

**INFORME**

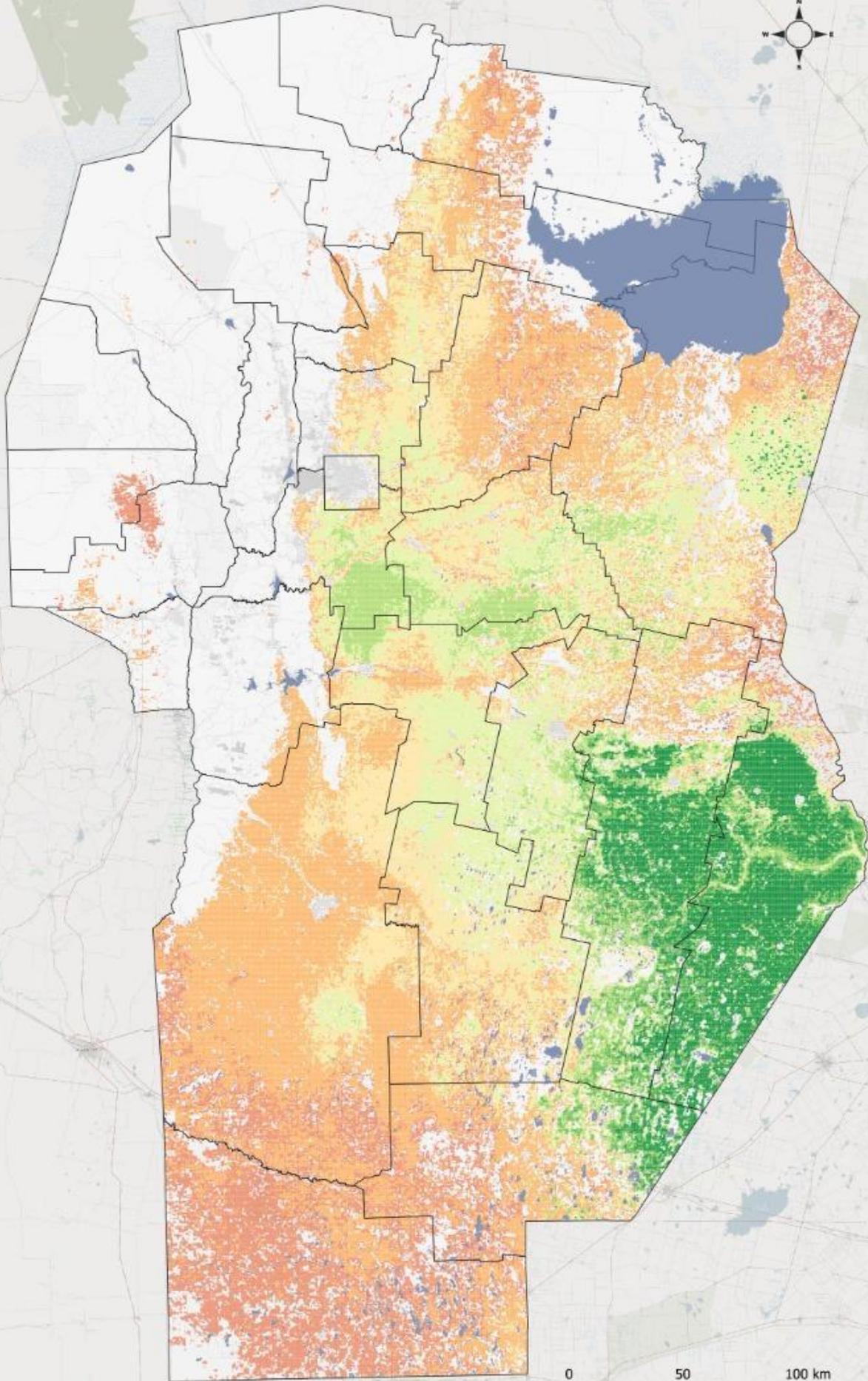
---

**ÁREA SEMBRADA,  
RINDES Y  
PRODUCCIÓN DE  
SOJA Y MAÍZ  
CAMPAÑA 2020/21**

---

**DE LA PROVINCIA DE  
CÓRDOBA**

# Estimación de Rendimiento de Maíz (qq/ha)



0 50 100 km



## Contenido

Campaña 2020/21 .....	4
Grandes datos a nivel provincial.....	4
1. Introducción .....	5
2. Relevamiento de datos a campo.....	6
3. Estimación de superficies sembradas por cultivo .....	8
3.1 Detalles metodológicos.....	8
3.1.1 Confección de mosaicos.....	8
3.1.2 Definición del área cultivable.....	9
3.1.3 Muestras de campo para entrenamiento y validación .....	10
3.1.4 Clasificación supervisada .....	11
3.2 Resultados y mapa de coberturas agrícolas estivales.....	11
4. Estimación de rendimientos por cultivo.....	14
4.1 Muestra utilizada .....	14
4.2 Grilla de predicción .....	14
4.3 Modelos de aprendizaje computacional aplicados .....	15
4.4 Resultados obtenidos .....	16
5. Estimación de la producción de Soja y Maíz .....	19
Referencias Bibliográficas.....	24
Anexo I. Variables utilizadas en el modelo de predicción de rendimientos .....	25

Campaña 2020/21

## Grandes datos a nivel provincial



### Metadata

Unidad espacial: la predicción del rendimiento de soja y maíz se hizo sobre el área cultivable de la provincia de Córdoba, basada en una grilla de una **resolución de 25 hectáreas** (500 metros por lado).

Cantidad de datos y origen: se trabajó con una muestra final de **2.007 observaciones**, aportadas por las Agencias Zonales del Ministerio de Agricultura y Ganadería de la Provincia e INTA principalmente, a través de "Rinde Collect", una aplicación móvil diseñada específicamente para el relevamiento por IDECOR

Modelos aplicados: **Random Forest** para la determinación de las áreas sembradas y **Quantile Regression Forest** para la determinación de los rendimientos por tipo de cultivo.

# 1. Introducción

Con el objetivo de estimar la producción de los principales cultivos agrícolas de la provincia de Córdoba, el presente informe detalla el proceso de relevamiento de campo, construcción de un mapa de coberturas agrícolas estivales para cuantificar su superficie y la posterior estimación de los rendimientos en quintales por hectárea de los cultivos soja y maíz para la campaña 2020/21.

Se detectó que 7.802.339 de hectáreas fueron destinadas a la siembra de cultivos estivales, representando esta superficie el 48,3% del total provincial. De los cultivos principales, la soja y el maíz se observaron distribuidos por toda el área cultivada, mientras que el maní se localizó principalmente sobre el centro y sur.

Sobre este universo de tierra cultivable, la mayor superficie estuvo dedicada al cultivo de soja, con un total identificado de 4.073.438 hectáreas, que representan el 52% de la superficie ocupada por los tres cultivos. En segundo lugar, se ubicó el maíz, con 3.120.337 de hectáreas que representan un 40% de la superficie agrícola total mapeada. Finalmente, la superficie destinada a otros cultivos ocupó 608.564 hectáreas (8% del área cultivable total).

Posteriormente, se avanzó en la estimación de los rendimientos de soja y maíz sobre una grilla regular de 25 hectáreas por celda (grilla que totaliza 362.493 celdas). Se trata de un esfuerzo que logra ampliar notablemente la información disponible en la actualidad, que permite conocer los rendimientos sólo a nivel departamental. En la determinación de los rendimientos se utilizaron modelos de aprendizaje computacional, entrenados a partir de 2.007 datos de campo y de 135 covariables territoriales que, agrupadas en grandes categorías, consideraron las siguientes temáticas: topografía, clima, edafología y cobertura vegetal, entre otras.

Los resultados arrojaron que el rendimiento promedio de la soja para la campaña 2020/21 fue de 33 quintales por hectárea (con un coeficiente de variación de 14%), mientras que para el maíz este valor se ubicó en 83 quintales por hectárea (coeficiente de variación de 16%). Sin embargo, más allá de los resultados promedio a nivel provincial, la principal innovación del presente estudio consiste en poder analizar la distribución espacial de los rendimientos a una escala mínima mapeable de 25 hectáreas reflejándose las múltiples particularidades del territorio provincial.

Finalmente, habiendo realizado una estimación de las superficies sembradas y de los rendimientos de cada tipo de cultivo, se calculó la producción. Para la soja, la producción total ascendió a 13.702.942 de toneladas, con máximos de 1.719.985 de toneladas en el Departamento Marcos Juárez; mientras que la producción total de maíz fue de 21.607.089 de toneladas netas<sup>1</sup>, con un pico máximo en el Departamento Río Cuarto, de 4.344.728 de toneladas netas.

---

<sup>1</sup> Es decir, descontando el maíz destinado a forraje y ajustando el área sembrada al área cosechada.

## 2. Relevamiento de datos a campo

Los datos de rinde a campo se obtuvieron mediante el uso de la **aplicación “Rindes Collect”** (Figura 1) la cual consiste en una importante innovación diseñada por IDECOR, sobre la base de la app Land Collect, con el objetivo de recolectar información relevante para el sector agropecuario.

Este desarrollo permite geoposicionar información de interés para el estudio de la dinámica sectorial, como la distribución espacial de los **cultivos**, la distribución espacial de la **fecha de siembra**, la cantidad aproximada de **hectáreas sembradas** por cada dato recopilado, la distribución espacial de **contingencias** o eventos relevantes como inundaciones, sequía, granizo, heladas e incendios, entre otra información crítica. La aplicación, sobre un punto a campo, permite asignar la siguiente información:

- Cultivo relevado.
- Cultivo antecesor.
- Fecha de siembra.
- Superficie sembrada.
- Rendimiento
- Eventos climáticos relevantes que afectaron el rendimiento.

