

Casos de estudio, ejemplos de aplicación de ciencias de datos en agricultura

Hugo Andrés Dorado

Científico de datos

hugo.doradob@gmail.com

Caso de estudio 1- Plátano Colombia

Socios



Datos



Objetivo

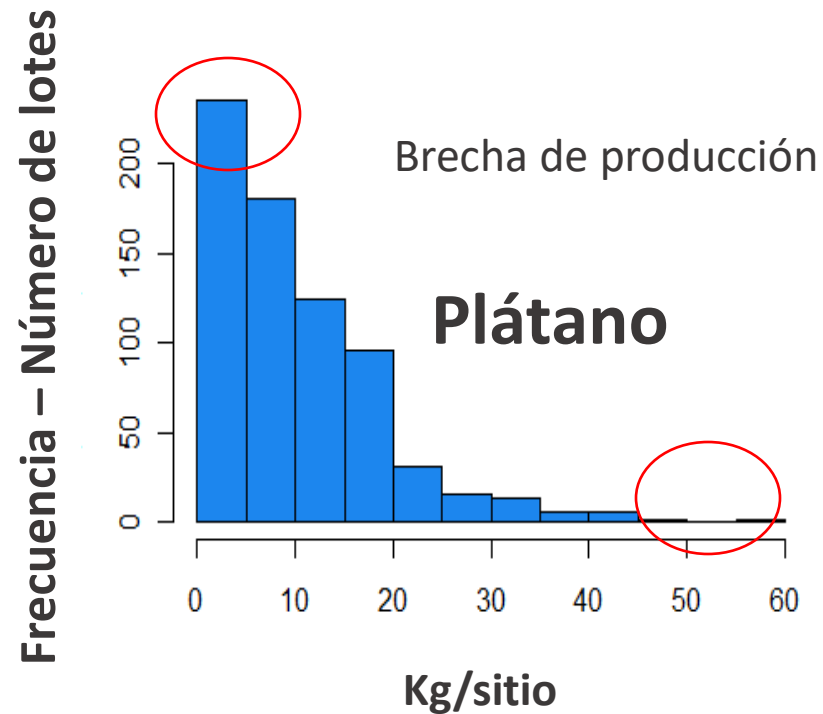
Identificar el potencial de variedades en distintas zonas agroecológicas de Colombia.

Aplicación

Proveer información a productores sobre cultivares con mayor potencial según sus condiciones sitio específicas.



