### ERROR DE GENERALIZACIÓN

Como predice lotes o años "no vistos"

### INTERNO, $k = 4$



### OPTIMIZACIÓN

Entrenamiento para identificar el mejor modelo:

#### Algoritmos

Técnica de "aprendizaje", por ej.:

- *Random forest* → Bosques Aleatorios CART
- *Gradient Boosting* → Bosques Secuenciales CART de residuales

#### Hiper-parametros

Valores de "configuración" (no vienen de los datos):

- *n<sub>tree</sub>* → Numero de "árboles" en el "random forest"
- *eta* → Tasa de aprendizaje en "XGBoost"
- *lambda* → regularización en "Elastic Net" o "LASSO"



# Modelo Bayesiano (Predictivo)

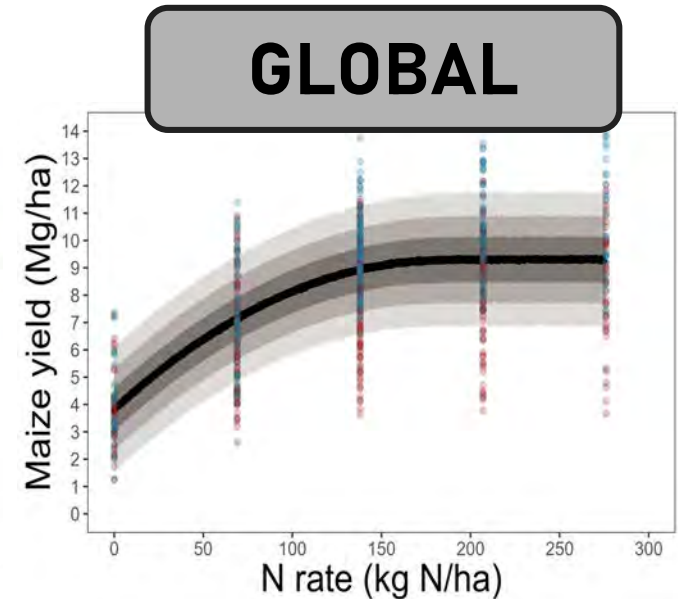
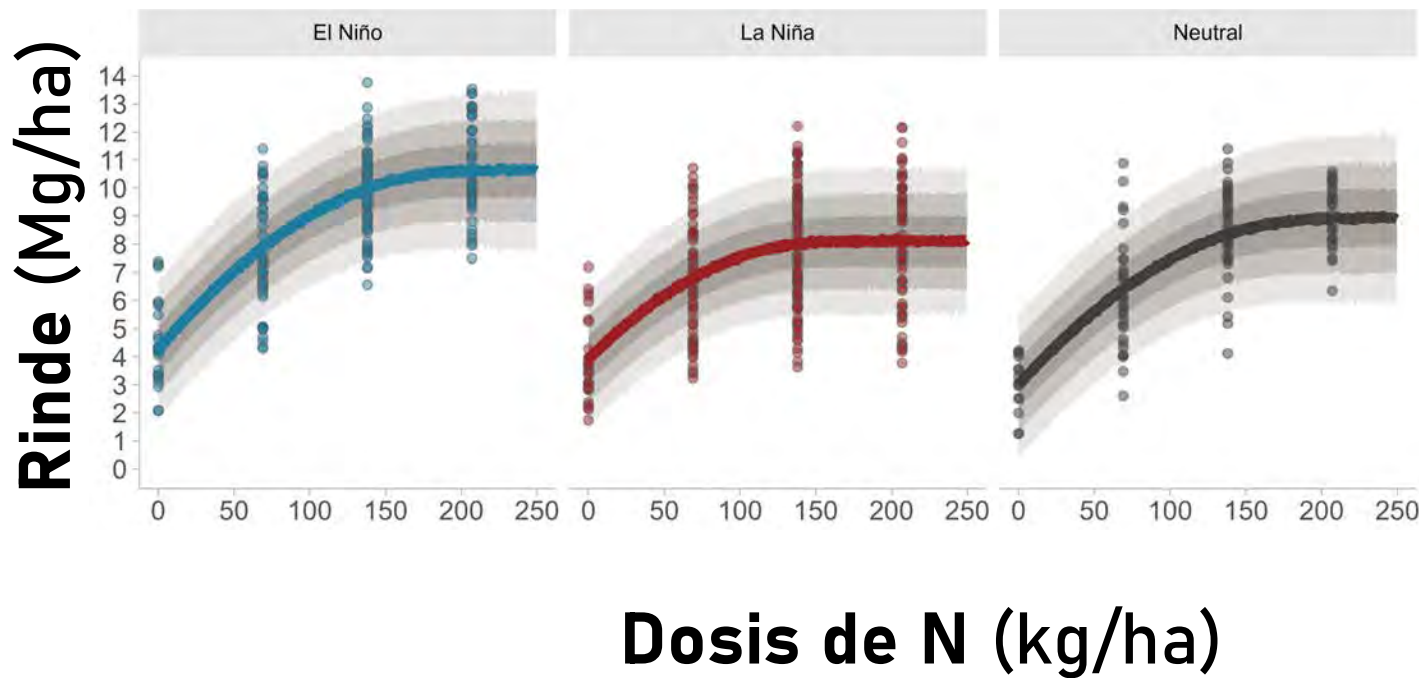
Ensayos Maiz Largo Plazo (1995-2022)