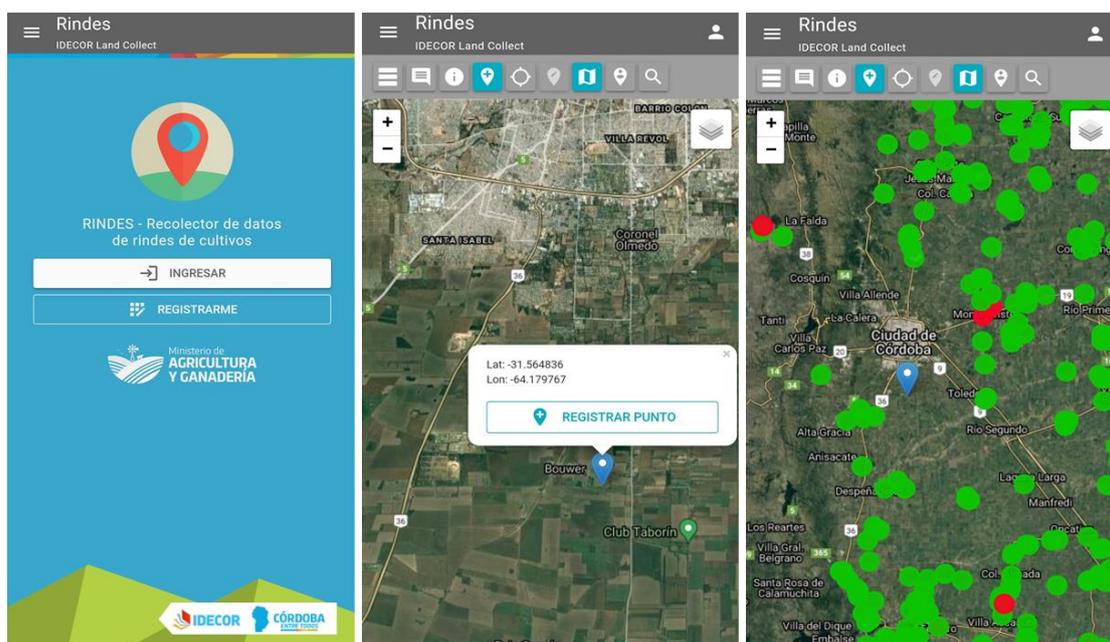


Figura 1. Capturas de pantalla de la aplicación “Rindes Collect”. Fuente: elaboración propia.

En el proceso de relevamiento participaron:

- Secretaría de Agricultura de la Provincia de Córdoba
- Agencias zonales dependientes del Ministerio de Agricultura y Ganadería de la Provincia de Córdoba.
- INTA, Agencias de Extensión.
- IDECOR, de la Secretaría de Ingresos Públicos de la Provincia de Córdoba.
- Agentes privados y otras instituciones.

A partir del proceso de relevamiento coordinado por IDECOR y el Ministerio de Agricultura y Ganadería se obtuvieron **en total 2.072 observaciones de campo**. De estas, luego de un proceso de depuración de observaciones espacialmente atípicas, 2.007 fueron utilizadas para la estimación de los rendimientos por cultivo, y 1.199 se usaron en la estimación de las superficies sembradas. En la figura 2 se presenta la distribución espacial de los datos recogidos y la participación de cada organismo.

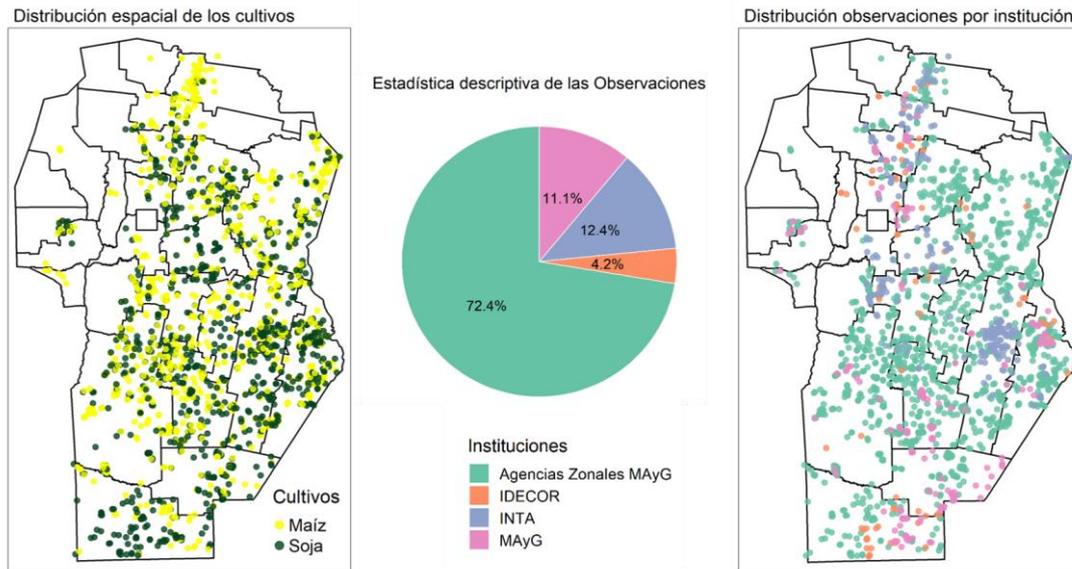


Figura 2. Distribución de datos de campo, por cultivo y por institución aportante. Fuente: elaboración propia.

También se solicitó a los participantes del proceso información sobre la fecha de siembra y si existió alguna contingencia o evento climático adverso. Esta información, cuya distribución espacial se puede observar en la Figura 3, es crítica para el tratamiento de la muestra utilizada en las estimaciones, a los fines de evitar sesgos en la estimación del rendimiento, además de constituir en sí misma información para identificar zonas de manejo según fecha de siembra y áreas afectadas por alguna eventualidad adversa:

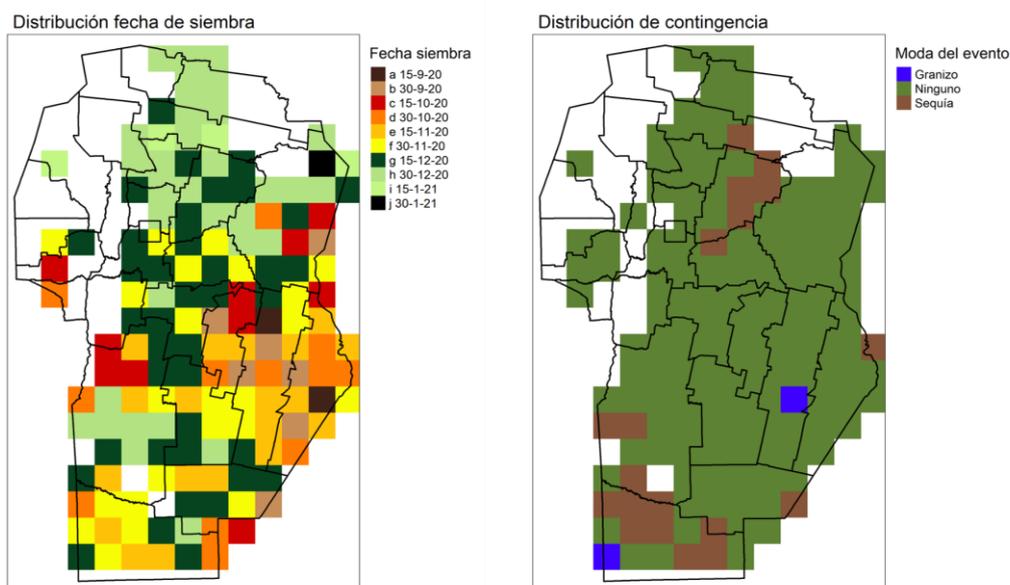


Figura 3. Aproximación de fechas de siembra (izq.) y de eventos climáticos adversos para el sector (der.). Fuente: elaboración propia.

### 3. Estimación de superficies sembradas por cultivo

El objetivo del mapa de coberturas agrícolas consistió en la determinación de la superficie cultivada para la posterior estimación del rendimiento de los principales cultivos agrícolas estivales en la provincia: maíz y soja; logrando de este modo satisfacer la demanda de información generada por el sector agropecuario de Córdoba, alcanzando un nivel de calidad que permita visibilizar las particularidades de la variabilidad espacial del territorio provincial a escala de 1 hectárea. Este mapa es considerado un insumo fundamental que provee las bases para la generación de políticas sectoriales y la generación de información pública para uno de los sectores productivos más dinámicos de la provincia. El mapa fue confeccionado siguiendo una metodología similar al Mapa de Cobertura del Suelo de la Provincia de Córdoba 2017/2018 (García et al., 2018), también desarrollado por IDECOR, aunque en este caso se trabajó con una Unidad Mínima Mapeable de 1 (una) hectárea.

Todo el procesamiento de imágenes fue realizado en Google Earth Engine (GEE) (<https://earthengine.google.com/>), que permite desarrollar análisis computacionales complejos y con grandes cantidades de datos a través del procesamiento de información geoespacial en la nube, independizando al usuario de la capacidad de procesamiento de su propio equipo. La plataforma GEE emplea, por defecto, el sistema de referencia de coordenadas (SRC) WGS 84 (EPSG 4326). Todos los procesos fueron realizados bajo este SRC y posteriormente, la etapa final de la cuantificación de la superficie de cada cultivo en hectáreas, fue realizada en el sistema de información geográfica QGIS v. 3.16 (QGIS Development Team, 2021) utilizando el SRC POSGAR 98/Argentina 4 (EPSG 22174).

#### 3.1 Detalles metodológicos

##### 3.1.1 Confección de mosaicos

El período de análisis comprende **seis épocas** distintas, con el objetivo de cubrir toda la extensión de la campaña 2020/21, identificando las diferentes etapas fenológicas de los cultivos.

Mosaico	Desde	Hasta
Período 1	2020-12-10	2020-12-25
Período 2	2021-01-05	2021-01-25
Período 3	2021-02-05	2021-02-25
Período 4	2021-03-05	2021-03-25
Período 5	2021-04-05	2021-04-25
Período 6	2021-05-05	2021-05-25

Tabla. 1. Fechas empleadas en la construcción de los mosaicos.

En función de los períodos definidos se emplearon, en promedio, 22 escenas de imágenes por período (cada selección por período se denomina colección). Todas las imágenes corresponden a los satélites **Sentinel 2A y 2B** pertenecientes al [programa Copernicus de la agencia espacial europea \(ESA\)](#), de resolución espacial 10 m, corregidas a reflectancia de superficie y bandas del espectro visible y de distintos rangos de infrarrojos, importantes en la determinación de la presencia de vegetación saludable y en la constitución de índices espectrales.

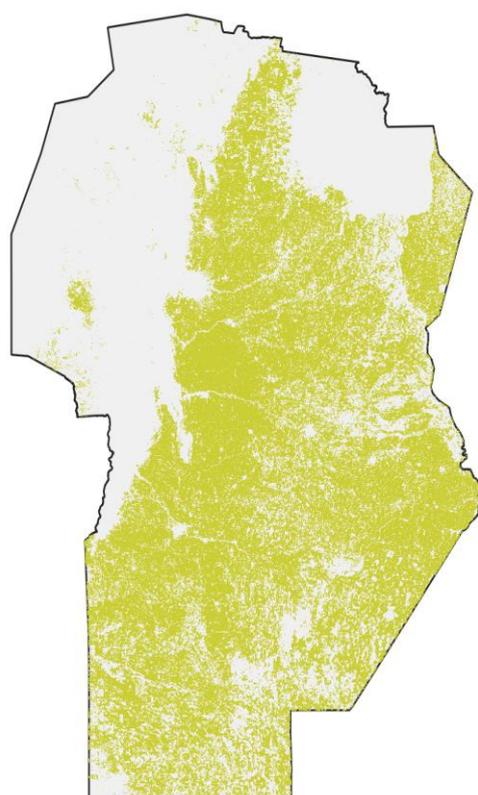
Sobre cada una de las 6 colecciones fue aplicada una máscara para nubes y sombras, seleccionando el valor de probabilidad umbral de 30 %. Adicionalmente, se construyeron los siguientes índices espectrales, de utilidad por su capacidad para identificar y separar coberturas:

- Índice de Vegetación de Diferencia Normalizada [NDVI, por sus siglas en inglés].
- Índice de Vegetación Ajustado al Suelo Mejorado [MSAVI].
- Índice Verde Mejorado [EVI].
- Índice de Diferencia de Agua Normalizado Modificado [MNDWI].