Caso de estudio Plátano en Colombia: Clima, Suelo y Manejo agronómico



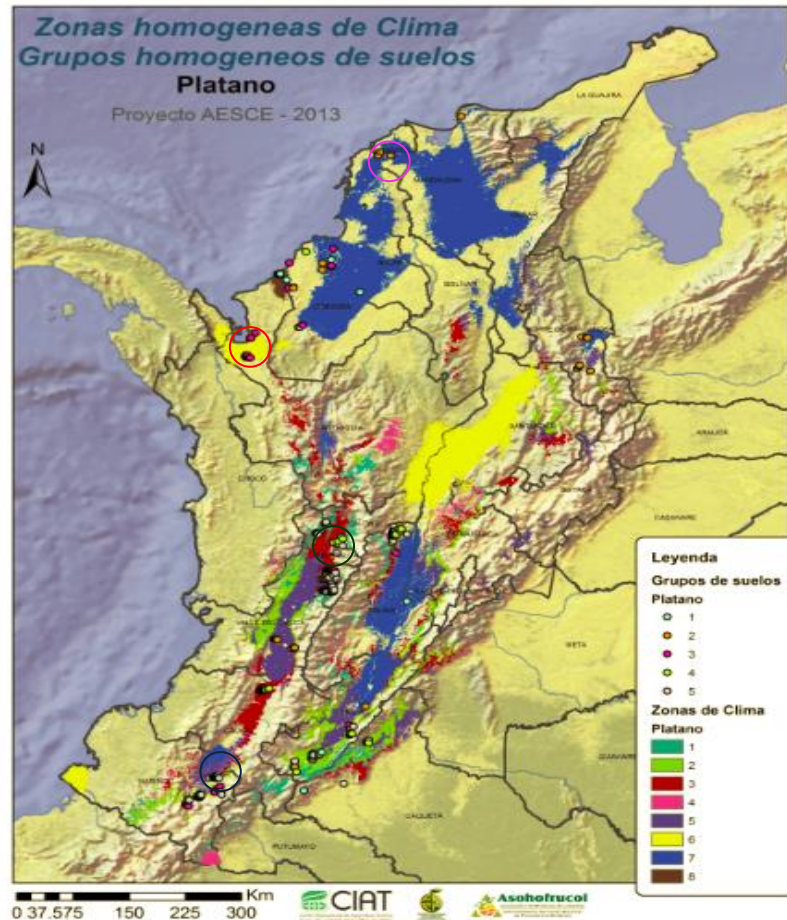
Hipótesis

Más de 800 lotes productivos – representando heterogeneidad de condiciones

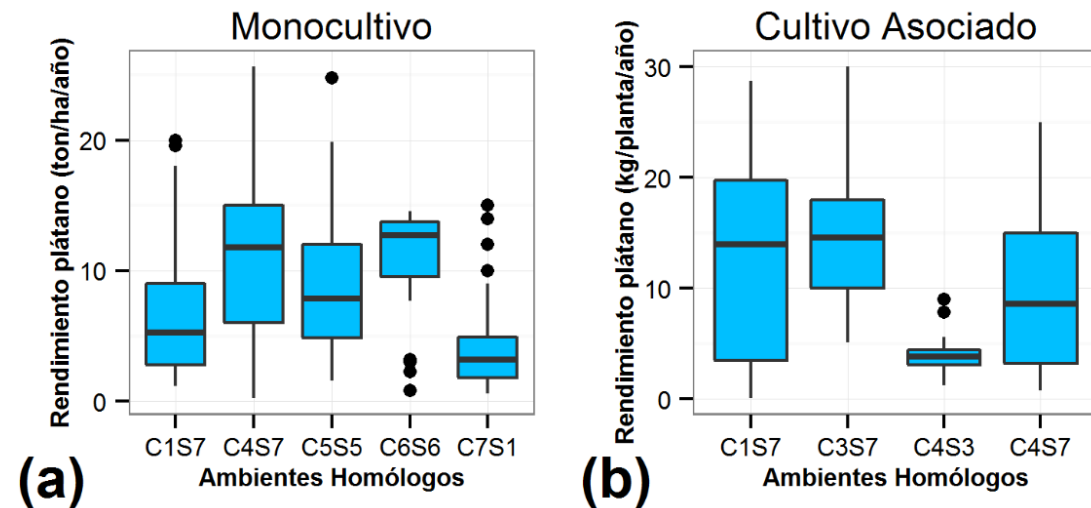


Caso de estudio Plátano en Colombia: Clima, Suelo y Manejo agronómico

Análisis factoriales: PCA- CATPCA. Modelos mixtos y BLUPS



Suelo y manejo: Caracterizado directamente de los lotes.
Clima: Worldclime

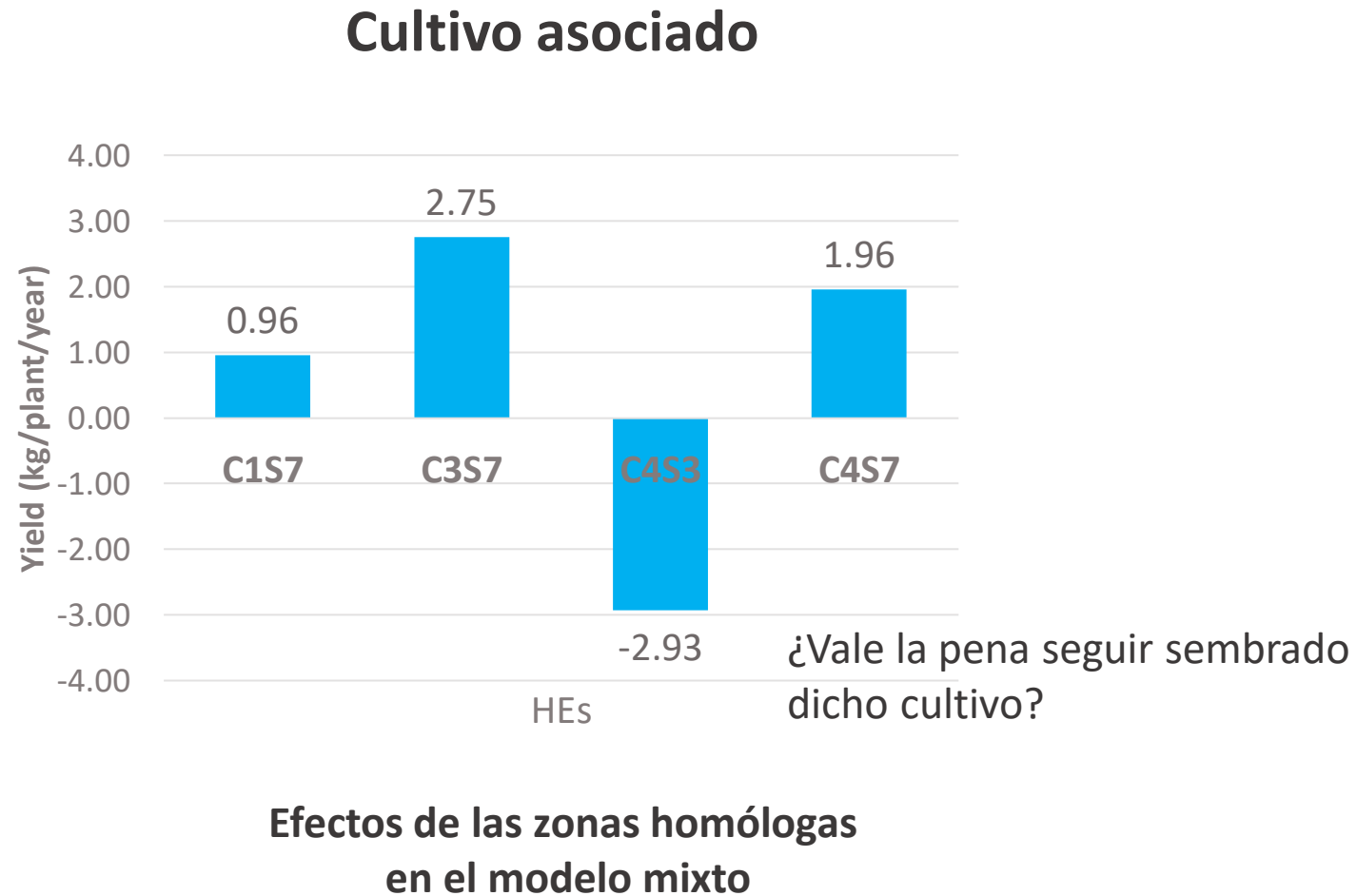


Productores con **las mismas condiciones** tenían una gran variación en rendimiento



Caso de estudio Plátano en Colombia: Clima, Suelo y Manejo agronómico

Análisis factoriales: PCA- CATPCA. Modelos mixtos y BLUPS

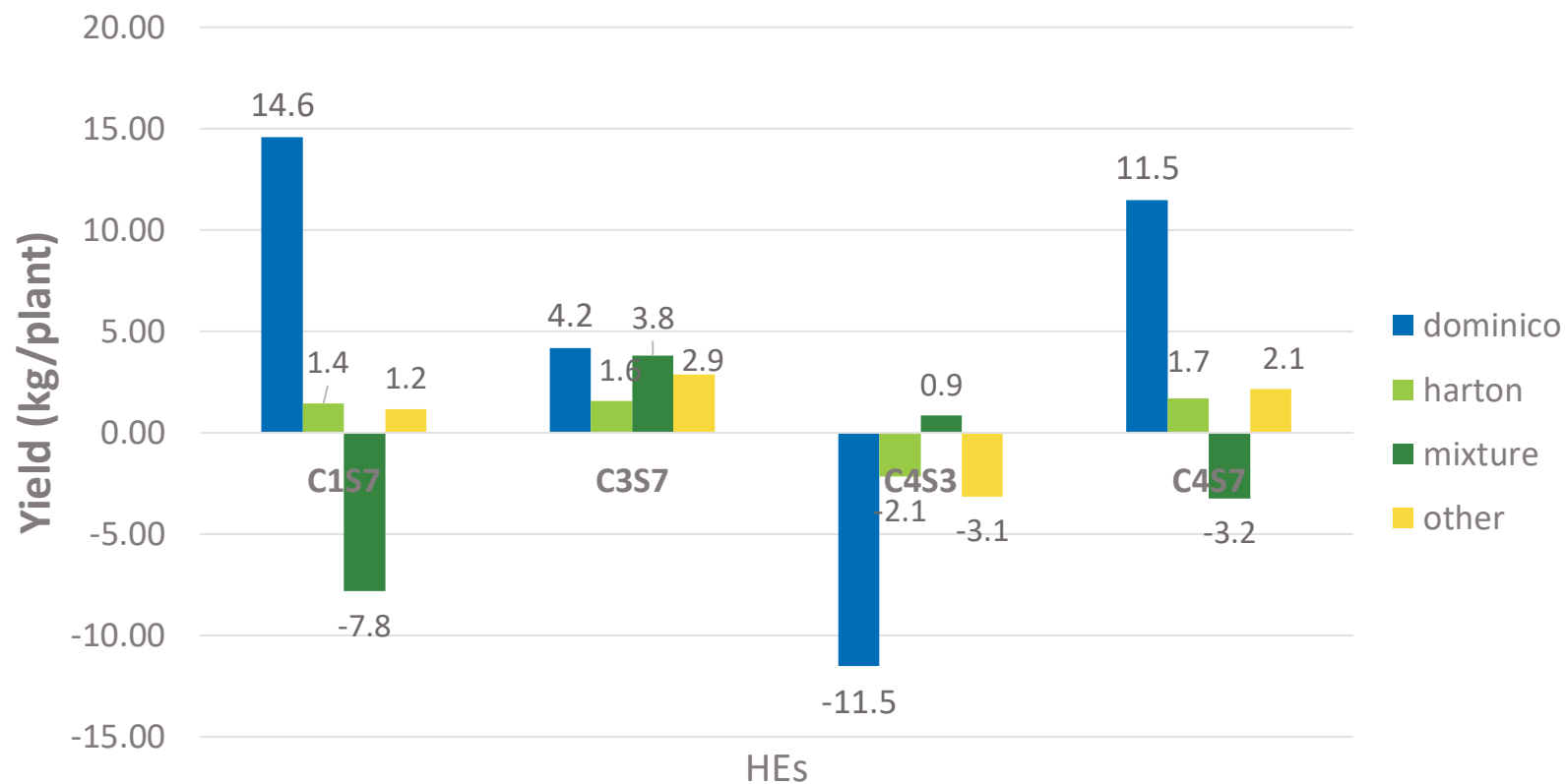




Caso de estudio Plátano en Colombia: Clima, Suelo y Manejo agronómico

Cultivo asociado

Modelos mixtos estimación de modelos mixtos



Efectos de las variedades en las diferentes zonas homólogas

Casos de estudio 2. Arroz - Colombia

Proyecto.



MINAGRICULTURA



CIAT

Preguntas de investigación.

¿Que efecto tiene el clima sobre el rendimiento en arroz?



FEDEARROZ

FONDO NACIONAL DEL ARROZ



IDEAM

Fuente de datos

- Encuesta nacional de arroz.
- Registros de cosecha.
- Experimentos de fecha de siembra.
- Estaciones meteorológicas (IDEAM).

Observaciones

1237

6000

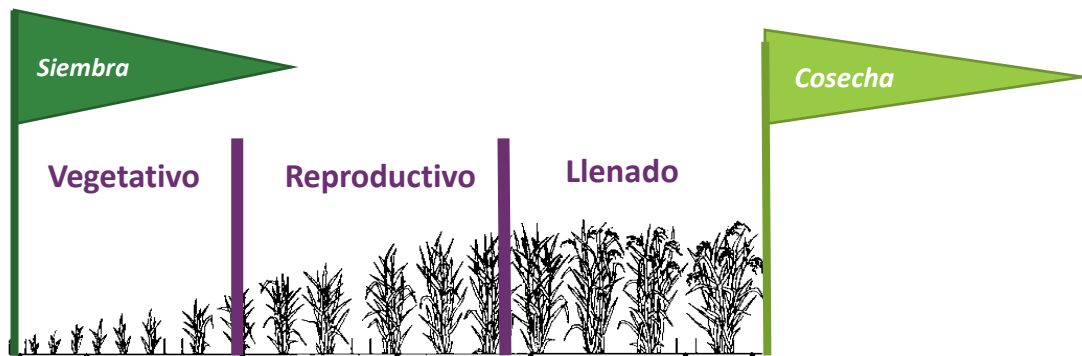
600

27

} 7837

Efectos del clima sobre variedades de arroz

Unidad de observación – Ciclo de cultivo



Variables climáticas por
etapa (X_1, X_2, X_3, \dots)
Y

Variables involucradas

- Fecha de siembra - Cosecha
- Variedad (F733, F60, Lagunas y F174)
- Localidad (Saldaña, Villavicencio)
- Temperatura promedio
- Energía solar acumulada
- Precipitación acumulada
- Número de días con precipitación mayor a 10 mm
- Promedio de humedad relativa
- Rendimiento en kg/ha

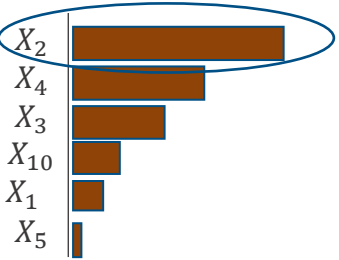
Efectos del clima sobre variedades de arroz

Metodología utilizada.