INTA Parana, Argentina

Modelo “-n”      Año “n”

Training      LOO

## Respuesta a N x ENSO



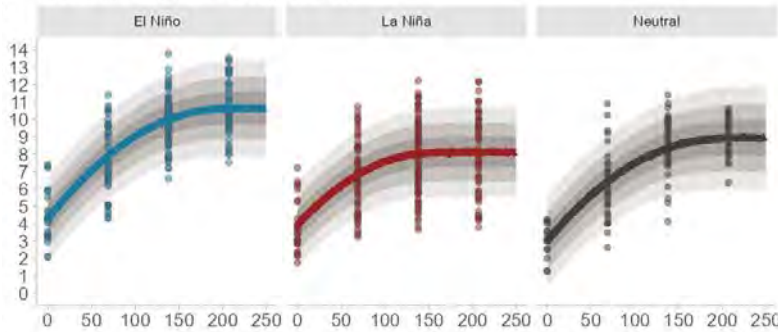
# RESPUESTA N MAIZ x ENSO



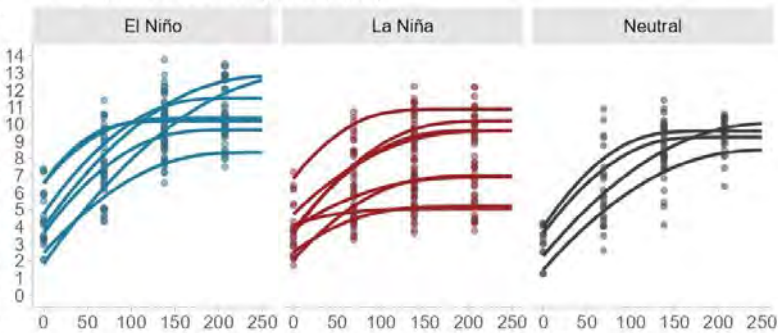
● El Niño    ● La Niña    ● Neutral    ● ENSO    ▲ Global

Rinde (Mg/ha)