Finalmente, se construyó un mosaico por período, utilizando las medianas de cada colección para cada banda, con la finalidad de obtener un valor acorde al estado fenológico de la mayoría de los píxeles en cada período del tiempo.

### 3.1.2 Definición del área cultivable

En función de los mosaicos de trabajo, se restringió el área de clasificación a zonas de uso agropecuario (Figura 4). Como punto de partida se emplearon los productos “Mapa de Coberturas de Suelo” de la campaña 17/18 y la versión en desarrollo del mismo producto para la campaña 20/21 para limitar el área que se dedica anualmente a cultivos agrícolas, descartando de este modo las tierras no cultivables.



#### Área de producción agrícola campaña 2020-2021

- **Superficie cultivada:**  
7.802.339 ha
- **Período analizado:**  
10/12/2020 - 25/05/2021
- **Fuente:**  
Imágenes satelitales Sentinel 2A y 2B, corregidas a reflectancia de superficie
- **Unidad mínima mapeable:** 1 ha
- **Resolución:** 10 m

Figura 4. Área de producción agrícola campaña 2020/2021. Fuente: elaboración propia

### 3.1.3 Muestras de campo para entrenamiento y validación

Los datos de entrenamiento utilizados para estimar las áreas de producción agrícola, según diferentes tipos de cultivos, consisten en un conjunto de observaciones obtenidas en primer lugar, de una campaña realizada ad-hoc mediante una aplicación diseñada específicamente por IDECOR para este fin denominada "Rindes Collect", a los que se sumaron datos de recetas fitosanitarias provistas por la Secretaría de Agricultura de la Provincia de Córdoba.

Una vez conformada la muestra de campo, se aplicaron series temporales de NDVI para el período de análisis en cada uno de los puntos, calculando la mediana de cada clase de cultivo y la distancia D (Chouakria, 2007) de cada serie a la mediana de su clase, de modo de seleccionar aquellos datos cuya distancia D fuera menor o igual a 0,05, con el fin de depurar la muestra al detectar observaciones espacialmente atípicas. Además, fue necesario realizar controles manuales, ya que en algunos sitios la fluctuación de las napas con diverso grado de tenor de sales afectó la homogeneidad y supervivencia de los cultivos.

La muestra final para el entrenamiento de los algoritmos corresponde a los datos obtenidos mediante la aplicación "Rindes Collect", haciendo un total **1.199 registros**, mientras que para la validación se utilizó el conjunto de datos proveniente de las recetas fitosanitarias, con un total de **4.254 observaciones**. La distribución espacial de los datos de entrenamiento y de validación puede apreciarse en la Figura 5.

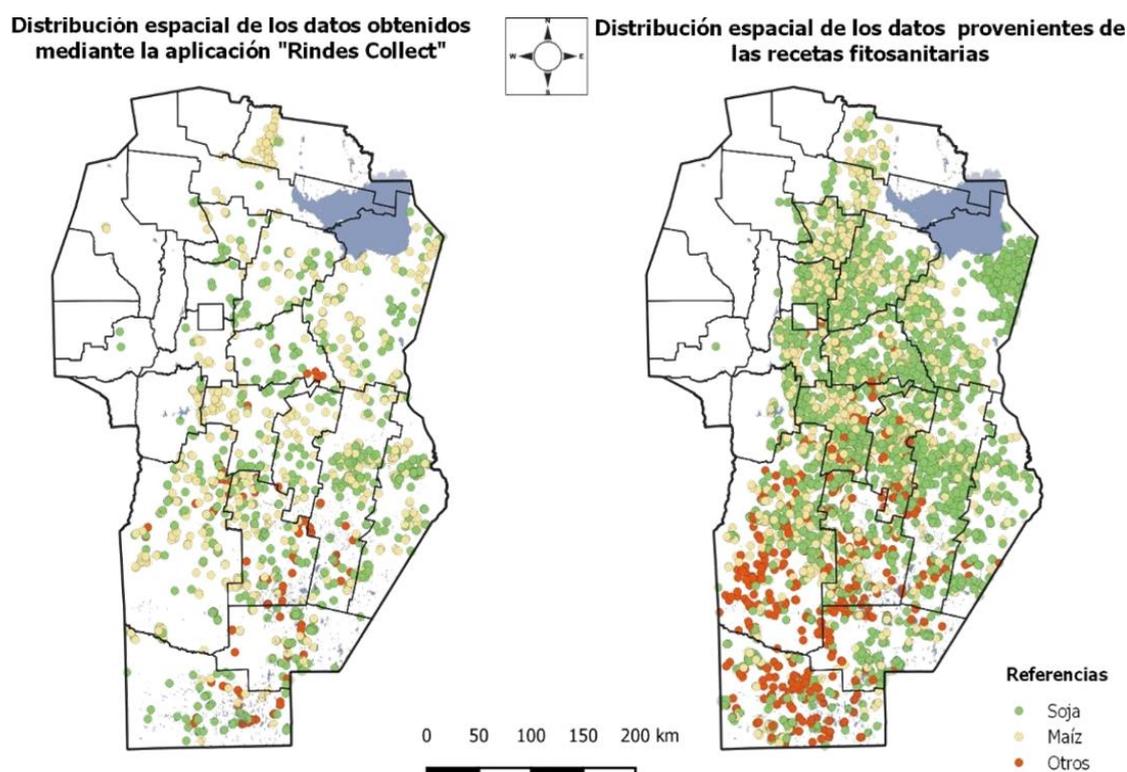


Fig.5. Distribución espacial de los datos provenientes de la aplicación "Rindes Collect" y de las Recetas Fitosanitarias clasificadas por cultivo. Fuente, que se utilizaron para el entrenamiento y validación del modelo: Elaboración propia

### 3.1.4 Clasificación supervisada

La determinación de las coberturas de cultivos se realizó mediante una clasificación supervisada iterativa restringida a la superficie establecida como área de cultivos. El algoritmo empleado como clasificador fue Random Forest (RF) (Breiman, 2001) consistente en la generalización de árboles de decisión (Breiman, 1984) mediante bootstrap y la selección contingente de covariables en cada nodo. El algoritmo asigna nuevas observaciones a cada clase a través de la votación mayoritaria de grupos de árboles de decisión. Los clasificadores son agrupados por bagging y entrenados por muestras seleccionadas mediante bootstrapping. El número de árboles de clasificación empleado fue de 100 y el número de covariables a considerar en cada nodo fue igual a la raíz cuadrada del número total de covariables empleadas.

Luego de probar la ejecución del modelo con diversos grupos de covariables conformados por las bandas espectrales y los índices de vegetación, se conformó el modelo con la combinación que mostró mejores valores en cuanto a los indicadores de precisión: Exactitud del Productor, del Usuario, Global y Kappa. Las covariables incluidas en el modelo fueron las bandas 3 (Verde 0,54-0,57  $\mu\text{m}$ ), banda 8 (Infrarrojo 0,78-0,90  $\mu\text{m}$ ), bandas 11 y 12 (SWIR 1 1,56-1,65  $\mu\text{m}$  y 2 2,10-2,28  $\mu\text{m}$ ). Nelson (2017) ha encontrado que la combinación óptima de bandas utilizando RF para separar coberturas boscosas, incluye una banda del espectro visible y dos bandas del rango del infrarrojo empleando Sentinel 2.

Posteriormente a la clasificación se realizó un proceso de segmentación para reducir la presencia de píxeles aislados. En este proceso se conservaron sólo aquellos clústeres de píxeles mayores a una hectárea. con el algoritmo SNIC (Simple Non-Iterative Clustering) (Achanta y Sússtrunk, 2017). Con el objetivo que los píxeles individuales formasen grupos con sus vecinos con comportamiento espectral y dinámica temporal semejantes, se utilizaron 3 mosaicos creados con la misma metodología que para la ejecución de la clasificación supervisada.

### 3.2 Resultados y mapa de coberturas agrícolas estivales

Los resultados indican que la mayor proporción de la superficie agrícola de la provincia estuvo dedicada al **cultivo de soja**, con un total de **4.073.438 hectáreas**; el **cultivo de maíz** abarcó un área de **3.120.337 hectáreas**, y se detectaron, además, 608.564 has correspondientes a otros cultivos. En total, se identificaron **7.802.339 has cultivadas** en la provincia de Córdoba, para la campaña 2020-21 (Fig. 6).