Random forest - importancia de variables (Breiman, 2001)

- ✓ Mejor desempeño en calidad de predicción (RMSE).
- ✓ Soportar relaciones no lineales.
- ✓ Equilibrar la influencia por correlación entre las variables.

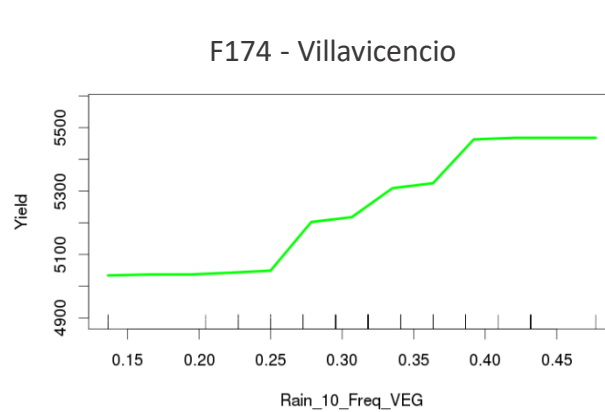
Resultados.



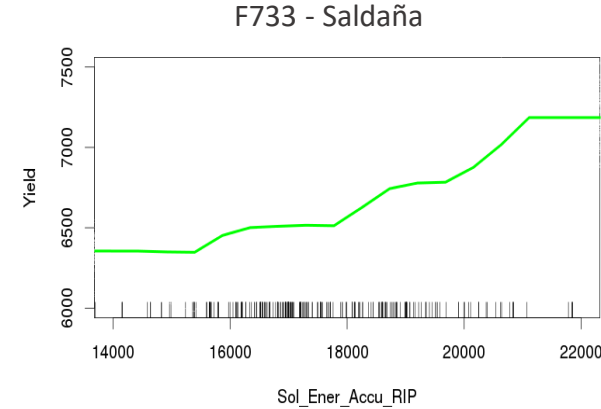
Modelo	Observaciones	Desempeño (R Cuadrado)	Variable mas relevante
Saldaña - F733	267	29.90%	Energia solar acumulada en fase de llenado de grano
Saldaña - F60	150	46.60%	Temperatura promedio en fase reproductiva
Saldaña -Lagunas	187	6.96%	Temperatura promedio en fase de llenado de grano
Villavicencio - F174	134	28.10%	Dias con lluvia en fase vegetativa

Efectos del clima sobre variedades de arroz

Dependencias parciales. – Condiciones climáticas asociadas a altos rendimientos



Precipitación en el 40% de los días de la fase vegetativa



Radiación solar acumulada en todo el ciclo de llenado de grano mayor a 21.000 cal/cm²

Utilidad de información.

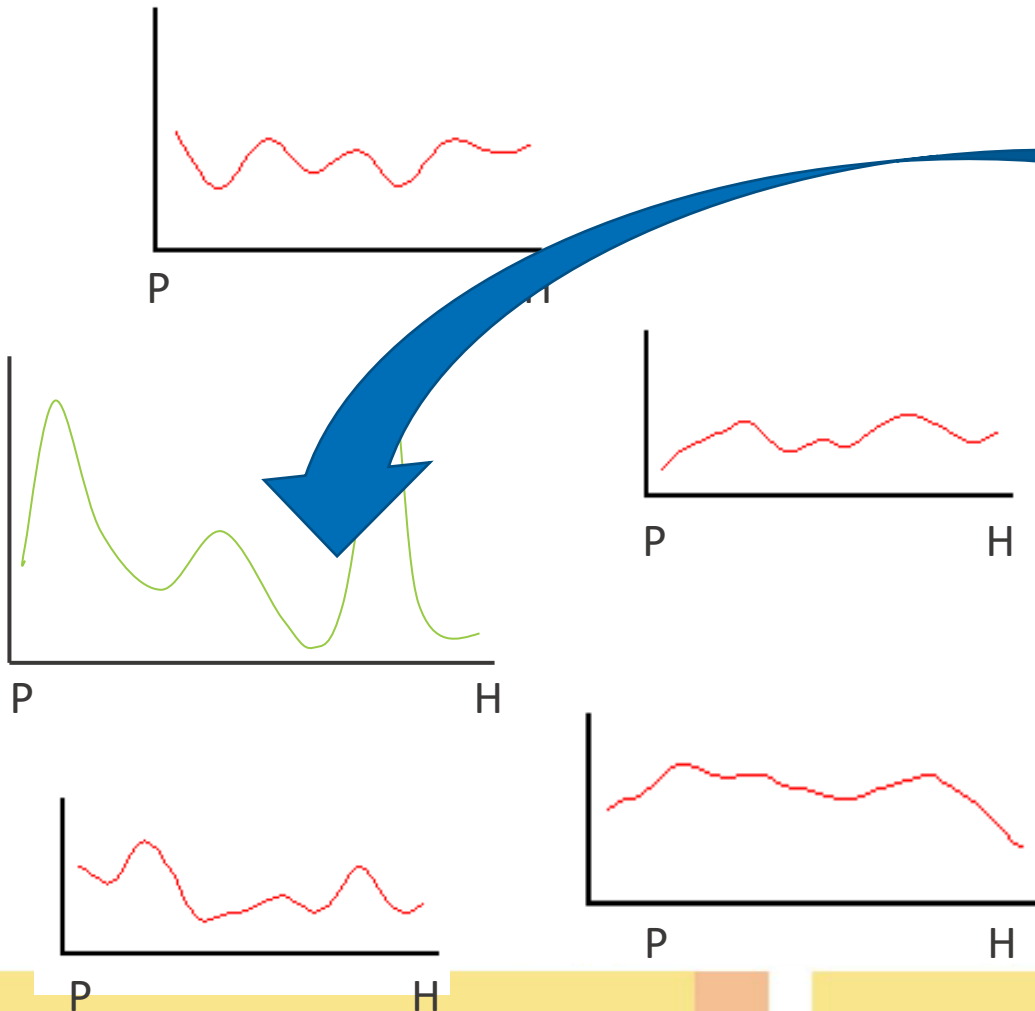
Productores: Plantear fechas de siembra estratégicamente en el año para acomodarse a las mejores condiciones según la variedad que se vaya a sembrar.

Mejoradores: Probar la efectividad y resistencia a factores climáticos de variedades desarrolladas a partir de datos de productores.

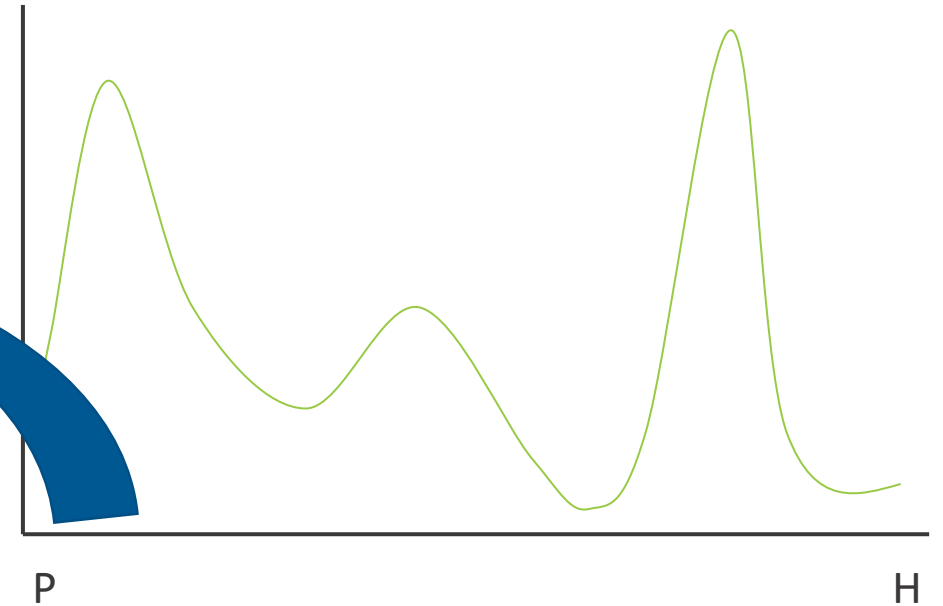
Perfiles históricos (Enfoque de clustering)

Climatic profiles

Weather behavior (P: Planting – H: Harvest)



Seasonal forecast



How will be the yield performance?