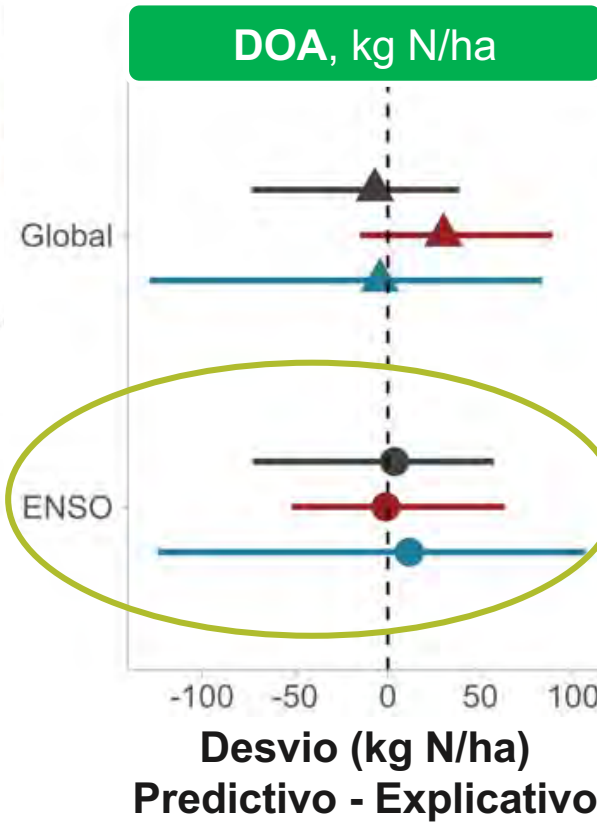
PREDICTIVO x ENSO (modelo)



EXPLICATIVOS x Año (Referencia)



Dosis de N (kg N/ha)



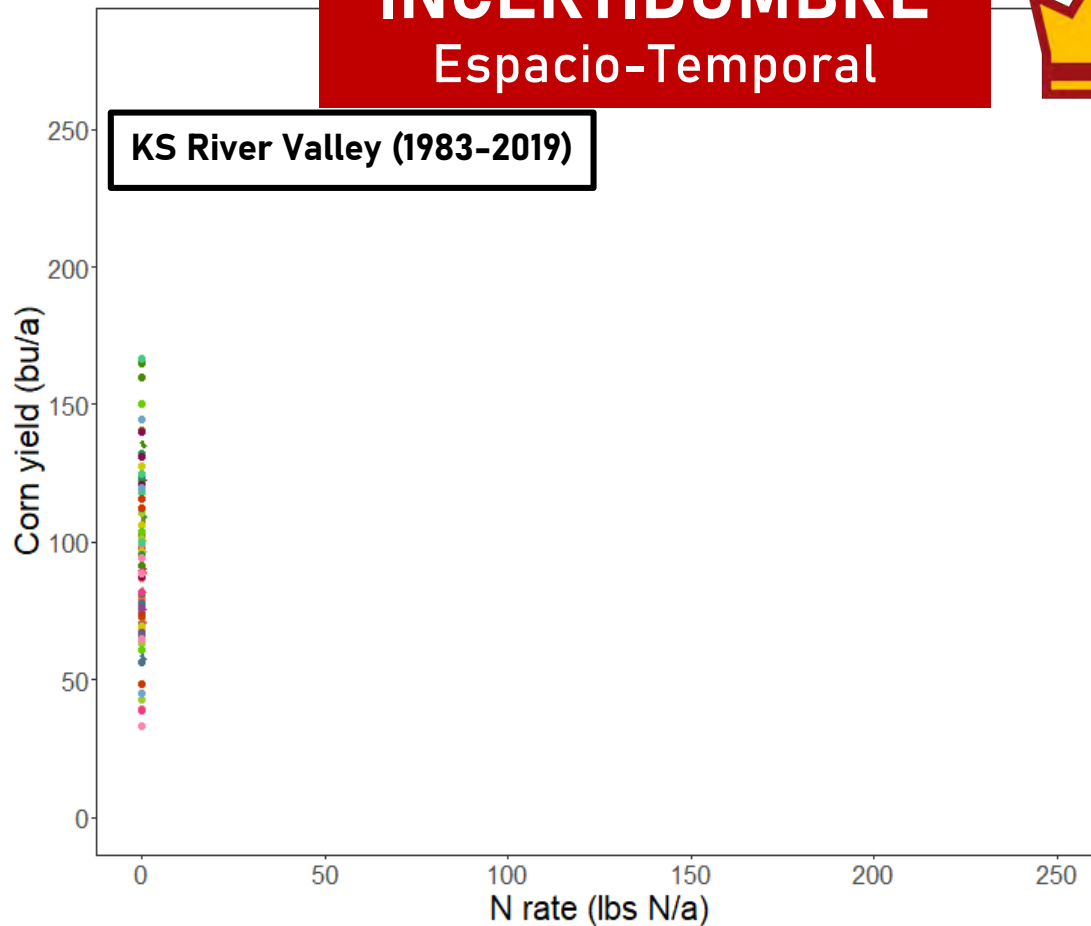
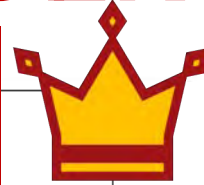
**Modelo x ENSO  
Reduce Sesgo!**