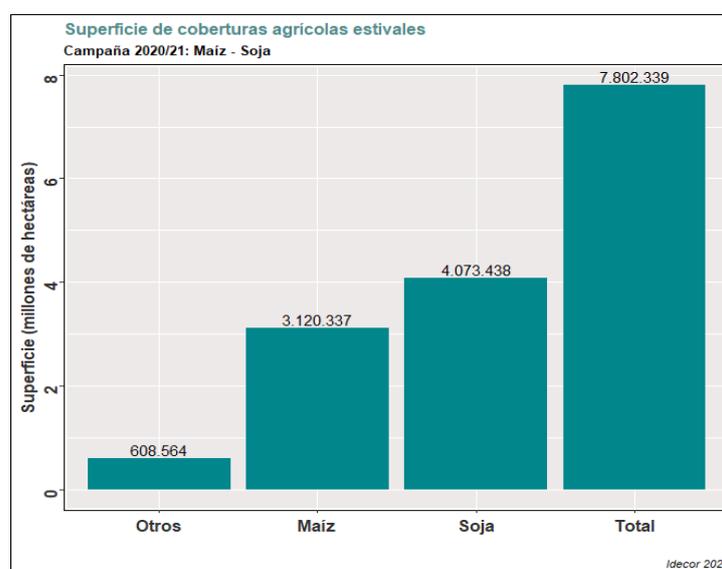
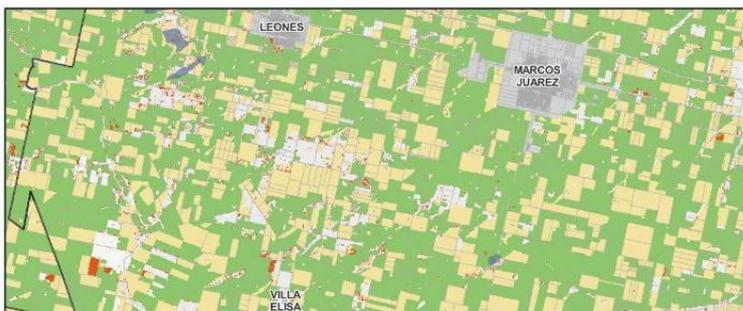
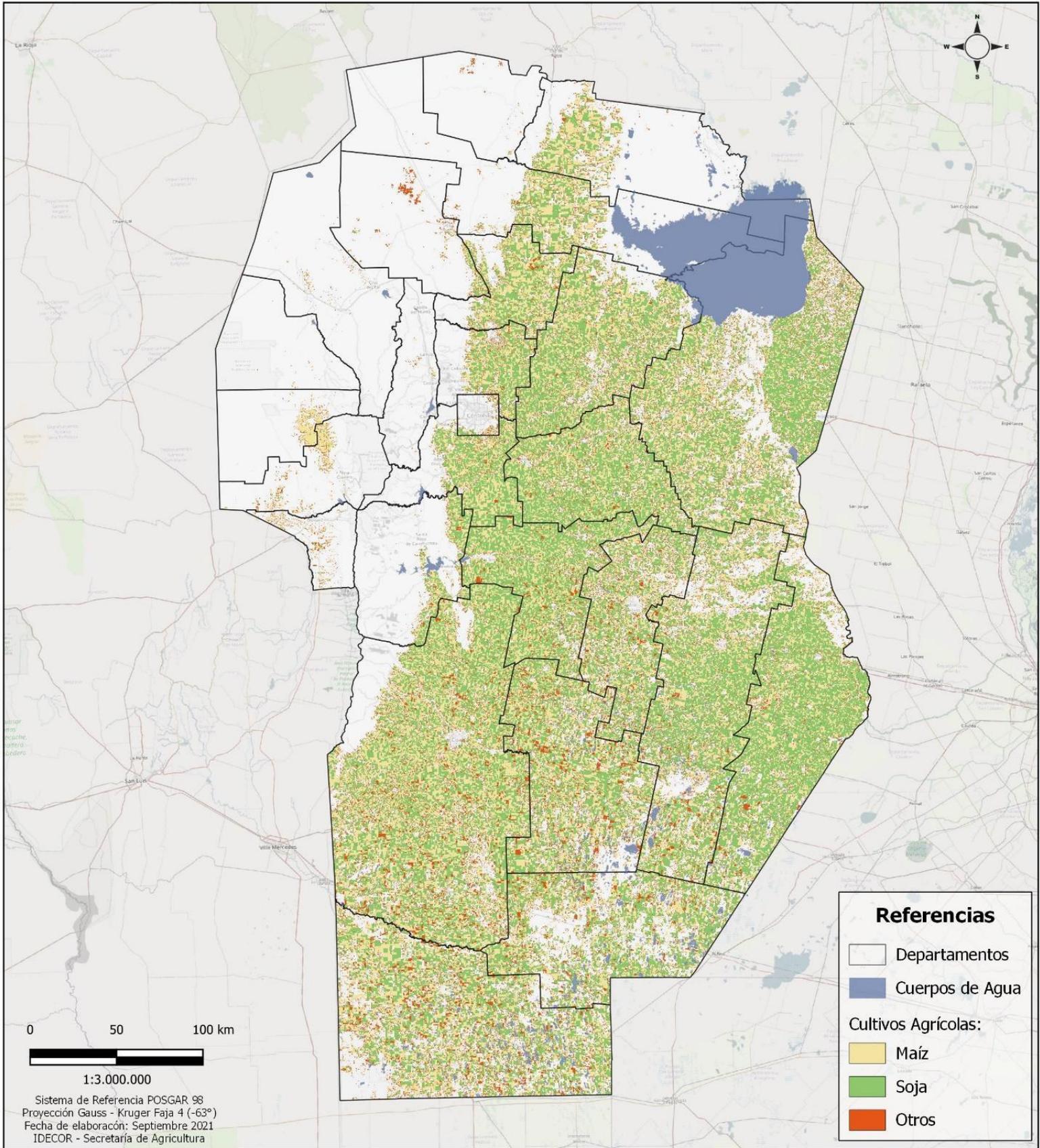


Figura 6. Superficie identificada en la campaña 2020/21 para los principales cultivos agrícolas estivales en la provincia de Córdoba. Fuente: Elaboración Propia

La evaluación del desempeño mostró una **exactitud global (EG) de 95,53%**, mientras que el valor alcanzado por el coeficiente de Kappa (Bishop et al., 1975) para la clasificación, fue de 0,921, indicando un alto grado de exactitud en la clasificación ejecutada a través del modelo y una baja probabilidad de que estos aciertos hayan sido debidos al azar. La Tabla 2 presenta la matriz de confusión obtenida.

Datos de coberturas		Observaciones				Exactitud del usuario
		Maiz	Soja	Otros	Totales	
	Maiz	1379	36	22	1437	<b>96,30%</b>
	Soja	48	2293	21	2362	<b>96,06%</b>
	Otros	17	46	392	455	<b>90,11%</b>
	Totales	1444	2375	435	4254	--
<b>Exactitud del productor</b>		<b>95,17%</b>	<b>97,57%</b>	<b>86,15%</b>	<b>Exactitud Global</b>	<b>95,53%</b>
<b>Coeficiente Kappa</b>						<b>0,921</b>

Tabla. 2: Matriz de confusión que presenta la exactitud alcanzada en el proceso de clasificación supervisada de la cobertura agrícola. Fuente: Elaboración propia



## 4. Estimación de rendimientos por cultivo

A partir de la determinación de las superficies sembradas para cada tipo de cultivo, el primer objetivo de cara a la estimación de la distribución espacial de la producción consiste en determinar la **distribución espacial de los rendimientos para cada cultivo**. Si bien existe información sobre los rendimientos agrícolas, el nivel de agregación que se maneja (a nivel departamental o total provincial) es una limitación para el diseño de políticas sectoriales. Predecir el **rendimiento en unidades espaciales más pequeñas (definidas en 25 hectáreas en el presente estudio)** permite observar, con un nivel de detalle mucho mayor, la variabilidad de la capacidad productiva de los suelos en el espacio. Además, si el relevamiento es sostenido en el tiempo, abarcando campañas consecutivas, se puede obtener información sobre la evolución espacio-temporal del rendimiento y la producción.

Conocer la producción total y su distribución en el espacio abre la posibilidad de construir múltiple información de vital importancia para la política pública del sector, como por ejemplo los márgenes brutos agrícolas a niveles de unidades espaciales más pequeñas que los departamentos (en este caso, unidades de 25 hectáreas).

Adicionalmente, los resultados de la estimación de rendimientos de los diferentes cultivos permiten analizar relaciones entre éstos con diferentes variables territoriales, conocer el consumo o demanda de nutrientes, establecer políticas de manejo de suelo zonales, inferir el potencial de producción, monitorear cambios temporales de la producción de granos, entre otras.

### 4.1 Muestra utilizada

De los 2.072 datos recopilados originalmente a través de la aplicación “Rindes Collect”, y luego de depurar algunas observaciones espacialmente atípicas, se obtuvo una muestra de 2.007 observaciones. Del análisis de las mismas, se observa que 947 correspondieron a maíz, con una superficie sembrada de 78.244 ha., y 1.060 datos a cultivos de soja, con una superficie sembrada de 100.491 ha.

Una vez conformada la muestra, se asignó a cada dato (punto georreferenciado) el valor correspondiente de 135 covariables territoriales (Anexo I), con el objetivo de entrenar diferentes algoritmos de aprendizaje computacional.

### 4.2 Grilla de predicción

Con el objetivo de lograr una **estimación territorial continua de los rendimientos agrícolas** se construyó una grilla de predicción regular para toda el área cultivable de la provincia, con una resolución espacial de **25 hectáreas (celdas de 500 metros de lado)**. Cada celda se consideró “cultivable” si el 50% o más de su superficie estaba ocupada por cultivos agrícolas. En total, la grilla cuenta con 362.493 celdas, sobre las cuales se imputaron 135 covariables territoriales que, agrupadas en grandes categorías, consideran temáticas como **topografía, clima, edafología y cobertura vegetal, entre otras**. Para el cálculo de dichas variables se utilizaron herramientas geomáticas (SIG, procesamiento de imágenes, bases de datos espaciales) y análisis geoespaciales avanzados.

### 4.3 Modelos de aprendizaje computacional aplicados

La predicción de los rendimientos en quintales por hectárea (qq/ha) en cada celda de la grilla surge de un **modelo de aprendizaje computacional denominado Quantile Regression Forest (QRF)** (Meinshausen, 2006). Este tipo de algoritmos se caracteriza por ser de comportamientos anidados (en contraposición a los modelos lineales clásicos que no permiten la interacción entre variables) y de estar compuesto por un conjunto de reglas de partición binaria recursivas, donde un conjunto de datos muestrales es sucesivamente particionado en función de diversas covariables.

El algoritmo QRF es una adaptación del algoritmo Random Forest (Breiman, 2001), que sigue la misma estrategia de ensamble (bagging), aunque se distingue en el procedimiento de agregación de las predicciones de cada árbol. A diferencia de Random Forest, QRF permite realizar predicciones no sólo de la media condicionada de la variable objetivo, sino también del resto de los cuantiles (Amat, 2020). Es decir, se amplía la información brindada por Random Forest y, por lo tanto, se puede proveer de información sobre la distribución total condicional de la variable objetivo, y no únicamente de la media condicionada (Meinshausen, 2006). Esto puede utilizarse para detectar outliers (o anomalías) en el set de datos por medio de la utilización de intervalos cuantílicos (Meinshausen, 2006). Por lo tanto, con el algoritmo QRF se puede obtener la mediana de las predicciones del bosque (cuantil 0,5) en lugar de la media, brindando una mayor robustez en presencia de outliers. Adicionalmente, QRF puede utilizarse para identificar áreas donde la variable explicada tenga mayor varianza (Amat, 2020), logrando una estimación territorial del nivel de dispersión en los resultados. La desviación estándar de la predicción en cada punto ( $y_i$ ) puede estimarse como puede apreciarse en la ecuación que sigue (Hengl et al., 2018):

$$\sigma(y_i) \approx \frac{(\hat{y}_{(q=0,841(y_i))}) - (\hat{y}_{(q=0,159(y_i))})}{2}$$

Los valores de  $\sigma(y_i)$  pueden expresarse en términos de la mediana y así conformar un aproximado del coeficiente de dispersión (CD) y su consiguiente distribución espacial. En la estimación de los rendimientos se analizaron dos posibles opciones de agregación del algoritmo Quantile Regression Forest: mediante la media de cada árbol del bosque (QRF-media) y a través de la mediana (QRF-mediana).