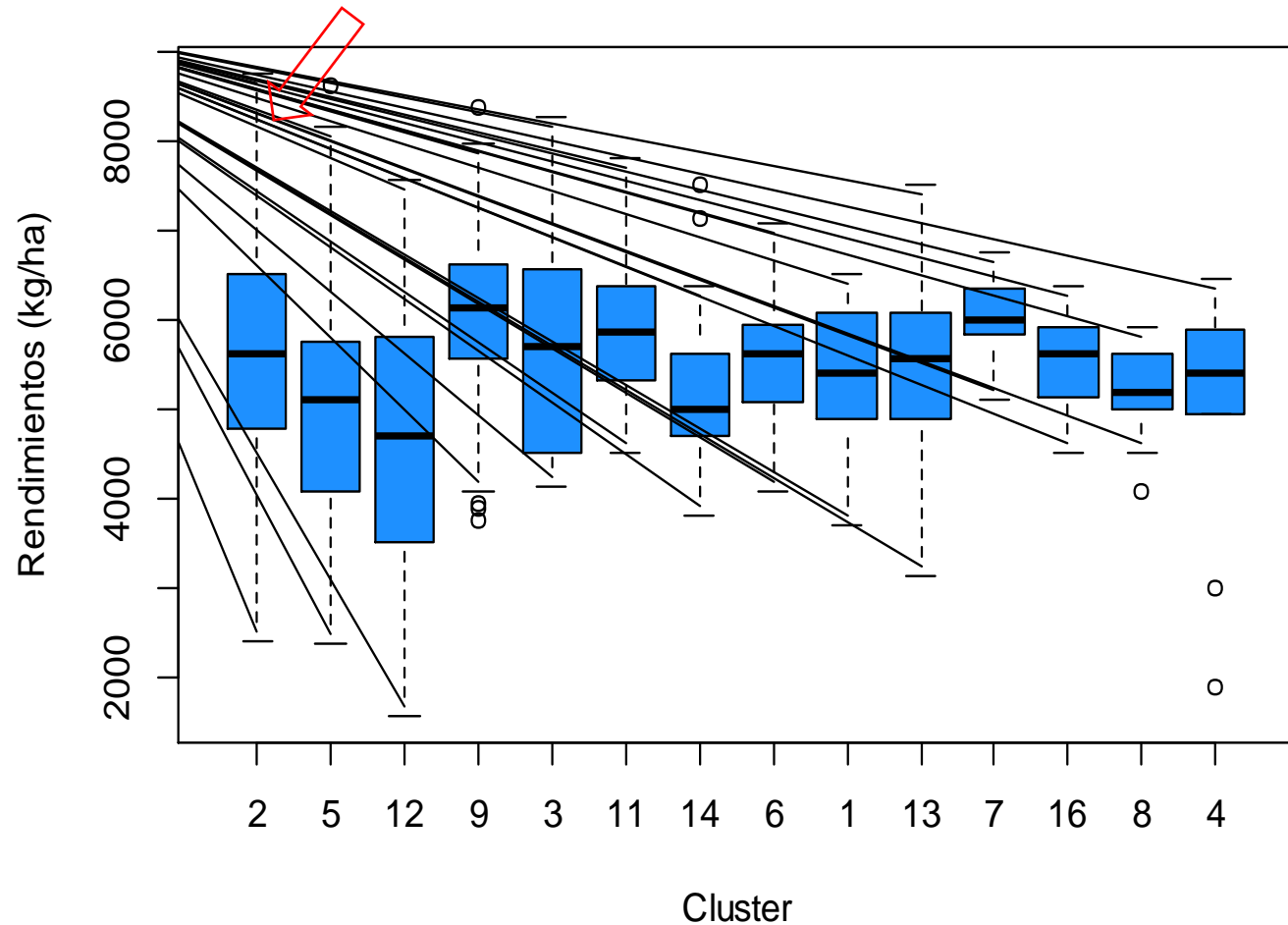
Which variety had more success in the past?

Perfiles históricos (Enfoque de clustering)

Region: Casanare: 2007 – 2014 – Irrigated
Weather stations from FEDEARROZ e IDEAM
(N= about 756 cropping events) – 17 clusters

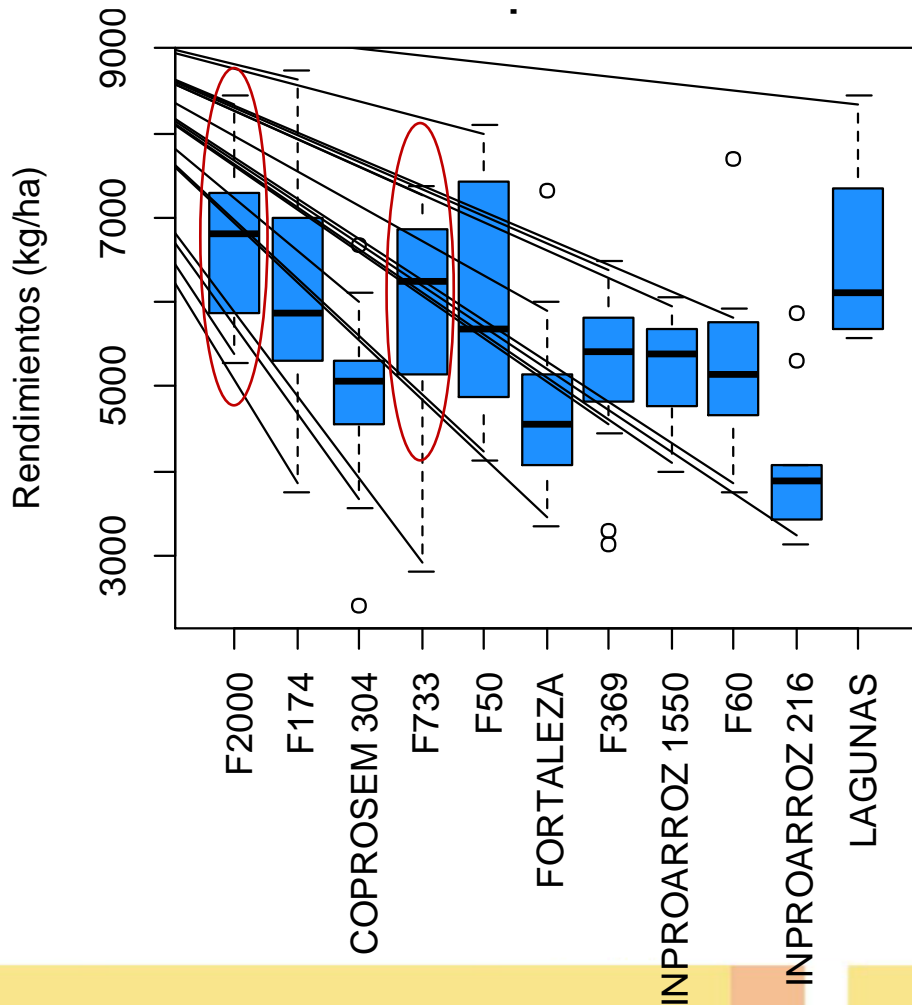
Cluster	Number of cropping events	Productivity (kg/Ha)
1	18	5,354
2	238	5,653
3	51	5,821
4	9	4,913
5	148	4,946
6	30	5,557
7	15	6,041
8	10	5,174
9	60	6,000
10	6	5,726
11	42	5,898
12	65	4,688
13	18	5,469
14	33	5,222
15	1	5,312
16	15	5,521
17	6	5,053
Total	765	5,438

Yield associated with each historical profile
(cluster) presented in the region



Perfiles históricos (Enfoque de clustering)

Performance of each variety within cluster 2 (seasonal forecast)



What to grow?

Rice variety	Number of cropping events	Productivity (Kg/Ha)
F2000	39	6,717
F174	27	5,979
COPROSEM 304	25	4,973
F733	23	5,981
F50	22	5,922
FORTALEZA	21	4,641
F369	18	5,244
INPROARROZ 1550	16	5,180
F60	12	5,189
INPROARROZ 216	9	4,104
LAGUNAS	7	6,598
Total	219	5,653

Entrega de resultados a productores y técnicos

Giras técnicas con productores



Reuniones y discusión con extensionistas y técnicos locales



Un paper para la comunidad científica

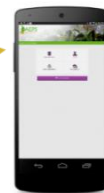
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0161620>



El caso de Maíz en Córdoba – Productividad- FENALCE; Cómo sembrar ?