**ENSO  
cambia el Error**

# EL MODELO FRECUENTISTA



## INCERTIDUMBRE Espacio-Temporal



(i) **Curva verdadera** (desconocida)

(ii) Datos

### Cuadratico

$$y = B_0 + B_1 x + B_2 x^2$$

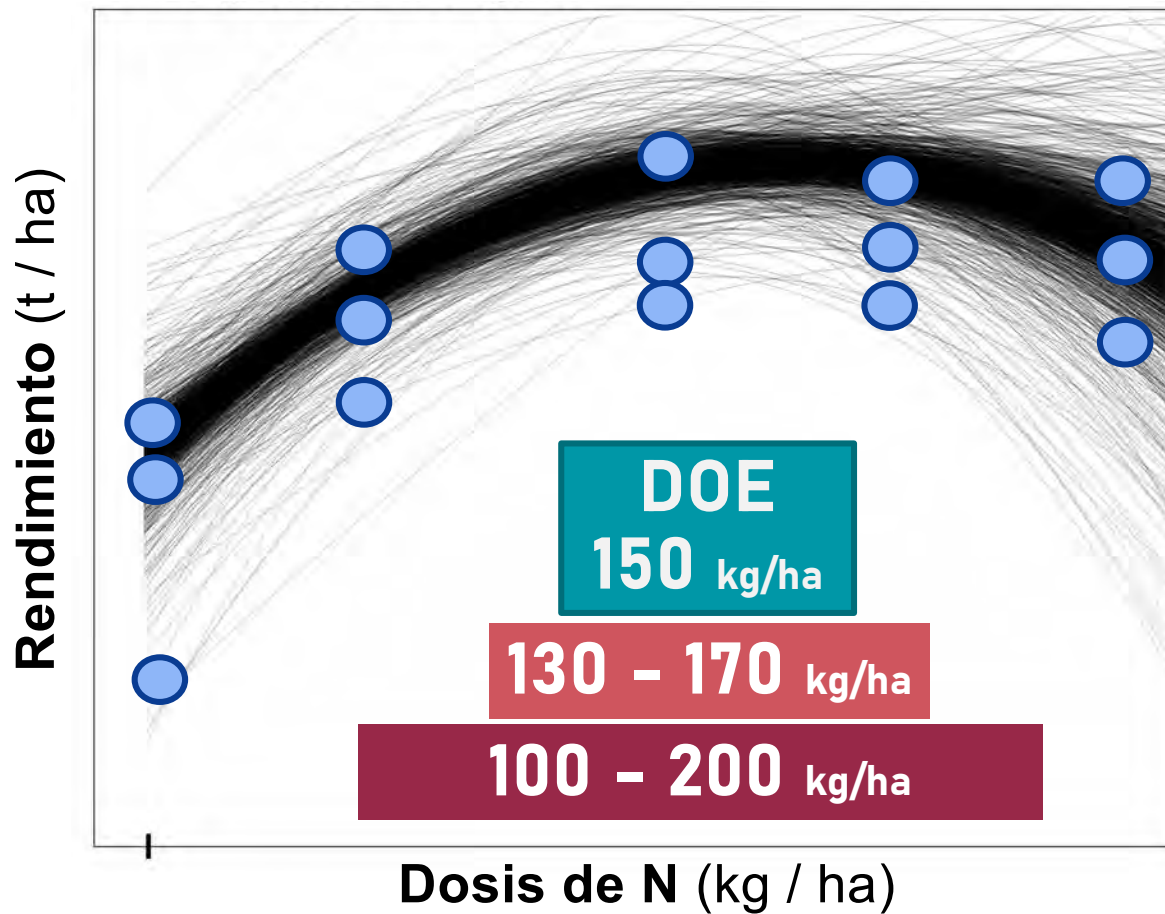
### Cuadratico-Plateau

$$y = B_0 + B_1 x + B_2 x^2 \text{ if } x < X_c$$

$$y = B_0 + B_1 X_c + B_2 x^2 \text{ if } x \geq X_c$$

# EL MODELO BAYESIANO

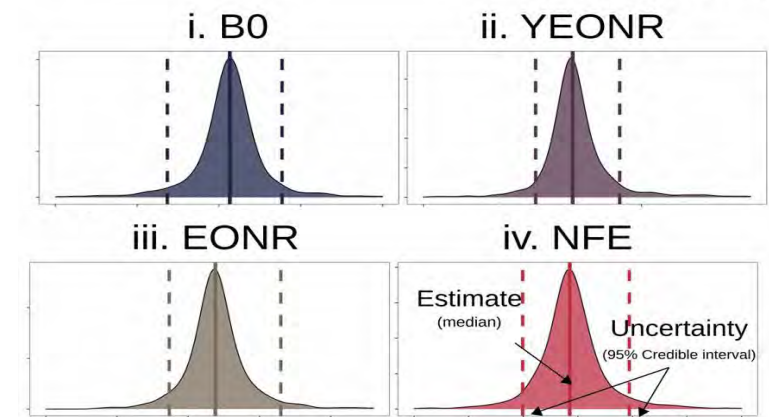
¿Qué trae de diferente?



(i) Probabilidad **previa**

(ii) **Evidencia** (datos)