Con el objetivo de controlar los resultados se agregaron, además de QRF, 2 algoritmos adicionales: Gradient Boosting Machine (GBM) y Support Vector Regression (SVR). La capacidad predictiva fuera de la muestra de cada uno de estos algoritmos se estimó mediante un proceso de validación cruzada en 10 grupos, consistente en subdividir la muestra en 10 grupos, sacar uno de ellos, estimar los 4 modelos utilizando los datos de los 9 grupos restantes y medir su capacidad predictiva en el grupo de datos extraído inicialmente. El procedimiento continúa de manera iterativa hasta que cada uno de los 10 grupos es evaluado fuera de la muestra. La capacidad predictiva, estimada de esta manera, se aproximó mediante las siguientes medidas y arrojando los resultados que se presentan en la Tabla 3.

- MAPE: Error relativo medio en valor absoluto.
- MedAPE: Error relativo mediano en valor absoluto.
- RMSE: Raíz cuadrada del error cuadrático medio.
- nRMSE: Raíz cuadrada del error cuadrático medio en términos porcentuales.
- R2: Coeficiente de determinación R cuadrado.
- CV: Coeficiente de variación.
- CD: Coeficiente de dispersión.

Modelos	MAPE	MedAPE	RMSE	nRMSE	R2	CV	CD
QRF-media	0,17	0,1	11,53	19,92	0,87	0,17	0,17
QRF-mediana	0,17	0,1	11,53	19,94	0,85	0,17	0,17
SVR	0,42	0,32	25,41	43,92	0,13	0,38	0,42
GBM	0,18	0,11	11,68	20,19	0,84	0,18	0,18

Cómo surge de la Tabla 3 de criterios de ajustes de los modelos, el algoritmo que minimiza las distintas medidas de cuantificación del error, es: **QRF-media con un MAPE = 0,17**, un RMSE = 11,53, CV y CD = 0,17 y un R2 = 0,87, razón por la cual fue el modelo elegido para efectuar la predicción a nivel provincial.

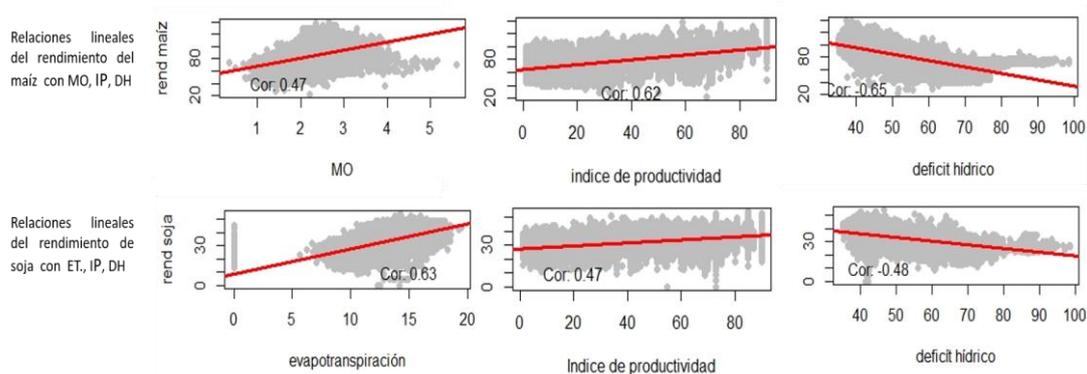
#### 4.4 Resultados obtenidos

A partir del modelo desarrollado se estimaron los rendimientos de soja y maíz en qq/ha para toda el área cultivable de la provincia de Córdoba, tal como puede apreciarse en los Mapas 2 y 3, resultando la siguiente estadística descriptiva de la predicción a nivel provincial:

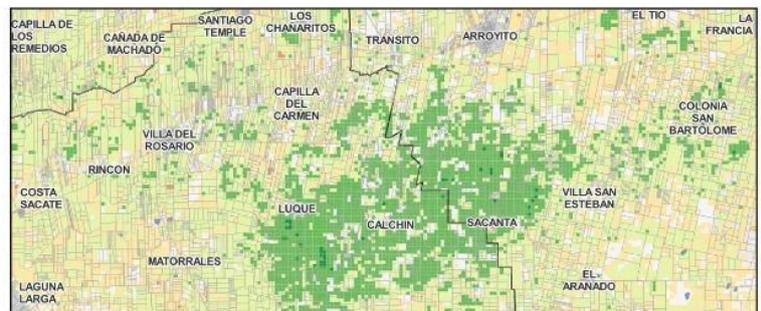
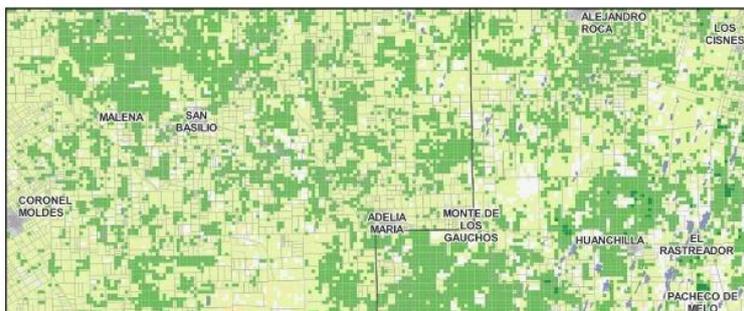
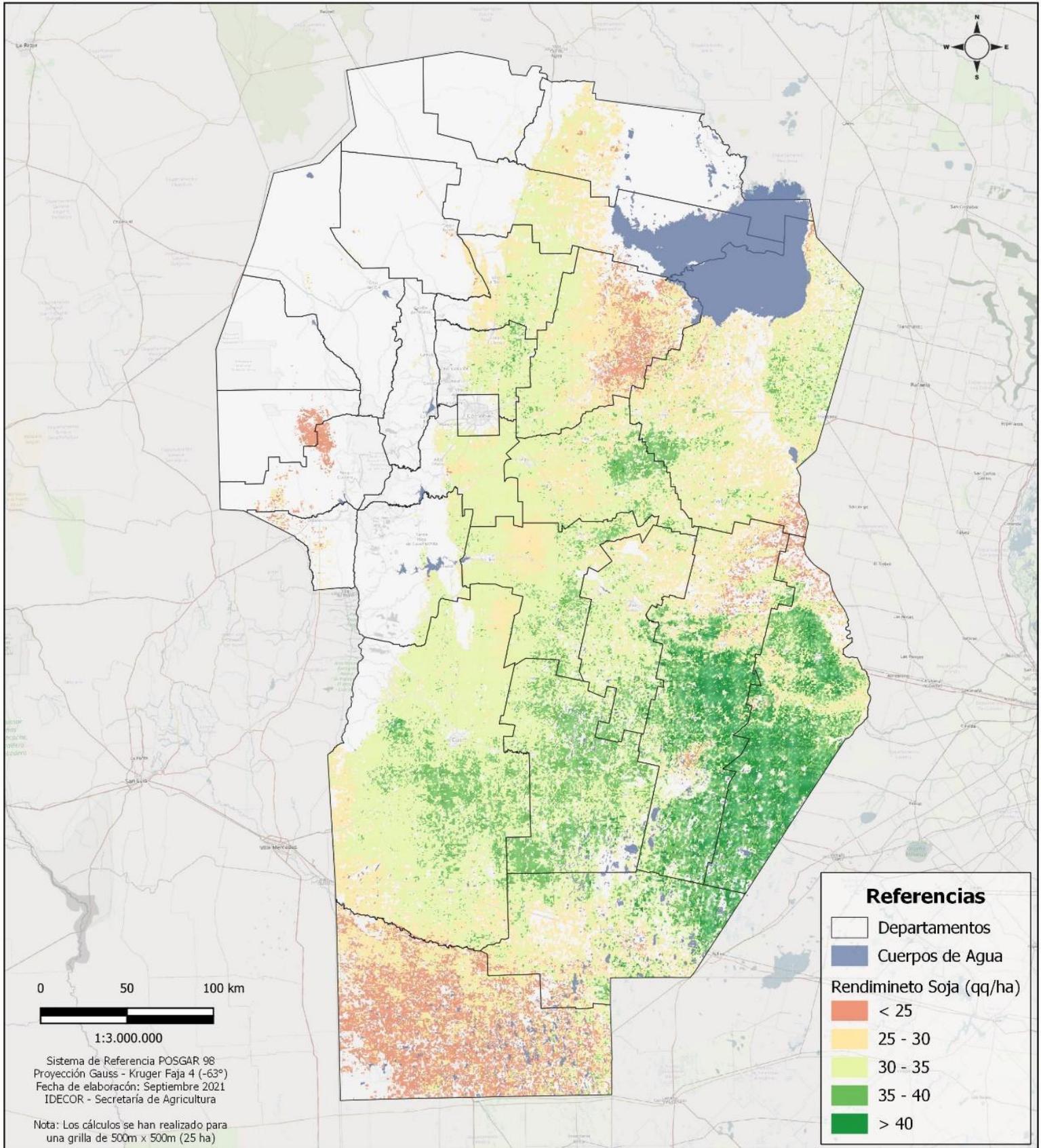
Cultivo	Celdas	Media	Mediana	Min	Max	CV(%)	P10	P90
Soja	362493	33	33	---	54	14	26	38
Maíz	362493	83	82	22	140	16	70	105

Del total de celdas predichas para los rendimientos de soja, menos del 10% resultaron menores a 26 qq/ha, mientras que para el maíz menos del 10% de las celdas resultaron menores a 70 qq/ha. Para los rendimientos más elevados, el 10% de las celdas de soja superaron los 38 qq/ha; entretanto, para el maíz el 10% de las celdas superaron los 105 qq/ha. Para la soja, la media provincial fue de 33 qq/ha, con un coeficiente de variación (cv%) del 14%, en cambio el maíz tuvo un rendimiento medio de 83 qq/ha y un coeficiente de variación (cv%) del 16%.