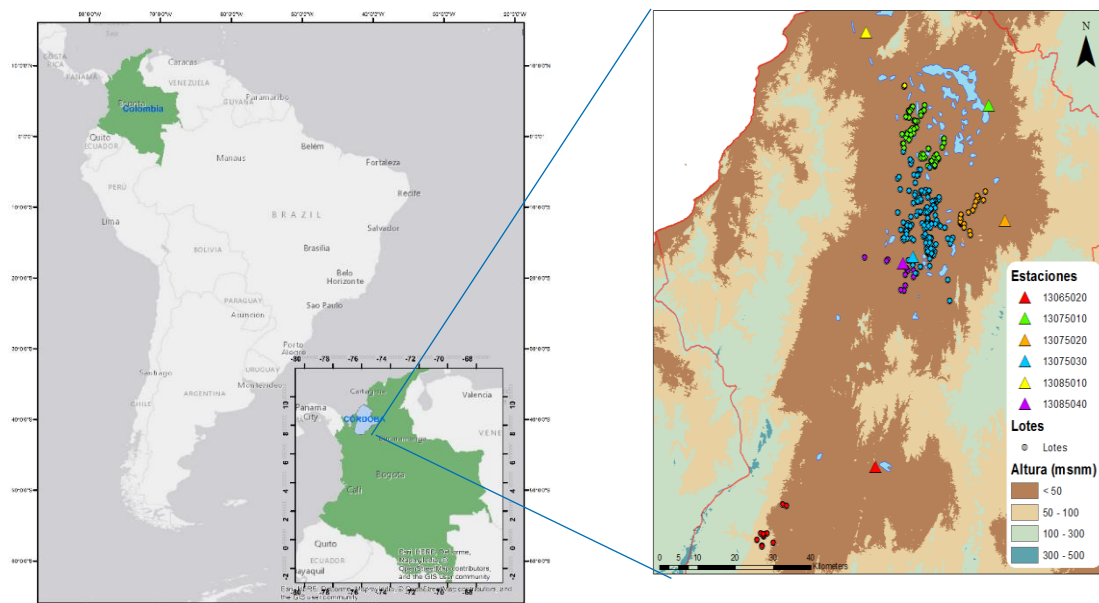
Manejo



Clima



Temperatura máxima
temperatura mínima
Precipitación acumulada
Radiación Solar
Humedad relativa



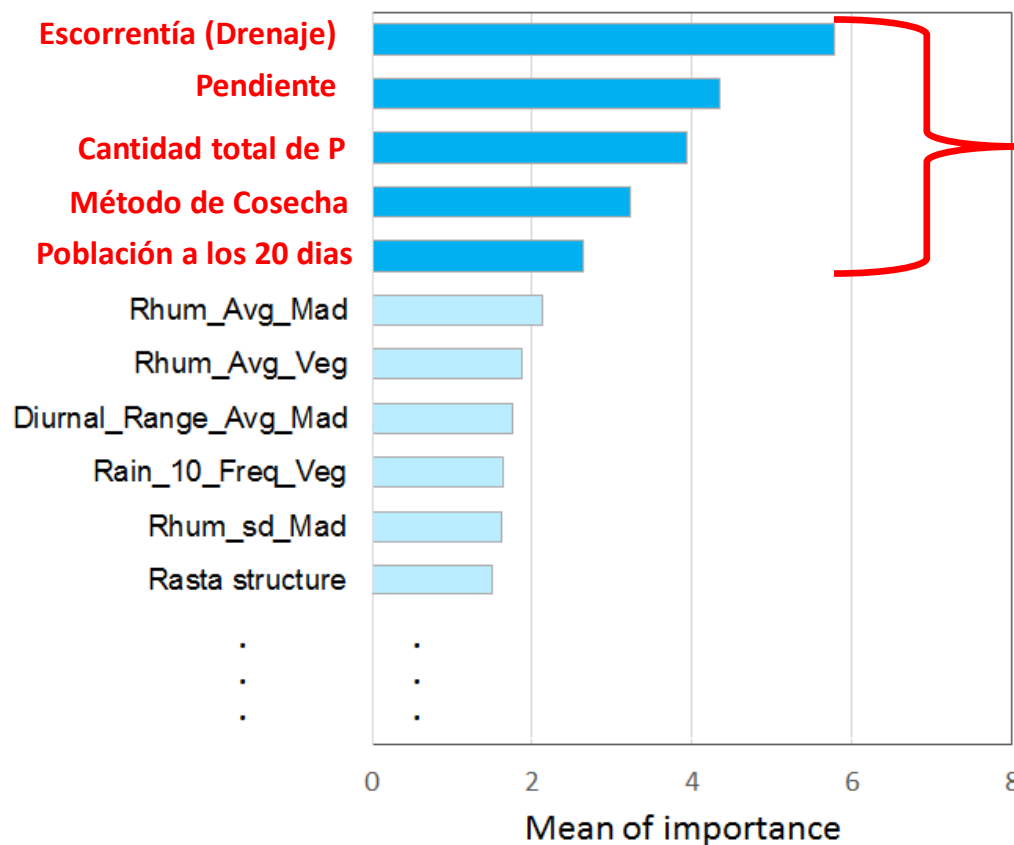
238 eventos ciclos de producción en dos años
Tiempo de datos (2014-2015)



Cómo sembrar ? El caso de Maíz en Córdoba - Productividad

Arboles condicionales

$R^2 = 45.79$

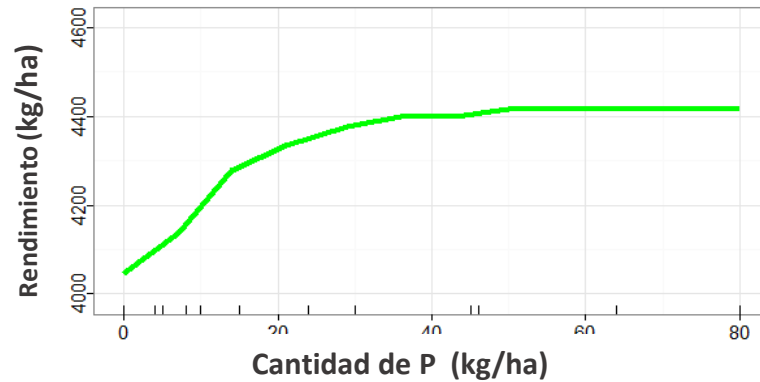


Los factores más importantes asociados con la variación en rendimiento de maíz en Córdoba

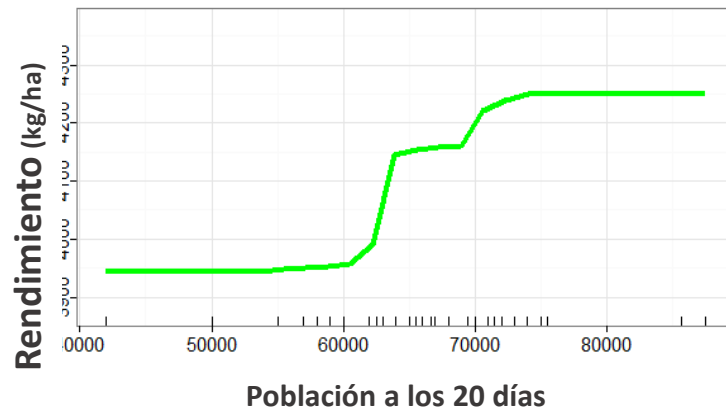




Cómo sembrar ? El caso de Maiz en Córdoba - Productividad



25 – 30 kg P /ha cantidad apropiada para maiz en Córdoba.



Población a los 20 días, al menos 65000 plantas/ha in Córdoba



Una agricultura climáticamente inteligente guiada por datos y que complementa conocimiento tradicional

El caso de Maíz Córdoba

Conocimiento tradicional – Paquete tecnológico

Profundidad efectiva suelo > 30 cm

pH 5.5-6.5

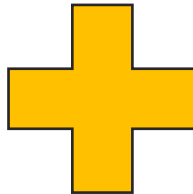
Distancia entre plantas 0.17m – 0.2m

Nutrientes requeridos N, P, K, Mg, S

Población a los 20 d: 50000 – 70000 DDE

Control arvenses: al menos 1 control (8 DAS – 2 DDS)

Control enfermedades: Al menos 1 (10 Días antes de floración)



Agricultura climáticamente inteligente guiada por datos (Big Data -minería de datos, aprendizaje automático y profundo)