(iii) Probabilidad **posterior**



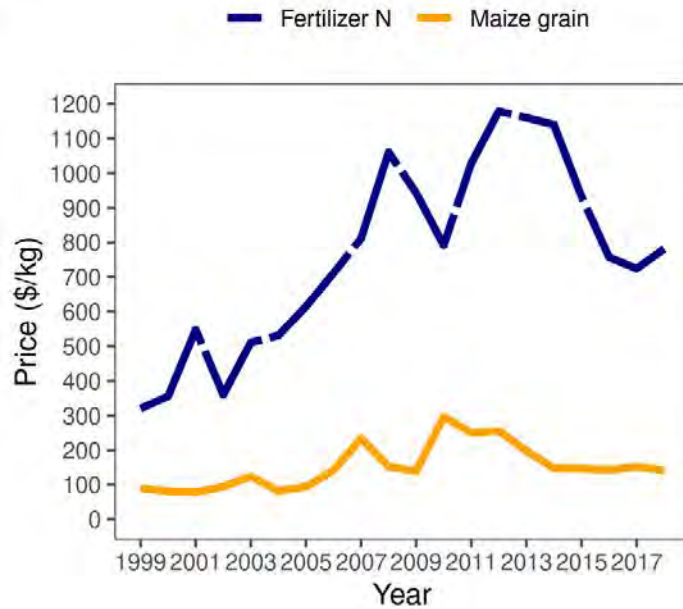




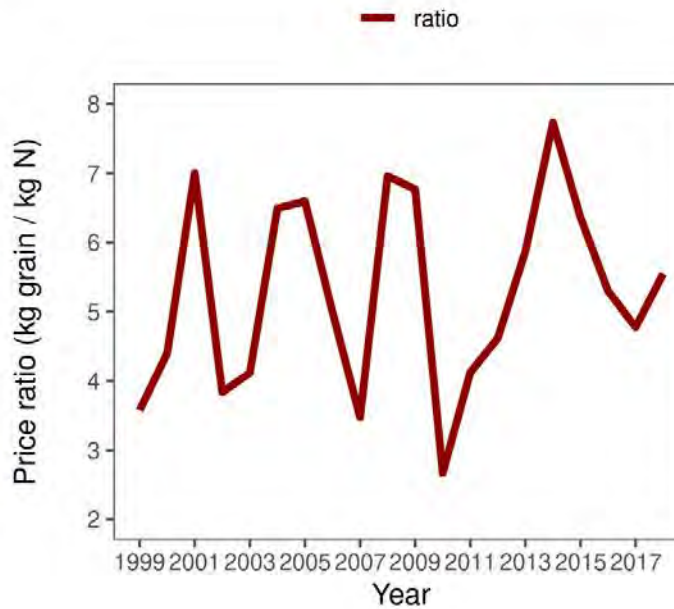
# PRECIOS

DOE -> incluir la variabilidad en el modelo

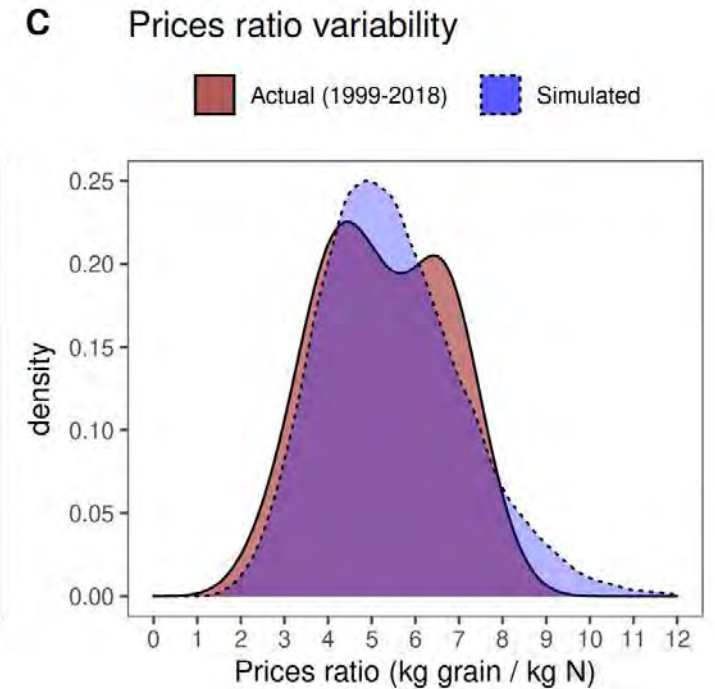
A



B



C



Correndo et al. (2021). *Agric. For. Meteorology*.