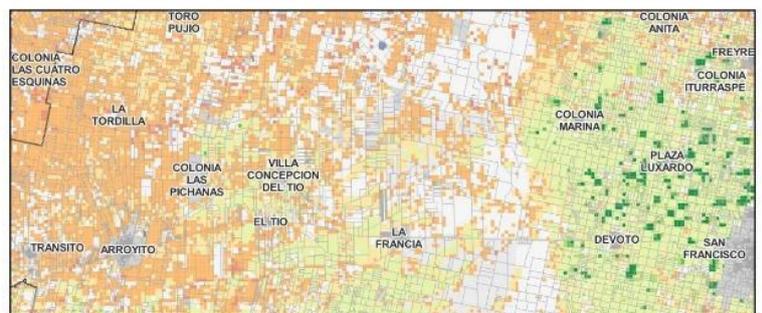
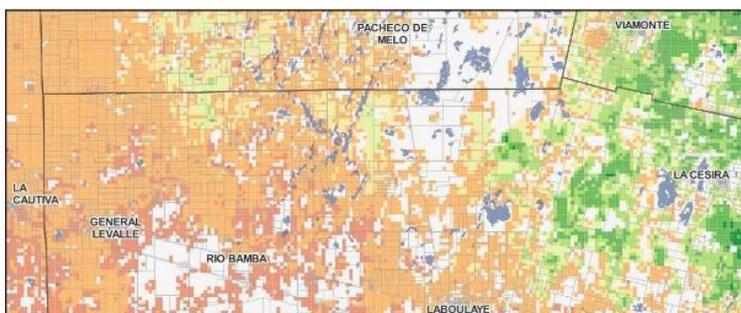
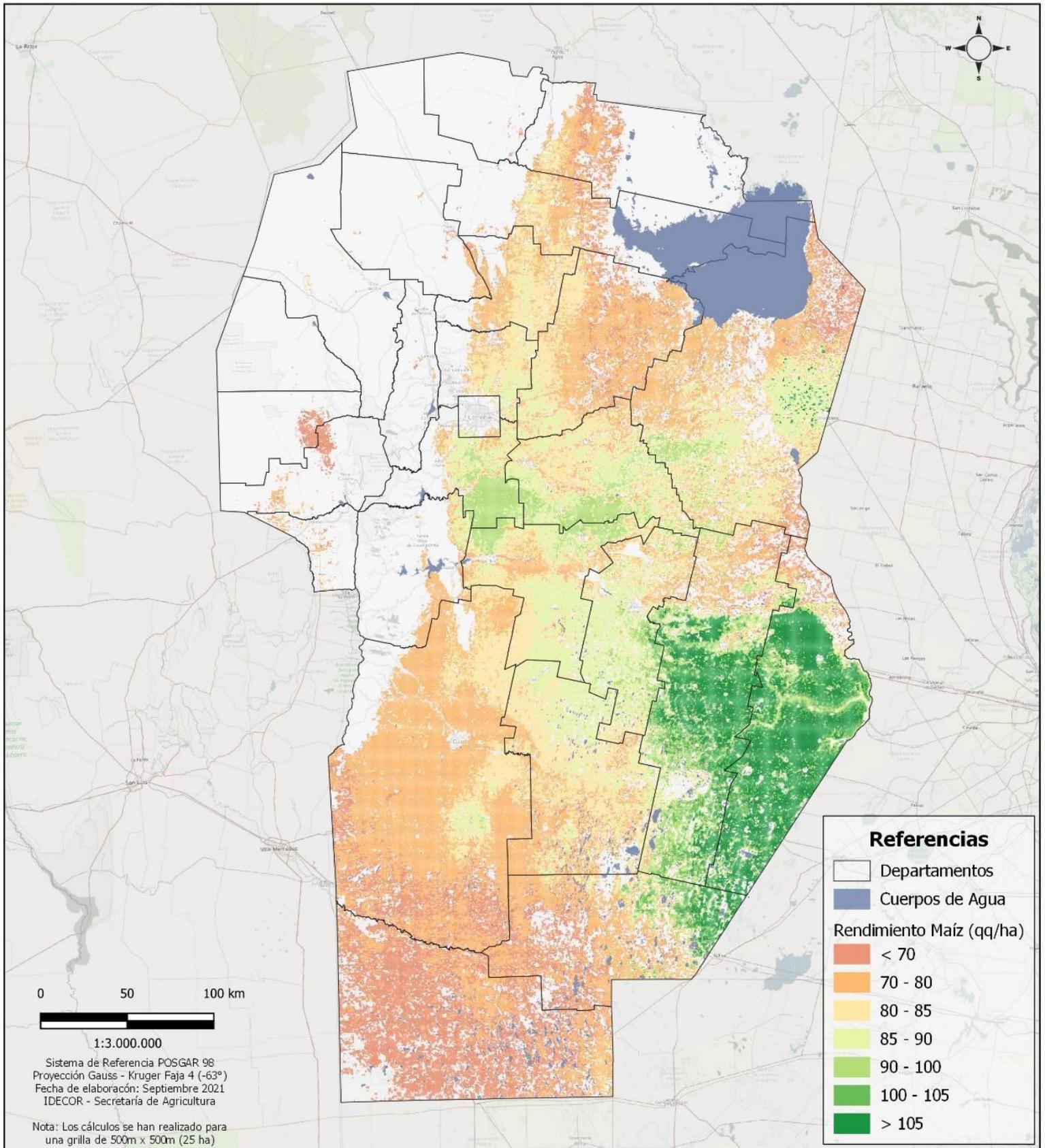
La magnitud de la información utilizada permite comenzar a analizar relaciones entre el rendimiento de cada cultivo con diferentes variables y fenómenos territoriales a una escala de análisis inédita en la provincia hasta el momento. Por ejemplo, puede resultar de interés analizar las correlaciones entre el rendimiento de los cultivos con Materia Orgánica, Índice de Productividad, Déficit Hídrico o la Evapotranspiración, entre muchas otras (Figura 7).



# Estimación de rendimiento Soja (qq/ha)



# Estimación de rendimiento Maíz (qq/ha)



## 5. Estimación de la producción de Soja y Maíz

A partir de la estimación de las superficies sembradas y de los rendimientos de cada tipo de cultivo, se realiza la estimación de la producción en cada celda de la grilla, para así llegar a la **producción total provincial de cada cultivo**. El proceso de cálculo y determinación se resume a continuación.

En primer lugar, se estimó el área sembrada a través del algoritmo de clasificación supervisada Random Forest, que clasifica los cultivos de soja, maíz, maní en una unidad espacial de una hectárea. Esta estimación fue agregada sobre la grilla regular de 25 hectáreas, tomando en cuenta la proporción de cada cultivo dentro de cada celda. De este modo fue posible obtener el área sembrada en términos de hectáreas, al multiplicar el porcentaje de cada cultivo por la superficie de la unidad espacial de celda (25 hectáreas, a excepción de las celdas que se encuentran sobre el límite administrativo de la provincia, en donde es menor).

En segundo lugar, por medio del algoritmo Quantile Regression Forest se estimó el rendimiento de soja y de maíz sobre la misma grilla regular. Luego, la producción de los cultivos resulta de multiplicar, en cada celda de la grilla, el área sembrada (porcentaje de cobertura de cada cultivo multiplicado por la superficie de la celda) por el rendimiento de cada cultivo en dicha celda.

Finalmente, para el cálculo de la producción neta de maíz, se descuentan de la producción bruta las hectáreas sembradas para forraje. Es decir, para calcular la producción de granos netos cosechados se multiplicó la producción bruta, calculada en base al área sembrada en un total de 25.994.663 de toneladas, por el siguiente coeficiente:

$$\text{coeficiente\_de\_cosecha} = \frac{\text{area\_cosechada}}{\text{area\_sembrada}}$$

El valor del coeficiente de cosecha aplicado en el cálculo de la producción neta de maíz surge de la información publicada por el Ministerio de Agricultura, Ganadería y Pesca de la Nación<sup>2</sup>. Para la cosecha 2020/21 en la Provincia de Córdoba el coeficiente informado es igual a 0,83. De esta manera, la producción total neta de granos de maíz asciende a 21.607.089 de toneladas.

Los resultados se pueden observar en los Mapas 4 y 5. Al contar con la producción agrícola a nivel de unidad mapeable de 25 hectáreas, la información puede agregarse en unidades espaciales más grandes, como por ejemplo unidades de carta de suelo, unidades censales o administrativas, pedanía o departamento, entre otras. A los fines del presente informe, **en Tablas 4 y 5 se presenta la producción en toneladas de soja y maíz, agregada a nivel departamental.**

Rendimiento, Superficie sembrada y Producción de **Soja** a nivel departamental

<sup>2</sup> Disponible en: <https://datosestimaciones.magyp.gob.ar/reportes.php?reporte=Estimaciones>

Departamentos	Rendimiento (qq/ha)	Superficie (ha)	Producción (tn)
CALAMUCHITA	32,73	35.597	116.514
CAPITAL	32,06	3.136	10.056
COLON	33,87	63.067	213.599
CRUZ DEL EJE	27,22	54	146
GENERAL ROCA	26,67	318.908	850.376
GENERAL SAN MARTIN	34,28	167.749	575.080
ISCHILIN	28,78	6.156	17.717
JUAREZ CELMAN	35,97	262.673	944.735
MARCOS JUAREZ	38,39	448.050	1.719.985
MINAS	0,0	0	0
POCHO	23,26	774	1.802
PTE ROQUE SAENZ PEÑA	34,42	231.608	797.283
PUNILLA	29,47	108	319
RIO CUARTO	33,61	644.116	2.164.712
RIO PRIMERO	30,26	250.026	756.696
RIO SECO	30,38	43.878	133.316
RIO SEGUNDO	33,37	249.959	834.161
SAN ALBERTO	24,45	1.770	4.327
SAN JAVIER	27,22	926	2.520
SAN JUSTO	32,08	444.648	1.426.307
SANTA MARIA	32,70	96.105	314.271
SOBREMONTÉ	29,32	133	390
TERCERO ARRIBA	33,61	252.225	847.779
TOTAL	32,29	85.676	276.668
TULUMBA	30,83	61.067	188.262
UNION	37,18	405.028	1.505.921
Córdoba Total	33,64	4.073.438	13.702.942

Tabla 4. Datos de Rendimiento, Superficie y Producción de **Soja** a nivel Departamento, Campaña 2020/21.  
Fuente: elaboración propia