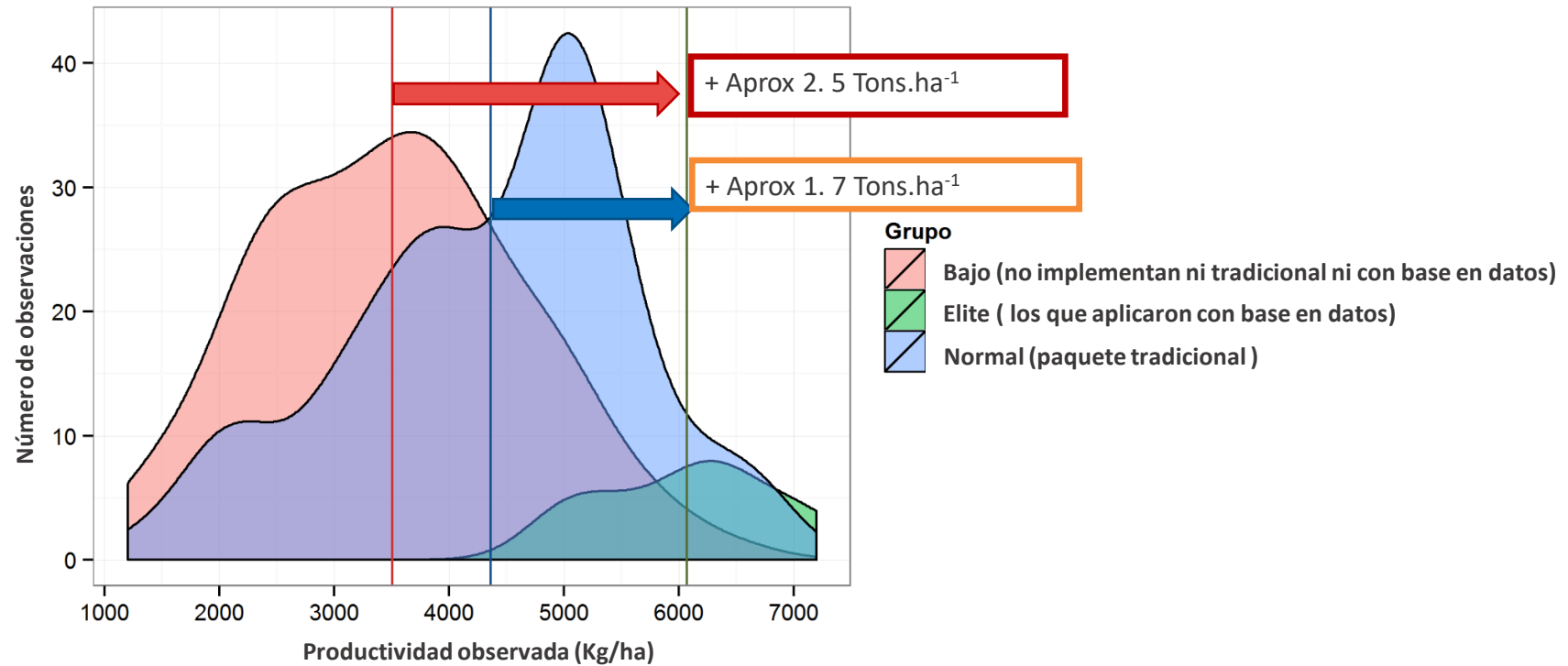
Fósforo total aplicado : 25 - 30 kg /ha

Población a los 20 d: Al menos 65000 DDE

Ambos acercamientos con base en observaciones , el Big Data revela cosas que desconocíamos



Cómo sembrar ? El caso de Maiz en Córdoba - Productividad



Distribuciones de rendimientos observados por los tres subgrupos de manejo en Córdoba- Colombia

Caso de estudio 3 - Maíz Chiapas México

Socios:



Datos:

- Daymet, Registros de Masagro (8 años, 4500 observaciones)



Objetivo:

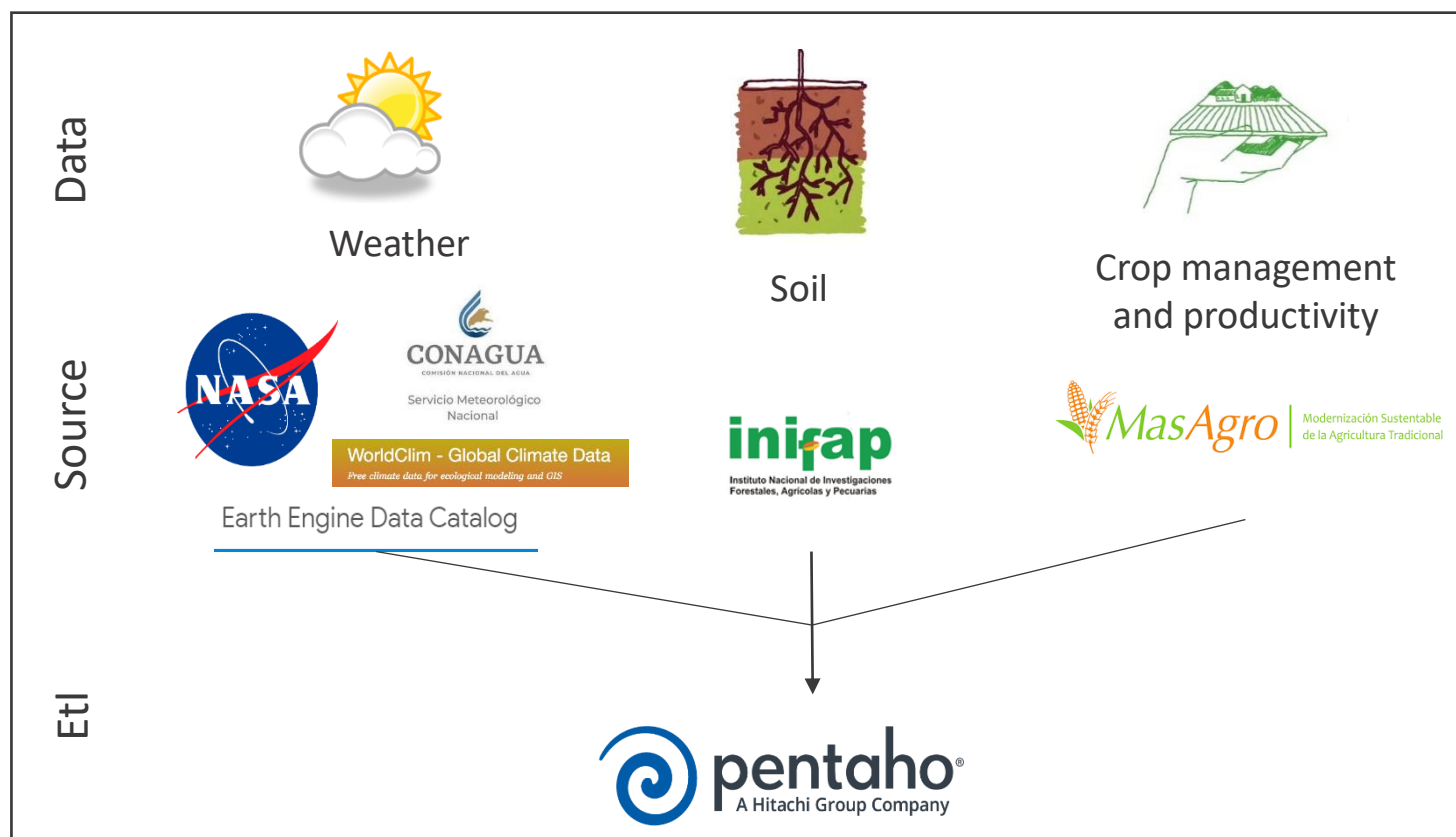
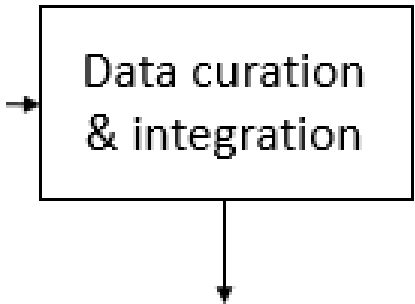
- Identificar los factores que afectan los rendimientos de maíz temporal.

Aplicación:

- Proveer información sitio específica a productores para mejorar rendimientos.

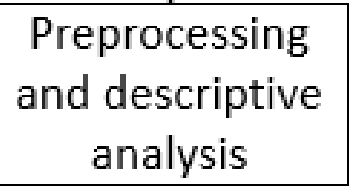


Maíz sistema temporal

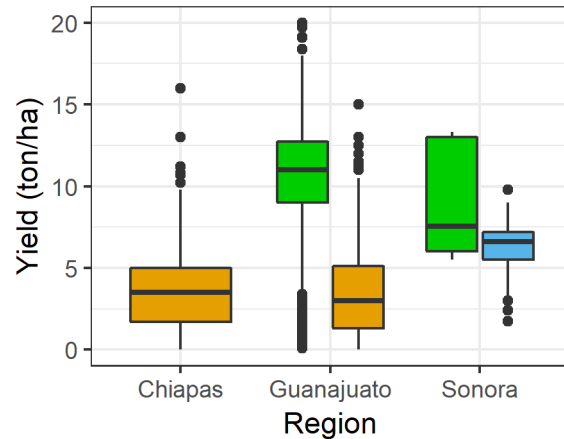


```
> str(BEM_Dataset)
'data.frame':   43790 obs. of  128 variables:
 $ id_de_la_bitacora
   : int  35 38 43 43 47 47 50 50 51 ...
 $ id_de_tipo_de_bitacora
   : int  58 63 64 73 74 81 82 87 88 89 ...
 $ row_number
   : int  1 1 1 1 1 1 1 1 1 1 ...
 $ before_id
   : num  3558 3863 3864 4373 4374 ...
```





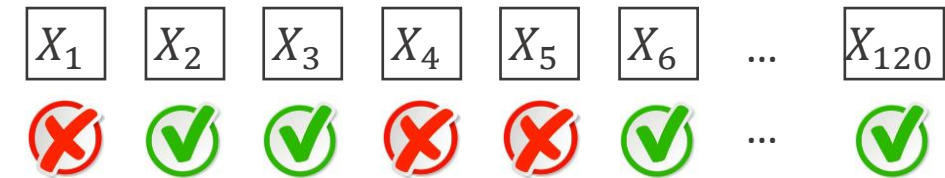
Exploratory data analysis



- Outliers identification
- Variables distribution
- Correlations and associations

Feature selection


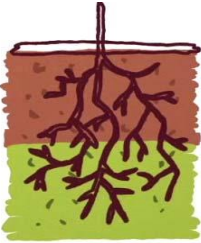

Removing non-informative and redundant variables



Transforming variables

- Group or remove observations poorly-represented in the dataset.
- Build variables based on indicators.
- Dimensionality reduction