# MAGIA Bayesiana



**NO EXISTE!**

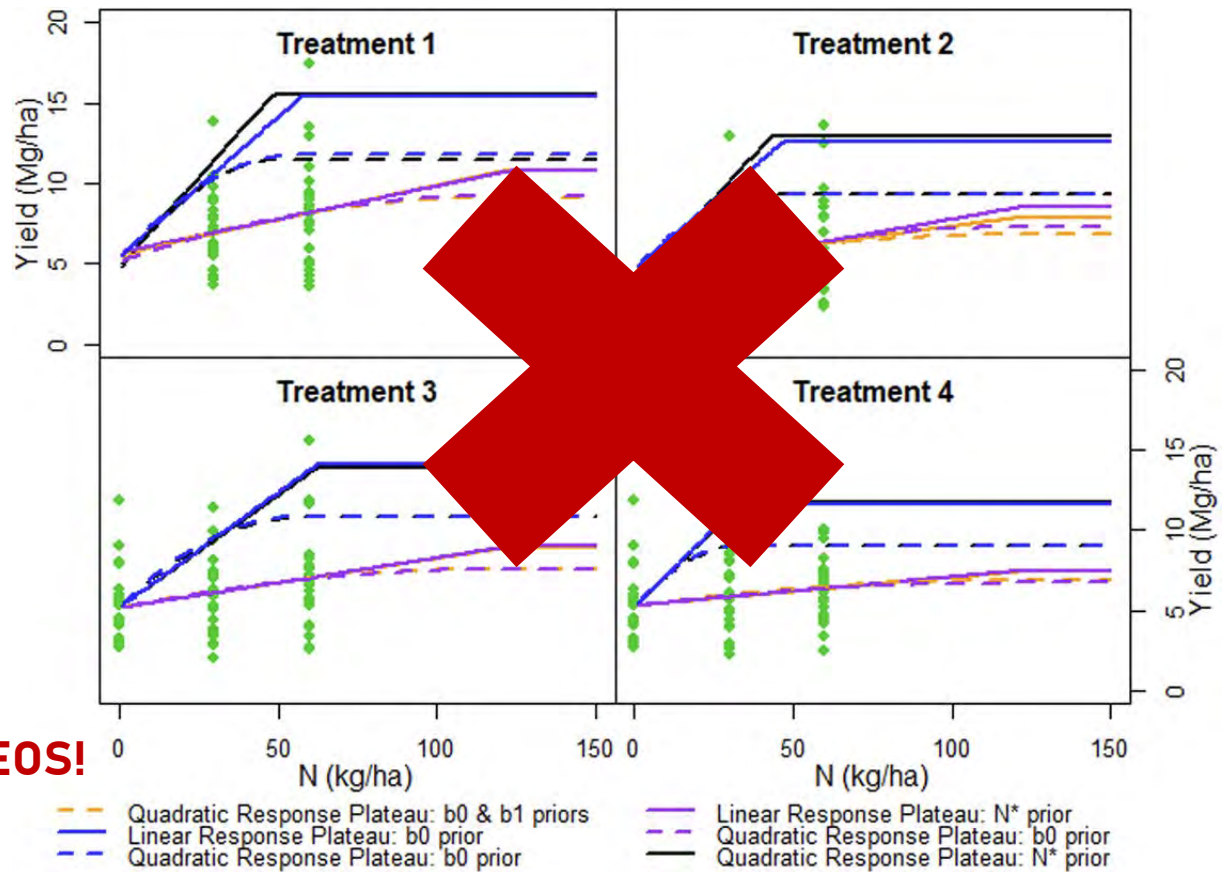
Si no tiene datos...

Junte y vuelva la próxima...

Se requiere **evidencia...**  
(solo priors no alcanza)

**NECESITAMOS REVISORES IDONEOS!**

Cho et al. (2023). Agrosystems, Geosci & Environ. <https://doi.org/10.1002/agg2.20358>



# MENSAJES

- Mucha tela por cortar...
- Deudas: **EDUCACION, REPRODUCIBILIDAD, ACCESIBILIDAD**
  - **PROGRAMACION** → R, Python, GitHub
  - **ESTADISTICA APLICADA**
- Identificar espejos de colores!
  - NADA reemplaza **diseño** & **objetivos claros**
  - Ni **Machine Learning**, ni Estadística Bayesiana hacen MAGIA!



En Dios confío...  
los demás que traigan datos!!

*W. Edwards Deming*

**MUCHAS GRACIAS!**