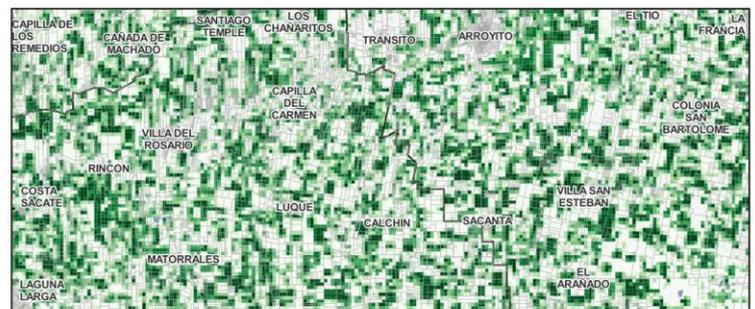
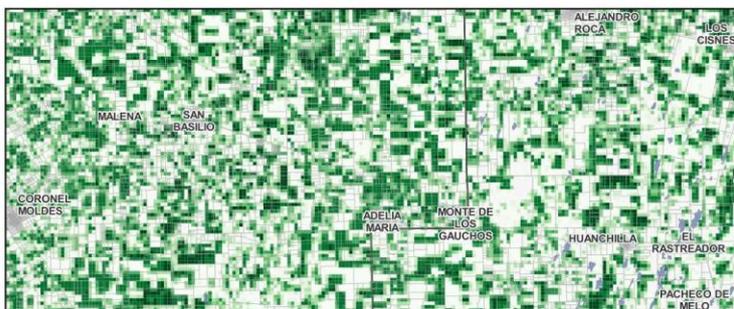
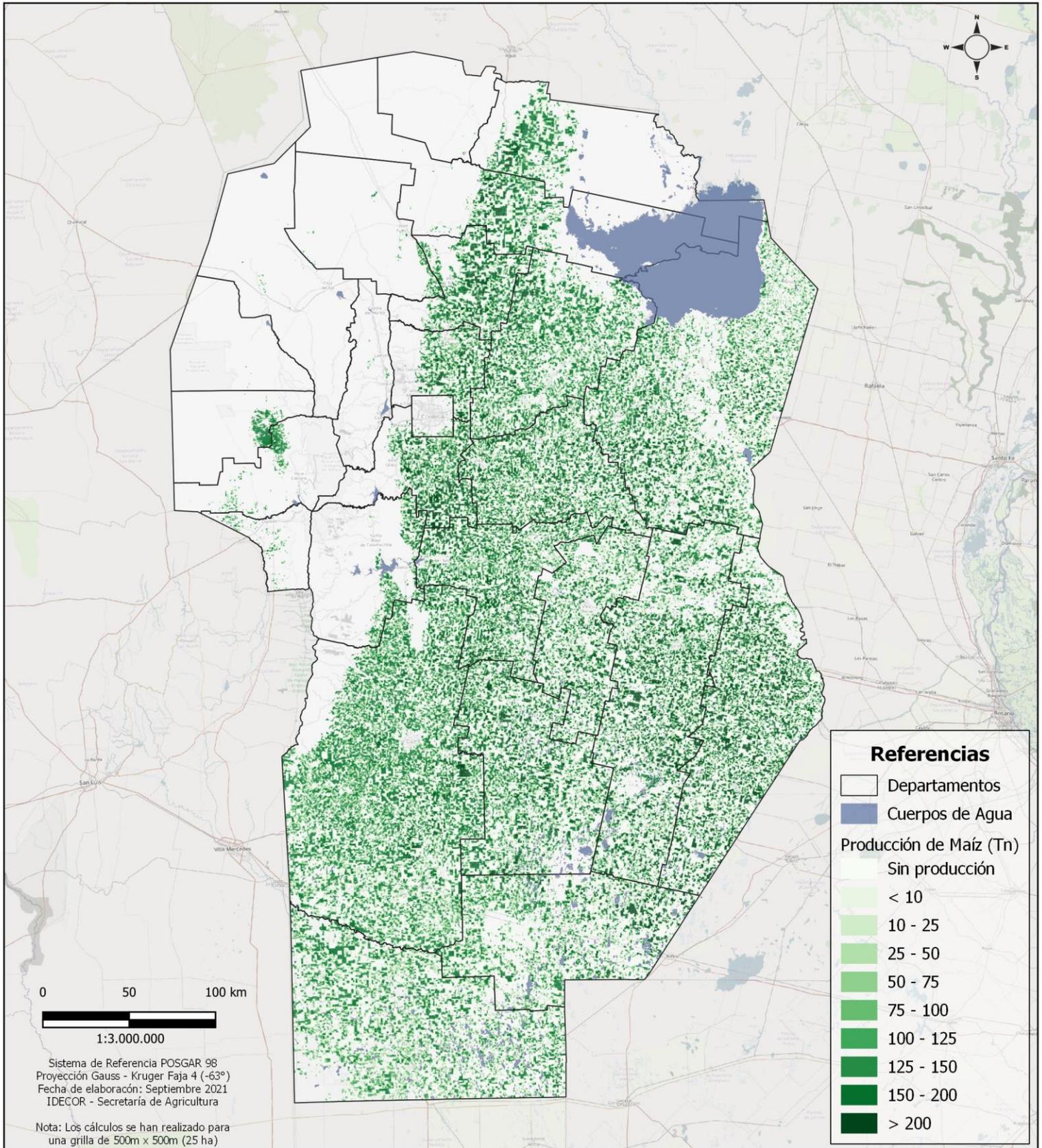
Rendimiento, Superficie sembrada y Producción de Maíz a nivel departamental

Departamentos	Rendimiento (qq/ha)	Superficie (ha)	coef. Cosechado	superficie silaje	Producción (tn)
CALAMUCHITA	81,32	31.409	0,90	3.059	230.545
CAPITAL	82,91	2.516	1,00	0	20.863
COLON	84,09	51.642	0,88	5.941	384.309
CRUZ DEL EJE	73,85	416	0,60	167	1.845
GENERAL ROCA	64,87	294.568	0,94	18.866	1.788.519
GENERAL SAN MARTIN	88,38	113.058	0,53	53.590	525.579
ISCHILIN	74,86	7.282	0,89	819	48.385
JUAREZ CELMAN	83,70	222.010	0,91	19.693	1.693.484
MARCOS JUAREZ	107,31	230.134	0,90	23.726	2.214.917
MINAS	67,71	65		65	0
POCHO	66,75	14.116	1,00	0	94.221
PTE ROQUE SAENZ PEÑA	82,10	166.380	0,92	12.875	1.260.246
PUNILLA	70,55	593	0,00	593	0
RIO CUARTO	77,96	597.625	0,93	40.337	4.344.728
RIO PRIMERO	78,98	195.802	0,76	46.632	1.178.122
RIO SECO	73,97	63.112	0,78	14.119	362.401
RIO SEGUNDO	87,01	159.391	0,86	22.021	1.195.262
SAN ALBERTO	67,41	14.003	1,00	0	94.400
SAN JAVIER	73,01	4.117	1,00	0	30.057
SAN JUSTO	82,53	288.100	0,85	42.254	2.028.884
SANTA MARIA	89,50	78.021	1,00	0	698.252
SOBREMONTA	68,68	196		196	0
TERCERO ARRIBA	86,38	166.474	0,88	19.177	1.272.403
TOTAL	81,85	89.752	0,83	15.097	611.013
TULUMBA	78,21	67.133	0,71	19.158	375.205
UNION	97,75	262.421	0,86	35.631	2.216.826
Córdoba Total	83,31	3.120.337	0,83	394.018	21.607.089

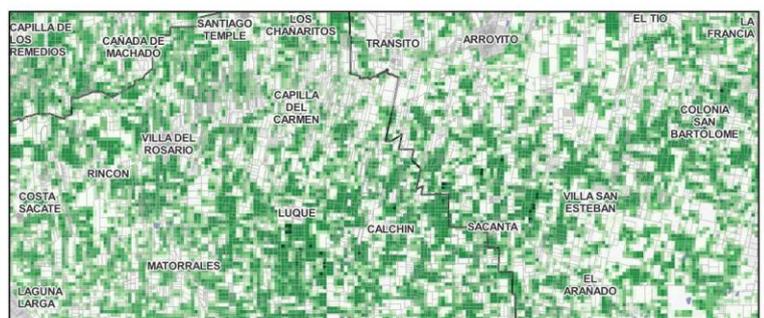
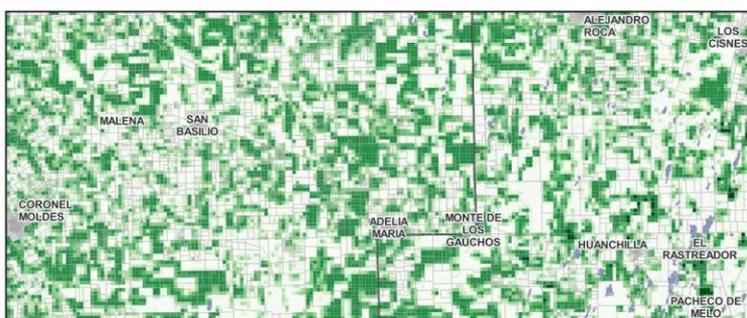
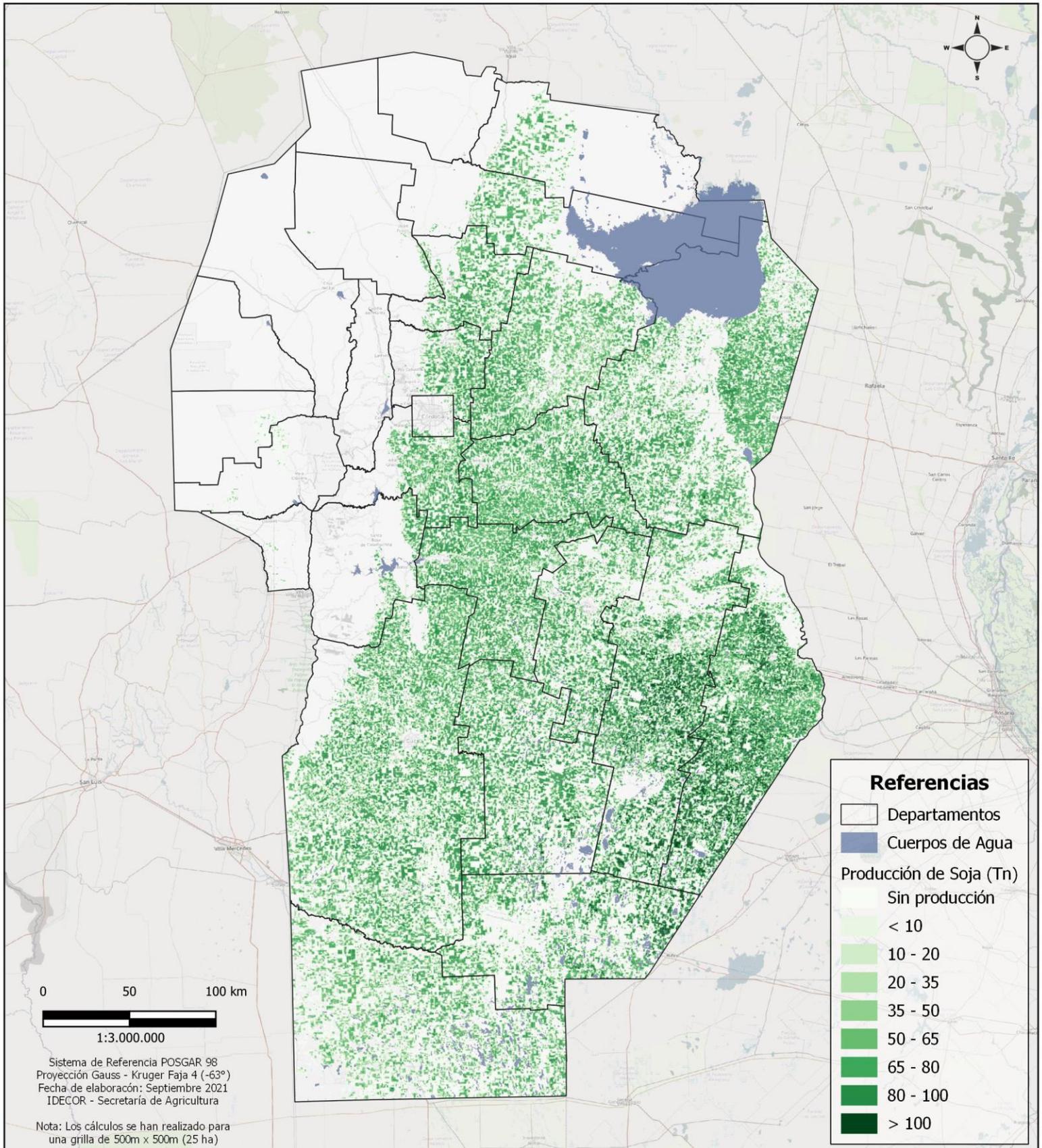
Tabla 5. Datos de Rendimiento, Superficie y Producción de **Maíz** a nivel departamento, Campaña 2020/21.