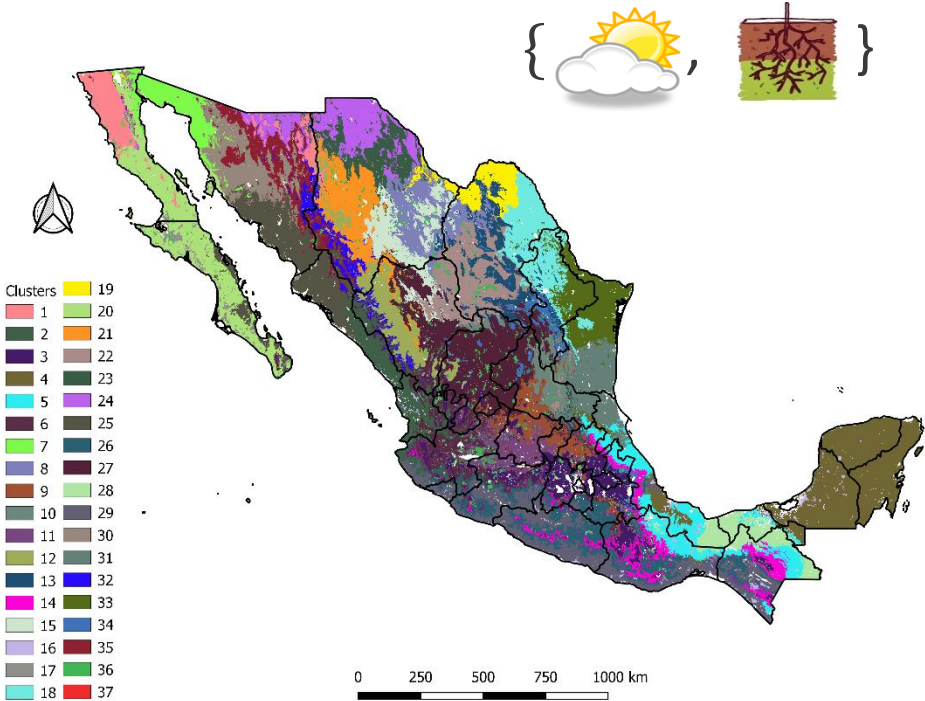
Tidy dataset

Weather (W)	Soil (S)	Crop management (M)	Yield/Profitability (Y, P)
<div><ul style="list-style-type: none">• Average minimum temperature• Average diurnal range• Accumulated solar energy• Frequency of days with maximum temperature above 34°C• Accumulated precipitation• Frequency of days with minimum temperature below 8°C• Average relative humidity• Standard deviation of the relative humidity</div> <div></div>	<div><ul style="list-style-type: none">• Clay content• Silt content• Soil organic content• Cationic exchange capacity• Basis saturation</div> <div></div>	<div><ul style="list-style-type: none">• Infiltration• Cultivar• Seed treatment• Type of tillage• Number of mechanical weeding• Number of applications of (fertilizations, foliar fertilizers, bio fertilizers, post-sowing herbicides, insecticides)• Total amount of nitrogen applied• Total amount of phosphorus applied• Total amount of potassium applied• .• .• .</div> <div></div>	

Training ML models

Clustering

Principal components analysis
K-means



Predictive models per location

$f(\text{Sun/Cloud}, \text{Soil Profile}, \text{Hand/Hill}) = \text{Productivity /profitability}$

Regression based on random forest

10 – Folds cross validation

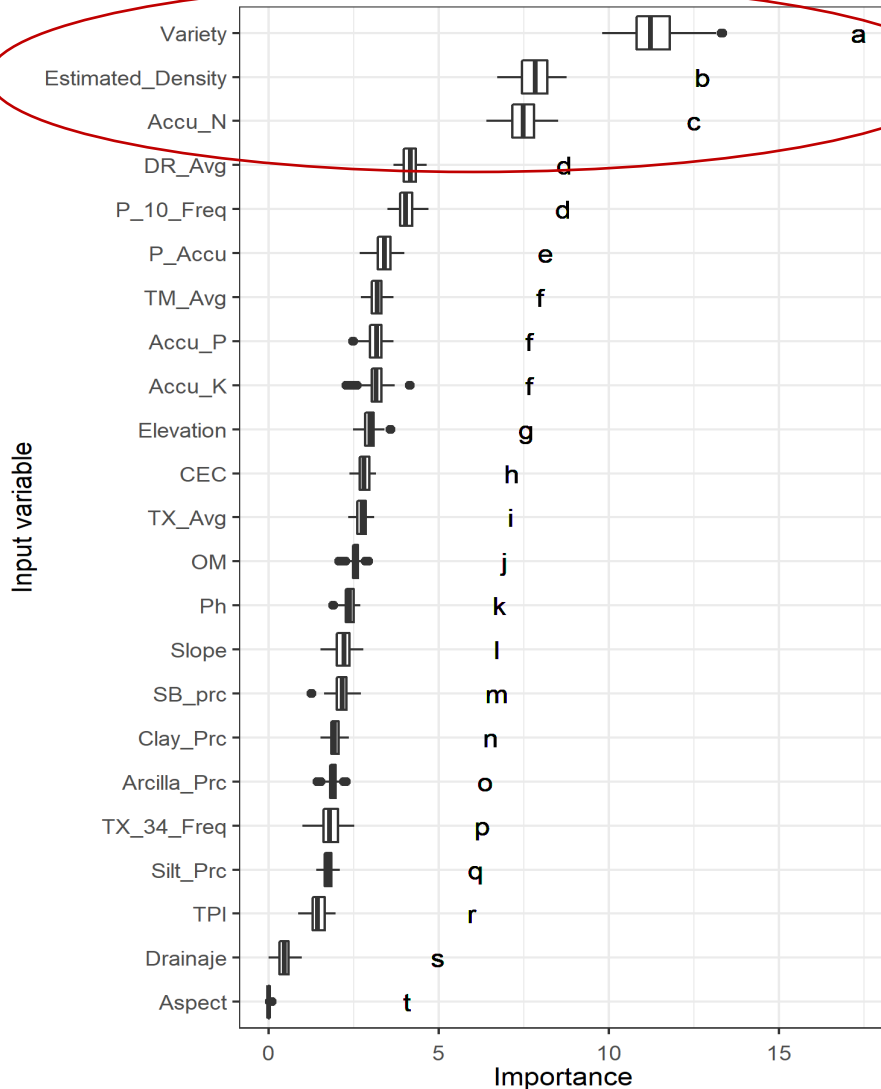
Model by region	R-squared (YIELD)	R-squared (PROFITABILITY)
Chiapas rainfed maize	75%	78%
Sonora irrigated wheat	65%	64%
Guanajuato irrigated maize	60%	77%
Guanajuato rainfed maize	81%	70%

We choose each model based on Ockam's razor blade principle

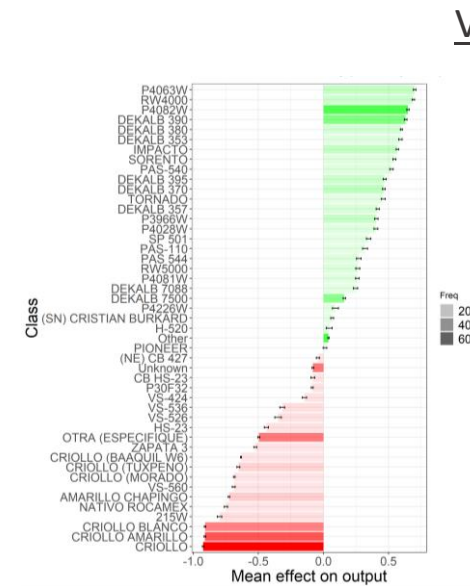
Training models (Explanatory)

Beyond the black box,...