Fuente: elaboración propia.

# Estimación Producción de Maíz



# Estimación Producción de Soja



## Referencias Bibliográficas

- Amat Rodrigo, J. (2020). Regresión cuantílica: Quantile regression forest. [https://www.cienciadedatos.net/documentos/53\\_regresion\\_cuantilica\\_quantile\\_regresion\\_forest.html](https://www.cienciadedatos.net/documentos/53_regresion_cuantilica_quantile_regresion_forest.html).
- Achanta R. and S. Sússtrunk, "Superpixels and Polygons Using Simple Non-iterative Clustering," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 4895-4904, doi: 10.1109/CVPR.2017.520.
- Bishop, Y.M., S.E. Fienberg, and P.W. Holland, 1975. Discrete Multivariate Analysis Theory and Practice, MIT Press, Cambridge, Massachusetts, 557 p.
- Breiman, L. (2001). Random forests. Machine learning, 45(1), 5–32.
- Chouakria, A. D., & Nagabhushan, P. N. (2007). Adaptive dissimilarity index for measuring time series proximity. Advances in Data Analysis and Classification, 1(1), 5-21.
- Friedman, JH (2001). Aproximación de funciones codiciosas: una máquina de aumento de gradiente. Anales de estadística, 1189-1232.
- García C., Álvarez P., Fleker L., Fuentes L., Morales H., Piunetto M., Teich I.(2018) Mapa de Cobertura y uso de suelo de la Provincia de Córdoba 2017-2018 - 1 ha. Infraestructura de Datos Espaciales de Córdoba (IDECOR), noviembre de 2020.
- Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B.M., Gräler, B., 2018. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ 6, e5518. <https://doi.org/10.7717/peerj.5518>
- Meinshausen, N. (2006). Quantile regression forests. Journal of Machine Learning Research, 7(Jun), 983–999.
- Nelson, M. (2017). Evaluating Multitemporal Sentinel-2 Data for Forest Mapping Using Random Forest. Master, Stockholm University, Stockholm. En <https://www.divaportal.org/smash/get/diva2:1138282/FULLTEXT01.pdf>
- QGIS Development Team. (2021). QGIS Geographic Information System. Open Source Geospatial Foundation (Version 3.16). En <http://qgis.osgeo.org>
- Zhang, L., Zhou, W. y Jiao, L. (2004). Máquina de vectores de soporte de wavelet. Transacciones IEEE sobre sistemas, hombre y cibernética, Parte B (Cibernética), 34 (1), 34-39.

## Anexo I. Variables utilizadas en el modelo de predicción de rendimientos

<b>Variables</b>	<b>Nombre.</b>	<b>Descripción</b>
<b>Climáticas</b>	t_med_anual	Temperatura media anual (1970-2000) World Clim
	rad_solar	Radiación solar media acumulada (1970-2000)
	pp_med_an	Precipitación media acumulada anual histórica (1958-2019)
	t_min_med	Temperatura mínima anual media (1958-2019)
	t_max_med	Temperatura máxima anual media (1958-2019)
	skin_t	Temperatura de la superficie de la Tierra (media mensual en °K). Una variable, para cada mes desde el 09/2020 al 04/2021
	rad_ss	Cantidad de radiación solar que alcanza la superficie de la Tierra por m2 (J/m2). Una variable, para cada mes desde el 11/2020 al 04/2021
	prec_t	Precipitación total acumulada en el mes de análisis (m). Una variable, para cada mes desde el 11/2020 al 04/2021
<b>Económicas</b>	arrenda_hist	Arrendamiento agrícola zonal - promedio campañas BCCBA
	rto_sj_hist	Rendimiento zonal de soja - promedio campañas de 2015 a 16/17 BCCBA
	rto_mz_hist	Rendimiento zonal de maíz - promedio campañas de 2015 a 16/17 BCCBA
	arrenda_2021	Arrendamiento agrícola zonal - 1ra estimación 20/21 BCCBA
	rto_mz1718	Rendimiento zonal de maíz. Cálculos finales campaña 17/18 BCCBA
	rto_sj1718	Rendimiento zonal de soja. Cálculos finales campaña 17/18 BCCBA
	vrm	Variación interanual de las medianas de escenarios de rinde de maíz (adimensional). Una variable, para cada mes desde el 02/2021 al 04/2021. CONAE (Saocom y otras)
	vrs	Variación interanual de las medianas de escenarios de rinde de soja (adimensional). Una variable, para cada mes desde el 02/2021 al 04/2021. CONAE (Saocom y otras)
<b>Posición</b>	geom	Geometría
	id	Id grilla (2019)
	x	Coordenada x del centroide
	y	Coordenada y del centroide
<b>Hídricas</b>	def_hidric	Déficit hídrico medio histórico (1958-2019)
	pdsi	Índice de Severidad de Sequía media histórica (1958-2019)

	rec_1median	Recurrencia 1984-2000 - mediana dentro de la celda
	rec_1stdev	Recurrencia 1984-2000 - desvío estándar dentro de la celda
	rec_2median	Recurrencia 2001-2009 - mediana dentro de la celda
	rec_2stdev	Recurrencia 2001-2009 - desvío estándar dentro de la celda
	rec_3median	Recurrencia 2010-2018 - mediana dentro de la celda
	rec_3stdev	Recurrencia 2010-2018 - desvío estándar dentro de la celda
	perc_agua	% agua en la celda (Año hidrológico 2020-2021)
	mndwi	Índice de agua de diferencia normalizada modificado (MNDWI). Una variable, para cada mes desde el 11/2020 al 04/2021
	vol_s_wat	Volumen de agua en profundidades de 0 a 289 cm (m3/m3). Una variable, para cada mes desde el 09/2020 al 04/2021
	hs	Volumen de agua del suelo de 0 a 50 cm de profundidad (%). Una variable, para cada mes desde el 09/2020 al 04/2021. SAOCOM, CONAE
<b>Suelo / Vegetación</b>	tipo_suelo	Posición - Orden de suelo
	drenaje	Posición - Limitante edafológica de drenaje
	alcalinidad	Posición - Limitante edafológica de alcalinidad sódica
	prof_efect	Posición - Limitante edafológica de profundidad efectiva
	salinidad	Posición - Limitante edafológica de salinidad
	textura	Posición - Limitante edafológica de textura
	fosforo	Contenido de fósforo en suelo (2021)
	nitrogeno18	Contenido de nitrógeno en suelo (FCA, BPA, otras, 2018)
	potasio18	Contenido de potasio en suelo (FCA, BPA, otras, 2018)
	cic18	Capacidad de intercambio catiónico (FCA)
	ph	pH del suelo (2021)
	ndvi_mediana	Mediana de NDVI (promedio histórico 2000-2020), MODIS
	ndvi_stdev	Desvío estándar de NDVI (promedio histórico 2000-2020), MODIS
	evapo_medi_an	Evapotranspiración media mensual acumulada de la serie (2001-2020)
	n2_cob16	% sup. Cultivos anuales de secano
	n2_cob17	% sup. Cultivos Irrigados
	cu_moda	Moda de CU de la celda

---

ip_median	Mediana del Índice de Productividad
ip_stdev	Desvío estándar del Índice de Productividad
cu_clase0	% Capacidad de Uso sin clasificar
cu_clase1	% Capacidad de Uso clase I
cu_clase2	% Capacidad de Uso clase II
cu_clase3	% Capacidad de Uso clase III
cu_clase4	% Capacidad de Uso clase VI
cu_clase5	% Capacidad de Uso clase V
cu_clase6	% Capacidad de Uso clase VI
cu_clase7	% Capacidad de Uso clase VII
cu_clase8	% Capacidad de Uso clase VIII
cu_clase9	% Capacidad de Uso clase áreas misceláneas
morganica	Contenido de materia orgánica en suelo (2021)
arcilla	Contenido de arcilla (2021)
perc_pivot	Porcentaje de superficie con pivotes en la grilla
evap_ftv	Evapotranspiración de la vegetación acumulada (m). Una variable, para cada mes desde el 11/2020 al 04/2021
lailveg	Superficie total de hojas verdes por m2 de suelo horizontal para vegetación baja (m2/m2). Una variable, para cada mes desde el 11/2020 al 04/2021
evap_t	Evaporación total acumulada en el mes de análisis (m). Una variable, para cada mes desde el 11/2020 al 04/2021
evi	Índice de vegetación mejorado (EVI). Una variable, para cada mes desde el 11/2020 al 04/2021. MODIS
msavi	Índice de vegetación ajustado al suelo modificado (MSAVI). Una variable, para cada mes desde el 11/2020 al 04/2021. MODIS
ndvi	Índice de vegetación de diferencia normalizada (NDVI). Una variable, para cada mes desde el 11/2020 al 04/2021. MODIS

---

**Topografía**

altura_median	Mediana de la altura (msnm)
altura_stdev	Desvío estándar de la altura (msnm)
pend_median	Mediana de la pendiente (%)
pend_stdev	Desvío estándar de la pendiente (%)

---



---

[mapascordoba.gob.ar](http://mapascordoba.gob.ar)    [idecor.cba.gov.ar](http://idecor.cba.gov.ar)  
[idecor@cba.gov.ar](mailto:idecor@cba.gov.ar)