Variables importance



Partial dependences

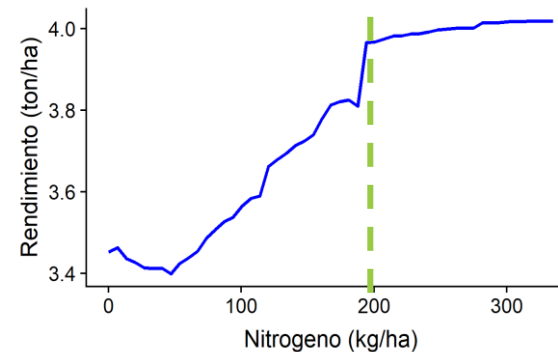


Variety

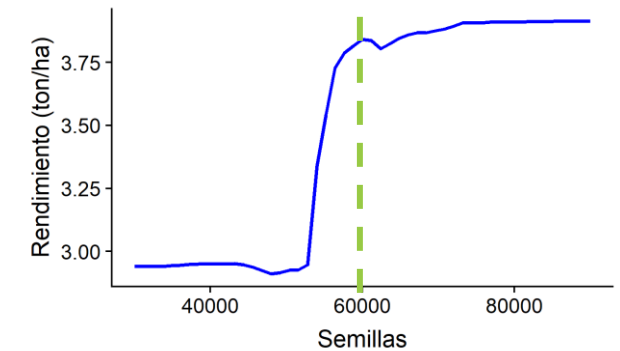
Best hybrid:
P4063W and RW4000

Best locals:
BAAQUIL W6 and TUXPEN0

Amount of Nitrogen (180 kg/ha)



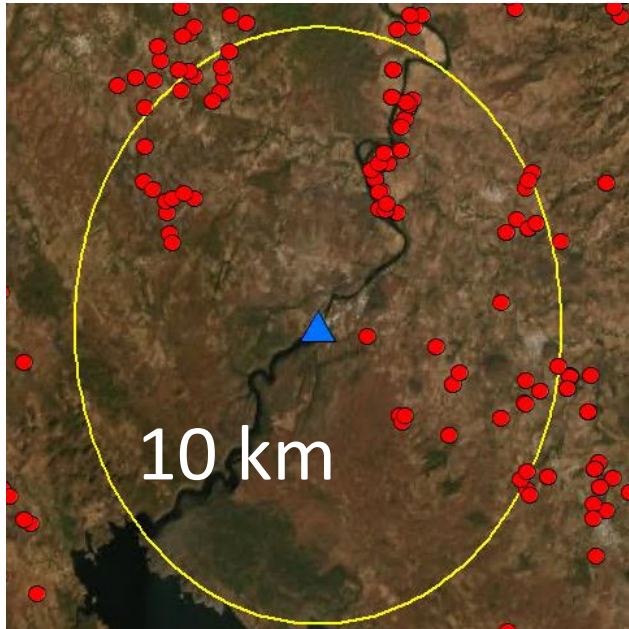
Estimated density (60000 seeds)



Yield in Rainfed Maize, Chiapas Mexico

Maíz Chiapas México - Modelo predictivo junto al pronóstico climático

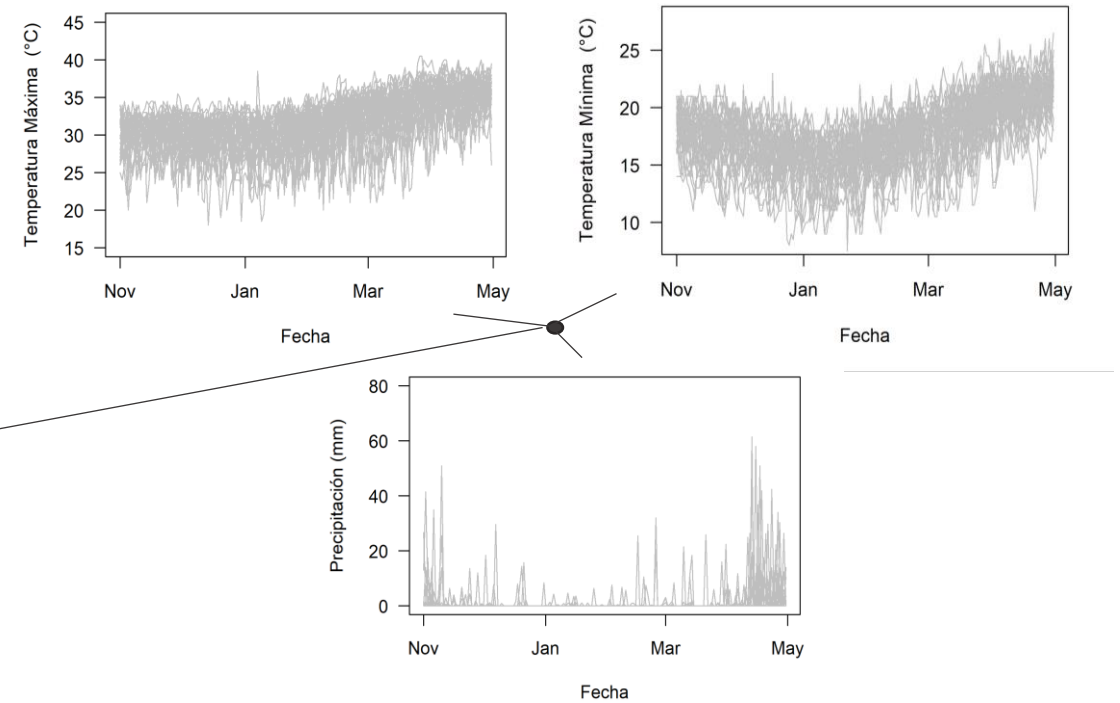
Parcelas cercanas a la estación



Total 93 ciclos de cosecha

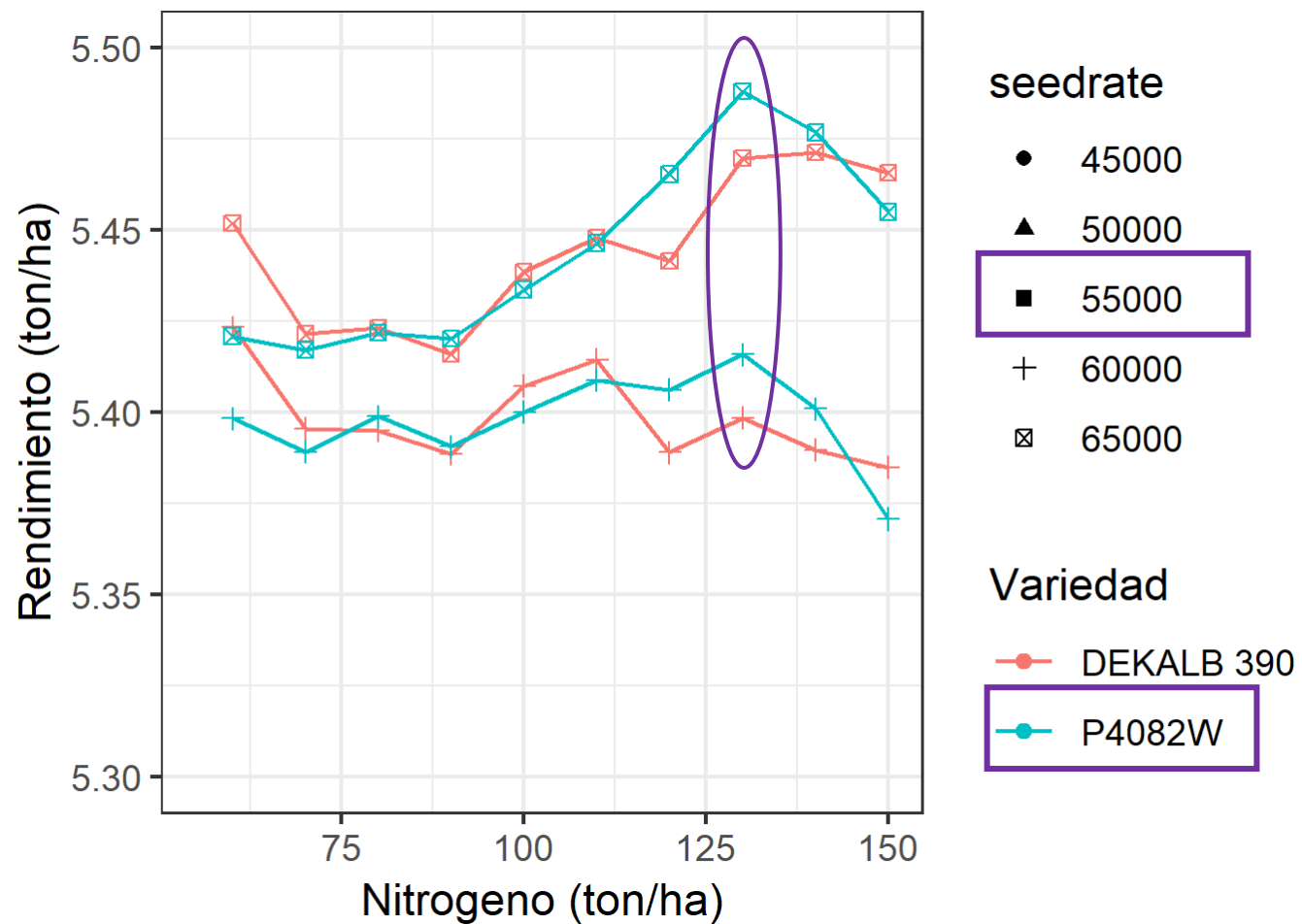
Variedad	Frecuencia
P4082W	20
DEKALB 390	13
OTRA (ESPECIFIQUE)	13
DEKALB 7500	7
DEKALB 380	6
P30F32	6
Otras	28

Pronóstico climático 100 simulaciones



$$f(C, S, M) = Y$$

Modelo predictivo junto al pronóstico climático para un productor



Gracias!

Hugo Andrés Dorado.

Científico de datos

hugo.doradob@gmail.com

Conocimiento generado a partir de proyectos de:

Alianza



Preguntas

- ¿Qué datos logra identificar en su área aplicación?
- ¿Qué ideas se le ocurre inspirado en estos ejemplos?